Journal of Autoimmunity 157 (2025) 103474

Contents lists available at ScienceDirect

Journal of Autoimmunity

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

Lymphoma driver mutations at the root of somatic evolution of nerve-damaging autoantibodies in myelin associated glycoprotein neuropathy

Shane Kelly [a,b], Mandeep Singh [a,b], Amanda Russell [a], Katherine J.L. Jackson [a], Timothy J. Peters [a], Andrew Carr [c], Anthony D. Kelleher [c,d], Matt Field [e,f], Matthew Silsby [g], Dan Suan [a,1], Christopher C. Goodnow [a,h,*,1]

[a] Garvan Institute of Medical Research, Darlinghurst, NSW, Australia
[b] St Vincent's Clinical School, UNSW Sydney, NSW, Australia
[c] Immunology and HIV Unit, St Vincent's Hospital, Darlinghurst, NSW, Australia
[d] The Kirby Institute, UNSW Sydney, Kensington, NSW, Australia
[e] Australian Institute of Tropical Health and Medicine and Centre for Tropical Bioinformatics and Molecular Biology, Smithfield, Cairns, QLD, Australia
[f] Menzies School of Health Research, Darwin, NT, Australia
[g] Department of Neurology, Westmead Hospital, Westmead, NSW, Australia
[h] Cellular Genomics Futures Institute & School of Biomedical Sciences, UNSW Sydney, Australia

ARTICLE INFO

Handling editor: C Selmi

Keywords:
MAG neuropathy
Demyelinating neuropathy
Autoantibody
Somatic mutation
MYD88
B lymphocyte
Antibody affinity

ABSTRACT

Background: In autoimmune disease it is not understood how self-reactive B cells escape immune tolerance checkpoints to produce pathogenic autoantibodies.
Objective: In patients with demyelinating polyneuropathy caused by IgM autoantibodies against myelin associated glycoprotein (MAG) and the sulphated trisaccharide CD57, we aimed to test the hypothesis that B cells making the autoantibody escaped tolerance by acquiring lymphoma driver somatic mutations.
Methods: Deep single-cell RNA, DNA, flow cytometric and antibody specificity analysis of blood from three patients with MAG neuropathy.
Results: MAG autoantibody-producing B cell clones exhibited extensive intraclonal immunoglobulin V(D)J hypermutation. In many of the sub-clonal branches, the replacement:silent ratio of V-region mutations was not different from that expected for unselected mutations, although in some branches the mutations either increased or eliminated binding to MAG and CD57 autoantigens. Prior to intraclonal V(D)J diversification, each clone had acquired a gain-of-function *MYD88*$^{p.L265P}$ mutation, and some branches had acquired additional somatic mutations in *CXCR4, IGLL5* and *BTG2*. Whilst all MAG-binding clones harboured the *MYD88*$^{p.L265P}$ mutation, the same mutation was also found in some control, polyclonal B cells. Deep sequencing of different blood cell subsets indicated *MYD88*$^{p.L265P}$ was confined to B cells.
Conclusion: In three MAG neuropathy patients we find evidence that the self-reactive B cells responsible for their disease acquired a classical lymphoma driver somatic mutation early in their clonal expansion.

1. Introduction

A key question in autoimmune disease is how do B cells making pathogenic autoantibodies evade the series of immune tolerance checkpoints that normally inhibit their maturation, survival, proliferation and plasma cell differentiation? [1] Here we address this question in the context of myelin associated glycoprotein (MAG) neuropathy.

MAG neuropathy is a form of chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) caused by IgM autoantibodies that bind to MAG, a 100 kDa transmembrane glycoprotein on the plasma membrane of Schwann cells that is critical in the formation and maintenance of neuronal myelin sheaths around axons [2,3]. The classical form of

* Corresponding author. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia.
E-mail address: c.goodnow@garvan.org.au (C.C. Goodnow).
[1] These senior authors contributed equally to this work.

https://doi.org/10.1016/j.jaut.2025.103474
Received 16 March 2025; Received in revised form 23 July 2025; Accepted 23 August 2025
Available online 6 September 2025

MAG neuropathy, accounting for three quarters of cases, presents with paraesthesia or sensory ataxia followed by slowly progressive, symmetrical weakness and later tremor [4,5]. Nerve conduction studies demonstrate a distal acquired demyelinating symmetric polyneuropathy ("DADS" phenotype), with disproportionately prolonged distal motor latencies. Nerve biopsies typically demonstrate characteristic widening and splitting of the myelin lamellae, associated with loss of myelinated axons. IgM, with or without C3d, may be deposited on myelin sheaths [6,7], although this is highly variable in larger cohorts of MAG neuropathy patients [4]. Evidence for the pathogenic nature of the autoantibody comes from studies where MAG neuropathy patient serum and purified IgM caused demyelination when injected with fresh complement into feline peripheral nerves [8,9]. In chickens, injection of anti-MAG antibodies caused segmental peripheral nerve demyelination and histological features similar to the human disease [10].

Atypical forms of MAG neuropathy include those with a "CIDP phenotype" and predominant weakness or small fibre neuropathy, and MAG autoantibody levels by ELISA may not differentiate these subtypes from the typical form of the disease [2,4,5]. 10–22 % of MAG neuropathy patients also test positive for autoantibodies against ganglioside antigens, including GD1b, GQ1b, GM1, GM2, GM3 and GM4 [4,5]. No other autoantibodies have been consistently reported in MAG neuropathy, nor is there an association between MAG neuropathy and systemic connective tissue diseases (*e.g.* lupus), anti-nuclear antibodies, nor antecedent infective triggers (c.f. Guillain Barre syndrome).

The "self" molecule targeted by MAG autoantibodies is CD57, a sulphated trisaccharide SO_4-3GlcAβ1–3Galβ1–4GlcNAc present on peripheral nerve myelin as part of N-linked or O-linked glycans on MAG, other myelin glycoproteins and in sulfoglucuronyl glycosphingolipids [11,12]. CD57 sulphated trisaccharide attached to a 110 kDa protein is also displayed abundantly on the surface of natural killer cells in the blood, where the antigen was originally called human natural killer 1 (HNK-1), and on a subset of effector CD8 and CD4 T cells in blood and tissues [13]. MAG neuropathy autoantibodies can be depleted by adsorption with $CD57^+$ T cell lysates [14], and serum antibody titres against the synthetic CD57 sulphated trisaccharide are better correlated with disease severity than antibody titres measured by ELISA against MAG glycoprotein [15]. Since tolerance checkpoints normally inhibit the maturation and survival of B cells making immunoglobulin against self-antigens on the plasma membrane of lymphocytes [16–19], B cells that express $CD57^-$and MAG-binding immunoglobulins would not be expected to accumulate as mature, recirculating B cells and would be expected to lose CD57 binding during immunoglobulin hypermutation in germinal centres [16,20].

It is not known how anti-MAG/CD57 B cells overcome these tolerance checkpoints in MAG neuropathy, but a clue comes from the observation that most patients with MAG neuropathy have a measurable monoclonal IgM paraprotein in their blood [2,5]. This homogeneous secreted IgM antibody binds the CD57 sulphated trisaccharide on MAG, whilst deglycosylation of the MAG antigen abolishes the IgM reactivity of serum from patients with MAG neuropathy [21]. The plasma concentrations of monoclonal IgM in MAG neuropathy are typically much lower than in the IgM-secreting lymphoplasmacytic lymphoma disease, Waldenstrom's macroglobulinemia, or in other forms of B cell non-Hodgkin's lymphoma, although these lymphomas develop in 12–35 % of patients with MAG neuropathy [2].

Were it not for the neuropathological effects of their monoclonal IgM, most MAG neuropathy patients would fit within the haematological criteria of IgM monoclonal gammopathy of undetermined significance (IgM-MGUS), due to the low paraprotein level and absence of end-organ damage. IgM-MGUS can be incidentally detected by serum electrophoresis and immunofixation screening in ~300–600 per 100,000 (0.3–0.6 %) people aged over 50 years [22,23], which is higher than the ~10 per 100,000 (0.01 %) incidence of CIDP in this age bracket [2]. Among people with MGUS, ~15 per 100,000 persons per year develop CIDP [24], whereas more than 1000 per 100,000 persons per year with IgM MGUS will progress to Waldenstrom's macroglobulinemia, B cell lymphoma, or chronic lymphocytic leukemia (CLL), with relative risks of 288, 10.6 and 4.3, respectively [25].

Waldenstrom's macroglobulinemia is unusual among B cell neoplasms because >90 % of these malignancies have acquired the same lymphoma driver mutation, a point mutation in *MYD88* changing the Leucine 265 codon to Proline [26]. $MYD88^{L265P}$ is also found in >50 % of IgM MGUS [27–29], 29 % of activated B cell type diffuse large B cell lymphomas (ABC DLBCL) with a lymphoplasmacytic gene expression program [30], and 6 % of the subtype of CLL with hypermutated *IGHV* genes [31,32].

MYD88 propagates signals downstream of Toll-like receptors (TLRs) and the IL-1 receptor by nucleating helical fibres (myddosomes) composed of MYD88, IRAK1/IRAK2 and IRAK4 to activate nuclear translocation of the NF-κB transcriptional activator [33,34]. $MYD88^{L265P}$ creates a gain-of-function in this pathway [30], but animal studies show that it is insufficient to drive sustained B cell clonal growth or lymphoma on its own because it is countered by the induced NF-κB inhibitor, *TNFAIP3* (A20), and the induced BCL2 inhibitor, *BCL2L11* (BIM) [35–37]. $MYD88^{L265P}$ corrupts immune tolerance checkpoints, driving mouse B cells with self-binding surface IgM to differentiate into lymphoplasmacytic cells secreting IgM autoantibodies *in vivo* if accompanied by increased expression of the surface IgM receptor subunit, CD79B [38].

The presence of $MYD88^{L265P}$ lymphoma driver mutations in bulk sequencing of bone marrow samples of 59–73 % of people with MAG neuropathy [39,40] poses two key questions in the pathogenesis of this autoimmune disease. First, is $MYD88^{L265P}$ present in the cells making the pathogenic autoantibody? If so, this leads to a second question: was $MYD88^{L265P}$ present at the start of evolution of a rogue clone making anti-MAG IgM, where dysregulated NF-κB might contribute to checkpoint escape, or was it acquired later? Since $MYD88^{L265P}$ is only found in the CLL subset that has undergone *IGHV* gene somatic hypermutation [31,32], self-reactive B cells causing MAG neuropathy may first need to escape tolerance, proliferate and undergo AID-mediated hypermutation before the *MYD88* mutation is acquired. Here, we address these questions by undertaking in-depth analysis of the autoantibody-forming clonal B cells in three patients with MAG neuropathy, using single cell DNA and RNA sequencing and by expressing representative antibodies from single cells.

2. Methods

2.1. Patient samples

Peripheral blood was collected from three patients (P1, P2 and P3; details in Supplementary Table 1) with demyelinating polyneuropathy associated with MAG autoantibodies >10,000 BTU by ELISA. Peripheral blood mononuclear cells (PBMCs) were prepared by Ficoll-Paque centrifugation and cryopreserved in RPMI 1640, 50 % foetal calf serum and 10 % DMSO. Written informed consent was obtained prior to blood collection. The study was approved by the relevant human research ethics committee (HREC/19/WSLHD/5449) and all research was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2. Flow cytometry and fluorescence activated cell sorting (FACS)

Thawed PBMCs were washed twice in PBS/1 % BSA and incubated with Fc block for 30 min on ice. Cells were washed and stained with fluorophore-labelled antibodies (Supplementary Table 2) as well as a Fixable Viability Dye (eBioscience). Approximately 1×10^7 PBMCs were sorted for P1 and P3, and 2×10^7 PBMCs for P2 (across two sorts), performed using a FACS Aria III (BD Biosciences). Single cell indexed sorting was performed, gated on clonal B cells and control memory B

cells; details of the gating strategy are discussed in the results section for each of the three patients. Cells were sorted into 2 μL of RLT buffer (Qiagen) in 96-well LoBind PCR plates (Eppendorf).

2.3. Parallel single cell genomic and transcriptomic sequencing

Genomic DNA (gDNA) and mRNA were isolated from the same single cells using G&T-seq [41,42]. The mRNA of each cell was processed for scRNA-seq using the Smart-seq2 protocol described below. gDNA was subjected to multiple displacement amplification (MDA) using the REPLI-g Single Cell Kit (QIAGEN) according to the manufacturer's instructions. MDA amplified products were purified using AMPure XP beads (Agencourt) and processed for either targeted capture or gene-specific PCR amplification.

2.4. Single cell RNA sequencing and gene expression analysis

mRNA from each cell was processed for scRNA-seq using the Smart-seq2 protocol [43]. Sequencing libraries were prepared using the Nextera XT Library Preparation Kit (Illumina). Cells were sequenced using an Illumina NextSeq 500 instrument with 150 bp paired-end reads to a median depth of 1 million reads per cell. scRNA-seq sequencing reads were aligned to GRCh38.p12 using STAR2.5.4b [44], and de-duplicated using SAMtools 0.1.19 [45]. Per-gene expression levels were called at count and transcript per million (TPM) levels using RSEM v1.3.0 [46]. Gene expression counts were normalized using Sanity [47] with default values, and differentially expressed genes (DEGs) were identified using limma [48] on the log-transformed recovered counts. Bonferroni correction was applied to each set of p-values and DEGs were defined as having a family-wise error rate (FWER) of <0.05.

2.5. Targeted capture sequencing for lymphoma driver mutations

Two target enrichment libraries (Roche NimbleGen) were designed for a panel of 178 genes found recurrently mutated in Waldenstrom's macroglobulinemia, CLL and non-Hodgkin's lymphoma [42]. For each gene, genome coordinates of their corresponding exons were obtained from the GRCh38 primary assembly. Design of probes from target regions and synthesis was performed by Roche NimbleGen using the SeqCap EZ format with a maximum of five matches to the human genome.

