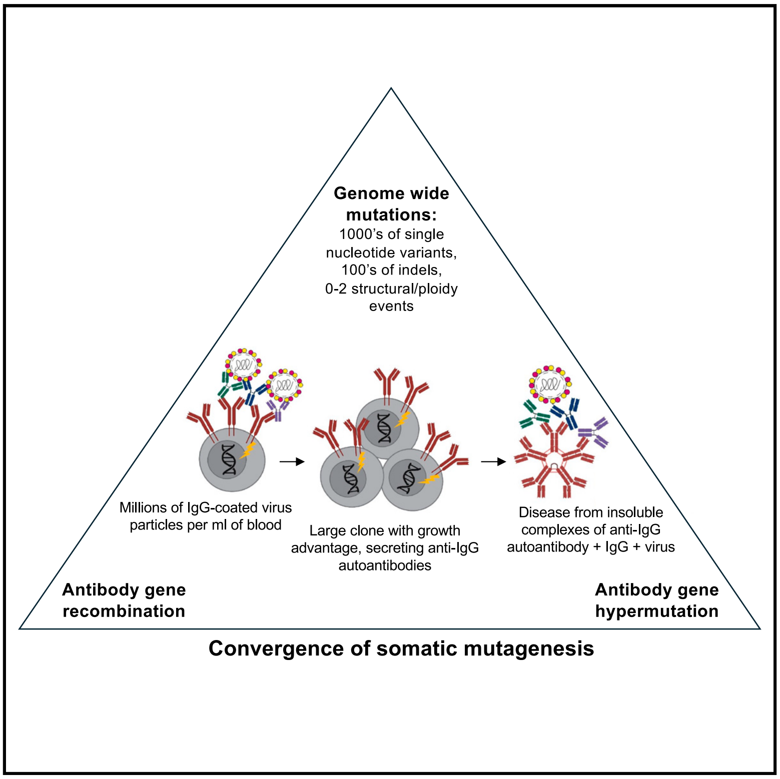
Article

Immunity

A triad of somatic mutagenesis converges in self-reactive B cells to cause a virus-induced autoimmune disease

Graphical abstract



Highlights

- Pathogenic B cell clones persist in HCV-cryovas patients despite HCV elimination
- Clones have acquired thousands of mutations, including 1–2 lymphoma driver mutations
- Ancestor and progeny self-reactive to multimerized IgG do not bind HCV E2 antigen
- IgM hypermutation increased affinity for IgG and introduced pathogenic insolubility

Authors

Clara Young, Mandeep Singh,
Katherine J.L. Jackson, ...,
Rowena A. Bull, Dan Suan,
Christopher C. Goodnow

Correspondence

d.suan@garvan.org.au (D.S.),
c.goodnow@garvan.org.au (C.C.G.)

In brief

Cryoglobulinemic vasculitis is a common complication of HCV infection. Young et al. reveal that three categories of somatic mutagenesis converge on a single B cell to generate pathogenic, autoantibody-producing clones in this virus-induced autoimmune disease, without evidence of molecular mimicry against the HCV E2 antigen as the initiating trigger.

Young et al., 2025, Immunity *58*, 412–430
February 11, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.immuni.2024.12.011

CellPress
OPEN ACCESS

Immunity

Article

A triad of somatic mutagenesis converges in self-reactive B cells to cause a virus-induced autoimmune disease

Clara Young,[1,2] Mandeep Singh,[1,2] Katherine J.L. Jackson,[1] Matt A. Field,[1,3,4] Timothy J. Peters,[1,2] Stefano Angioletti-Uberti,[5] Daan Frenkel,[6] Shyamsundar Ravishankar,[1] Money Gupta,[7,8] Jing J. Wang,[9] David Agapiou,[8] Megan L. Faulks,[1] Ghamdan Al-Eryani,[1] Fabio Luciani,[7,8] Tom P. Gordon,[9] Joanne H. Reed,[1,11] Mark Danta,[2] Andrew Carr,[10] Anthony D. Kelleher,[8,10] Gregory J. Dore,[2,8] Gail Matthews,[2,8] Robert Brink,[1,2] Rowena A. Bull,[7,8] Dan Suan,[1,2,12,*] and Christopher C. Goodnow[1,2,12,13,*]

[1]Garvan Institute of Medical Research, Darlinghurst, NSW, Australia
[2]St Vincent's Clinical School, UNSW Sydney, Sydney, NSW, Australia
[3]Australian Institute of Tropical Health and Medicine and Centre for Tropical Bioinformatics and Molecular Biology, Smithfield, Cairns, QLD, Australia
[4]Menzies School of Health Research, Darwin, NT, Australia
[5]Department of Materials, Imperial College London, London, UK
[6]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK
[7]School of Biomedical Sciences, UNSW Sydney, Sydney, NSW, Australia
[8]The Kirby Institute, UNSW Sydney, Sydney, NSW, Australia
[9]Department of Immunology, Flinders University and SA Pathology, Bedford Park, Adelaide, SA, Australia
[10]Immunology and HIV Unit, St Vincent's Hospital, Sydney, NSW, Australia
[11]Westmead Institute for Medical Research, Westmead, Sydney, NSW, Australia
[12]These authors contributed equally
[13]Lead contact
*Correspondence: d.suan@garvan.org.au (D.S.), c.goodnow@garvan.org.au (C.C.G.)

https://doi.org/10.1016/j.immuni.2024.12.011

SUMMARY

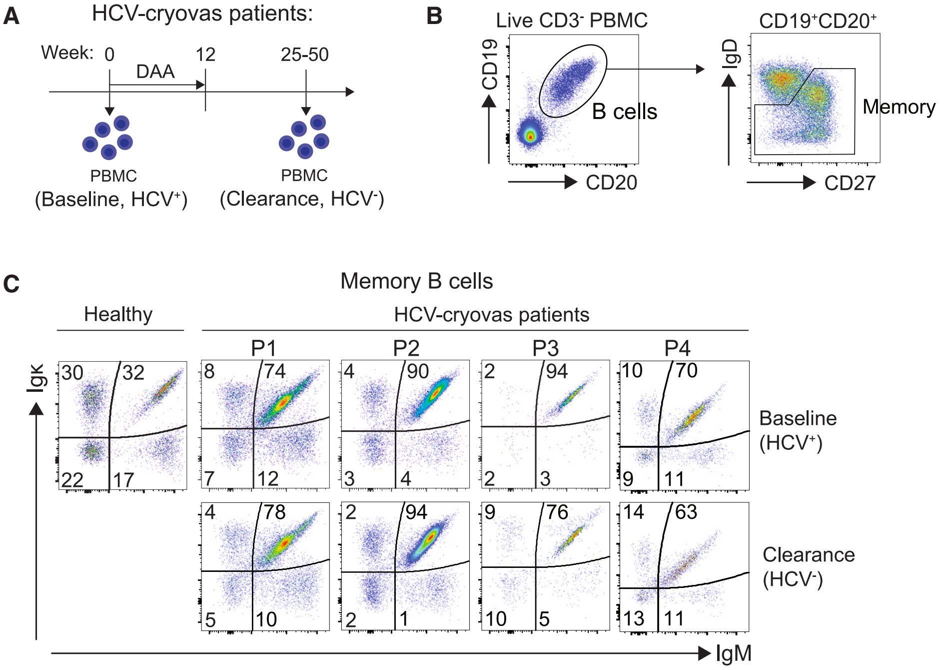
The unexplained association between infection and autoimmune disease is strongest for hepatitis C virus-induced cryoglobulinemic vasculitis (HCV-cryovas). To analyze its origins, we traced the evolution of pathogenic rheumatoid factor (RF) autoantibodies in four HCV-cryovas patients by deep single-cell multi-omic analysis, revealing three sources of B cell somatic mutation converged to drive the accumulation of a large disease-causing clone. A method for quantifying low-affinity binding revealed recurring antibody variable domain combinations created by V(D)J recombination that bound self-immunoglobulin G (IgG) but not viral E2 antigen. Whole-genome sequencing revealed thousands of somatic mutations, numerically comparable to chronic lymphocytic leukemia and normal memory B cells, but with 1–2 corresponding to driver mutations found recurrently in B cell leukemia and lymphoma. V(D)J hypermutation created autoantibodies with compromised solubility in complex with self-IgG. In this virus-induced autoimmune disease, infection promotes a catastrophic confluence of somatic mutagenesis in the descendants of a single B cell.

INTRODUCTION

Viral infection is postulated to trigger many autoimmune diseases, yet their connection is poorly understood. Here, we explore the pathogenesis of one of the strongest known associations between an autoimmune disease and an infection, namely the development of cryoglobulinemic vasculitis (cryovas) following infection with hepatitis C virus (HCV).

An estimated 58 million people are living with HCV infection,[1] and the development of direct-acting antiviral (DAA) therapy has been a pivotal impetus for the global hepatitis C elimination strategy.[2] HCV-cryovas is an autoimmune disease complicating HCV infection, characterized by small vessel leukocytoclastic vasculitis and mediated by a rheumatoid factor (RF) cryoglobulin: an IgM autoantibody that binds multiple monomers of plasma immunoglobulin G (IgG), forming insoluble precipitates at temperatures below 37°C.[3–6] Approximately half of HCV-infected individuals develop a circulating cryoglobulin. A subgroup develops clinical vasculitis: skin purpura and arthralgias develop in >15% of HCV-infected individuals, while neuropathy or glomerulonephritis occur in 9% and 5%, respectively.[7] In individuals diagnosed with cryovas prior to the advent of HCV treatments, >80% had HCV infection and were viremic with HCV

© 2025 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



D

Patient	Heavy					Light			
	V	D	J	CDR3	CDR3 length	V	J	CDR3	CDR3 length
P1	IGHV1-69*01	IGHD4-23*01	IGHJ4*02	`CAREGKGHDYGGNPFDYW`	16	IGKV3-20*01	IGKJ1*01	`CQQYGSSPWTF`	9
P2	IGHV4-59*01	IGHD6-19*01	IGHJ5*02	`CATSLGGSGWYFHPW`	13	IGKV3-15*01	IGKJ1*01	`CQQYNNWPPWTF`	10
P3	IGHV3-7*01	IGHD3-22*01	IGHJ3*02	`CARGDYYDSSGYFTDAFDIW`	18	IGKV3-15*01	IGKJ1*01	`CQQYNNWPPWTF`	10
P4	IGHV1-69*01	IGHD6-19*01	IGHJ6*02	`CAREGGGIAVAGYYGMDVW`	17	IGKV3-20*01	IGKJ1*02	`CQQYGSSHTF`	8
	IGHV4-59*01	IGHD4-23*01	IGHJ5*02	`CASGGHDSPKWLDPW`	13	IGKV3-15*01	IGKJ1*01	`CQQYNNWPPWTF`	10

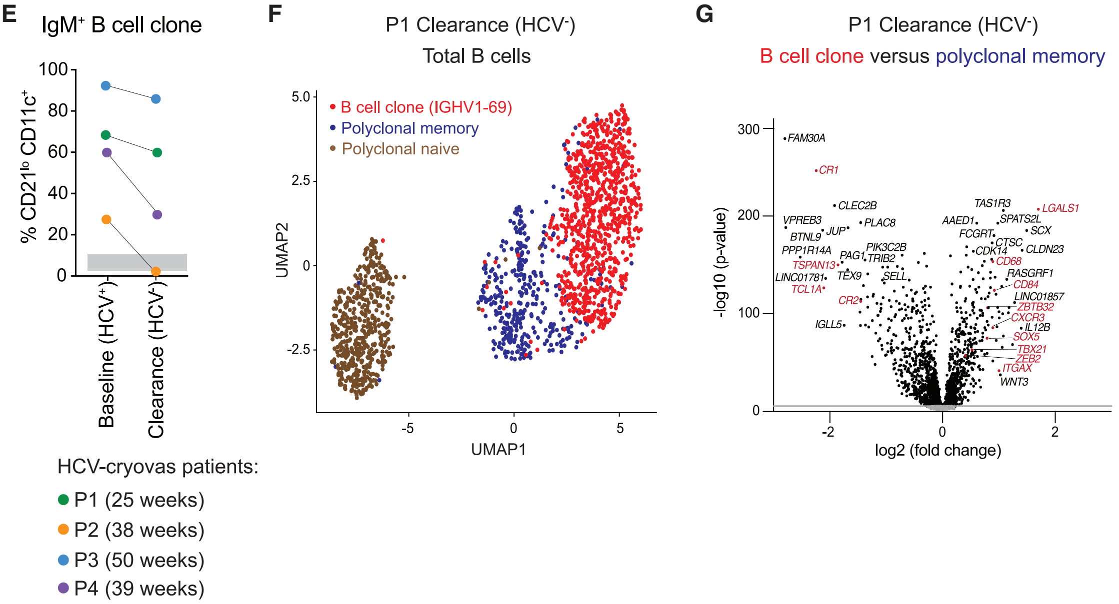


Figure 1. IgM⁺ B cell clonal expansions persist in the blood of cryovas patients following clearance of HCV, related to Figure S1

(A) Timeline of treatment and sample acquisition.