For targeted capture the quality of MDA efficiency from each single cell was first assessed with 3–5 individual PCR reactions of different chromosomal regions, using primer sequences [49] and PCR with *Taq* polymerase kit (Invitrogen) and the following conditions: 95 °C for 3min; [95 °C for 30s, 60 °C for 15s, 72 °C for 1min] x35 cycles; 72 °C for 5min. Only single cell samples that contained more than 30 % PCR amplicons among all PCR reactions were selected for targeted capture. 1 μg of amplified DNA was used for Illumina library preparation with the KAPA Hyper Plus Kit (Roche) with 9 cycles of PCR. Targeted capture was performed using the previously described protocol [50] with the following modifications. Two rounds of hybridisation were performed for 24 h each at 47 °C. Following each round of hybridisation and capture, PCR was performed using KAPA HiFI HotStart ReadyMix (Roche) with the following PCR cycling conditions: 98 °C for 3 min; 98 °C for 20 s, 65 °C for 15 s, 72 °C for 30 s 5 cycles (first round) or 20 cycles (second round); 72 °C for 1min. Enrichment of post-capture libraries was measured using qPCR and sequencing performed on an Illumina NextSeq instrument with 150bp paired end reads to a depth of 10 million reads per cell. For P1 a total of 46 single cells (22 red branch, 12 yellow branch and 12 control cells), with two control pools of 50 $CD27^+$ memory B cells and 50 T cells, respectively, were sequenced. For P2 a total of 72 single cells (18 red branch, 17 green branch, 17 yellow branch, 6 blue branch and 14 controls cells) were sequenced. For P3 a total of 71 single cells (18 red branch, 18 yellow branch, 15 blue branch, 6 green branch and 14 control cells) and one pool of 50 $CD27^+$ memory B cells were sequenced.

2.6. Whole genome sequencing (WGS)

In P2 and P3, WGS was performed on 10,000 $CD19^+IgM^+kappa^+$ cells, and 5000 $CD27^+IgD^-IgM^-$ control memory B cells. WGS was not performed on P1 as we had limited sample for this patient, which was prioritised for single cell sequencing. Cells were sorted into 100 μl Buffer ATL in a 1.5 ml Eppendorf tube, which was processed on the same day as cell sorting. It was processed as per QIAamp DNA Micro Kit (Qiagen) instructions. gDNA was transferred to one well per sample of a 96 well plate and made to a concentration of 10 ng in 35 μl of water. Samples were processed as per the kit instructions (KAPA Hyper Plus Kit 96 Reactions, Roche). Sequencing was performed on an Illumina NovaSeq 6000 utilising a v1.5 (300 cycle) reagent kit. The pool was normalized to 1 nM, and sequencing was run on a single lane of an S4 flowcell utilising the XP workflow. Run conditions utilised were 2 × 150bp paired-end, dual indexed (8bp) sequencing.

Germline variants were called on whole genome sequencing data using a modified version of a previously described workflow [51]. Validation of variants of interests identified from genomic sequencing was performed via visual inspection of read alignments in IGV and confirming the location and impact of variants within ENSEMBL (emsembl.org). Finally, mutations were then checked against published mutations at ciobioportal.org.

2.7. Somatic variant analysis

Using G&T-Seq single cell data, somatic mutations between clonal and control B cells were determined using DeepSNVMiner [52] in combination with a modified version of an existing pipeline [53] optimised for rare somatic variant detection. Variants were reported as somatic if: i) at least two clonal cells contained read depth≥15, ii) at least two control cells contained read depth≥15, iii) the variant was present in two or more clonal cells, and iv) the variant was not present in any control B cells. A further variant quality score filter of >20 was applied for each variant and annotated using variant effect predictor [54] in order to prioritise missense and nonsense SNPs and frameshift indels. All mutations were confirmed by Sanger sequencing.

Read depth for each clone across all amplicons was calculated using SAMtools depth [45]. The mean cover across all clones as well as the portion of bases with >1 or >15 reads was calculated. A read depth≥15 was considered a conservative cutoff to reliably detect heterozygous variants. Single cells across all patients had an average of 81.46 % and 60.69 % target bases covered by > 1 read or >15 reads, respectively.

2.8. Somatic variant validation by sanger sequencing

Sanger sequencing was performed on PCR products generated from amplified gDNA from MDA reactions to validate the variants called. gDNA was diluted 1/100 and amplified using primers (Supplementary Table 3) using *Taq* polymerase (Invitrogen) with the following conditions: 98 C for 20 s; [98 C for 10 s, 60 C for 30 s, 72 C for 1 min] x 45 cycles; 72 C for 5 min. Positive PCR products were purified with AMPure XP beads (Beckman Coulter) and Sanger sequenced by the Garvan Molecular Genetics facility (Garvan Institute, Australia).

2.9. Quantitation of $MYD88^{L295P}$ in bulk-sorted cell populations

PBMCs from P2 and P3, along with four age- and sex-matched healthy controls, were stained with fluorophore-labelled antibodies, and the following cell subsets sorted on an Aria III (BD Biosciences): $CD4^+$ T cells, $CD8^+$T cells, NK cells, $CD34^+$ hematopoietic stem cells, monocytes, IgM^+kappa^+ B cells and $CD27^+IgM^+lambda^+$ memory B cells. DNA was prepared from the bulk sorted subsets using the QIAmpDNA micro kit (Qiagen). DNA from 75 to 100000 pooled cells (75

was the minimum for the rare CD34$^+$ hematopoietic stem cells) was amplified by PCR using Q5 2x High Fidelity Mastermix (New England Biolabs) with barcoded versions of the L265P primers (Supplementary Table 3) to allow read identification following pooling of PCR products. Products were purified with AMPure XP beads (Beckman Coulter) and pooled in equal quantities. Amplicon libraries were sequenced using Miseq 250bp paired end reads (Ramaciotti, UNSW).

Barcoded datasets corresponding to the different sorted cell populations were demultiplexed as part of Illumina's FASTQ generation pipeline. Read pairs for each cell population were mapped to the targeted region of MYD88 using bwa aln and sampe [55] and samtools [45] to generate BAM files. BAM files were converted to variant call files (VCFs) using bcftools mpileup and call functions [56]. Variant counts and frequencies were calculated from VCFs in R using packages vcfR [57] and tidyverse [58].

2.10. Single cell immunoglobulin analysis and clonal tree generation

Immunoglobulin heavy and light chain nucleotide sequences for each single cell were reconstructed from the single cell Smart-seq2 data using the VDJPuzzle software package (https://bitbucket.org/kirbyvisp/vdjpuzzle/src/master/; [59]. To build lineage trees, sequences of interest (a clone, or a set of sequences sharing IGHV, IGHJ and CDR3 length) were extracted from VDJPuzzle output (sequence_alignment column) and FASTA formatted files were generated for input to linearham [60]. Tree building used full length IGH nucleotide sequences and linearham was called with default parameters plus the addition of the –all-clonal-seqs flag to generate a single tree for each sequence set. Trees were visualised in R (version 4.3.0; R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria.<https://www.R-project.org/>.) using ggtree (version 3.10.1, [61]) and tidyverse packages (version 2.0.0; [58]).

To calculate expected replacement:silent ratios for each antibody sequence in the absence of selection, the 'mutability' of each germline sequence framework region (FR) and complementary-determining region (CDR) was calculated by summing the trinucleotide probabilities for the sequence, derived from non-productive sequences, indicating the likelihood of a mutation at the central position of the 3-mer. This was performed for FR1, CDR1, FR2, CDR2 and FR3. The relative mutability of the CDRs (and FRs) for each germline gene was calculated by (CDR1_mut + CDR2_mut)/(total_seq_mut). This was used to indicate the likelihood of a random mutation in a sequence occurring in the CDR or FR. The 'replacement frequency' was calculated by taking the germline sequences for each CDR/FR region and considering each position. The number of R (replacement, non-synonymous) and S (silent, synonymous) changes were calculated (stop codons were counted as R) for each position and then summed across the sequence. The replacement frequency was then calculated as the proportion of R/total mutational outcomes for the region. This Rf value was used to indicate the likelihood of a random mutation being a replacement, while (1-Rf) was the likelihood of a silent mutation. To determine the expected R:S ratio for a sequence the mutability and Rf values for the utilised germline were used to calculate the expected number of R and S changes in the CDR and FR given the total number of mutations in the sequence. The R:S ratio was then calculated as for the observed mutations: (R_CDR + 1)/(S_CDR + 1) where the +1 is a pseudo count added to deal with sequences that have 0 observed silent mutations.

2.11. IgG antibody expression

B cell receptor sequences of clonal populations, from both mutant and selected corresponding predicted germline sequences, were commercially synthesised and purified as IgG1 antibodies (Biointron and Genscript) using the amino acid sequences assembled from the scRNA-seq data for individual cells using VDJPuzzle [59]. Where it was not possible to predict the germline amino acid in heavy and light chain CDR3 sequences (marked "X" in the Figures and Supplementary Table 4), the CDR3 sequence of the mutated sequence was used in the "germline" reverted ancestral sequence. Additionally, the B cell receptor sequence of one non-clonal IgM kappa B cell from P2 was also expressed as an IgG kappa antibody and purified to serve as a negative "antibody control".

2.12. Enzyme-linked immunosorbent assay (ELISA)

Expressed IgG1 antibodies were tested in triplicate in serial dilutions from 8 μg/mL to 0.125 μg/mL, in parallel with patient sera, using a commercial MAG ELISA plate (Bulhmann). They were also tested in duplicate at 16 μg/mL on commercial GanglioCombi MAG ELISA plates (Bulhmann), whose individual wells are coated with one of six antigens: MAG, GM1, GM2, GD1a, GD1b, GQ1b. Serum and recombinant antibodies were prepared at the desired concentrations with the kit diluent. 100 μL of sample and kit controls were added to each well and incubated on ice for 2 h. Negative control wells containing only buffer were run with each plate as per manufacturer's instructions. Plates were washed four times with 200 μL of kit wash buffer (Wash Buffer Concentrate10x, diluted at 1:10). 100 μL of HRP anti-IgM conjugate was added to each well that contained serum samples. For those wells that contained the expressed IgG antibodies, HRP anti-IgG conjugate was added. Plates were incubated on ice for 2 h and washed four times with 200 μL of wash buffer. 100 μL of kit TMB Chromogen was added to each well and the plate incubated in the dark for 30 min. 100 μL of HRP Stop Solution was then added to each well. The absorbance of each well was read within 30 min at 450 nm using a spectrophotometer microplate reader (BMG Labtech).

3. Results

3.1. Extensively branched clonal tree producing MAG autoantibody in P1

P1 is an 89-year-old male who presented twenty years earlier with impaired lower limb sensation and poor balance (Supplementary Table 1). Nerve conduction studies showed a demyelinating sensorimotor polyneuropathy. He was diagnosed with CIDP and commenced on monthly intravenous immunoglobulin, without immunosuppression. At age 89, he presented with worsening neuropathy and mobility problems. Testing revealed two IgM kappa paraproteins (both 1 g/L) on serum immunoelectrophoresis. Total serum IgM was elevated at 9.59 g/L (normal range [NR] 0.3–2.3 g/L), and he had a mild lymphopenia of 0.9 $\times 10^9$/L (NR 1.0-4.0 $\times 10^9$/L) with no evidence of lymphadenopathy or hepatosplenomegaly. MAG serology was positive by ELISA at >70,000 BTU (NR < 1000 BTU) and by immunoblot.

Flow cytometric analysis of peripheral blood from P1 revealed a prominent population comprising 20 % of CD19$^+$ B cells with high levels of surface IgM almost exclusively paired with kappa light chains (Fig. 1A). Since polyclonal B cells randomly use either kappa or lambda light chains in a 60:40 ratio, the skewed IgM$^+$kappa$^+$ subset represented a candidate monoclonal population. The IgM$^+$kappa$^+$ cells comprised 12.3 % of IgD$^-$CD27$^+$CD20$^+$ memory B cells (Fig. 1A). Among CD19$^+$ cells, 23 % had a CD20^{dim} phenotype consistent with lymphoplasmacytic cells, and 29 % of CD20^{dim} cells were IgM$^+$kappa$^+$ cells (Fig. 1A). By contrast, among more differentiated plasmablasts defined by expressing CD38 and CD27 in addition to being CD20^{dim}, only 1 % were IgM$^+$kappa$^+$ cells (Fig. 1A).

Single cells were flow sorted into individual wells of 96 well plates, selecting in a 2:1 ratio the CD19$^+$IgM$^+$kappa$^+$ candidate monoclonal cells and control polyclonal memory B cells. Control CD19$^+$CD20$^+$CD27$^+$IgD$^-$IgM$^+$lambda$^+$ B cells were initially selected, but due to their low frequency, we also sorted CD19$^+$CD20$^+$CD27$^+$IgD$^-$IgM$^-$ cells. Due to limited sample availability, four plates were acquired and processed in parallel. From each single

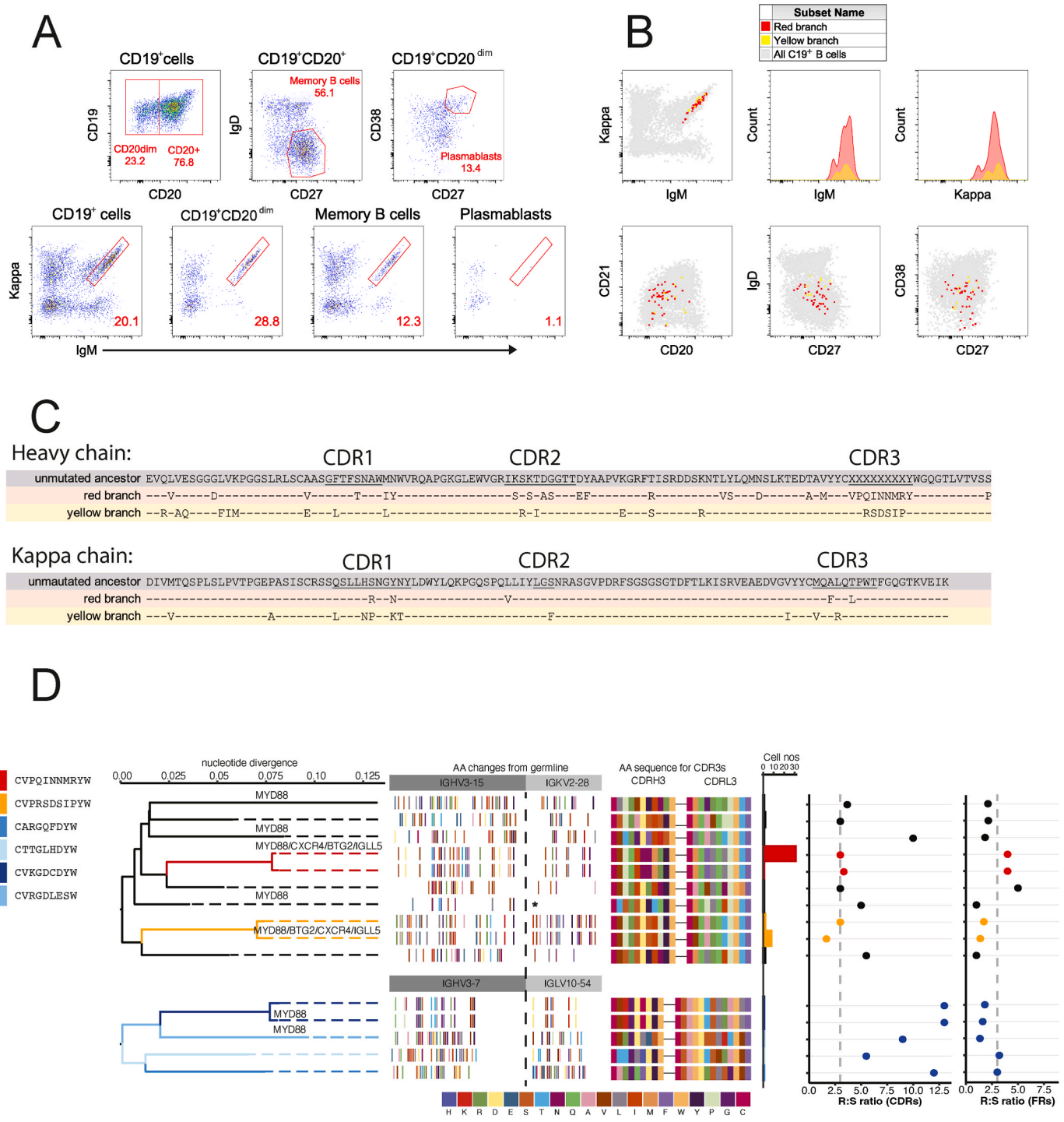


Fig. 1. IgM-producing memory B cell clones with lymphoma driver mutations and extensive intraclonal V(D)J sequence divergence in P1.
(A) Flow cytometric analysis of peripheral blood CD19$^+$ B cells showing % within key gated subsets (denoted in red in top row), and % IgM$^+$kappa$^+$ cells within those subsets (red in bottom row).
(B) Indexed sorting flow cytometric data showing unsorted CD19$^+$ cells (grey) and individual cells identified by immunoglobulin mRNA sequence as belonging to expanded subclonal branches in red and yellow.
(C) Immunoglobulin heavy chain (top) and kappa light chain (bottom) amino acid sequences for germline V and J elements, and somatic substitutions in the most expanded subclonal branches (red and yellow). Dashes denote identity with germline. X's denote unable to infer germline CDR3 sequence.
(D). Clonal trees reconstructed by nearest-neighbour clustering of shared V(D)J mutations in paired heavy- and light-chain nucleotide sequences of individual B cells. Solid lines show nucleotide divergence of each subclonal branch. Amino acid substitutions in V regions of each subclonal branch and CDR3 sequence divergence are displayed in the centre of the figure. Lymphoma driver mutations in the indicated genes are annotated above each subclonal branch where these mutations were detected in one or more cells. Due to allelic dropout during gDNA amplification, failure to detect a driver mutation in a branch cannot be interpreted as the absence of a mutation in that branch. Number of cells identified in each subclonal branch is denoted under "Cell nos". The right panel shows the observed replacement to silent (R:S) ratio of mutations in heavy and light chain V element CDRs and FRs is shown for each branch of the tree by a coloured dot, and the expected R:S ratio in the absence of selection denoted by a grey vertical line. Asterisks (*) denote a truncated sequence with a complete heavy chain sequence but lacking a light chain sequence.

cell, gDNA and mRNA were prepared using the G&T-seq protocol [41, 42], employing MDA of the DNA genome and Smart-seq2 cDNA library sequencing of the mRNA transcriptome.

The complete V(D)J nucleotide sequence of the immunoglobulin heavy and light chain mRNAs made in each cell was reconstructed from large numbers of overlapping short read cDNA fragments using VDJPuzzle (Supplementary Tables 4 and 5) [59]. As expected, the control memory B cells were polyclonal with unrelated V(D)J sequences (data not shown). By contrast, 53 % of IgM$^+$kappa$^+$ cells were members of a single clone sharing identical heavy chain *IGHV3-15*07* and *IGHJ4*02* and light chain *IGKV2-28*01* and *IGKJ1*01* elements and identical CDR3H length, but forming a clonal tree with very divergent branches due to numerous somatic point mutations (Fig. 1C and D).

A major subclonal branch, comprising 43 IgM$^+$kappa$^+$ cells and denoted in red in Fig. 1B–D and referred to herein as the red branch, had identical *IGHM* and *IGK* mRNA and protein sequences to one another, with 9.5 % and 4.35 % of VJ nucleotides in heavy chain and light chain, respectively, changed by somatic hypermutation relative to the germline V and J sequences (Fig. 1C–Supplementary Table 5). It was not possible to infer a germline D element because of extensive sequence divergence from all germline D elements. Excluding the CDR3$_H$ region, the red branch had 19 and 5 replacement mutations in the heavy and light chains, and 4 and 5 silent mutations, respectively.

A second, highly divergent subclonal branch comprising 11 IgM$^+$kappa$^+$ cells, denoted in yellow in Fig. 1B–D and referred to herein as the yellow branch, made identical *IGHM* and *IGK* mRNA and protein sequences to one another with 7.7 % and 6.9 % hypermutated VJ nucleotides and 14 and 11 amino acid substitutions relative to the heavy chain and light chain germline sequences, respectively (Fig. 1C–Supplementary Table 4). Again, it was not possible to infer a germline D element for the yellow branch because of extensive sequence divergence from all germline D elements. The CDR3$_H$ sequence was extremely divergent between the red branch (CVPQINNMRYW) and the yellow branch (CVP**RSDSIPY**W, differences in bold). Remarkably, despite the extensive antibody hypermutation of both the red and yellow branches, they shared only one nucleotide mutation (Supplementary Table 5). The two branches thus appear to have descended from different daughter cells of a single B cell at the onset of clonal expansion and subsequent somatic hypermutation. In both branches, the ratio of replacement to silent mutations (R:S ratio) in CDR1 and CDR2 of heavy and light chains was not different from the ratio predicted for unselected mutations (Fig. 1D, right panel).

Additional cells were also identified either as singletons or pairs expressing *IGHV3-15*07* IgM and *IGKV2-28*01* kappa mRNA sequences descended from the same unmutated ancestor B cell as the red and yellow branches (Fig. 1D–Supplementary Table 4). Each of these divergent branches had heavily mutated *V(D)J* sequences highly divergent from the more abundant red or yellow branches. The CDR R:S ratio was consistent with unselected mutations in most of these sequences (Fig. 1D, right panel).

Because the single cells were index sorted, it was possible to analyse the cell surface protein expression for each cell in the two most abundant subclonal branches, compared with the corresponding data acquired in parallel for millions of unsorted CD19$^+$ cells in the sample (Fig. 1B). Cells in the two branches had similar flow cytometric profiles: high surface expression of IgM and kappa, expression of CD27 but not CD38, and most were CD20^{dim}. This profile is consistent with lymphoplasmacytic cells that have yet to further differentiate into CD38$^+$ plasmablasts, as is also observed for infiltrating bone marrow clonal B cells in IgM-MGUS and Waldenstrom's macroglobulinemia [62].

These flow results were reinforced by comparison of the full transcriptome of the sorted single cells (Supplementary Fig. 1). The most differentially expressed mRNA sequences between the clonal and polyclonal cells were the clone-specific *IGHV3-15* and *IGKV2-28* elements, followed by constant region sequences for IgM and kappa light chains, which were highly expressed in each of the clonal cells. By contrast, many polyclonal memory B cells lacked *IGHM* because they have switched to IgG or IgA (as revealed in VDJPuzzle analysis), and lacked *IGKC* mRNA and instead expressed *IGLC2* or *IGLC3* lambda light chain mRNAs due to allelic exclusion. An early marker of plasma cell differentiation, mRNA encoding Immunoglobulin J-chain, was strongly expressed in the majority of clonal cells but only a small subset of polyclonal cells (Supplementary Fig. 1A and 1B). *PMAIP3*, encoding the pro-apoptotic protein NOXA, and *DUSP22* encoding a JNK-activating phosphatase, were also increased in most of the clonal cells (Supplementary Fig. 1A and 1B).

Single cell *V(D)J* mRNA sequencing of the control memory B cells revealed five of the cells were members of an expanded clone, denoted in blue in Fig. 1D, expressing *IGHV3-7* IgM and *IGLV10-54* lambda light chains unrelated to the large *IGHV3-15*07/IGKV2-28*01* clone described above. This IgM lambda B cell clone comprised many subclonal branches that were highly divergent due to V(D)J mutations (Fig. 1D–Supplementary Table 4). Mutations in the CDRs of the V-regions in this clone exhibited a high R:S ratio compared to that expected by chance (Fig. 1D, right panel).

3.2. Antibody mutations in P1 increase binding to MAG and CD57

To investigate the binding specificity of the *IGHV3-15*07/IGKV2-28*01* antibodies made by the most expanded clone in patient P1, antibodies with variable domains corresponding to the hypermutated sequence of the predominant red and yellow subclonal branches were synthesised and purified as IgG. Parallel antibodies were synthesised with V and J sequences of light and heavy chain reverted to the unmutated common ancestor, although the CDR3$_H$ sequences remained unchanged from the hypermutated antibodies because the ancestral D and N region sequences could not be determined. The sequence of a polyclonal control IgM kappa memory B cell from P2 was expressed in parallel to serve as a negative control IgG antibody. The mutated and unmutated ancestor IgG antibodies were tested by ELISA for binding to immobilised MAG protein (Fig. 2A) or synthetic CD57 SO_4-3GlcAβ1–3Galβ1–4GlcNAc trisaccharide (Fig. 3A, B). As expected, P1's sera bound both MAG (Fig. 2D) and CD57 (Fig. 3C) by IgM ELISA. Antibody corresponding to the yellow branch bound to MAG and CD57, and both binding activities were decreased in the unmutated yellow ancestor, which was only marginally higher than the negative control IgG expressed in parallel. Antibody corresponding to the dominant red branch bound more weakly to MAG protein and CD57 than the yellow branch antibody, but this binding was higher than the unmutated red or yellow ancestors or the control IgG (Figs. 2 and 3). While it was more feasible to express and purify a series of antibodies as monomeric IgG instead of pentameric IgM, it is possible that stronger binding to MAG and CD57 would have been observed if expressed as IgM. We did not synthesise and test the antibody from the blue clone for binding to MAG or CD57, since we only identified this small lambda clone later and had already determined the larger IgM kappa yellow branch bound MAG, consistent with the serum IgM kappa paraprotein. Compared to their CD57 binding, neither the red or yellow branch antibodies, nor the unmutated ancestors, bound detectably greater than the negative control IgG to five ganglioside antigens coating adjacent wells: GM1, GM2, GD1a, GD1b or Gq1b (Supplementary Figure 4). Collectively, the results above indicate that the completely different sets of antibody mutations acquired in red and yellow subclonal branches increased binding to the self-antigens MAG and CD57 SO_4-3GlcAβ1–3Galβ1–4GlcNAc, despite exhibiting an R:S ratio no greater than would be expected in the absence of selection.

3.3. Shared MYD88$^{p.L265P}$ mutation in divergent clonal branches making anti-MAG antibody in P1

Amplified genomic DNA from 22 single cells from the red branch, 12 single cells from the yellow branch, and 12 single polyclonal

$CD20^{+}CD27^{+}$ cells underwent targeted capture sequencing of all the exons of 178 lymphoma driver genes [42]. As a reference for P1's germline sequence, genomic DNA from two control samples of 50 pooled sorted $CD3^{+}$ T cells and 50 pooled sorted $CD19^{+}IgD^{-}IgM^{-}CD27^{+}$ memory B cells was processed and sequenced by targeted capture at the same time. Comparison with the reference samples revealed non-synonymous somatic mutations in four lymphoma driver genes: $MYD88^{p.L265P}$, $CXCR4^{p.R334*}$, $BTG2^{p.A45T}$ and $IGLL5^{p.G19S}$ (Supplementary Fig. 2A).

To validate each of these mutations, targeted primers were used to PCR amplify and Sanger sequence the relevant regions of each gene from amplified gDNA of each sorted single cell. The incomplete and fragmentary nature of MDA of gDNA from single cells fails to amplify one or other allele of a given genomic target in more than half the cells [42]. The resulting "allelic drop-out" yielded 25–50 % of cells where no allele was amplified, and a high proportion of cells that appeared homozygous for either the variant or reference sequence because of failure to amplify the other allele. Among informative cells where at least one allele was PCR amplified and Sanger sequenced, $MYD88^{p.L265P}$ was confirmed in 12 of 26 successfully sequenced red branch cells, 2 of 5 yellow branch cells and in at least one cell from three other divergent branches of the $IgM^{+}kappa^{+}$ clonal tree (Fig. 1D and Supplementary Fig. 2A). There are two interpretations for the presence of $MYD88^{p.L265P}$ in five highly divergent subclonal branches of the IgM^{+} kappa $^{+}$ clone, including the yellow and red branches shown to make MAG- and CD57-binding autoantibody: (1) the mutation arose independently five times during the somatic evolution of the clone; or (2) the mutation arose once in the unmutated ancestor of all the branches, before its progeny diverged by V(D)J hypermutation and gained stronger binding to MAG and CD57.

$MYD88^{p.L265P}$ was also present in two divergent branches of the unrelated *IGHV3-7* IgM^{+} lambda $^{+}$ blue clone (Fig. 1D), but not detected in 55 successfully sequenced polyclonal $CD27^{+}$ B cells (Supplementary Fig. 2A). The unmutated ancestor of this clone may have independently acquired $MYD88^{p.L265P}$, or the driver mutation may have arisen once in a hematopoietic progenitor and different progeny of this progenitor underwent V(D)J recombination to rearrange different heavy and light chain V and J elements and differentiate into the unrelated B cell ancestors of the two observed clonal expansions.

The $CXCR4^{p,R334*}$, $BTG2^{p,A45T}$ and $IGLL5^{p,G19S}$ mutations detected by targeted capture sequencing in the large autoantibody-producing clone were validated by PCR and Sanger sequencing in 7, 7 and 4 cells respectively in the most abundant red branch, and each mutation was found in the yellow branch, but not in polyclonal memory B cells (Fig. 1D and Supplementary Fig. 2A). An identical $CXCR4^{R334*}$ mutation in the germline is responsible for 52 % of cases of the autosomal dominant disease, WHIM syndrome (warts, hypogammaglobulinaemia, immunodeficiency and myelokathexis), where it truncates the CXCR4 cytoplasmic tail to render the G-protein coupled receptor overactive by eliminating serine phosphorylation sites needed for receptor internalisation and desensitization upon binding CXCL12 chemokine [63,64]. $CXCR4^{R334*}$ gain-of-function mutation prevents CXCR4 internalisation and dramatically exaggerates AKT activation in response to CXCL12 [65,66]. Truncating *CXCR4* somatic mutations accompany $MYD88^{L265P}$ in 29 % of cases of Waldenstrom's macroglobulinemia, although >50 % are $CXCR4^{S338*}$ whereas $CXCR4^{R334*}$ only accounts for 1.7 % [67]. The presence of this uncommon CXCR4 somatic mutation in the red and yellow subclonal branches is unlikely to reflect independent occurrence twice. The more likely explanation is that a single $CXCR4^{R334*}$ mutation occurred before the onset of clonal expansion and V(D)J somatic hypermutation.