(B) Gating strategy for memory B cells.

(legend continued on next page)

RNA.[8–10] In ~30% of individuals with HCV-cryovas, the IgM autoantibody is polyclonal (type 3 cryoglobulinemia), while in ~70%, the IgM is monoclonal (type 2 cryoglobulinemia).[7,11] Individuals with acute HCV-cryovas are often treated with immunosuppression, but clearance of the virus through DAA treatment results in gradual remission of vasculitis in 70%–90% of individuals.[12–16]

Antigenic mimicry between viral and self-antigens is one explanation for the association between infection and autoimmunity. Favoring this possibility, HCV-neutralizing antibodies and HCV-crovas autoantibodies often start with identical sequences at four of six complementarity determining regions (CDRs). One antibody heavy-chain variable element, IGHV1-69, is used by 90% of potent HCV-neutralizing antibodies recognizing the envelope glycoprotein, E2.[17–23] These antibodies employ diverse kappa light-chain variable elements, with 22% employing IGKV3-20. Identical pairing of IGHV1-69 with IGKV3-20 occurs in 60% of pathogenic IgM RFs in HCV-cryovas, comprising the public "Wa" antibody idiotype.[24–26] Expanded B cell clones displaying IGVH1-69/IGKV3-20 IgM receptors are found in many individuals with HCV-cryovas as a non-leukemic, monoclonal B cell lymphocytosis.[27–32] These clonal expansions have usually acquired V(D)J mutations required for binding to self-IgG, but an unresolved question is whether or not the unmutated precursor antibodies bind HCV E2 or IgG.[31–35]

For reasons yet to be determined, individuals with HCV are 2.4 times more likely to develop B cell lymphoma than the general population, and this increases to 35 times in those with cryovas.[36,37] These lymphomas often display membrane IgM with the same public Wa idiotype and RF activity as the pathogenic secreted cryoglobulins.[21,38,39] One hypothesis for HCV-cryovas is that clones making the pathogenic autoantibody have escaped immune tolerance checkpoints and accumulated to larger numbers by acquiring somatic mutations corresponding to recurrent "driver" mutations found in B cell lymphomas and leukemias.[40–42] B cell tolerance checkpoints inhibit self-reactive B cell survival, proliferation, and plasma cell differentiation,[42,43] and driver mutations dysregulate these processes.[44] Such mutations have been found in patients developing cryovas as a complication of Sjogren's syndrome,[45] where there is no known infectious trigger.

To investigate how HCV infection may trigger autoimmune cryovas, here we traced the steps in the evolution of the pathogenic RF autoantibodies by performing in-depth single-cell RNA, DNA, and protein analysis of self-reactive B cell clones in four HCV-cryovas patients.

RESULTS

Large IgM$^+$ memory B cell clones with public autoantibody idiotypes accumulate before and after HCV treatment

In four HCV-cryovas patients aged 54–66 years, blood was analyzed at baseline when HCV RNA was detectable (HCV$^+$) and again 25–50 weeks later following DAA therapy-induced viral clearance (HCV$^-$) (Figure 1A; Table S1). All four experienced remission of vasculitis symptoms 1–26 weeks after DAA commencement. Patient P1 experienced a relapse of HCV-cryovas followed by remission again at 124 weeks following DAA commencement (Table S1).

Circulating B cell clones were identified in each patient both before and after HCV clearance, defined as CD19$^+$ CD20$^+$IgD$^{lo/-}$ memory B cells (Figure 1B) expressing surface IgM and restricted to kappa light chain (IgM$^+$IgK$^+$; Figure 1C). At the baseline time point (HCV$^+$), the IgM$^+$IgK$^+$ B cell clone comprised 70%–94% of the memory B cell compartment, which remained largely unchanged in all four patients following clearance of HCV (HCV$^-$) (Figure 1C). No patients exhibited an elevated total lymphocyte count (Table S1) or an overall increased frequency of B cells, with the exception of patient P2, who demonstrated a slightly elevated B cell frequency at the clearance time point (Figure S1A).

Bulk and single-cell immunoglobulin RNA sequencing of the patients' blood confirmed the expanded IgM$^+$ B cell clones at the clearance time point were the same clones detected at the baseline time point (Figures S1C and S1D). Their IgM corresponded to recurring "public" idiotypes of HCV-associated RF immunoglobulin heavy/light-chain combinations encoded by IGHV1-69/IGKV3-20 (Wa idiotype), IGHV4-59/IGKV3-15 (Bla idiotype), or IGHV3-7/IGKV3-15 (Po idiotype) (Figure 1D). The IgM$^+$ B cell expansion of patient P4 was bi-clonal, composed of a dominant IGHV1-69/IGKV3-20 clone (35% of total B cells) and a smaller IGHV4-59/IGKV3-15 clone (3% of B cells) (Figures 1D, S1C, and S1D). Using the IGHV1-69 anti-idiotypic antibody, G6,[46] blood IgM$^+$IGHV1-69$^+$ cells decreased 30% 39 weeks after commencing DAA therapy in patient P4 but increased 20% 13 weeks post-DAA therapy in P1 (Figure S1B).

CD2l^{lo}CD11c$^+$ age-associated memory B cell clones persist in the absence of virus

At the baseline time point, most IgM$^+$ clonal cells in patients P1, P3, and P4 (and 28% in P2) displayed a CD21^{lo}CD11c$^+$CD19^{hi} age-associated memory B cell profile, consistent with previous studies,[28,31,32,34] undergoing a small decrease after HCV

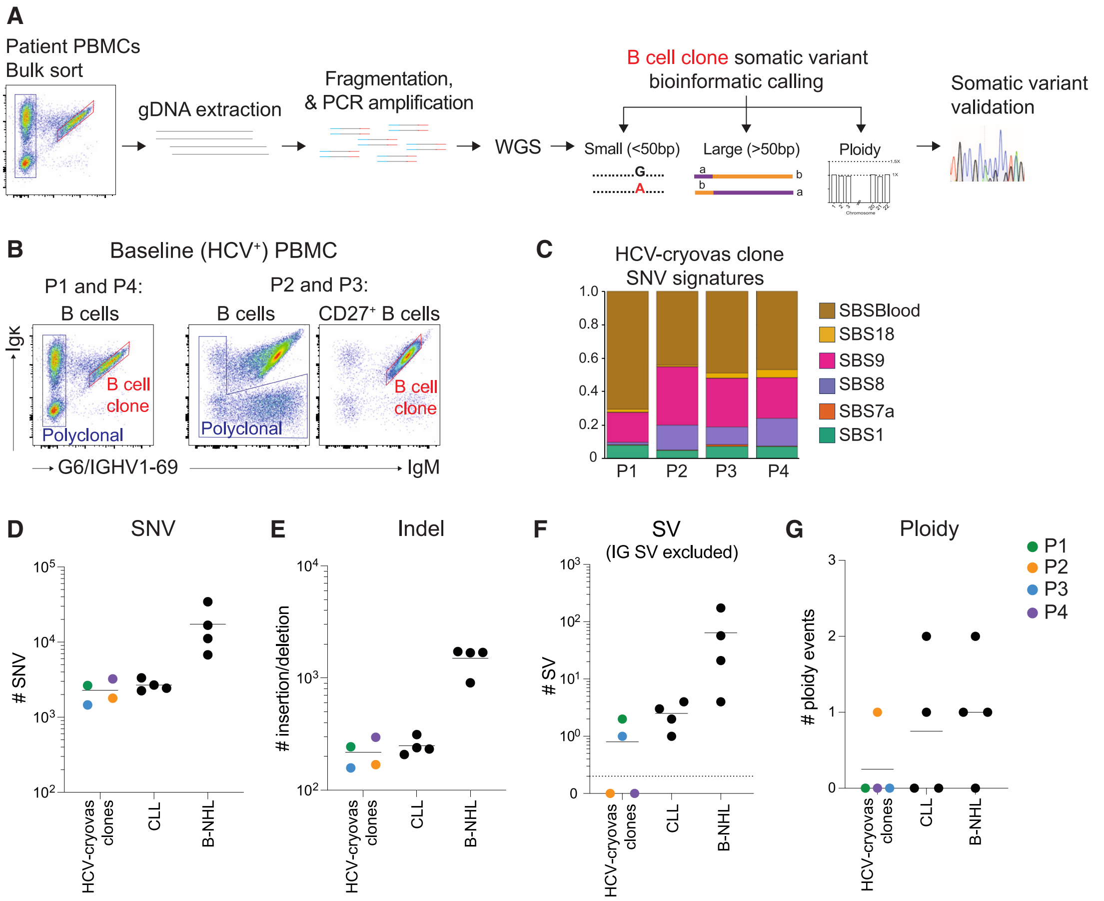
(C) Percentage of memory B cells expressing membrane IgM heavy chain and kappa (Igκ) light chains in each patient (P1–P4) before (HCV$^+$) and after virus clearance (HCV$^-$).

(D) Immunoglobulin heavy-chain VDJ and light-chain VJ gene rearrangements in each patient's clone.

(E) Frequency of CD21^{lo}CD11c$^+$ cells within the B cell clone of each patient at baseline and clearance. Gray shading: range of CD21^{lo}CD11c$^+$ cells among memory B cells of three age-matched healthy controls. Gating strategy shown in Figure S1F.

(F) Single-cell mRNA sequencing analysis of immunoglobulin receptor and global gene expression in CD19$^+$CD20$^+$ B cells sorted from the blood of patient P1 at the clearance (HCV$^-$) time point, with dimension reduction by uniform manifold approximation and projection (UMAP). Individual cells (denoted by dots) were annotated as the IGHV1-69 B cell clone ($n = 849$, red) or polyclonal cells based on their immunoglobulin sequence. Polyclonal naive ($n = 490$, brown) and memory B cells ($n = 360$, dark blue) were annotated according to landmark genes.

(G) Volcano plot of differentially expressed mRNAs between single cells corresponding to the clonal IGHV1-69 B cells ($n = 849$) and the polyclonal memory B cells ($n = 360$) of patient P1 at the clearance time point. Differentially expressed genes with family-wise error rate (FWER) < 0.05 are shown in black. Landmark mRNAs increased (e.g., *ITGAX* and *TBX21*) or decreased (e.g., *CR2*) in CD21^{lo} CD11c$^+$ age-associated memory B cell genes are shown in red.