Missense *IGLL5* mutations occur frequently in DLBCL, although their functional consequences are unknown. The function of IGLL5 is unknown but it is homologous to the pre-B cell receptor subunit, Lambda5 (*IGLL1* gene), and can be secreted to form pathological amyloids [68]. An identical $IGLL5^{p.G19S}$ mutation to P1, within the predicted signal peptide sequence, has been found in CLL [69] and myeloma [70] and proposed to result from off-target mutagenesis by activation-induced cytidine deaminase (AICDA).

3.4. Extensively branched clonal tree producing MAG autoantibody in P2

P2 was a 71-year-old male with a ten year history of IgM-MGUS, serum IgM kappa paraprotein of 5 g/L, and a small kappa-restricted B cell monoclonal population present in peripheral blood and on bone marrow biopsy, without features of myeloma (Supplementary Table 1). He had a normal blood lymphocyte count of 1.5×10^{9}/L (NR 1.0- 4.0 $\times$ 10^{9}/L), mild secondary hypogammaglobulinaemia with a serum IgG of 4 g/L (NR 6–17 g/L) and a history of recurrent cellulitis, but no history of anaemia, renal impairment, bone fractures or hypercalcaemia. At age 71 he presented with predominant ataxia and nerve conduction studies were consistent with a demyelinating sensorimotor polyneuropathy. Anti-MAG serology was positive both by ELISA at >70,000 BTU (NR <1,000 BTU) and by immunoblot. CT scanning showed no evidence of lymphoproliferative disease. After progression of his neuropathy four months later, he was treated with rituximab. He did not receive corticosteroid treatment or other systemic immunosuppression. Peripheral blood was collected twice prior to rituximab treatment.

Flow cytometry of blood cells revealed a prominent $IgM^{+}kappa^{+}$ population with high surface IgM, accounting for 76.4 % of $CD19^{+}$ cells (Fig. 4A). Most $IgM^{+}kappa^{+}$ cells also expressed IgD, so that only 15.9 % of $IgD^{-}CD27^{+}$ memory B cells were $IgM^{+}kappa^{+}$ cells (Fig. 4A). Single cell sorting and G&Tseq allowed VDJPuzzle reconstruction of the immunoglobulin heavy and light chain mRNA sequences in 97 sorted $CD19^{+}IgM^{+}kappa^{+}$ cells and two sorted $CD20^{+}CD27^{+}IgM^{+}IgD^{-}$ memory B cells (Supplementary Tables 4 and 5). This revealed 95 % of $IgM^{+}kappa^{+}$ cells were clonally related descendants from a single B cell, sharing the same *IGHV3-30, IGHJ5, IGKV3-15* and *IGKJ4* elements and identical CDR3H length, although the germline D element could not be determined because of extensive somatic hypermutation. These cells formed a clonal tree with numerous highly divergent branches due to antibody somatic mutations (Fig. 4C and D), including accumulations of cells with identical sequence in three branches – herein referred to as the red, green and yellow branches - with 9.79 %, 9.76 % and 9.62 % nucleotide substitutions, respectively, relative to the germline heavy chain V and J elements and 2.49 %, 5.13 % and 2.48 % mutations in their light chain V and J elements. Amongst the numerous replacement mutations in each of these branches, only two in the heavy chain were common to all three branches, whilst no light chain mutations were common to all three branches (Fig. 4C). Nevertheless the R:S ratio in the CDRs of the three main branches was much higher than expected for unselected mutations (Fig. 4D, right panel). Cells from the red branch accounted for twice as many cells as those in the green branch, which in turn were twice as frequent as yellow branch cells (Fig. 4D and Supplementary Table 4).

Indexed sort analysis comparing flow cytometric data acquired for each sequenced clonal cell with all unsorted $CD19^{+}$ cells revealed the red and yellow branch cells corresponded to the $CD19^{+}CD20^{+}CD27^{+}CD38^{-}$ memory B cell population with high surface IgM and kappa, and most of these co-expressed surface IgD (Fig. 4B). By contrast, cells in the green branch formed a distinct subset with lower surface IgM and disproportionately lower surface kappa light chain staining, the latter explained by absence of surface IgD (Fig. 4B). Their lack of IgD resulted in the green branch of the clone appearing among the sorted polyclonal control $IgD^{-}CD27^{+}$ memory B cells, revealed as a subclonal branch upon single cell V(D)J sequencing.

These three dominant sub-clonal branches were also present in blood collected 4 months later just prior to rituximab treatment (Supplementary Fig. 3A). At this timepoint red branch cells were still the dominant subset, whereas yellow branch cells were less frequent than at time of diagnosis. From this second sample the sort gate for clonal cells was restricted to $CD27^{bright}$ and widened to admit cells with lower surface expression of IgM and kappa, due to the identification of the

$IgM^{dim}kappa^{dim}$ green subclonal branch in the first indexed sort. In contrast to the red and yellow branch cells, those in the green branch lacked IgD but expressed CD38 (Supplementary Fig. 3A). Sequencing cells from the wider gate revealed a small accumulation of cells in a fourth branch of the clone, denoted as the blue branch (Fig. 4C and D), which had lower surface IgM similar to the green branch but without the disproportionate decrease in kappa staining because blue branch cells expressed IgD (Supplementary Fig. 3A). The immunoglobulin sequence of blue branch cells was highly divergent from the other branches, although the numerous CDR mutations had an R:S ratio comparable to that expected for unselected mutations (Fig. 4D, right panel). From this second sample, single polyclonal control memory B cells were sorted in parallel as $CD19^{+}CD20^{+}CD27^{+}IgM^{+}lambda^{+}$ cells.

The mRNA transcriptome of the single sorted cells at this second timepoint revealed higher expression of mRNA encoding J-chain in the clonal cells overall and in red or green branch cells, compared to polyclonal memory B cells (Supplementary Fig. 3B and 3C). *CD79B*, whose overexpression cooperates with MYD^{L265P} to disrupt tolerance checkpoints and promote autoantibody secretion [38], and *IGLL5* were also elevated in the clonal cells (Supplementary Fig. 3B and 3C). Like the clonal cells in P1, *PMAIP1* and *DUSP22* were also increased. Consistent with flow cytometric expression of surface IgD on red but not green branch cells, *IGHD* mRNA was expressed more highly in red branch cells (Supplementary Fig. 3D and 3E). By contrast *RASSF6*, encoding a TP53-stabilizing tumour suppressor [71], was selectively elevated in green branch cells (Supplementary Fig. 3D and 3E).

3.5. Divergent effects of antibody mutations on binding to MAG in P2

Immunoglobulin sequences of the four divergent $IgM^{+}kappa^{+}$ subclonal branches in P2 were synthesised and purified as IgG, along with those from a polyclonal control $IgM^{+}kappa^{+}$ memory B cell to serve as a negative control IgG. In addition, IgG with the variable domain sequence of the unmutated ancestors of the two most abundant branches (red and green branch) was also synthesised although we could not revert the extensive mutations in $CDR3_H$ because these obscured identification of the germline D element or N regions. Each of the expressed IgG antibodies were tested by ELISA for binding to MAG protein (Fig. 2B) or to synthetic CD57 SO_4-3GlcAβ1–3Galβ1–4GlcNAc trisaccharide (Fig. 3A and B), and for binding to a series of gangliosides (Supplementary Fig. 4). P2's sera bound both MAG (Fig. 2D) and CD57 (Fig. 3C) by IgM ELISA. IgG corresponding to the most prevalent red branch cells bound MAG and the sulphated trisaccharide strongly, as did its unmutated ancestor. IgG corresponding to the highly divergent blue branch with a low R:S ratio in the CDRs bound comparably strongly, whereas the yellow branch antibody bound half as strongly to both antigens. Curiously, immunoglobulin corresponding to the green branch did not bind MAG or the synthetic trisaccharide, whereas its unmutated ancestor bound strongly to both (Fig. 2B and 3B).

Compared to their CD57 binding, the red, yellow or blue branch antibodies did not bind detectably greater than the negative control IgG to five ganglioside antigens coating adjacent wells: GM1, GM2, GD1a, GD1b or Gq1b (Supplementary Fig. 4). By contrast, the unmutated ancestor antibodies of the green and red branch weakly bound gangliosides GD1a, GD1b and GM2, and this cross-reactivity was abolished in the fully mutated green and red branch antibodies (Supplementary Fig. 4). Thus, in P2 the extensive V(D)J mutations accumulating in the divergent clonal sub-branches did not increase autoantigen binding, despite a high R:S ratio in the CDRs. In the second largest subclonal branch (green), the V-region mutations had abolished autoreactivity with MAG, CD57, and gangliosides GM2, GD1a and GD1b. A caveat, however, is that we were unable to test an unmutated ancestor where

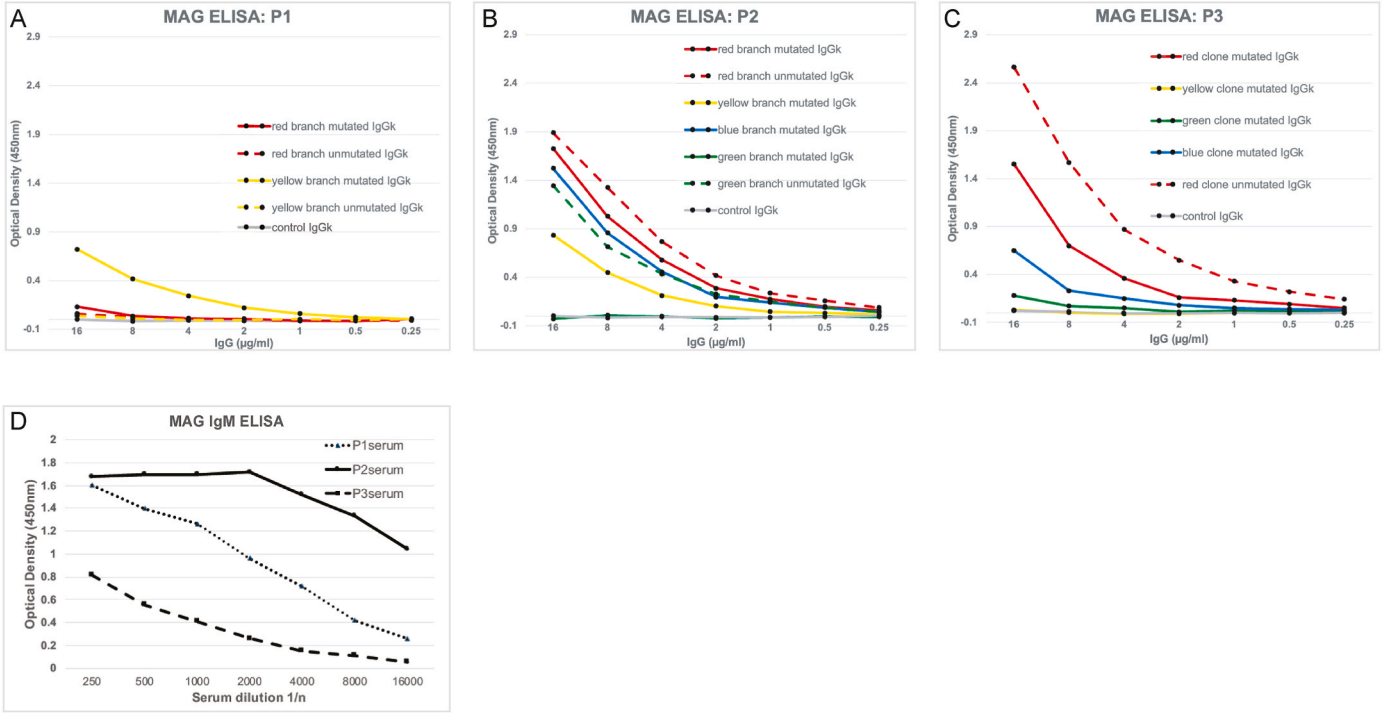


Fig. 2. Antibodies corresponding to expanded subclonal branches in P1, P2 and P3 bind to MAG.
Immunoglobulin V(D)J sequences of the most abundant clonal branches, their unmutated ancestral sequences, or from an unrelated polyclonal IgM kappa memory B cell serving as a negative control (grey) were synthesised as IgG and tested by ELISA at the indicated concentrations for binding to immobilised MAG. The mean OD value for the kit negative control was 0.018. All OD values are expressed after subtraction of blank well controls. Each point shows the mean of triplicates. Data representative of 3 independent experiments. Panels A, B and C represent separate experiments.
(A–C) Binding of the mutated and unmutated ancestor antibodies for the largest subclonal branches from P1 (A), P2 (B), and P3 (C).
(D) Binding of IgM in each patient's sera to MAG at the indicated serum dilutions.

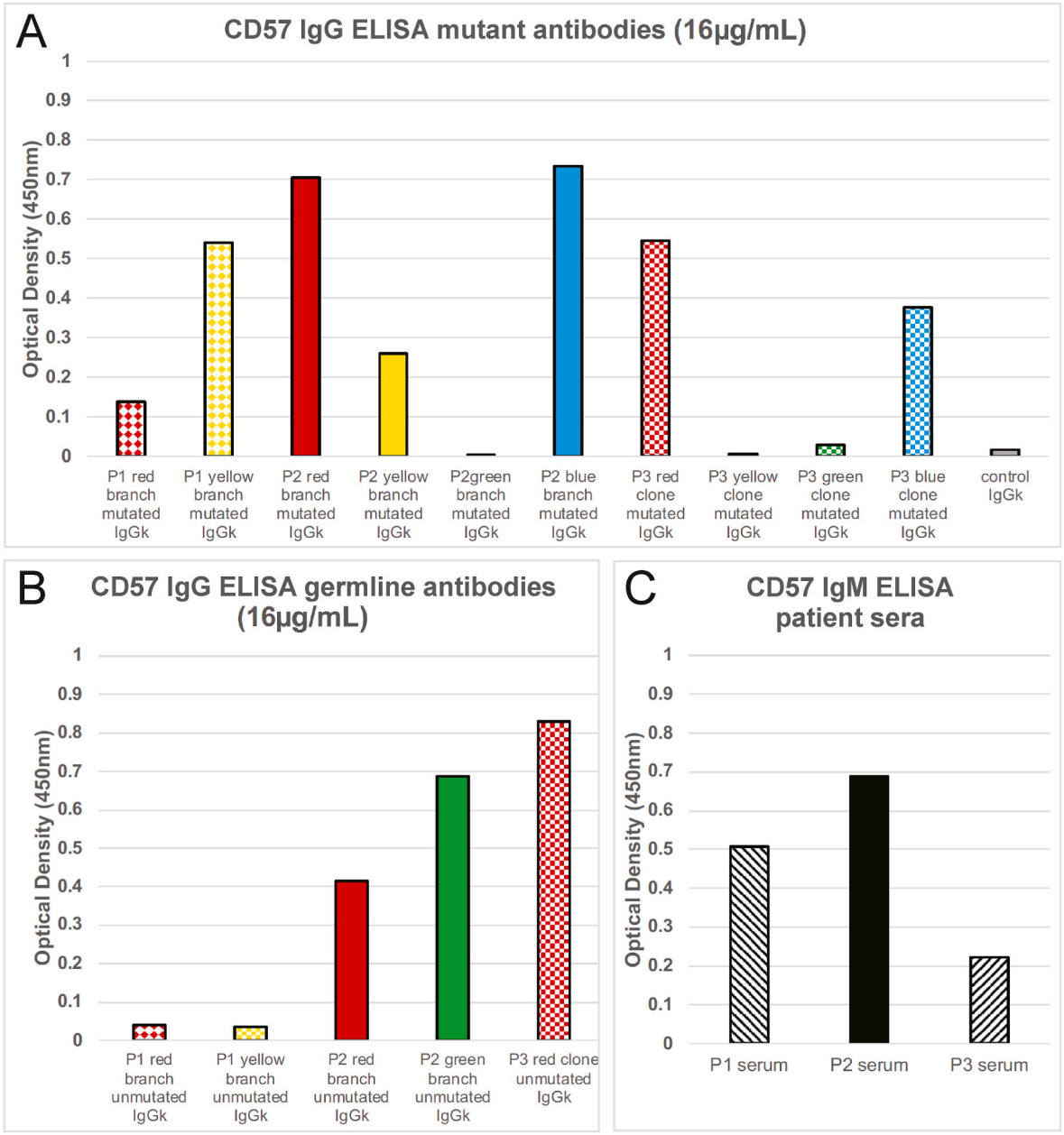


Fig. 3. Antibodies corresponding to expanded subclonal branches in P1, P2 and P3 bind to CD57.
Immunoglobulin V(D)J sequences of the most abundant clonal branches, their unmutated ancestral sequences, or from an unrelated polyclonal IgM kappa memory B cell serving as negative control (grey) were synthesised as IgG and tested in parallel by ELISA for binding to immobilised CD57 SO_4-3GlcAβ1-3Galβ1-4GlcNAc synthetic trisaccharide. The mean OD value for the kit negative control was 0.028. All OD values are expressed after subtraction of blank well controls. Value shows the mean of duplicates.
(A) Binding of 16 µg/mL IgG corresponding to the mutated clonal branch antibodies from P1, P2 and P3 and the negative control antibody.
(B) Binding of 16 µg/mL IgG corresponding to unmutated ancestor antibodies from P1, P2 and P3, measured alongside those in (A) so that OD's can be directly compared between A and B.
(C) Binding of 1:50 dilution of serum from each patient, developed with second-stage antibody to IgM.

the $CDR3_H$ mutations were reverted, because the sequences were so divergent that an ancestral sequence could not be deduced.

3.6. *MYD88$^{p.L265P}$* mutation in divergent clonal branches making anti-MAG antibody in P2

Amplified genomic DNA from 18, 16, 6 and 17 cells from the red, green, blue and yellow branches, respectively, and 14 polyclonal $CD20^+CD27^+IgM^+lambda^+$ memory B cells, underwent targeted capture sequencing of all the exons of 178 lymphoma driver genes [42]. As a reference for P2's germline sequence, DNA from 5000 sorted pooled $CD19^+CD20^+CD27^+IgM^-IgD^-$ memory B cells was analysed by WGS. This revealed *MYD88$^{p.L265P}$* in each of these subclonal branches and a trio of *IGLL5* missense mutations, *IGLL5$^{p.K4R}$*, *IGLL5$^{p.H36R}$* and *IGLL5$^{p.M42L}$*, that co-occurred *in cis* in individual sequence reads from multiple cells only in the red branch. *MYD88$^{p.L265P}$* was confirmed by PCR and Sanger sequencing to be present in 58 cells from 15 divergent subclonal branches, including multiple cells in each of the red, green, yellow and blue branches (Fig. 4D and Supplementary Fig. 2A). This result has two interpretations: (1) *MYD88$^{p.L265P}$* arose independently 15 times during

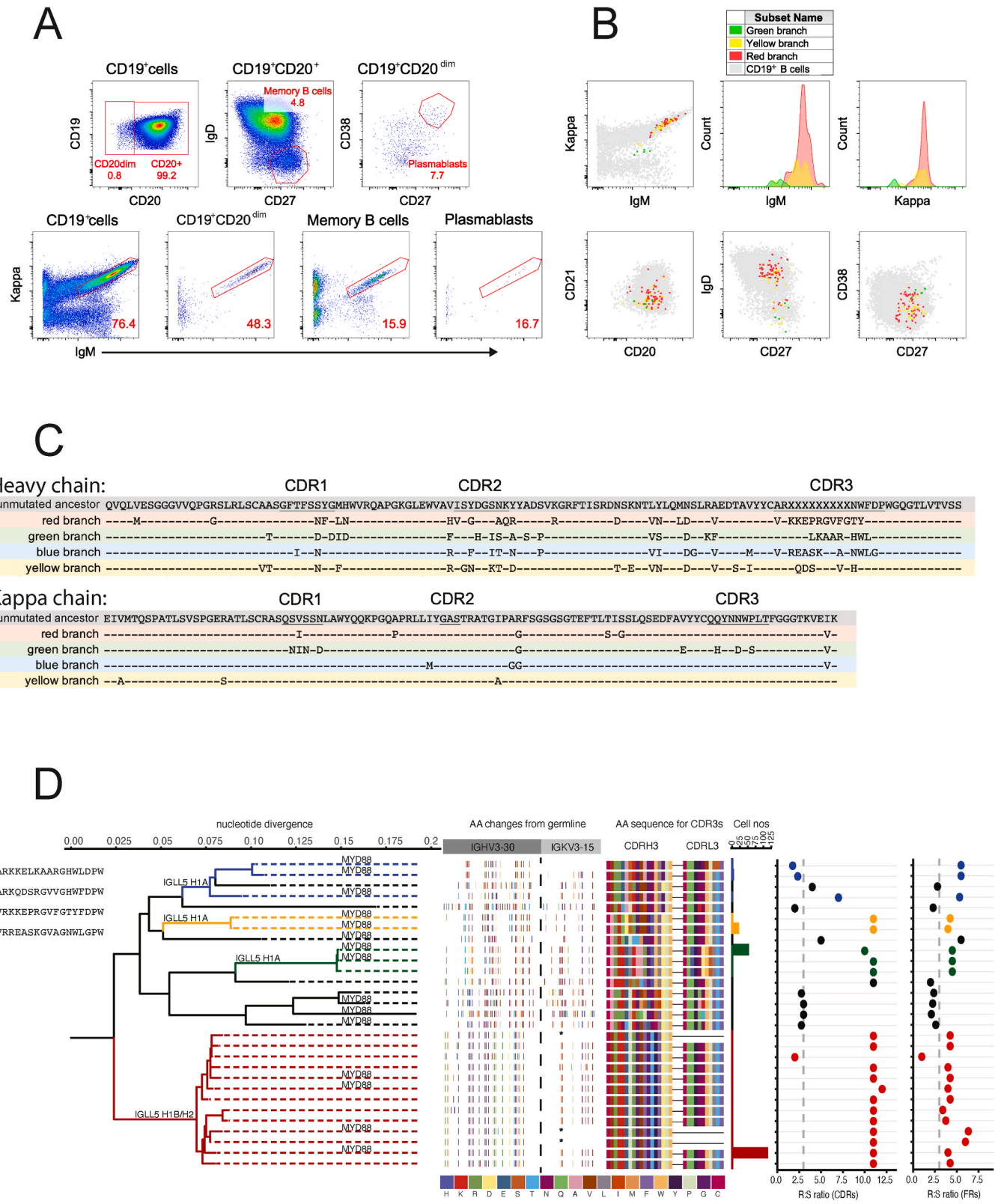


Fig. 4. IgM-producing memory B cell clone with lymphoma driver mutations and extensive intraclonal V(D)J sequence divergence in P2.
(A) Flow cytometric analysis of peripheral blood CD19^{+} B cells analysed as in Fig. 1A.
(B) Indexed sorting flow cytometric data showing unsorted CD19^{+} cells (grey) and individual cells identified by immunoglobulin mRNA sequence as belonging to expanded subclonal branches in red, yellow and green.
(C) Immunoglobulin heavy chain (top) and kappa light chain (bottom) amino acid sequences for germline V and J elements, and somatic substitutions in the most expanded subclonal branches denoted red, green, blue and yellow. Dashes denote identity with germline. X's denote unable to infer germline CDR3 sequence. (D). Clonal tree prepared as in Fig. 1D. Asterisks (*) denote a truncated sequence with a complete heavy chain sequence, but lacking a light chain sequence.

the somatic evolution of the clone; or (2) the mutation arose once in the unmutated ancestor of all the branches, before its progeny diverged extensively by V(D)J hypermutation.

Validation and further analysis of the *IGLL5* mutations by PCR and sanger sequencing DNA from single cells revealed three mutant haplotypes, denoted H1A, H1B and H2, which segregated across different branches of the clonal tree (Fig. 4D and 5, Supplementary Fig. 2A). Haplotype H1A comprised three intronic single nucleotide substitutions *in cis*, which were present in cells in the green, yellow and blue branches. Haplotype H1B carried *in cis* these same three intronic mutations plus one additional intronic nucleotide substitution and the three exonic missense mutations *IGLL5*$^{p.K4R}$, *IGLL5*$^{p.H36R}$ and *IGLL5*$^{p.M42L}$, and was only present in cells in the red branch. Haplotype H2 carried five distinct intronic nucleotide substitutions and one synonymous exonic substitution, and was also limited to cells in the red branch (Fig. 4D, Supplementary Fig. 2A). The high frequency of allelic drop-out during MDA of DNA from single cells established that the H1B and H2 haplotypes comprised separate alleles. These results indicate an initial round of *IGLL5* somatic mutation created three intronic substitutions in the ancestral H1A allele either before V(D)J recombination or in the ancestral B cell before the clonal branches diverged by antibody hypermutation. A second round of *IGLL5* mutation occurred in the red branch, adding three missense mutations and an intronic mutation to the ancestral H1A allele and introducing six synonymous mutations in the other allele.

In polyclonal memory B cells unrelated to the expanded clone from patient P2, *MYD88*$^{p.L265P}$ was also present in 6 of 32 informative single cells (Supplementary Fig. 2B). Each of these *MYD88*$^{p.L265P}$ mutant memory B cells had unrelated antibody H and L chain sequences to one another or to the expanded clone, including one with lambda light chain. It is possible *MYD88*$^{p.L265P}$ arose independently in each of these six memory B cells and in the ancestor of the expanded clone. Alternatively a single *MYD88*$^{p.L265}$ mutation may have arisen in a hematopoietic progenitor preceding V(D)J recombination, and was inherited by these six normal B cells and by the B cell that founded the large, divergent clone.

3.7. *MYD88*$^{p.L265}$ *mutant memory B cell clones in patient P3*

P3 was an 85-year-old male who presented thirty years earlier with paraesthesia in his hands (Supplementary Table 1). A sural nerve biopsy at initial diagnosis showed demyelination with IgM deposits in the perineurium and perivascular regions. He had been treated with daily low-dose prednisolone for almost thirty years and plasma exchange every seven weeks. He had also received nine years of azathioprine in the past. At the time we encountered his case the patient exhibited slow but steady progression of neuropathy. Progress nerve conduction studies showed a demyelinating sensorimotor peripheral neuropathy with no responses in the lower limbs consistent with axonal loss due to the chronicity of his neuropathy. Serum immunoelectrophoresis showed three IgM kappa paraproteins (2 g/L, 3 g/L and a beta-migrating paraprotein that could not be accurately quantitated) and the MAG ELISA was positive at 10,340 BTU (NR < 1000 BTU), but was negative by immunoblot.