H

Patient	Clone	Variant	Variant impact	Genome coordinates (hg38)			Del length (bp)
				Chr	Start	End	
P1	IGHV1-69	*KLF2* p.Cys279Del	In-frame deletion	19	16325973	16325975	3
P2	IGHV4-59	Trisomy 12 (+12)	Trisomy	12	-	-	-
P3	IGHV3-7	11q22-23 del (*BIRC3-PAFH1B2*)	Large deletion	11	102332947	117166632	14833685
P4	IGHV1-69	*NOTCH1* p.Pro2514fs*4	Frameshift	9	136496197	136496198	2
		TRAF3 p.Ser233fs*30	Frameshift	14	102889604	102889607	4

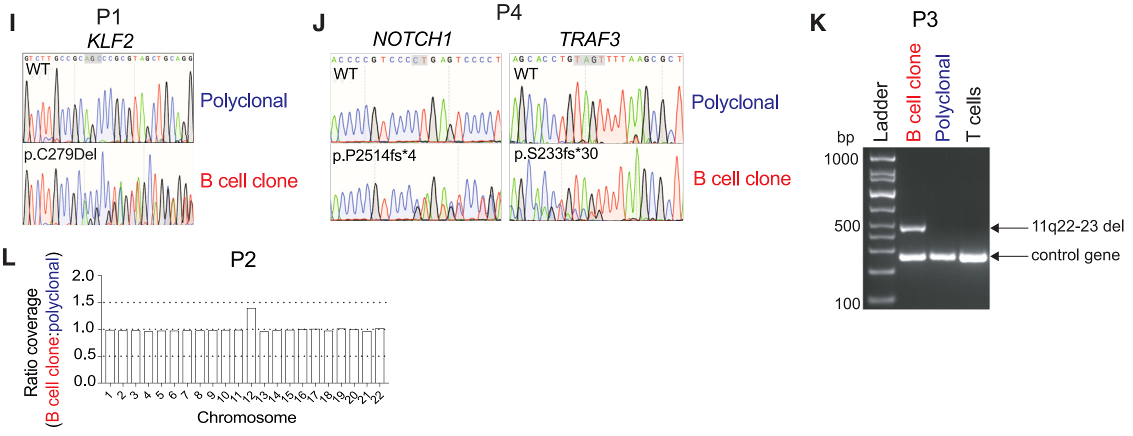


Figure 2. Somatic genome-wide and lymphoma driver mutations in the IgM+ clone of each patient, related to Figure S2

(A) Workflow to detect genome-wide somatic mutations.

(B) Gating strategy for bulk sorting of clonal B cells and polyclonal memory B cells from patient blood at the baseline (HCV+) time point.

(legend continued on next page)

clearance (Figures 1E, S1E, and S1F). Single-cell repertoire and gene expression mRNA sequencing (RAGE-seq)[47] in patient P1 confirmed the CD21^{lo}CD11c$^+$ profile in the persisting IgM$^+$IGHV1-69$^+$ B cell clone with upregulation of landmark genes including *ITGAX* (CD11c), *LGALS1* (Galectin-1), *CXCR3*, *SOX5*, *ZBTB32*, *TBX21* (T-bet), and *ZEB2* (Figure 1H), indicating the virus is not required for sustaining the age-associated memory B cell phenotype of the expanded clones. Patient P2 was the only case where the CD21^{lo}CD11c$^+$ subset declined into the normal range, despite the clone persisting or increasing in this patient at the clearance time point (Figure 1C).

HCV-cryovas clonal expansions accumulate thousands of somatic mutations across the genome

To identify genome-wide somatic mutations that might confer a clonal advantage, thousands of cells from each patient's IgM$^+$IgK$^+$ B cell clone at baseline were bulk sorted with estimated clone purity >90%, and genomic DNA was processed for short-read whole-genome sequencing (WGS) (Figure 2A). Paired analysis of WGS on DNA from pools of sorted control B cells from each patient (Figure 2B) was performed to remove germline variants. A three-tiered bioinformatic approach detected somatic mutations: (1) single-nucleotide variants (SNVs) and small (<50 bp) indels; (2) large chromosomal structural variants (SVs), including >50 bp deletions, translocations, and inversions; and (3) somatic ploidy events (Figure 2A). As an internal benchmark, the same bioinformatic pipeline called somatic mutations in tumor versus normal WGS pairs from four IGHV-mutated B cell chronic lymphocytic leukemias (CLLs) and four B cell non-Hodgkin's lymphomas (B-NHLs) representative from a large published analysis.[48]

Across the genome, the HCV-cryovas clones carried a mean of 2,281 SNVs (range 1,459–6,771) and 217 indels (range 169–296; Figures 2D and 2E; Table S2). A comparable mutation burden of mean 2,678 SNVs and 248 indels was called in the four CLLs, consistent with independent analysis of a large CLL set,[48] whereas mutations were 7–8 times higher in B-NHLs.[48] The majority of SNVs in the HCV-cryovas clonal B cells corresponded to mutation signatures associated with blood stem cell aging (SBSblood) and non-canonical activation-induced cytidine deaminase activity (SBS9) (Figure 2C), as observed in IGHV-mutated CLL[48,49] and in normal memory B cells.[50,51] The HCV-cryovas B cell clones carried a mean of 0.75 (range 0–2) chromosomal SVs not involving immunoglobulin genes (Figure 2F; Table S3), lower than the burden in CLL (mean 2.5, range 1–4) and much lower than B-NHL (mean 64, range 4–174) (Figure 2F). With respect to ploidy events, only the HCV-cryovas P2 clone had an event detected consisting of a whole chromosome 12 gain (Figure 2G; Table S4). A similar burden of ploidy events was called in CLL and B-NHL (range 0–2), including the identical chromosome 12 gain in two of the CLL samples (Table S4). Collectively, these analyses suggest the overall somatic mutation burden of the HCV-cryovas clones in these four patients is similar to CLL and lower than B-NHL.

Each clone acquires at least one lymphoma driver somatic mutation

Cross-referencing the somatic mutations in each HCV-cryovas clone with databases of lymphoma driver mutations found recurrently in human B cell lymphomas and leukemias[45] revealed a single putative driver mutation in patients P1, P2, and P3 and two putative driver mutations in P4 (Figure 2H). In patient P1, a heterozygous 3 bp exonic deletion present in 35% of reads, *KLF2*$^{p.Cys279Del}$, deleted a cysteine (Figures 2I and S2B; data not shown). Loss-of-function *KLF2* somatic mutations occur in 20%–42% of splenic marginal zone lymphoma (SMZL) and ~30% of diffuse large B cell lymphomas (DLBCLs) associated with hepatitis B virus infection (Figure S2B).[52–55] The deleted cysteine was absolutely conserved between species and paralogues (Figure S2C) and critical for coordinating a zinc ion to fold the C2H2 zinc finger for DNA binding. A *KLF2*$^{p.Cys274Tyr}$ somatic mutation in the other zinc-coordinating cysteine in this finger, from an SMZL patient, confers loss of function.[53] We confirmed both *KLF2*$^{p.Cys279Del}$ and *KLF2*$^{p.Cys274Tyr}$ mutations diminish the ability of KLF2 to repress an nuclear factor κB (NF-κB) reporter in transfected cells (Figure S2D).

In patient P4, the clone had two driver somatic mutations: a 2-nt exonic deletion, *NOTCH1*$^{p.Pro2514fs*4}$, and a 4-nt exonic deletion, *TRAF3*$^{p.Ser233fs*30}$, present in 45% and 39% of reads, respectively (Figures 2J, S2E, and S2F; data not shown). *NOTCH1*$^{p.Pro2514fs*4}$ occurs in 4%–12% of CLL patients and in monoclonal B cell lymphocytosis (MBL), conferring gain of function by truncating the C-terminal PEST domain (Figure S2E).[56–62] The truncating *TRAF3*$^{p.Ser233fs*30}$ mutation eliminates the coiled-coil domain required for heterodimerization with TRAF2 and for stabilizing TRAF3 trimers[63] and eliminates the MATH and TRAF-C domain that recruits NIK for ubiquitination by BIRC3.[64,65] Similar *TRAF3* truncating mutations and small chromosomal deletions eliminating *TRAF3* occur in 10% of SMZL, 2% of CLL, 10% of myelomas, 9% of DLBCL, and 45% of canine B cell lymphomas.[58,66–70]

In patient P3's clone, *BIRC3* was inactivated by a 15 Mb deletion within chromosome 11 (11q22-23 del) with breakpoints in *BIRC3* and *PAFH1B2* (Figures 2K and S2I; Table S3). The presence of the deletion in clonal B cells but not polyclonal memory B cells or T cells was confirmed by PCR with primers flanking the breakpoint (Figure 2K). Between 18% and 30% of CLL cases carry 11q deletions.[58,71] Patient P3's 11q22-23 deletion results

(C) Proportion of somatic SNVs in each patient's clone corresponding to the indicated mutational signatures.

(D–G) Number (#) of somatic SNVs (D), indels (E), and SVs at non-immunoglobulin (IG) loci (F) and ploidy events, including chromosome gain/loss (G), within the WGS of the HCV-cryovas B cell clones (P1–P4) compared with 4 representative IGHV-mutated CLL samples and 4 B-NHLs.[48] Dotted line in (F) indicates a change from logarithmic scale to a linear scale on the y axis.

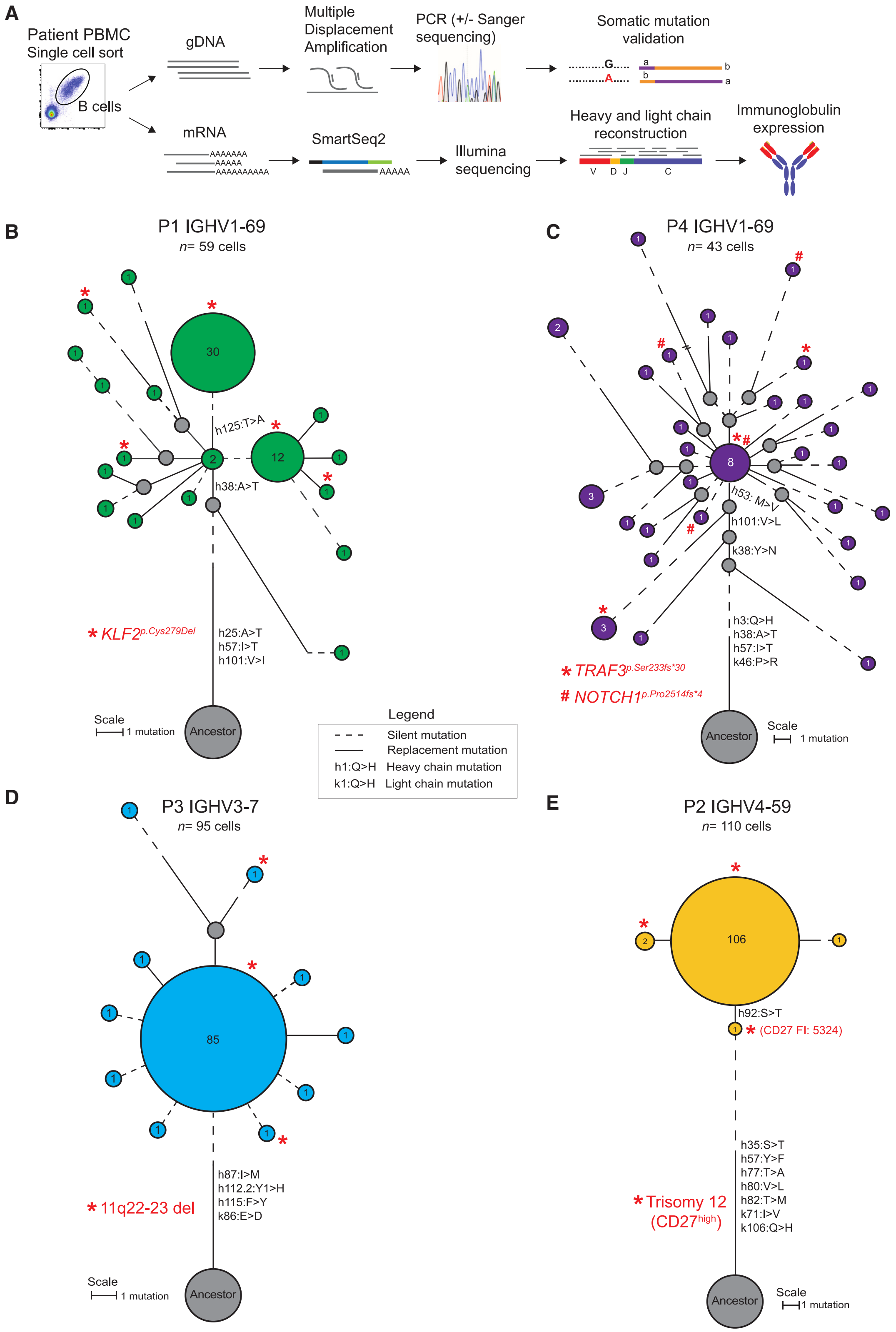
(H) Summary of lymphoma driver mutations detected.

(I) Confirmation of *KLF2*$^{p.C279del}$ somatic mutation in patient P1's clone and absence in polyclonal B cells by PCR and Sanger sequencing.

(J) Confirmation of the *NOTCH1*$^{p.P2514fs*4}$ and *TRAF3*$^{p.S233fs*30}$ somatic mutations in patient P4's clone by PCR and Sanger sequencing.

(K) Confirmation of the 11q22-23 deletion in the clone (red) and not in polyclonal B cells (blue) or T cells of patient P3 by PCR amplification across the breakpoint (top arrow). Primers to a control gene (*SEC23IP*) were included as a positive control (bottom arrow).

(L) Ratio of average WGS read depth per chromosome in patient P2's clone relative to polyclonal memory B cells.



(legend on next page)

in a truncating mutation, *BIRC3*$^{p.N442*}$, and heterozygous loss of 95 protein-coding genes, including genes relevant to B cell function: *ATM*, *BIRC2*, *POU2AF1*, *BTG4*, *CUL5*, *NPAT*, *PPP2R1B*, and caspases 1, 4, and 5 (Figure S2I). *BIRC3* loss-of-function mutations occur in 11% of SMZL, 5% of CLL, and 8% of MBL, and loss of *BIRC3* occurs in 83% of CLL cases with an 11q deletion.[56,58,69,72]

Patient P2's clone had chromosome 12 trisomy (Figure 2L; Table S4), a cytogenetic abnormality found in 10%–20% of CLL[58,71] (Table S4) and 18% of MBL.[73] WGS read coverage was increased 1.4–1.5× for chromosome 12 from sorted clonal cells compared with polyclonal memory B cell counterparts from the baseline blood sample (Figure 2L) and after HCV clearance (Figure S2G; Table S4). GSEA of differentially expressed mRNAs in clonal B cells revealed enrichment of gene sets corresponding to many chromosome 12 cytogenetic bands (Figure S2H), as observed for CLL samples with trisomy 12.[74,75]

Lymphoma driver mutations arise prior to clonal tree branching

Amplification and sequencing of mRNA and genomic DNA from single sorted memory B cells (Figure 3A) were used to analyze intraclonal heterogeneity of the HCV-cryovas clonal B cells at the baseline time point. Immunoglobulin mRNA V(D)J sequences revealed prior and ongoing somatic hypermutation among the clonal cells, enabling the generation of a clonal tree for each patient (Figures 3B–3E). Analysis of DNA for a subset of these single B cells placed the driver mutations within the clonal trees. As controls, sorted single memory B cells that were not part of the expanded clone were analyzed in parallel, with each displaying a polyclonal, unique antibody V(D)J rearrangement (Figure S3).

In patient P1 (Figure 3B), the *KLF2*$^{p.Cys279Del}$ mutation was detected in cells across the clonal tree, including branches that precede the predominant expansion of cells with the heavy-chain Thr125Ala (h125:T>A) mutation that increases binding to IgG (see below). The mutant *KLF2* allele was amplified in 63% of clonal B cells tested ($n = 7/11$ cells) and none of the polyclonal memory B cells ($n = 0/16$ cells) (Figure S3A). Only ~50% of allelic sequences present in genomic DNA of single cells are amplified during multiple displacement amplification, explaining why the mutant allele was not detected in all cells from the clonal tree. Thus, the *KLF2* mutation was acquired before the clone acquired potent RF activity and accumulated to the size circulating when P1 began DAA treatment.

Similarly, in patient P4, the *NOTCH1*$^{p.Pro2514fs*4}$ and *TRAF3*$^{p.Ser233fs*30}$ mutations were both found in single cells across the clonal tree but not in polyclonal memory B cells (Figures 3C and S3A). *TRAF3*$^{p.Ser233fs*30}$ was already present in an early branch preceding acquisition of the heavy-chain M53V mutation shared by 88% of clonal B cells (Figure 3C).

By contrast, the pathogenic clones in patients P2 and P3 (Figures 3D and 3E) displayed almost no intraclonal V(D)J diversity. PCR amplification of single cells from patient P3 detected the 11q22-23 deletion in 60% of individual clonal B cells ($n = 18/30$ cells) and none of three polyclonal B cells sampled ($n = 0/3$ cells) (Figure S3A), consistent with carriage by all of the clonal cells. In patient P2, flow cytometric measurement of CD27 (encoded on chromosome 12) revealed increased CD27 on the majority of clonal cells compared with polyclonal memory B cells (Figure S3B). CD27 was not increased on clonal cells of the other patients (data not shown). In P2, CD27 fluorescence intensity >2 SD above the mean of non-expanded polyclonal B cells was observed on 55% of clonal cells ($n = 33/60$ cells), including an ancestral cell yet to acquire the heavy-chain Ser92Thr (h92:S.T) mutation shared by all other cells in the circulating clone (Figures 3D and S3B). Collectively, these results indicate that a lymphoma driver gene mutation was acquired before each clone accumulated to the large size circulating at the baseline time point.

Accumulation of V(D)J mutations confers pathogenic autoantibody activity

Next, we tested the hypothesis that the clonal B cells were the source of RF cryoglobulins in the HCV-cryovas patients. We selected the immunoglobulin sequence shared by at least 50% of individual clonal cells; Figures 3B–3E, S4A, and S4B) and expressed it as pentameric secretory IgM. In parallel, unmutated ancestor IgM antibodies were synthesized corresponding to each patient's most frequent clone but with the V(D)J somatic mutations reverted to their ancestral sequence (Figures S4A and S4B).

The clonal IgM antibodies from patients P1, P2, and P4 bound IgG in ELISA and thus demonstrated RF activity (Figure 4A) and cryoprecipitated with IgG below 25°C (Figure 4B). By contrast, IgM corresponding to the unmutated ancestor of each clone displayed minimal or no detectable RF or cryoglobulin activity (Figures 4A and 4B). These results indicate the 5–8 non-synonymous V(D)J mutations acquired by the clones in patients P1, P2, and P4 conferred IgG self-reactivity and pathogenic insolubility of the IgM autoantibody when complexed with IgG.

In patient P1, 51% of clonal cells had diverged by acquiring a heavy-chain Thr125Ala mutation (Figure 3B). When this single mutation was reverted to express the immediate precursor antibody, the RF activity measured by ELISA was markedly reduced (Figures 4C and S4C). Since cells in both branches already carried *KLF2*$^{p.Cys279Del}$, increased self-reactivity to IgG was acquired after the driver mutation.

Two heavy-chain mutations, Ala38Thr and Ile57Thr, were independently acquired early in the clonal tree by the IGHV1-69^{+} B cell clones of patients P1 and P4 (Figure S4A). Reversion

Figure 3. Clonal trees inferred from single-cell mutation analysis, related to Figure S3

(A) Schematic of mutation detection in genomic DNA and immunoglobulin V(D)J mRNA.

(B–E) Clonal trees depicting the acquisition of individual V(D)J mutations by manual nearest neighbor analysis. Gray circles indicate inferred unmutated ancestor ("ancestor") and inferred intermediate cells. Colored circles denote one or more sequenced cells sharing the identical immunoglobulin nucleotide sequence, with the size of the circle proportional to the number of cells (cell number indicated in center). Red asterisk or hash denotes one or more cells identified with the indicated driver mutation. Length of branches corresponds to the number of V(D)J mutations (see scale): solid lines, number of replacement mutations; dotted lines, number of silent mutations. Amino acid change and position (IMGT numbering) are indicated with lower case "h" for heavy chain and "k" for the kappa light chain.

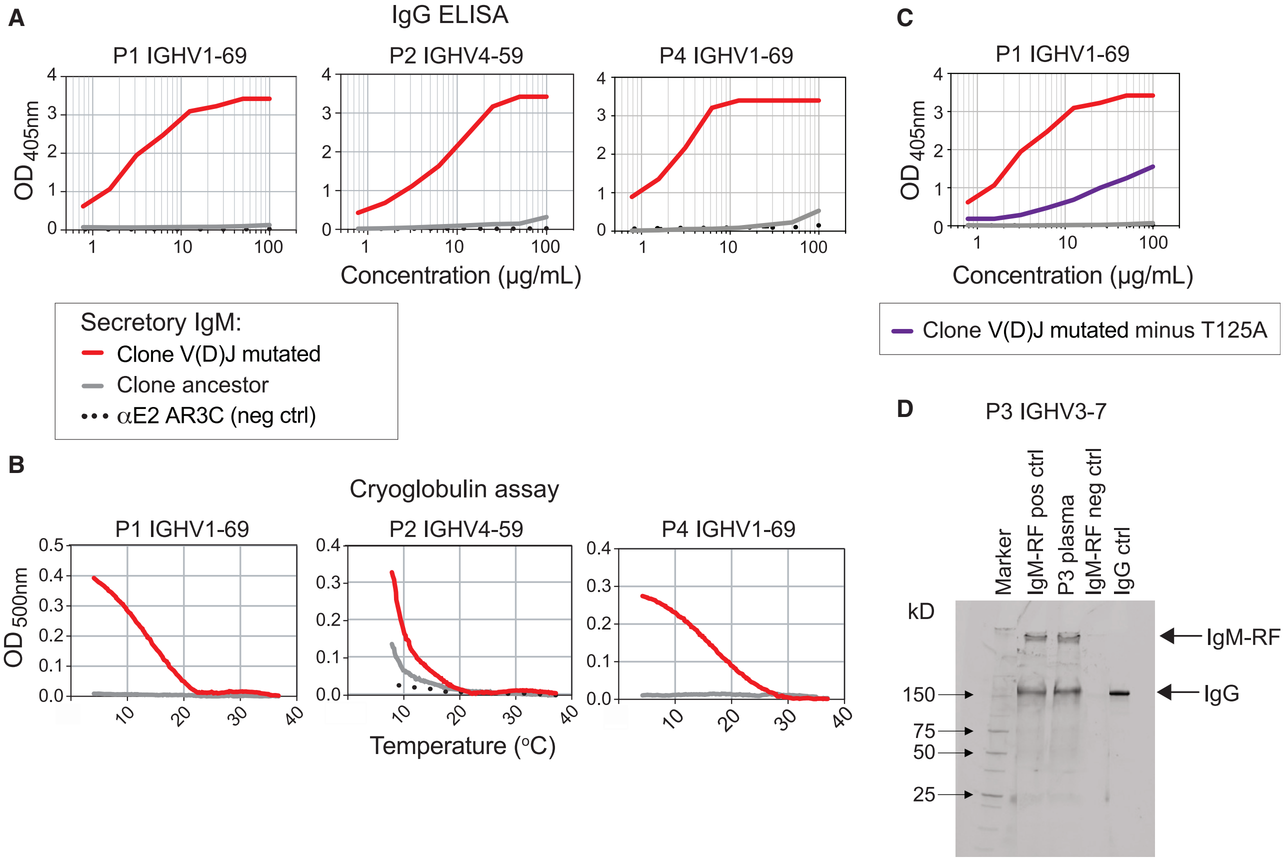


Figure 4. IgG binding and cryoprecipitation of expressed pentameric IgM, related to Figure S4

(A) Mutated (red) and unmutated ancestor (gray) IgM antibodies were expressed and tested by ELISA for binding to human IgG. Negative control (black) was AR3C expressed as IgM. Representative of 2–4 independent experiments, run in duplicate. OD, optical density.