Flow cytometry of blood cells revealed a prominent IgM$^+$kappa$^+$ population representing 64.3 % of CD19$^+$ B cells, many with higher surface IgM and kappa than kappa-negative or IgM-negative cells in the sample (Fig. 6A). Single cell sorting and G&T-seq analysis was performed on CD19$^+$IgM$^+$kappa$^+$ B cells and CD19$^+$CD20$^+$CD27$^+$IgM$^+$lambda$^+$ control memory B cells. Paired immunoglobulin heavy and light chain full-length sequences assembled from scRNA-seq data for individual cells revealed four unrelated IgM$^+$kappa$^+$ clonal populations with different V and J gene usage and different CDR3$_H$ lengths (Fig. 6C). In order of frequency, which again

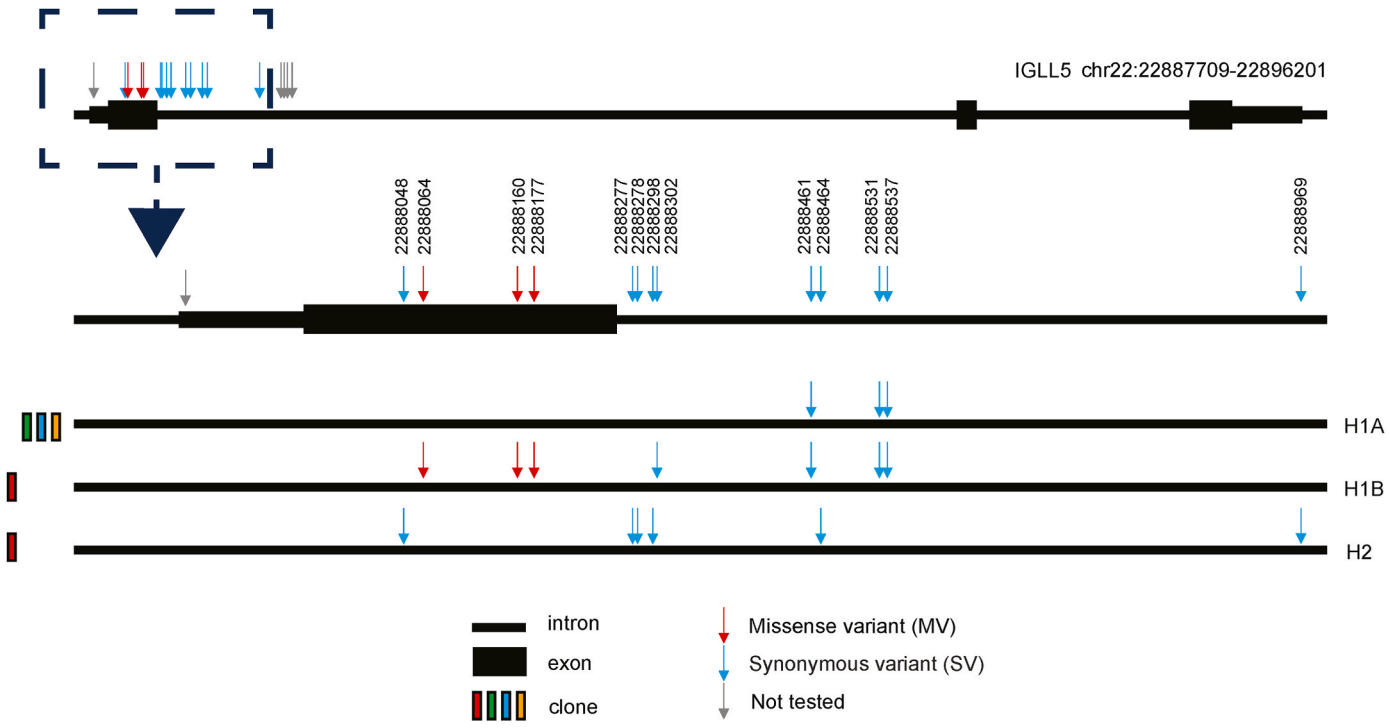
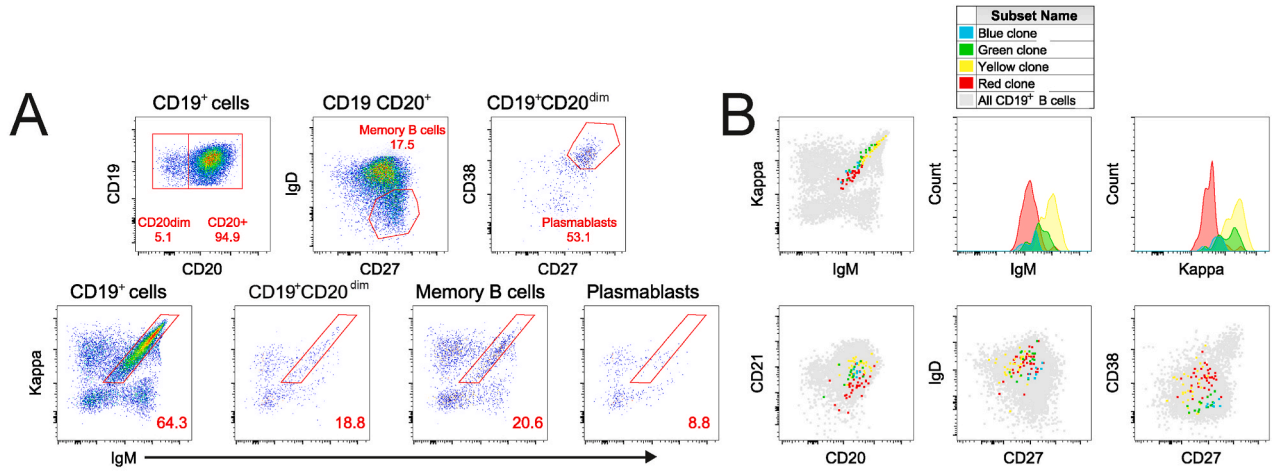


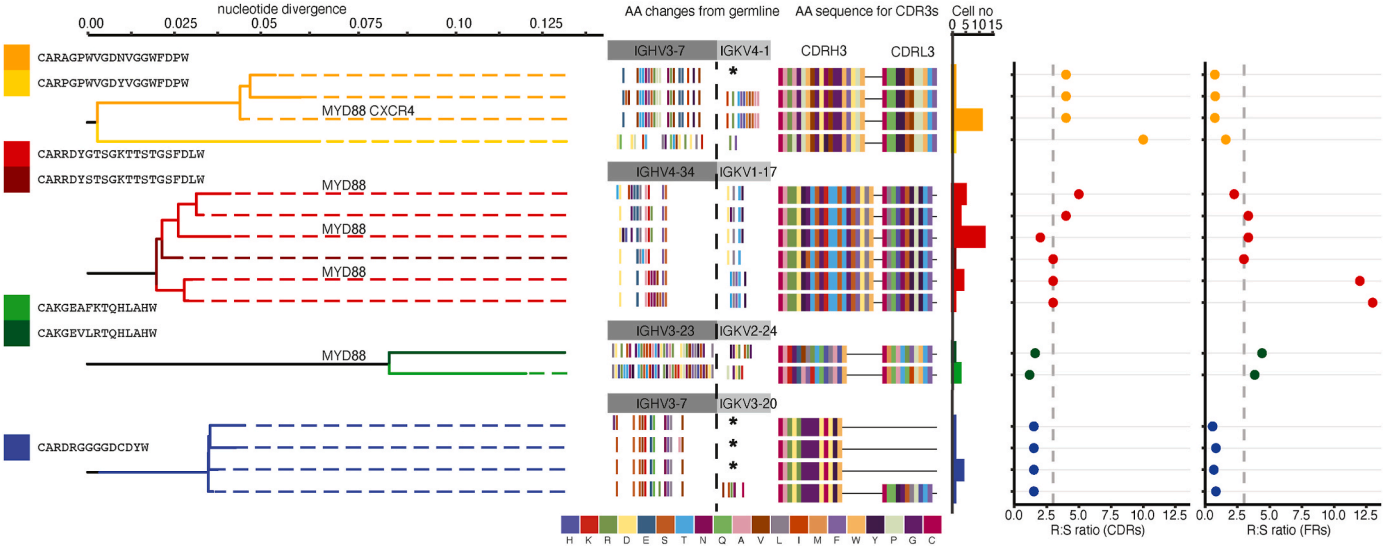
Fig. 5. Three somatic variant *IGLL5* haplotypes in different subclonal branches of P2 validated by single cell Sanger sequencing.
The transcribed region of the *IGLL5* gene is shown at the top and the genomic region surrounding exon 1 (dashed box) is expanded below, with hg38 genomic coordinates for the nucleotides substituted in individual cells. The bottom three lines show the location of variant nucleotides linked *in cis* in individual cells, representing variant haplotypes. The H1A variant haplotype was found in cells from the yellow, green and blue subclonal branches but not in cells from the red branch. Red branch cells had two allelic *IGLL5* variant haplotypes: H1B carried the three variants defining haplotype H1A plus four additional variants (1 synonymous and three missense); H2 lacked any of the variants in H1A but carried six unique synonymous variant nucleotides.



C

	CDR1	CDR2	CDR3
unmutated VH	QVQLQQWGAGLLKPSETLSLTCAVYGGSFSGYYWSWIRQPPGKGLEWIGEINHSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARXXYCTXXXXXXXXXFDLWGRGTLVTVSS		
red clone	----------D-H----------E----------A-K-------R----------F-S-ARRDYGTSGKTTSTGSFDL----------		
unmutated VL	DIQMTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKRLIYAASSLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYPYTFGQGTKLEIK		
red clone	----------D----------L----------T----------		
unmutated VH	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYWMSWVRQAPGKGLEWVANIKQDGSEKYYVDSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCARXXXXGDXXXXWFDPWGQGTLVTVSS		
yellow clone	----------V--T-------NF----------P------R---V-EE----------V--------ARAGPWVGDNVGGWFDP----------		
unmutated VL	DIVMTQSPDSLAVSLGERATINCKSSQSVLYSSNNKNYLAWYQQKPGQPPKLLIYWASTRESGVPDRFSGSGSGTDFTLTISSLQAEDVAVYYCQQYYSTPPWTFGQGTKVEIK		
yellow clone	----I----------R--------T-F-----H-------MV-S----------A----------V----------		
unmutated VH	EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISGSGGSTYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAKXXAFXXXXXXWGQGTLVTVSS		
green clone	---------D--H----------TN---T-------R------T---T--R----A--R---T--K-DPN--VF-H-SG--VD---L---AKGEAFKTQHLAH----------		
unmutated VL	DIVMTQTPLSSPVTLGQPASISCRSSQSLVHSDGNTYLSWLQQRPGQPPRLLIYKISNRFSGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCMQATQFPWTFGQGTKVEIK		
green clone	----------D----------I----------		
unmutated VH	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYWMSWVRQAPGKGLEWVANIKQDGSEKYYVDSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCARXXXXGGDFDYWGQGTLVTVSS		
blue clone	---------S----------N--E---Q----------F----------ARDRGGGGDCDY----------		
unmutated VL	EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSPXTFGQGTRLEIK		
blue clone	----------V----------RN----------C----------L-H----------		

D



(caption on next page)

Fig. 6. Four independent IgM$^+$ kappa$^+$ clones in P3, three harbouring *MYD88*-mutated cells.
(A) Flow cytometric analysis of peripheral blood CD19$^+$ B cells in P3 analysed as in Fig. 1A.
(B) Indexed sorting flow cytometric data from P3 analysed as in Fig. 1B.
(C) Immunoglobulin heavy chain (top) and kappa light chain (bottom) amino acid sequences for germline V and J elements in the independent red, yellow, green and blue clones, and somatic substitutions in the most expanded subclonal branches. Dashes denote identity with germline. X's denote unable to infer germline CDR3 sequence.
(D). Clonal trees of the four independent clones, prepared as in Fig. 1D. Asterisks (*) denote a truncated sequence with a complete heavy chain sequence, but lacking a light chain sequence.

have been assigned colours for reference: the red clone (*IGHV4-34*01, IGHJ2*01, IGKV1-17*01, IGKJ2*01*) comprised 32 % of IgM$^+$kappa$^+$ cells; the yellow clone (*IGHV3-7*01, IGHJ5*02, IGKV4-1*01, IGKJ1*01*) comprised 24 %; the green clone (*IGHV3-23*01, IGHJ4*02, IGKV2-24*01, IGKJ1*01*) comprised 12 %; and the blue clone (*IGHV3-7*01, IGHJ4*02, IGKV3-20*01, IGKJ5*01*) was 5.7 % of IgM$^+$kappa$^+$ cells. The percentage nucleotide substitution was 5.47 %, 5.53 %, 14.89 %, and 3.84 % for the red, yellow, green and blue heavy chains, respectively, and 2.03 %, 4.55 %, 2.13 %, and 2.47 % for the corresponding light chains. The R:S ratio in CDRs was comparable to that expected for unselected mutations for the main accumulating population of each clone (Fig. 6D, right panel). Indexed sort flow cytometric data revealed the individual cells corresponding to each of the expanded clones were IgD$^+$CD27$^+$CD20$^+$CD38$^{low/-}$ (Fig. 6B). Cells in the red clone had low surface IgM and kappa, lower than most non-clonal memory B cells in the same sample (Fig. 6A and B).

The immunoglobulin sequences corresponding to the four independent clones were synthesised and purified as IgG and tested for binding to MAG protein (Fig. 2C) and synthetic CD57 (Fig. 3A and B) by ELISA, along with patient serum. The serum bound both MAG (Fig. 2D) and CD57 (Fig. 3C) by IgM ELISA. IgG corresponding to the largest red clone bound both antigens strongly, and this binding was higher in the V-region unmutated ancestor (Fig. 2C and 3A, B). Antibody corresponding to the mutated blue clone also bound MAG and CD57. By contrast the yellow and green clones did not bind either antigen appreciably relative to the negative control IgG. None of the antibodies bound appreciably to the ganglioside autoantigens GM1, GM2, GD1a, GD1b or Gq1b (Supplementary Fig. 4).

Targeted capture DNA sequencing of all the exons of 178 lymphoma driver genes was performed on 18 cells from the red clone, 18 cells from the yellow clone, 15 cells from the green clone, 6 cells from the blue clone, and 14 polyclonal control CD19$^+$C20$^+$ CD27$^+$IgM$^+$lambda $^+$ memory B cells. *MYD88*$^{p.L265P}$ was identified and confirmed by PCR and Sanger sequencing in cells from three divergent branches of the red clone. It was also present in yellow and green clone cells, but not detected in the two informative blue clone cells (Fig. 6D, Supplementary Fig. 2A). *MYD88*$^{p.L265P}$ was also confirmed in 1 of 8 informative polyclonal control cells expressing a lambda light chain (Supplementary Fig. 2B). A *CXCR4*$^{p.E345*}$ mutation was also identified by targeted capture sequencing and confirmed by PCR and Sanger sequencing in the yellow clone but not in the other clones (Fig. 6D, Supplementary Fig. 2A). Thus patient P3 had two circulating clones of memory B cells making anti-MAG antibody: a large clone (red) with *MYD88*$^{p.L265P}$ and a small clone (blue) with too few cells to exclude the presence of this mutation. *MYD88*$^{p.L265P}$ was also carried by two intermediate sized clones that did not make MAG-binding antibody.

The presence of *MYD88*$^{p.L265P}$ in at least four unrelated memory B cells in P3 could either be explained by four independent *MYD88* mutation events during B cell clonal expansion, or a single event in a hematopoietic progenitor prior to V(D)J recombination. We attempted to phase variants around *MYD88*^{L265P} mutations in different cells by identifying heterozygous inherited single nucleotide polymorphisms (SNPs) within 5 kb. P3 inherited a heterozygous SNP (rs4988457) in the third intron of *MYD88*, 505 bp upstream of the *MYD88*^{L265} codon (Supplementary Fig. 5A). PCR and Sanger sequencing for the inherited SNP thus distinguished which *MYD88* allele had undergone MDA during G&T-seq from each single informative cell, because one allele had the variant nucleotide at *MYD88*^{L265} and the other had the reference nucleotide. From four informative cells in the red clone making anti-MAG antibody, two had the *MYD88*^{L265} mutated allele amplified and this amplicon carried the rs4988457 reference nucleotide, whereas in the other two cells the amplicon lacked the *MYD88* mutation and carried the variant nucleotide at rs4988457 (Supplementary Fig. 5B). This same pattern linking the *MYD88*^{L265} mutation *in cis* with the reference rs4988457 allele was observed in the yellow clone and in the single polyclonal memory B cell carrying *MYD88*^{L265} (Supplementary Fig. 5B). By contrast, in the green clone the *MYD88*^{L265} mutation was carried *in cis* on the allele with the variant nucleotide at rs4988457 (Supplementary Fig. 5B). Thus at least two independent mutation events involving different alleles gave rise to *MYD88*^{L265P} in the different expanded B cell clones circulating in patient P3. We were unable to perform the same analysis in patients P1 and P2 as they lacked an inherited heterozygous SNP linked to *MYD88*^{L265P}.

3.8. *MYD88*^{L265P} *mutations appear restricted to B cell lineage in P2 and P3*

As noted above, the presence of *MYD88*^{L265P} in six clonally unrelated memory B cells and in 15 branches of a large clone in patient P2, and occurrence of *MYD88*^{L265P} on the same allele in three unrelated B cells in patient P3, raised the possibility that these mutations arose once in a hematopoietic progenitor prior to V(D)J recombination, and were seeded out into divergent progeny. Evidence for *MYD88*^{L265P} in B cell precursors in the bone marrow of patients with Waldenstrom's macroglobulinemia [72] supports the possibility it may arise before V(D)J recombination. We therefore tested for the presence of *MYD88*^{L265P} in other hematopoietic lineages in P2 and P3, compared to four healthy controls with no known history of lymphoproliferative disease.

The *MYD88*^{L265P} genomic segment was PCR amplified and deep sequenced in DNA from bulk-sorted populations of CD4$^+$ T cell, CD8$^+$ T cell, NK cell, CD34$^+$ hematopoietic stem cells and monocytes, in addition to IgM$^+$kappa$^+$ B cells and IgM$^+$lambda$^+$ memory B cells. The variant *MYD88*^{L265P} nucleotide was present in 0.02–0.05 % of reads in every cell type including kappa$^+$ or lambda$^+$ memory B cells from healthy controls HC1, HC3 and HC4, and this low variant allele frequency is assumed to represent the error rate generated by PCR amplification and sequencing (Supplementary Fig. 6). By contrast, variant frequencies of 47 % and 28 % were observed in the CD27$^+$IgM + kappa + B cell populations and at 1.22 % and 0.098 % in CD27$^+$IgM + lambda + memory B cells in patients P2 and P3, respectively (Supplementary Fig. 6), consistent with the single cell analysis identifying *MYD88*^{L265P} in the expanded clones and in unrelated polyclonal B cells in these two patients (Supplementary Fig. 2). Variant allele frequencies of 0.02-0.06 %, which are within the assumed error rate, were observed in all other blood cell types tested from P2 and P3, indicating that *MYD88*^{L265P} was limited to clonal and polyclonal B cells in these patients and was either absent or in <0.1 % of hematopoietic stem cells or other blood cell lineages.

Unexpectedly, one of the healthy controls, HC2, was also discovered to have a high frequency of *MYD88*^{L265P} variant B cells, with allele frequencies of 6 % and 18 % in kappa$^+$ and lambda$^+$ memory B cells, respectively (Supplementary Fig. 6). HC2 was a 60 year old male with no history of an autoimmune, lymphoproliferative or haematological disorder and this finding may correspond to the 0.3–0.6 % people aged over

50 years with IgM MGUS [22,23]. Despite high frequencies of *MYD88*^{L265P} in the two subsets of memory B cells, the mutation was either absent or in <1 % of hematopoietic stem cells or other blood cell lineages in this incidental case.

4. Discussion

Here we sought to understand how self-reactive B cells evade immune tolerance checkpoints to produce damaging autoantibodies in a prototypic antibody-mediated demyelinating disease, MAG neuropathy. By performing deep single cell RNA and DNA analysis of three patients, our experiments aimed to answer two questions. First, is *MYD88*^{L265P} present in the cells making the pathogenic autoantibody? In all three patients, we demonstrated *MYD88*^{L265P} was present in multiple cells whose immunoglobulin V(D)J sequences, when expressed as IgG, bound specifically to MAG protein and to synthetic CD57 sulphated trisaccharide.

Given this affirmative answer, the follow-on question was when did the *MYD88* mutation arise during the evolution of these autoantibody-producing clones? The results reveal unexpectedly diverse clonal evolutionary trees responsible for making MAG- and CD57-binding IgM autoantibodies in each patient. In P1 the B cell clone had diverged at the outset of V(D)J-hypermutation, yielding many subclonal branches accumulating non-overlapping antibody mutations. Taking the simplest interpretation, that a single *MYD88*^{L265P} mutation event explains its presence in five different branches, we conclude the NF-κB dysregulating mutation was present in the ancestral B cell before the onset of V(D)J-hypermutation and clonal expansion. The unmutated ancestor B cell also appears to have acquired, prior to clonal expansion, a *CXCR4* gain-of-function somatic mutation like those found frequently in Waldenstrom's macroglobulinemia, and mutations in *BTG2* and *IGGL5* of unknown functional consequence but reported previously in B cell neoplasms. In P2 the B cell clone had diverged through antibody gene hypermutation into an even larger number of divergent branches, with 15 branches shown to carry *MYD88*^{L265P}, indicating the NF-κB dysregulating mutation was present at the beginning of this antibody divergence. In addition to *MYD88*^{L265P} the ancestral B cell had also already acquired three somatic mutations in the first intron of one *IGGL5* allele, with the MAG/CD57 autoantibody-producing progeny acquiring many additional *IGGL5* mutations in the two most abundant branches. In P3, *MYD88*^{L265P} also appeared to have been present at the start of clonal expansion of the largest clone making MAG/CD57 autoantibodies, being found in three divergent subclonal branches.

The presence of *MYD88*^{L265P} in MAG/CD57 self-reactive B cells at the outset of extraordinary clonal proliferation and IgM autoantibody secretion raises the possibility that exaggerated MYD88 signalling mimics the effects of lipopolysaccharide (LPS) or other TLR-dependent B cell mitogens that activate MYD88. Classic experiments demonstrate that LPS cooperates with surface immunoglobulin stimulation by foreign antigens to induce IgM antibody secreting cells, but LPS is unable to do this in B cells recognising self-antigens on autologous erythrocytes or rendered tolerant to foreign gammaglobulin [73,74]. Tolerant B cells have lost the synergistic proliferative response to low concentrations of TLR ligands plus surface immunoglobulin signals, although high concentrations of LPS or a TLR9 agonist induce proliferation without autoantibody secretion; the latter being actively repressed by surface immunoglobulin signalling from binding to self antigen [75–77]. Enforced expression of *MYD88*^{L265P} protects tolerant B cells from self antigen-induced apoptosis and drives their differentiation into lymphoplasmacytic cells secreting IgM autoantibodies *in vivo* [38]. However, this effect is countered by *MYD88*^{L265P}-induced downregulation of surface IgM and only unleashed proportional to dysregulation of the surface IgM receptor subunit, CD79B, and resulting increases in surface IgM [38]. Consistent with cooperation with surface IgM signalling, the *MYD88*^{L265P}-bearing clonal B cells making anti-MAG/CD57 in P1 and P2 had higher surface IgM than their normal memory B cell counterparts, with evidence of increased *CD79B* mRNA in P2. There was similarly high IgM expression on the *MYD88*^{L265P}-bearing yellow clone making a non-MAG binding IgM in P3.

By contrast, surface IgM and kappa light chains were lower than on normal memory B cells in the anti-MAG *IGHV4-34* antibody-producing clone in P3, although the cells expressed high levels of IgD. This IgM^{low}IgD^{hi} phenotype mirrors tolerant human B cells expressing *IGHV4-34* IgM encoding IgM against the Ii carbohydrate autoantigen, which are normally excluded from the memory B cell compartment [78]. While *MYD88*^{L265P} was present in the anti-MAG *IGHV4-34* clone here and in two previously described cases of *IGHV4-34* anti-MAG IgM-MGUS [39], this driver mutation was absent from *IGHV4-34* B cell clonal expansions in 27 cases of IgM-MGUS with primary cold agglutinin disease (PCAD) [79]. Instead, B cells clones causing PCAD often acquire gain-of-function *CARD11* mutations [80], involving an essential signalling protein between surface immunoglobulin and NF-kB required for persistence of IgM^{low} tolerant B cells [81]. Unlike *MYD88*^{L265P}, enforced expression of *CARD11* gain of function (GOF) mutations are not countered by surface IgM downregulation and drive proliferation and differentiation of self-reactive B cells into lymphoplasmacytic cells secreting IgM autoantibodies *in vivo* [82]. Unlike the anti-MAG clone in P3, which paired *IGHV4-34* with *IGKV1-17,* 83 % of IgM-MGUS clones in PCAD pair *IGHV4-34* with *IGKV3-20* or *IGKV3-15*, which potentially increases IgM binding affinity for the I autoantigen and the temperature at which it fixes complement to autologous erythrocytes [79]. Collectively, these observations imply cooperation between surface IgM specificity and different NF-kB activating driver mutations: *MYD88*^{L265P} preferred in IgM-MGUS clones with high surface IgM, and *CARD11*^{GOF} preferred in clones with low surface IgM.

Dependence of *MYD88*^{L265P} on crosstalk with surface IgM may explain the 85 % median decrease in serum IgM and alleviation of neuropathy in *MYD88*^{L265P}-bearing Waldenstrom's macroglobulinemia patients treated with ibrutinib or zanubrutinib, two drugs that inhibit the surface IgM signal-initiating enzyme, Bruton's tyrosine kinase (BTK) [83,84]. A physical complex of MYD88^{L265P} and phosphorylated BTK exists in Waldenstrom's cells, and MYD88-IRAK-BTK signalling contributes to NF-kB activation and survival in vitro [85]. Similarly, an ibrutinib-sensitive super-complex of surface IgM, BTK-interacting proteins, TLR9, MYD88^{L265P}, NF-kB activating proteins, and the MTORC1 complex exists in ABC DLBCL cells [86]. Collectively these observations provide a rationale for trialling BTK inhibitors to treat MAG neuropathy [87,88].

It is unclear if surface IgM binding to CD57 is a cooperating factor driving clonal expansion of the *MYD88*^{L265P} mutant clones. The R:S ratio for mutations in CDRs in P1 and P3, and in the blue branch of P2, was not higher than expected for unselected mutations. A high R:S ratio consistent with selection was observed in the other clonal branches of P2, but these mutations in one branch abolished binding to CD57 and self-gangliosides, and in another branch abolished binding to self-gangliosides. In the P2 red branch, the high R:S ratio of acquired mutations was not accompanied by any increase in binding to MAG or CD57, but instead the mutations decreased binding to CD57 (Fig. 3). Viewing all three patients and their divergent branches and clones, there is no consistent pattern of selection to improve binding to CD57 during extensive hypermutation. If CD57-binding is not selected, it is possible that surface IgM binding to another self- or foreign antigen, or polyreactivity with multiple antigens, drives the mutation and accumulation of diverse subclonal branches in cooperation with *MYD88*^{L265P}. While the mutated antibodies from each patient did not display polyreactivity with respect to the five gangliosides we tested, we did not test for reactivity against more diverse antigens. Alternatively, it is possible that *MYD88*^{L265P} frees the hypermutating progeny from selection for binding any antigen to surface IgM, explaining the divergent patterns of CD57 binding and low R:S ratios.

The study here has several limitations. First, while MAG neuropathy is a prototypic demyelinating disease where the pathogenic role of the

monoclonal IgM is established, it may not be representative for the pathogenesis of other forms of CIDP where no monoclonal IgM can be detected in serum. Second, because single cell DNA and RNA sequencing and antibody expression studies are labour-intensive and costly, we did not have the resources to study more than three patients, so we cannot be certain these findings are typical of the broader MAG neuropathy population and studies on larger number of patients are warranted. Thirdly, we were unable to deduce the ancestral sequences for CDR3H because of extensive mutation, and were unable to test the specificity of antibodies where these mutations were reverted. Finally, we did not express the BCR sequences as IgM, as it is technically difficult to do so, nor did we express the sequences as membrane IgM and test binding to multimerised CD57 to recapitulate B cell recognition of the antigen on the plasma membrane of NK cells or T cells. Low affinity binding to CD57 may therefore have gone undetected by ELISA, as we described recently for the unmutated ancestors of pathogenic rheumatoid factors [89].

The findings here nevertheless enable reconstruction of a series of evolutionary steps involving somatic mutation resulting in human autoimmune disease. In the three MAG neuropathy patients analysed here, the first step appears to be acquisition of a lymphoma driver mutation, $MYD88^{L265P}$. This step may occur preceding or soon after the second step, V(D)J recombination, which produces a B cell expressing surface IgM that binds CD57 self-glycan. A third step is clonal proliferation of the CD57-binding B cell followed by extensive V(D)J hypermutation in the progeny, accompanied by other molecular abnormalities such as *CXCR4* and *IGLL5* mutations. As an ensemble, these mutations appear sufficient to drive accumulation of large numbers of clonal progeny exhibiting lymphoplasmacytic differentiation, and accumulation of high titres of anti-MAG IgM in the blood. When the secreted IgM reaches a sufficient concentration in the circulation it triggers clinically significant demyelination and clinical neuropathy. It is conceivable that a similar sequence of events may contribute to other autoimmune diseases that - like MAG neuropathy and MGUS - become more frequent in people older than fifty.

CRediT authorship contribution statement

Shane Kelly: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Mandeep Singh:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Amanda Russell:** Writing – review & editing, Supervision, Project administration, Investigation, Formal analysis, Data curation. **Katherine J.L. Jackson:** Writing – review & editing, Supervision, Software, Formal analysis, Data curation. **Timothy J. Peters:** Writing – review & editing, Supervision, Software, Formal analysis, Data curation. **Andrew Carr:** Writing – review & editing, Supervision, Resources, Funding acquisition, Formal analysis. **Anthony D. Kelleher:** Writing – review & editing, Visualization, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Matt Field:** Writing – review & editing, Supervision, Software, Formal analysis. **Matthew Silsby:** Writing – review & editing, Resources, Data curation. **Dan Suan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Christopher C. Goodnow:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interests

Nil to declare.

Acknowledgements

This research was supported by the National Health and Medical Research Council (NHMRC) grants; APP2010134 & APP1113904, the UNSW Cellular Genomics Futures Institute, the Bill & Patricia Ritchie Foundation, the Croall Foundation, an Australian Government Research Training Program (RTP) scholarship, and a Garvan Support Award from the Garvan Institute of Medical Research. The authors wish to thank Mohammad Qashlan for his assistance with the graphic design of the figures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2025.103474.

Data availability

Data will be made available on request.