(B) Expressed IgM antibodies in (A) were mixed with human polyclonal IgG, and aggregation measured by visible wavelength optical density in a spectrometer cooled from 37°C to 4°C at 0.1°C/min. Representative of 2-3 independent experiments.

(C) ELISA for expressed IgM from patient P1 as described in A but comparing IgM corresponding to the predominant clonal sequence without the IGHV1-69 T125A mutation (dark purple).

(D) Heat-aggregated IgG was mixed with plasma from patient P3 at baseline time point, or positive or negative control plasma samples, and precipitates analyzed by SDS-PAGE under non-reducing conditions to detect intact IgM and IgG (arrows).

of Ala38Thr alone had no impact on IgG binding (Figure S4C). It occurred within a predicted activation-induced cytidine deaminase (AID) hotspot comprising overlapping DGYW and WRCY motifs and may have co-occurred by chance. By contrast, reversion of Ile57Thr alone abolished IgG binding to IgM in patients P1 and P4 (Figure S4C), consistent with convergent evolution to increase self-reactivity. Ile57Thr arose early in the clonal tree in both patients (h57:I>T, Figures 3B and 3C).

Synthesized IGHV3-7 IgM of patient P3 lacked demonstrable RF in ELISA (Figure S4D) nor cryoglobulin activity in the rapid cooling spectrophotometer assay (data not shown). However, when the plasma of patient P3 was mixed with heat-aggregated IgG, IgM RF activity was detected (Figure 4D). This precipitating IgM RF was gel purified and subject to *de novo* sequencing by mass spectrometry, identifying heavy- and light-chain CDR3 amino acid sequences identical to the synthesized IgM. Thus, the IGHV3-7/IGKV3-15 antibody expressed on the IgM$^+$ B cell clone of patient P3 corresponds to the cryoglobulin autoantibody present in blood but is too low affinity to bind IgG in ELISA.

HCV E2 reactivity in anti-E2 ancestors but not autoantibody ancestors

Apparent lack of self-IgG binding by the unmutated ancestral autoantibodies in the four patients posed a dilemma for understanding how clonal expansion was initiated. We sought to distinguish between two possible explanations: (1) the ancestor IgM bound HCV E2, and this initiated B cell proliferation, with daughter cells acquiring IgG binding through V(D)J mutations, or (2) the ancestor IgM bound IgG, albeit with too low affinity for detection by ELISA. To resolve these alternatives, we extended a protocol that permits detection of low-affinity IgM binding to influenza hemagglutinin[76] by expressing membrane IgM on HEK293 cells (Figure 5A). Binding to multimerized HCV E2 or IgG antigen was tested by flow cytometry using fluorescent dextran molecules carrying 20–25 biotin acceptor sites (5–6 streptavidin molecules) loaded with biotinylated antigen (Figure 5A). A valuable characteristic of this method arises from each transfected cell expressing a different amount of membrane IgM, spanning a 500-fold range comparable to the range

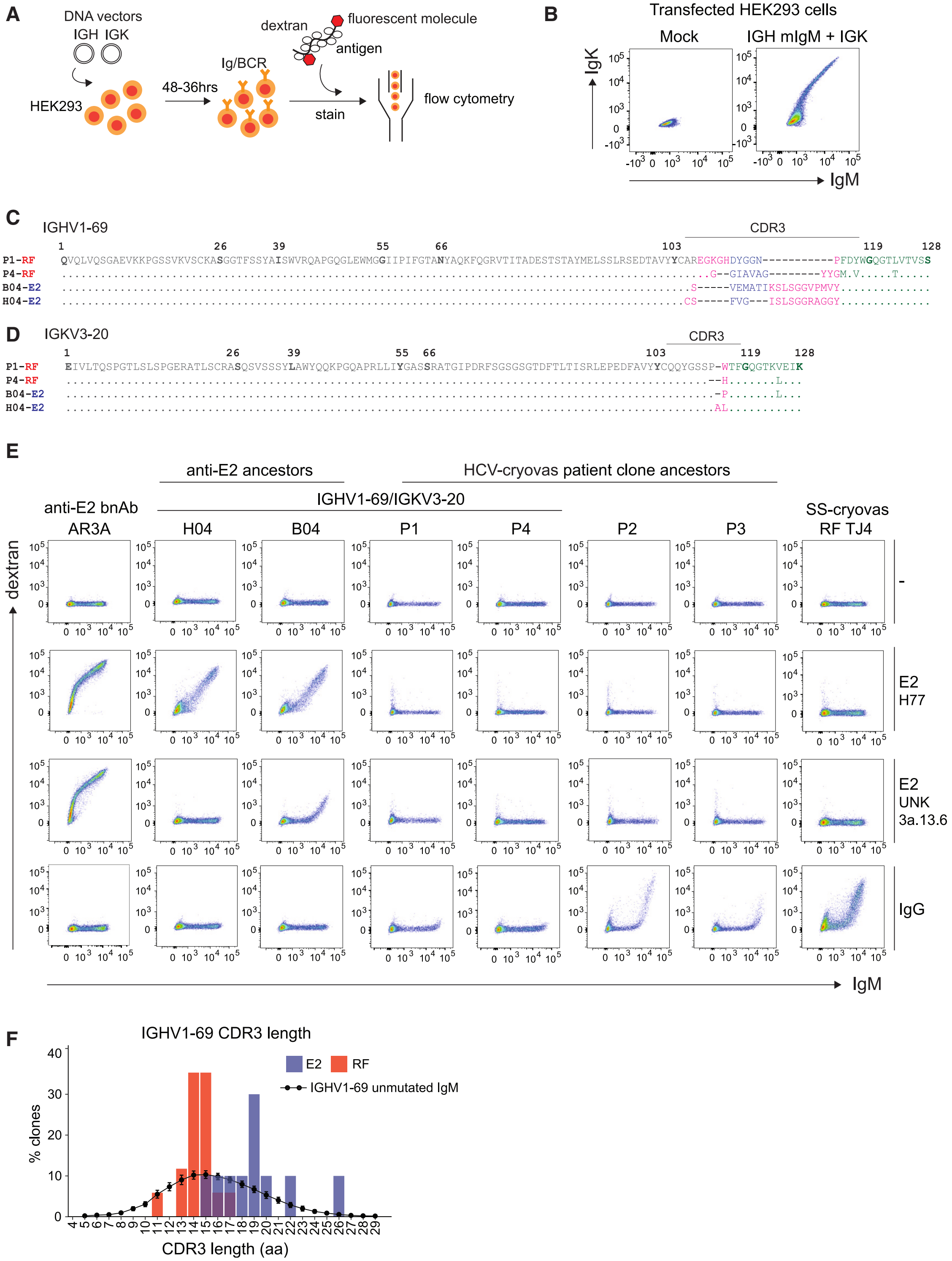


Figure 5. Cryoglobulin clones ancestral IgM bind multimerized IgG but not HCV E2, related to Figure S5

(A) Experimental protocol. HEK293F cells were transiently transfected with vectors encoding membrane IgM corresponding to patient or control antibodies. Fluorescent dextran molecules bearing an estimated 5–6 streptavidin molecules were mixed with biotinylated proteins (e.g., IgG), and binding to the transfected cells was analyzed by flow cytometry.

(B) Flow cytometric analysis of transfected or mock-transfected HEK293F cells stained with fluorescently labeled anti-human IgM and anti-human kappa light-chain antibodies, demonstrating the wide range of cell surface IgM density from cell to cell.

(legend continued on next page)

of membrane IgM concentrations on human B cells (Figures 5B and S5A).

Multimerized HCV E2 corresponded to genotype 1a (isolate H77) and genotype 3a (isolate UNK 3a.13.6), which comprise >80% of Australian HCV infections.[77] All four patients were infected with HCV genotype 1, patients P2 and P4 matching H77 (Table S1). As positive and negative controls, HEK293 cells were transfected to express membrane IgM corresponding to a hypermutated, pathogenic IGHV1-69/IGKV3-20 IgM cryoglobulin, TJ4, from a Sjogren's syndrome patient with cryovas,[45] or a hypermutated, broadly neutralizing IGHV1-69/IGKV3-20 HCV E2 antibody, AR3A.[19] In parallel, E2-dextran binding was tested on transfected cells expressing the unmutated ancestor IgM of two clonally unrelated IGHV1-69/IGKV3-20 anti-HCV E2 antibodies, B04 and H04, isolated from circulating memory B cells of a patient chronically infected with HCV (R.A. Bull et al., unpublished data). The heavy- and light-chain amino acid sequences of the ancestral B04 and H04 antibodies were identical to the RF ancestors of patients P1 and P4 with the exception of the heavy- and light-chain CDR3 loops (Figures 5C and 5D).

Cells with each different antibody sequence were transfected, stained, and analyzed in parallel to hold key variables constant, including the extent of antigen multimerization or sensitivity of the flow cytometer. None of the transfected cells bound fluorescent dextran devoid of antigen (Figure 5E, top row). As expected, cells expressing the HCV E2 antibody AR3A[19] bound E2 dextran (Figure 5E, far left). Cells expressing ancestral B04 and H04 bound fluorescent dextran decorated with E2 H77 (genotype 1a) (Figure 5E, second row). A higher threshold membrane IgM was needed for ancestral B04-expressing cells to bind E2 dextran for isolate UNK3a.13.6 (genotype 3a) (Figure 5E, third row). Neither isolate of HCV E2 dextran bound to cells expressing the ancestral IgM from patients P1 to P4 (Figure 5E). Additionally, none of the mutated IgM RF cryoglobulins from patients P1 to P4 bound HCV E2 (Figures S5C and S5D).

The results above did not support the hypothesis that clonal expansion of the autoantibody-forming cells was driven by binding to HCV E2, neither in the unmutated ancestor nor in their mutated progeny. By contrast, HEK cells expressing high membrane densities of the IgM ancestor from P2 and P3 consistently bound IgG dextran. For P1 and P4, only cells with the very highest densities of the ancestor IgM bound IgG dextran, and this was at the limit of sensitivity in some experiments (Figures 5E bottom row, 6A second row, S6A, and S6B; Table S6). Reciprocally, no IgG-dextran binding was detectable on cells expressing membrane IgM corresponding to the ancestral anti-HCV B04 and H04 antibodies (Figure 5E, bottom row).

The ancestral B04 and H04 antibodies have longer heavy-chain CDR3 loops than the ancestral P1 and P4 antibodies (Figure 5C). We, therefore, compared $CDR3_H$ lengths in published sequences for IGHV1-69/IGKV3-20 hypermutated antibodies known to bind either HCV E2 (10 sequences) or self-IgG (17 sequences) and for 1,343 IGHV1-69 unmutated IgM antibodies of unknown specificity expressed in the circulating B cell repertoire of healthy adults (Figures 5F and S5B). The naive IgM repertoire spanned a broad range of $CDR3_H$ lengths with mode 15 aa. IgG-binding autoantibodies were skewed toward the shorter part of this range (mode 14–15 aa, range 11–17), whereas E2-binding antibodies were skewed toward longer $CDR3_H$ lengths (mode 19 aa, range 15–26) (Figure 5F). $CDR3_H$ length, which becomes fixed by V(D)J recombination at the pre-B cell stage, appears to be one of the determining factors for self or virus antigen to bind to IGHV1-69/IGKV3-20 antibodies.

Next, we tested whether the unmutated IgM ancestors would recognize IgG bound to a multivalent array of virus antigen. When soluble monomeric IgG1 corresponding to the anti-E2 broadly neutralizing antibody AR3C was mixed with dextran-bearing multimerized biotinylated E2 virus protein, the IgG1-virus antigen complexes bound to cells expressing high membrane IgM corresponding to the ancestral IgM cryoglobulins (Figure S5E). Thus, while the ancestral autoantibodies had no detectable binding to the virus E2 antigen, they bound self-IgG when it decorated multimers of viral antigen.

IgG self-reactivity is low affinity at clonal initiation

Transfected cells expressing the mutated cryoglobulin autoantibodies or the unmutated IgM ancestors bound IgG dextran above a threshold amount of membrane IgM that differed between the patients. Above the threshold, binding of IgG dextran increased steeply with small increases in membrane IgM (Figures 5E, 6A, and 6B). The ancestor with the lowest threshold for IgG-dextran binding, from patient P2, was the only ancestral antibody with detectable cryoglobulin activity (Figure 4B). To test if the different thresholds reflected differing affinities for IgG, we transfected cells to express membrane IgM with known affinities for IgG previously determined by biolayer interferometry.[45] The mutated IGHV1-69/IGKV3-20 cryoglobulin, TJ5, and its immediate clonal precursor lacking one replacement mutation, TJ4, bind IgG with monovalent K_D of 4.8×10^{-6} M and 11.2×10^{-6} M, respectively. Consistent with a 2.3-fold higher K_D, the threshold IgM per cell needed to bind IgG dextran was 1.9-fold higher for TJ4 than for TJ5 (Figure 6B).

(C and D) Amino acid alignments of heavy- (C) and kappa light-chain (D) variable domains of the unmutated ancestor IgM in patients P1 and P4 and the unmutated ancestor IgM for the anti-HCV E2 antibodies B04 and H04. IMGT numbering, with CDR flanking residues in bold. The V gene segment (dark gray), D gene segment (navy blue), J gene segment (green), and likely N-additions are shown in pink. Periods (.) indicate identity, and dashes (-) indicate gaps.

(E) Binding of empty dextran (top row), HCV E2 H77 dextran (second row), and HCV E2 UNK3a.13.6 dextran E2 dextran (third row) and IgG dextran (bottom row) to transfected HEK293 cells expressing membrane IgM corresponding to the mutated anti-HCV E2 antibody AR3A (far left column), the unmutated ancestors of two anti-HCV E2 antibodies (B04 and H04), the unmutated ancestors of patients P1–P4, or a mutated IgM RF cryoglobulin from a Sjogren's syndrome-cryovas patient, TJ4 (far right).

(F) Percent of IGHV1-69/IGKV3-20 antibodies with the indicated amino acid length of heavy-chain CDR3, for antibodies that bind self-IgG (RF, red, $n = 17$) and antibodies that bind HCV E2 (E2, blue, $n = 10$). The black line denotes corresponding length distribution for 1,343 IGHV1-69 IgM antibodies of unknown specificity, representing the naive repertoire (<2% SHM) from PBMCs of 61 healthy donors. Points show mean ± SD.

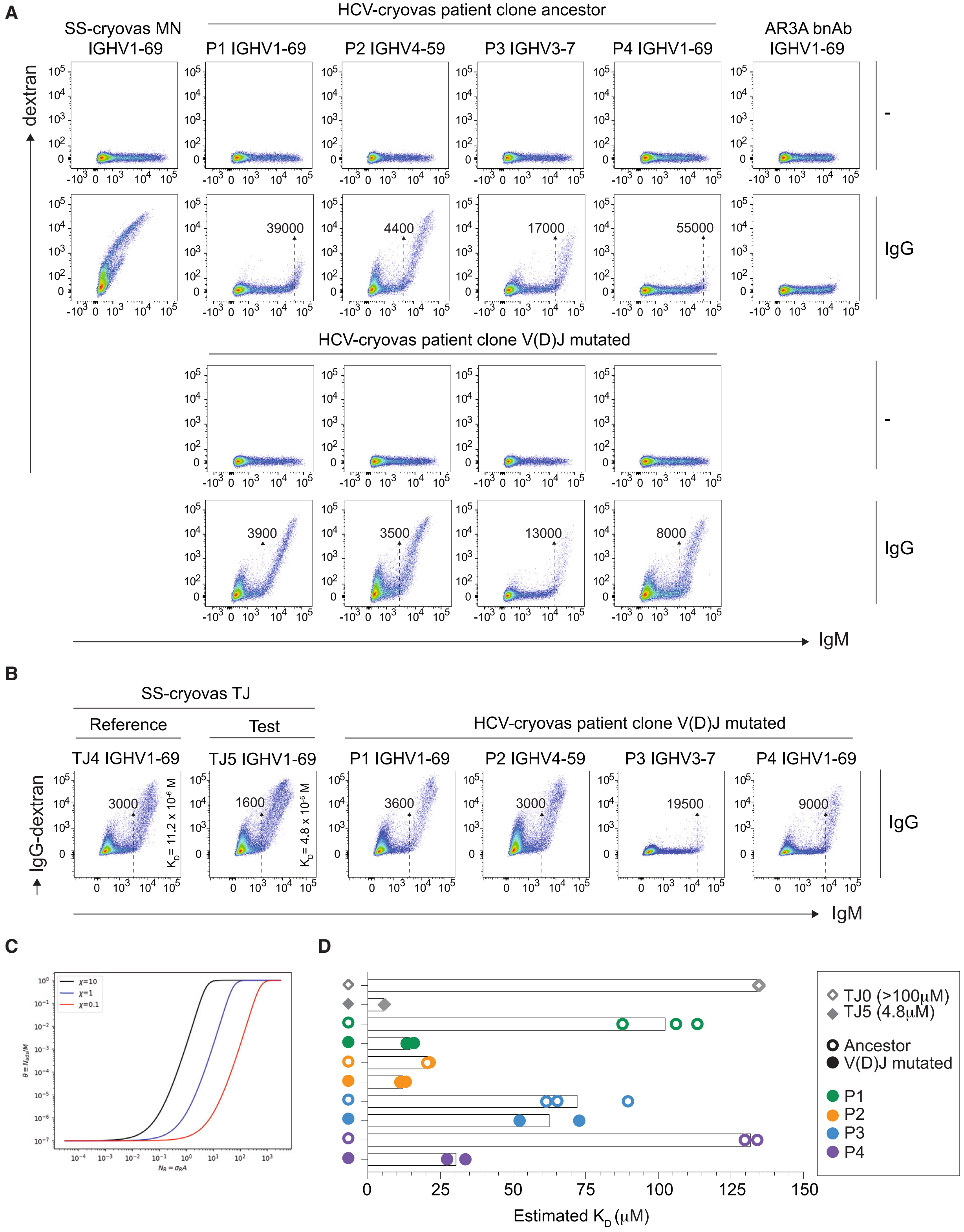


Figure 6. Estimating antibody affinities for IgG, related to Figure S6

(A) Binding of empty dextran or IgG dextran to HEK293 cells expressing membrane IgM corresponding to the ancestor or the predominant V(D)J mutated clone from patients P1 to P4. Control cells (far left and right columns) express membrane IgM corresponding to a mutated IGHV1-69/IGKV3-20 RF cryoglobulin from a

(legend continued on next page)

The shapes of the IgG-dextran-binding curves correspond to the superselective binding reaction predicted by mathematical modeling of low-affinity interactions between multivalent ligand in solution and surfaces bearing cognate receptors[78–82] (STAR Methods). The model predicts that, holding other variables constant, an increase in antibody K_D shifts the threshold amount of membrane IgM to a proportionally higher value (Figure 6C; STAR Methods). As such, if the K_D value of a reference IgM is known, an unknown K_D can be extrapolated from the membrane IgM binding threshold and subsequent shift in the binding curve (Figure 6C). To test the equation linking IgM threshold to K_D, we used the equation and the visually determined membrane IgM threshold for IgG-dextran binding for TJ4 to calculate TJ5 IgG affinity to be $K_D = 5.7 \times 10^{-6}$ M, and the unmutated ancestor of the TJ4 and TJ5 RF antibodies, TJ0, to be $K_D = 135 \times 10^{-6}$ M (Figure 6D; Table S6) (see STAR Methods), which were within range of their biolayer interferometry measurements of $K_D = 4.8 \times 10^{-6}$ M and $K_D > 100 \times 10^{-6}$ M, respectively.

We used the equation and visually called thresholds to extrapolate the affinities for IgG in three independent experiments that used TJ4 as the reference (Figures 6B, 6D, S6A, and S6B; Table S6). In one of the three experiments, the threshold could not be called for the lowest affinity ancestor, P4 (Figure S6B; Table S6), indicating the limit of sensitivity for the assay. The K_D of the four V(D)J mutated pathogenic HCV-cryovas autoantibodies was estimated (from highest to lowest affinity): P2, 12×10^{-6} M; P1, 15×10^{-6} M; P4, 31×10^{-6} M; and P3, 63×10^{-6} M (Figure 6D; Table S6). The estimated IgG affinities of their ancestors were: P2, 21×10^{-6} M; P3, 72×10^{-6} M; P1, 103×10^{-6} M; and P4, 132×10^{-6} M.

An alternative, objective method was also used to calculate the membrane IgM threshold for onset of IgG-dextran binding. The IgM and dextran fluorescent values for individual transfected cells expressing membrane IgM were fit to an equation using a basin-hopping Monte Carlo procedure using original code (“FluorescenceFit”) (Figure S6C) (see STAR Methods). The mean K_D estimates determined using the fitting algorithm (Figure S6D) were similar to the K_D estimates determined by visual inspection (Figure 6D). However, no K_D could be objectively estimated for ancestral P1 and P4 despite IgG binding to the highest membrane IgM-expressing cells (Figures 6A, S6A, and S6B), likely limited by insufficient cells for the Monte Carlo procedure to fit a curve with high confidence. The flow cytometric methods developed here provide a practical way to quantify low binding affinities with K_D in the 1–100 μM range.

DISCUSSION

These findings address the question of why autoimmune cryovas arises in HCV-infected individuals and why individuals with HCV-cryovas have a 35-fold increased risk of B cell lymphoma. In the four patients analyzed in-depth here, three different sources of somatic mutation in B cells converged upon damaging combinations that allowed a pathogenic autoantibody-producing clone to accumulate in large numbers and cause clinical vasculitis. One source was V(D)J recombination, fixing a particular IGHV/IGKV pair and short $CDR3_H$ length conferring low-affinity binding to self-IgG but not to HCV E2. The second somatic contributor was genome-wide accumulation of indel, SV, and trisomy mutations, one or two of which conferred a growth or survival advantage on the self-reactive B cells, enabling their accumulation. The third pathogenic category was V(D)J hypermutation, which created clonal trajectories to produce insoluble cryoglobulin complexes of autoantibody and autoantigen.