References

[1] C.C. Goodnow, et al., Self-tolerance checkpoints in B lymphocyte development, Adv. Immunol. 59 (1995) 279–368.
[2] N. Latov, Diagnosis and treatment of chronic acquired demyelinating polyneuropathies, Nat. Rev. Neurol. 10 (8) (2014) 435–446.
[3] N. Latov, et al., Plasma-cell dyscrasia and peripheral neuropathy with a monoclonal antibody to peripheral-nerve myelin, N. Engl. J. Med. 303 (11) (1980) 618–621.
[4] L. Magy, et al., Heterogeneity of polyneuropathy associated with Anti-MAG antibodies, J Immunol Res 2015 (2015) 450391.
[5] J. Svahn, et al., Anti-MAG antibodies in 202 patients: clinicopathological and therapeutic features, J. Neurol. Neurosurg. Psychiatry 89 (5) (2018) 499–505.
[6] A.P. Hays, S.S. Lee, N. Latov, Immune reactive C3d on the surface of myelin sheaths in neuropathy, J. Neuroimmunol. 18 (3) (1988) 231–244.
[7] M. Takatsu, et al., Immunofluorescence study of patients with neuropathy and IgM M proteins, Ann. Neurol. 18 (2) (1985) 173–181.
[8] A.P. Hays, et al., Experimental demyelination of nerve induced by serum of patients with neuropathy and an anti-MAG IgM M-protein, Neurology 37 (2) (1987) 242–256.
[9] H.J. Willison, et al., Demyelination induced by intraneural injection of human antimyelin-associated glycoprotein antibodies, Muscle Nerve 11 (11) (1988) 1169–1176.
[10] A.H. Tatum, Experimental paraprotein neuropathy, demyelination by passive transfer of human IgM anti-myelin-associated glycoprotein, Ann. Neurol. 33 (5) (1993) 502–506.
[11] D.K. Chou, et al., Structure of sulfated glucuronyl glycolipids in the nervous system reacting with HNK-1 antibody and some IgM paraproteins in neuropathy, J. Biol. Chem. 261 (25) (1986) 11717–11725.
[12] H. Voshol, et al., Structure of the HNK-1 carbohydrate epitope on bovine peripheral myelin glycoprotein P0, J. Biol. Chem. 271 (38) (1996) 22957–22960.
[13] D. Focosi, et al., CD57+ T lymphocytes and functional immune deficiency, J. Leukoc. Biol. 87 (1) (2010) 107–116.
[14] S. Sato, et al., Shared antigen between the myelin-associated glycoprotein (MAG) and a cell line from human T cell leukemia (HSB-2), J. Neuroimmunol. 7 (5–6) (1985) 287–298.
[15] E. Delmont, et al., Relevance of anti-HNK1 antibodies in the management of anti-MAG neuropathies, J. Neurol. 266 (8) (2019) 1973–1979.
[16] D.L. Burnett, et al., Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination, Science 360 (6385) (2018) 223–226.
[17] S.B. Hartley, et al., Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death, Cell 72 (3) (1993) 325–335.
[18] S.B. Hartley, et al., Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens, Nature 353 (6346) (1991) 765–769.
[19] D.A. Nemazee, K. Burki, Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes, Nature 337 (6207) (1989) 562–566.
[20] J.H. Reed, et al., Clonal redemption of autoantibodies by somatic hypermutation away from self-reactivity during human immunization, J. Exp. Med. 213 (7) (2016) 1255–1265.
[21] M.E. Shy, et al., Specificity of human IgM M-proteins that bind to myelin-associated glycoprotein: peptide mapping, deglycosylation, and competitive binding studies, J. Immunol. 133 (5) (1984) 2509–2512.
[22] R.A. Kyle, et al., Prevalence of monoclonal gammopathy of undetermined significance, N. Engl. J. Med. 354 (13) (2006) 1362–1369.
[23] D. Suan, et al., Prevalence of paraproteinaemia in older Australians, Intern. Med. J. 42 (2) (2012) 165–169.

[24] J.P. Bida, et al., Disease associations with monoclonal gammopathy of undetermined significance: a population-based study of 17,398 patients, Mayo Clin. Proc. 84 (8) (2009) 685–693.
[25] R.A. Kyle, et al., Long-term Follow-up of monoclonal gammopathy of undetermined significance, N. Engl. J. Med. 378 (3) (2018) 241–249.
[26] S.P. Treon, et al., MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia, N. Engl. J. Med. 367 (9) (2012) 826–833.
[27] O. Landgren, L. Staudt, MYD88 L265P somatic mutation in IgM MGUS, N. Engl. J. Med. 367 (23) (2012) 2255–2256, author reply 2256-2257.
[28] M. Varettoni, et al., Prevalence and clinical significance of the MYD88 (L265P) somatic mutation in Waldenstrom's macroglobulinemia and related lymphoid neoplasms, Blood 121 (13) (2013) 2522–2528.
[29] L. Xu, et al., MYD88 L265P in waldenstrom macroglobulinemia, immunoglobulin M monoclonal gammopathy, and other B-cell lymphoproliferative disorders using conventional and quantitative allele-specific polymerase chain reaction, Blood 121 (11) (2013) 2051–2058.
[30] V.N. Ngo, et al., Oncogenically active MYD88 mutations in human lymphoma, Nature 470 (7332) (2011) 115–119.
[31] X.S. Puente, et al., Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia, Nature 475 (7354) (2011) 101–105.
[32] L. Wang, et al., SF3B1 and other novel cancer genes in chronic lymphocytic leukemia, N. Engl. J. Med. 365 (26) (2011) 2497–2506.
[33] S. Akira, K. Takeda, Toll-like receptor signalling, Nat. Rev. Immunol. 4 (7) (2004) 499–511.
[34] S.C. Lin, Y.C. Lo, H. Wu, Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling, Nature 465 (7300) (2010) 885–890.
[35] G. Knittel, et al., B-cell-specific conditional expression of Myd88p.L252P leads to the development of diffuse large B-cell lymphoma in mice, Blood 127 (22) (2016) 2732–2741.
[36] T. Sewastianik, et al., Human MYD88L265P is insufficient by itself to drive neoplastic transformation in mature mouse B cells, Blood Adv. 3 (21) (2019) 3360–3374.
[37] J.Q. Wang, et al., Consequences of the recurrent MYD88(L265P) somatic mutation for B cell tolerance, J. Exp. Med. 211 (3) (2014) 413–426.
[38] J.Q. Wang, et al., Synergistic cooperation and crosstalk between MYD88(L265P) and mutations that dysregulate CD79B and surface IgM, J. Exp. Med. 214 (9) (2017) 2759–2776.
[39] J.S. Allain, et al., IGHV segment utilization in immunoglobulin gene rearrangement differentiates patients with anti-myelin-associated glycoprotein neuropathy from others immunoglobulin M-gammopathies, Haematologica 103 (5) (2018) e207–e210.
[40] J.M. Vos, et al., High prevalence of the MYD88 L265P mutation in IgM anti-MAG paraprotein-associated peripheral neuropathy, J. Neurol. Neurosurg. Psychiatry 89 (9) (2018) 1007–1009.
[41] I.C. Macaulay, et al., G&T-seq: parallel sequencing of single-cell genomes and transcriptomes, Nat. Methods 12 (6) (2015) 519–522.
[42] M. Singh, et al., Lymphoma driver mutations in the pathogenic evolution of an iconic human autoantibody, Cell 180 (5) (2020) 878–894 e19.
[43] S. Picelli, et al., Full-length RNA-seq from single cells using Smart-seq2, Nat. Protoc. 9 (1) (2014) 171–181.
[44] A. Dobin, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (2013) 15–21.
[45] H. Li, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.
[46] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.
[47] J. Breda, M. Zavolan, E. van Nimwegen, Bayesian inference of gene expression states from single-cell RNA-seq data, Nat. Biotechnol. 39 (8) (2021) 1008–1016.
[48] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.
[49] M.L. Leung, et al., Highly multiplexed targeted DNA sequencing from single nuclei, Nat. Protoc. 11 (2) (2016) 214–235.
[50] T.R. Mercer, et al., Targeted sequencing for gene discovery and quantification using RNA CaptureSeq, Nat. Protoc. 9 (5) (2014) 989–1009.
[51] M.A. Field, Detecting pathogenic variants in autoimmune diseases using high-throughput sequencing, Immunol. Cell Biol. 99 (2) (2021) 146–156.
[52] T.D. Andrews, et al., DeepSNVMiner: a sequence analysis tool to detect emergent, rare mutations in subsets of cell populations, PeerJ 4 (2016) e2074.
[53] A.R. Hamzeh, T.D. Andrews, M.A. Field, Detecting causal variants in Mendelian disorders using whole-genome sequencing, Methods Mol. Biol. 2243 (2021) 1–25.
[54] W. McLaren, et al., The ensembl variant effect predictor, Genome Biol. 17 (1) (2016) 122.
[55] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.
[56] P. Danecek, et al., Twelve years of SAMtools and BCFtools, GigaScience 10 (2) (2021).
[57] B.J. Knaus, N.J. Grunwald, Vcfr: a package to manipulate and visualize variant call format data in R, Mol Ecol Resour 17 (1) (2017) 44–53.
[58] H. Wickham, et al., Welcome to the tidyverse, J. Open Source Softw. 4 (43) (2019) 1686.
[59] S. Rizzetto, et al., B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle, Bioinformatics 34 (16) (2018) 2846–2847.
[60] A. Dhar, et al., A Bayesian phylogenetic hidden Markov model for B cell receptor sequence analysis, PLoS Comput. Biol. 16 (8) (2020) e1008030.
[61] S. Xu, et al., Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data, Imeta 1 (4) (2022) e56.
[62] B. Paiva, et al., Multiparameter flow cytometry for the identification of the Waldenstrom's clone in IgM-MGUS and Waldenstrom's Macroglobulinemia: new criteria for differential diagnosis and risk stratification, Leukemia 28 (1) (2014) 166–173.
[63] P.A. Hernandez, et al., Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease, Nat. Genet. 34 (1) (2003) 70–74.
[64] L.E. Heusinkveld, et al., WHIM syndrome: from pathogenesis towards personalized medicine and cure, J. Clin. Immunol. 39 (6) (2019) 532–556.
[65] Y. Cao, et al., The WHIM-like CXCR4(S338X) somatic mutation activates AKT and ERK, and promotes resistance to ibrutinib and other agents used in the treatment of Waldenstrom's Macroglobulinemia, Leukemia 29 (1) (2015) 169–176.
[66] D.H. McDermott, et al., AMD3100 is a potent antagonist at CXCR4(R334X) , a hyperfunctional mutant chemokine receptor and cause of WHIM syndrome, J. Cell Mol. Med. 15 (10) (2011) 2071–2081.
[67] Z.R. Hunter, et al., The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis, Blood 123 (11) (2014) 1637–1646.
[68] H. Itagaki, et al., Recurrent pericardial effusion with pericardial amyloid deposition: a case report and literature review, Cardiovasc. Pathol. 46 (2020) 107191.
[69] S. Kasar, et al., Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution, Nat. Commun. 6 (2015) 8866.
[70] B.S. White, et al., A multiple myeloma-specific capture sequencing platform discovers novel translocations and frequent, risk-associated point mutations in IGLL5, Blood Cancer J. 8 (3) (2018) 35.
[71] J.A. Kuleape, et al., DNA damage triggers the nuclear accumulation of RASSF6 tumor suppressor protein via CDK9 and BAF53 to regulate p53 target gene transcription, Mol. Cell Biol. 42 (2) (2022) e0031021.
[72] S. Rodriguez, et al., Preneoplastic somatic mutations including MYD88(L265P) in lymphoplasmacytic lymphoma, Sci. Adv. 8 (3) (2022) eabl4644.
[73] J.M. Chiller, W.O. Weigle, Termination of tolerance to human gamma globulin in mice by antigen and bacterial lipopolysaccharide (endotoxin), J. Exp. Med. 137 (3) (1973) 740–750.
[74] J.R. Schmidtke, F.J. Dixon, Immune response to a hapten coupled to a nonimmunogenic carrier. Influence of lipopolysaccharide, J. Exp. Med. 136 (2) (1972) 392–397.
[75] C.C. Goodnow, R. Brink, E. Adams, Breakdown of self-tolerance in anergic B lymphocytes, Nature 352 (6335) (1991) 532–536.
[76] L. Rui, et al., ERK signaling is a molecular switch integrating opposing inputs from B cell receptor and T cell cytokines to control TLR4-driven plasma cell differentiation, J. Immunol. 177 (8) (2006) 5337–5346.
[77] L. Rui, et al., Resistance to CpG DNA-induced autoimmunity through tolerogenic B cell antigen receptor ERK signaling, Nat. Immunol. 4 (6) (2003) 594–600.
[78] A.E. Pugh-Bernard, et al., Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance, J. Clin. Investig. 108 (7) (2001) 1061–1070.
[79] A. Malecka, et al., Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity, Haematologica 101 (9) (2016) e361–e364.
[80] A. Malecka, et al., The mutational landscape of cold agglutinin disease: CARD11 and CXCR4 mutations are correlated with lower hemoglobin levels, Am. J. Hematol. 96 (8) (2021) E279–E283.
[81] J.E. Jun, et al., Identifying the MAGUK protein Carma-1 as a central regulator of humoral immune responses and atopy by genome-wide mouse mutagenesis, Immunity 18 (6) (2003) 751–762.
[82] Y.S. Jeelall, et al., Human lymphoma mutations reveal CARD11 as the switch between self-antigen-induced B cell death or proliferation and autoantibody production, J. Exp. Med. 209 (11) (2012) 1907–1917.
[83] S.P. Treon, et al., Ibrutinib in previously treated Waldenstrom's macroglobulinemia, N. Engl. J. Med. 372 (15) (2015) 1430–1440.
[84] J. Trotman, et al., Zanubrutinib for the treatment of patients with Waldenstrom macroglobulinemia: 3 years of follow-up, Blood 136 (18) (2020) 2027–2037.
[85] G. Yang, et al., A mutation in MYD88 (L265P) supports the survival of lymphoplasmacytic cells by activation of Bruton tyrosine kinase in Waldenstrom macroglobulinemia, Blood 122 (7) (2013) 1222–1232.
[86] J.D. Phelan, et al., A multiprotein supercomplex controlling oncogenic signalling in lymphoma, Nature 560 (7718) (2018) 387–391.
[87] F. Castellani, et al., The Bruton tyrosine kinase inhibitor ibrutinib improves anti-MAG antibody polyneuropathy, Neurol Neuroimmunol Neuroinflamm 7 (4) (2020).
[88] H. Yasuda, et al., Anti-myelin-associated-glycoprotein neuropathy successfully treated with tirabrutinib, Heliyon 8 (10) (2022) e10928.
[89] C. Young, M. Singh, K.J.L. Jackson, et al., A triad of somatic mutagenesis converges in self-reactive B cells to cause a virus-induced autoimmune disease, Immunity 58 (2) (2025) 412–430.e10.