Out of hundreds of different pairs of IGHV and IGKV frequently produced by V(D)J recombination, the pathogenic cryoglobulins in the four patients were specified by three that occur in >95% of HCV-induced cryoglobulins and monoclonal RFs.[24–26] Antigenic mimicry between HCV and IgG could hypothetically explain this bias, since the IGHV1-69/IGKV3-20 pair found in 60% of cryoglobulins is frequent in HCV-neutralizing IgG antibodies. This hypothesis was not supported here since the IgM ancestor in each patient bound self-IgG but not E2, and IGHV1-69/IGKV3-20 pairs that bind HCV have a longer $CDR3_H$ than those binding IgG. Crystallographic studies of HCV-neutralizing IGHV1-69 antibodies bound to E2 show 40%–60% of buried surface area derives from $CDR3_H$.[19,22,23] By contrast, crystallography of an IgG-bound IGHV1-69/IGKV3-20 autoantibody showed $CDR3_H$ contributes only 10% of buried surface area and proposed that longer $CDR3_H$ loops may prevent binding to IgG because they could no longer fit within the IgG C_H2-C_H3 cleft and would fill the binding pocket for a protruding IgG C_H3 Leu432-His435 loop.[83] The higher frequency of IGHV1-69/IGKV3-20 RFs may simply reflect the higher frequency of this pair in the human B cell repertoire (5%) compared with the frequency of IGHV3-7/IGKV3-15 and IGHV4-59/IGKV3-15 pairs (1% and 2%, respectively).[84]

Among thousands of somatic SNV, indel, and SV mutations acquired genome-wide in each patient’s clone, only one or two corresponded to putative driver mutations found recurrently in human B cell lymphomas and leukemias. Two complementary

Sjogren’s syndrome-cryovas patient (MN) or a mutated IGHV1-69/IGKV3-20 broadly neutralizing HCV E2 antibody (AR3A). The membrane IgM fluorescence intensity threshold for IgG-dextran binding determined by visual inspection is shown by the dotted black arrow.

(B) Comparison of IgG-dextran binding by cells expressing V(D)J mutated IgM from P1 to P4 with cells expressing membrane IgM corresponding to two clonally related IGHV1-69/IGKV3-20 RF autoantibodies with known affinities (TJ4 and TJ5). The IgG-dextran-binding threshold determined by visual inspection is shown by the dotted black arrow.

(C) Mathematical modeling of absorption curves (e.g., fluorescence intensity, represented by θ) and surface receptor density (σ_r) for different bond strengths (χ) (or affinities) for a fixed number of ligands (N_L). The model shows the effect of increasing (or decreasing) bond strength (e.g., $\chi = 0.1$; red, $\chi = 1.0$; blue, and $\chi = 10$; black), causing the binding curve to rigidly shift to the left (or right). This relationship can be described by Equation 1 (see STAR Methods).

(D) Estimated K_D for IgG of each patient’s ancestor and mutated antibodies from 3 independent experiments (circles), including the representative experiment shown in (B), which used RF antibody TJ4 with known K_D as the reference antibody, Equation 1, and threshold IgM fluorescence intensities determined by visual inspection (see STAR Methods). RF antibodies TJ5 and TJ0 with known K_D were included as validation antibodies. Only one data point for TJ0 and two data points are shown for P4 ancestor, as sufficient membrane IgM- for IgG-dextran binding was only achieved in 1 and 2 out of 3 experiments, respectively (see Figure S6B; Table S6). Bar columns represent the mean estimated K_D for each antibody.

bodies of evidence support the conclusion that these mutations provided a growth or survival advantage explaining the large size of these clones. First, the loss-of-function mutations in *KLF2*, *TRAF3*, and *BIRC3* and gain-of-function mutations in *NOTCH1* and chromosome 12 occur frequently in neoplastic B cell clones and in their asymptomatic precursor, MBL, but were not observed in normal memory B cells.[50] Second, the mutated genes regulated key biochemical events in B cell growth and survival. The 11q22-23 deletion results in dual loss of *BIRC3* and *BIRC2*, while the *TRAF3* indel mutation creates a loss of function. TRAF3, BIRC3, and BIRC2 are critical negative regulators of the NF-κB signaling pathway that mediates the response to B cell growth and survival factors CD40L, B-cell activating factor (BAFF), or Toll-like receptor (TLR) ligands.[64,68,85] The *KLF2* mutation diminished its activity as a negative regulator of NF-κB.[86–88] The gain-of-function *NOTCH1* mutation prevents degradation of an intracellular NOTCH fragment that promotes expression of B cell proliferation genes, cyclin D and *MYC*.[89,90] Trisomy 12 increased expression of chromosome 12 genes in the clonal B cells. Among ch12 genes are drivers of cell cycle (*CCND2*, *CDKN1B*, *CDK2*, and *CDK4*), B cell lymphomagenesis (*BCL7A*, *P2RX7*, *BTG1*, *KRAS*, *KMT2D*, and *STAT6*), and B cell signaling (*PIK3C2G* and *IRAK4*).

Acquisition of lymphoma driver mutations in the autoimmune clones provides an explanation for the 35-fold higher incidence of B cell lymphoma in HCV patients with cryovas compared with the 2.4-fold increased incidence in HCV patients overall.[36,37] In ∼70% of individuals with HCV-cryovas, the IgM autoantibody is monoclonal and associated with higher serum cryoglobulin concentrations, indicating a clonal advantage.[7,11] Based on the acquisition of a lymphoma driver mutation in each of the four monoclonal HCV-cryovas cases studied here, and the absence of driver mutations found in normal memory B cells,[50] people with monoclonal HCV-cryovas have a B cell clone that has already taken a key step along the mutation pathway to malignant lymphoma. Current international guidelines do not provide specific advice on monitoring HCV-cryovas patients for lymphoproliferative disease after DAA treatment.[91] Our findings highlight the need for longitudinal studies to test if patients remain at elevated risk of B cell lymphoma after viral clearance.

The findings here inform a specific hypothesis to explain why HCV so frequently creates a perfect mutation storm in a single B cell. Plasma IgG is effectively monovalent for binding to "public idiotype" monoclonal RFs.[92–94] While the plasma IgG concentration (10^{-4} M) is above the K_D of these antibodies, it may not crosslink membrane IgM receptors to trigger B cell tolerance checkpoints. Transgenic mouse experiments tracing B cells displaying membrane IgM with low affinity for self-IgG (AM14, $K_D = 2 \times 10^{-6}$ M) have shown circulating plasma IgG does not trigger B cell tolerance checkpoints in these cells.[95] However, when a small amount of anti-DNA or anti-RNA IgG is multimerized on nucleic acid-containing apoptotic cell nanoparticles, it acquires the capacity to bind stably to the AM14 IgM-expressing B cells measured by flow cytometry and to stimulate their proliferation, V(D)J hypermutation, and RF secretion.[96–98] Similarly, we showed here that multimerization of anti-HCV IgG on E2 dextran conferred binding to cells expressing membrane IgM corresponding to public idiotype RFs. Our data support a model whereby IgG-coated virus forms multivalent Fc particles that trigger the recruitment of low-affinity, self-reactive B cells into the immune response.

In chronic HCV infection, large amounts of viral RNA circulate in plasma primarily within IgG-bound 35–65 nm particles.[99–101] Cryoglobulinemia is also a common feature of other chronic viral infections in humans, albeit with a lower prevalence, for example, in chronic HIV infection (∼20%)[102] and chronic HBV infection (<5%).[103,104] One explanation for the differences in cryoglobulinemia prevalence may be plasma viral loads, with higher viral loads providing a greater antigenic stimulus in the form of multivalent IgG-coated virus. Average viral loads for patients with chronic infection are 10^6 copies/mL in HCV infection,[105] 10^4–10^5 copies/mL in HIV infection,[106] and $10^{3.7}$ copies/mL in chronic HBV infection.[107] Over time, circulating IgG-virus nanoparticles may activate naive B cells emerging from V(D)J recombination with public idiotype RFs, promoting each into proliferation and further somatic mutagenesis. The result is a catastrophic confluence where the normal processes of genome-wide mutagenesis, V(D)J recombination, and V(D)J hypermutation converge by chance on rare combinations that produce an autoantibody in sufficient quantity and quality to trigger clinical vasculitis in 5%–15% of infected individuals.

Given that each of the sources of somatic mutation is common to all memory B cells, it is conceivable that similar processes contribute to the pathogenesis of other autoimmune diseases where the environmental trigger is less understood.

Limitations of the study

Future studies of larger numbers of B cell clones will be needed to determine the prevalence of the findings here for HCV-cryovas and to determine to what extent the somatic variant burden in the pathogenic clones from HCV-cryovas patients differs from those in normal memory B cells, malignant B cells, or in self-reactive memory B cell clones in other autoimmune diseases. Second, we have not studied HCV-infected patients without cryoglobulinemia or HCV-infected patients with cryoglobulinemia but without vasculitis, so it is unclear at what stage of HCV infection these somatic mutation events first arise. While our data focus on the molecular events in B cells that contribute to disease pathogenesis in four patients with clinical vasculitis, our data do not explain what other viral and host factors help determine the subset of HCV-infected patients that develop cryovas. Third, we have not directly demonstrated that the individual driver mutations present in each patient B cell clone provide a growth or survival advantage *in vivo*. Fourth, while we did not detect binding of ancestor antibodies to the HCV E2 protein, this does not exclude the possibility of very low-affinity interaction with E2 below the sensitivity of the multimerized dextran flow cytometry assay, nor with other HCV proteins.

RESOURCE AVAILABILITY

Lead contact

Requests for resources and reagents should be directed to Chris Goodnow (c.goodnow@garvan.org.au).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- Single-cell RNA-seq (10×) and WGS data deposited at NCBI GEO and NCBI with data accession numbers listed in the key resources table. Antibody sequences are in Table S5. Bulk immunoglobulin sequencing, single-cell RNA-seq (Smart-seq2), and RAGE-seq V(D)J data will be shared upon request. This paper analyzes existing, publicly available data with accession numbers listed in the key resources table.
- Original code has been deposited on Github with DOI listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Daniel Lingwood for membrane IgM assay advice, Heather Machado for advice on somatic mutations in memory B cells, and Daniele Cultrone for assistance with the NF-κB luciferase assay. This work was supported by the Bill and Patricia Ritchie Foundation, the Croall Foundation, the National Health and Medical Research Council, Australia grants APP2010134 and APP1113904, and the UNSW Cellular Genomics Futures Institute.

AUTHOR CONTRIBUTIONS

Conceptualization: C.Y., D.S., and C.C.G.; resources: M.D., A.C., A.D.K., G.J.D., G.M., R.A.B., and D.S.; methodology: C.Y., M.S., K.J.L.J., M.A.F., S.A.-U., D.F., G.A.-E., J.H.R., D.S., and C.C.G.; investigation: C.Y., M.S., J.J.W., D.A., and M.L.F.; formal analysis: C.Y., K.J.L.J., T.J.P., M.A.F., S.A.-U., S.R., M.G., and C.C.G.; visualization: C.Y., M.S., K.J.L.J., T.J.P., S.A.-U., and C.C.G.; software: S.A.-U.; data curation: K.J.L.J., T.J.P., M.A.F., and S.A.-U.; supervision: F.L., R.B., D.S., and C.C.G.; funding acquisition: D.S. and C.C.G.; writing – original draft: C.Y., D.S., and C.C.G.; writing – review & editing: C.Y., S.A.-U., A.C., G.J.D., D.S., and C.C.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient samples
- METHOD DETAILS
  - Flow cytometry
  - Fluorescence activated cell sorting (FACS)
  - Bulk immunoglobulin sequencing
  - Single cell RNA sequencing
  - Single cell immunoglobulin analysis and clonal tree generation
  - Single cell RNA-seq gene expression analysis
  - Whole genome sequencing
  - SNV mutational signatures
  - Small (<50bp) somatic variant (SNV and indel) analysis
  - Large (>50bp) structural somatic variant analysis
  - Somatic ploidy analysis
  - Somatic variant validation
  - NF-κB luciferase assay
  - Soluble IgM antibody expression
  - Soluble IgM quantification by ELISA
  - IgG (RF) ELISA
  - Cryoglobulin assay
  - RF peptide sequencing and data analysis
  - Recombinant HCV E2 (including biotinylated)
  - Surface IgM expression and antigen staining
  - IGHV1-69/IGKV3-20 anti-HCV E2 antibodies
  - IGHV1-69 CDR3 amino acid lengths
  - Estimating K_D using the IgM onset of binding
  - Modelling of multivalent antigen binding to membrane IgM
  - Fitting procedure for calculating the IgM binding threshold
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2024.12.011.

Received: January 30, 2024
Revised: September 22, 2024
Accepted: December 18, 2024
Published: January 15, 2025

REFERENCES

1. Polaris Observatory HCV Collaborators (2022). Global change in hepatitis C virus prevalence and cascade of care between 2015 and 2020: a modelling study. Lancet Gastroenterol. Hepatol. *7*, 396–415. https://doi.org/10.1016/S2468-1253(21)00472-6.
2. Dore, G.J., Martinello, M., Alavi, M., and Grebely, J. (2020). Global elimination of hepatitis C virus by 2030: why not? Nat. Med. *26*, 157–160. https://doi.org/10.1038/s41591-019-0706-x.
3. Meltzer, M., Franklin, E.C., Elias, K., McCluskey, R.T., and Cooper, N. (1966). Cryoglobulinemia—a clinical and laboratory study. II. Cryoglobulins with rheumatoid factor activity. Am. J. Med. *40*, 837–856. https://doi.org/10.1016/0002-9343(66)90200-2.
4. Desbois, A.C., Cacoub, P., and Saadoun, D. (2019). Cryoglobulinemia: An update in 2019. Joint Bone Spine *86*, 707–713. https://doi.org/10.1016/j.jbspin.2019.01.016.
5. Cacoub, P., and Saadoun, D. (2021). Extrahepatic manifestations of chronic HCV infection. N. Engl. J. Med. *384*, 1038–1052. https://doi.org/10.1056/NEJMra2033539.
6. LoSpalluto, J., Dorward, B., Miller, W., and Ziff, M. (1962). Cryoglobulinemia based on interaction between a gamma macroglobulin and 7S gamma globulin. Am. J. Med. *32*, 142–147. https://doi.org/10.1016/0002-9343(62)90191-2.
7. Cacoub, P., Renou, C., Rosenthal, E., Cohen, P., Loury, I., Loustaud-Ratti, V., Yamamoto, A.-M., Camproux, A.-C., Hausfater, P., Musset, L., et al. (2000). Extrahepatic manifestations associated with hepatitis C virus infection: a prospective multicenter study of 321 patients. Medicine *79*, 47–56.
8. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science *244*, 359–362. https://doi.org/10.1126/science.2523562.
9. Agnello, V., Chung, R.T., and Kaplan, L.M. (1992). A role for hepatitis C virus infection in type II cryoglobulinemia. N. Engl. J. Med. *327*, 1490–1495. https://doi.org/10.1056/NEJM199211193272104.
10. Misiani, R., Bellavita, P., Fenili, D., Borelli, G., Marchesi, D., Massazza, M., Vendramin, G., Comotti, B., Tanzi, E., and Scudeller, G. (1992). Hepatitis C virus infection in patients with essential mixed cryoglobulinemia. Ann. Intern. Med. *117*, 573–577. https://doi.org/10.7326/0003-4819-117-7-573.
11. Lunel, F., Musset, L., Cacoub, P., Frangeul, L., Cresta, P., Perrin, M., Grippon, P., Hoang, C., Piette, J.C., Huraux, J.-M., et al. (1994). Cryoglobulinemia in chronic liver diseases: role of hepatitis C virus and liver damage. Gastroenterology *106*, 1291–1300. https://doi.org/10.1016/0016-5085(94)90022-1.
12. Bonacci, M., Lens, S., Londoño, M.-C., Mariño, Z., Cid, M.C., Ramos-Casals, M., Sánchez-Tapias, J.M., Forns, X., and Hernández-Rodríguez, J. (2017). Virologic, clinical, and immune response outcomes of patients with hepatitis C virus–associated cryoglobulinemia treated

with direct-acting antivirals. Clin. Gastroenterol. Hepatol. *15*, 575–583.e1. https://doi.org/10.1016/j.cgh.2016.09.158.

13. Cacoub, P., Si Ahmed, S.N.S., Ferfar, Y., Pol, S., Thabut, D., Hezode, C., Alric, L., Comarmond, C., Ragab, G., Quartuccio, L., et al. (2019). Long-term efficacy of interferon-free antiviral treatment regimens in patients with hepatitis C virus–associated cryoglobulinemia vasculitis. Clin. Gastroenterol. Hepatol. *17*, 518–526. https://doi.org/10.1016/j.cgh.2018.05.021.
14. Gragnani, L., Visentini, M., Fognani, E., Urraro, T., De Santis, A., Petraccia, L., Perez, M., Ceccotti, G., Colantuono, S., Mitrevski, M., et al. (2016). Prospective study of guideline-tailored therapy with direct-acting antivirals for hepatitis C virus-associated mixed cryoglobulinemia. Hepatology *64*, 1473–1482. https://doi.org/10.1002/hep.28753.
15. Saadoun, D., Pol, S., Ferfar, Y., Alric, L., Hezode, C., Ahmed, S.N.S., de Saint Martin, L., Comarmond, C., Bouyer, A.S., and Musset, L. (2017). Efficacy and safety of sofosbuvir plus daclatasvir for treatment of HCV-associated cryoglobulinemia vasculitis. Gastroenterology *153*, 49–52.e5.
16. Sise, M.E., Bloom, A.K., Wisocky, J., Lin, M.V., Gustafson, J.L., Lundquist, A.L., Steele, D., Thiim, M., Williams, W.W., Hashemi, N., et al. (2016). Treatment of hepatitis C virus–associated mixed cryoglobulinemia with direct-acting antiviral agents. Hepatology *63*, 408–417. https://doi.org/10.1002/hep.28297.
17. Law, M., Maruyama, T., Lewis, J., Giang, E., Tarr, A.W., Stamataki, Z., Gastaminza, P., Chisari, F.V., Jones, I.M., Fox, R.I., et al. (2008). Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. Nat. Med. *14*, 25–27. https://doi.org/10.1038/nm1698.
18. Keck, Z.-Y., Li, T.-K., Xia, J., Gal-Tanamy, M., Olson, O., Li, S.H., Patel, A.H., Ball, J.K., Lemon, S.M., and Foung, S.K.H. (2008). Definition of a conserved immunodominant domain on hepatitis C virus E2 glycoprotein by neutralizing human monoclonal antibodies. J. Virol. *82*, 6061–6066. https://doi.org/10.1128/JVI.02475-07.
19. Tzarum, N., Giang, E., Kong, L., He, L., Prentoe, J., Augestad, E., Hua, Y., Castillo, S., Lauer, G.M., Bukh, J., et al. (2019). Genetic and structural insights into broad neutralization of hepatitis C virus by human VH1-69 antibodies. Sci. Adv. *5*, eaav1882. https://doi.org/10.1126/sciadv.aav1882.
20. Colbert, M.D., Flyak, A.I., Ogega, C.O., Kinchen, V.J., Massaccesi, G., Hernandez, M., Davidson, E., Doranz, B.J., Cox, A.L., and Crowe, J.E., Jr. (2019). Broadly neutralizing antibodies targeting new sites of vulnerability in hepatitis C virus E1E2. J. Virol. *93*, e0207018.
21. Chan, C.H., Hadlock, K.G., Foung, S.K., and Levy, S. (2001). VH1-69 gene is preferentially used by hepatitis C virus–associated B cell lymphomas and by normal B cells responding to the E2 viral antigen. Blood *97*, 1023–1026.
22. Flyak, A.I., Ruiz, S., Colbert, M.D., Luong, T., Crowe, J.E., Jr., Bailey, J.R., and Bjorkman, P.J. (2018). HCV broadly neutralizing antibodies use a CDRH3 disulfide motif to recognize an E2 glycoprotein site that can be targeted for vaccine design. Cell Host Microbe *24*, 703–716.e3. https://doi.org/10.1016/j.chom.2018.10.009.
23. Weber, T., Potthoff, J., Bizu, S., Labuhn, M., Dold, L., Schoofs, T., Horning, M., Ercanoglu, M.S., Kreer, C., Gieselmann, L., et al. (2022). Analysis of antibodies from HCV elite neutralizers identifies genetic determinants of broad neutralization. Immunity *55*, 341–354.e7. https://doi.org/10.1016/j.immuni.2021.12.003.
24. Kunkel, H.G., Agnello, V., Joslin, F.G., Winchester, R.J., and Capra, J.D. (1973). Cross-idiotypic specificity among monoclonal IgM proteins with anti-γ-globulin activity. J. Exp. Med. *137*, 331–342. https://doi.org/10.1084/jem.137.2.331.
25. Victor, K.D., Randen, I., Thompson, K., Forre, O., Natvig, J.B., Fu, S.M., and Capra, J.D. (1991). Rheumatoid factors isolated from patients with autoimmune disorders are derived from germline genes distinct from those encoding the Wa, Po, and Bla cross-reactive idiotypes. J. Clin. Invest. *87*, 1603–1613. https://doi.org/10.1172/JCI115174.
26. Andrews, D.W., and Capra, J.D. (1981). Complete amino acid sequence of variable domains from two monoclonal human anti-gamma globulins of the Wa cross-idiotypic group: suggestion that the J segments are involved in the structural correlate of the idiotype. Proc. Natl. Acad. Sci. USA *78*, 3799–3803. https://doi.org/10.1073/pnas.78.6.3799.
27. Carbonari, M., Caprini, E., Tedesco, T., Mazzetta, F., Tocco, V., Casato, M., Russo, G., and Fiorilli, M. (2005). Hepatitis C virus drives the unconstrained monoclonal expansion of VH1–69-expressing memory B cells in type II cryoglobulinemia: a model of infection-driven lymphomagenesis. J. Immunol. *174*, 6532–6539. https://doi.org/10.4049/jimmunol.174.10.6532.
28. Charles, E.D., Brunetti, C., Marukian, S., Ritola, K.D., Talal, A.H., Marks, K., Jacobson, I.M., Rice, C.M., and Dustin, L.B. (2011). Clonal B cells in patients with hepatitis C virus-associated mixed cryoglobulinemia contain an expanded anergic CD21low B-cell subset. Blood *117*, 5425–5437. https://doi.org/10.1182/blood-2010-10-312942.
29. Charles, E.D., Green, R.M., Marukian, S., Talal, A.H., Lake-Bakaar, G.V., Jacobson, I.M., Rice, C.M., and Dustin, L.B. (2008). Clonal expansion of immunoglobulin M+ CD27+ B cells in HCV-associated mixed cryoglobulinemia. Blood *111*, 1344–1356.
30. Franzin, F., Efremov, D.G., Pozzato, G., Tulissi, P., Batista, F., and Burrone, O.R. (1995). Clonal B-cell expansions in peripheral blood of HCV-infected patients. Br. J. Haematol. *90*, 548–552. https://doi.org/10.1111/j.1365-2141.1995.tb05582.x.
31. Terrier, B., Joly, F., Vazquez, T., Benech, P., Rosenzwajg, M., Carpentier, W., Garrido, M., Ghillani-Dalbin, P., Klatzmann, D., Cacoub, P., et al. (2011). Expansion of functionally anergic CD21−/low marginal zone-like B cell clones in hepatitis C virus infection-related autoimmunity. J. Immunol. *187*, 6550–6563. https://doi.org/10.4049/jimmunol.1102022.
32. Visentini, M., Cagliuso, M., Conti, V., Carbonari, M., Casato, M., and Fiorilli, M. (2011). The VH1-69–expressing marginal zone B cells expanded in HCV-associated mixed cryoglobulinemia display proliferative anergy irrespective of CD21low phenotype. Blood *118*, 3440–3441. https://doi.org/10.1182/blood-2011-05-353821.
33. Charles, E.D., Orloff, M.I.M., Nishiuchi, E., Marukian, S., Rice, C.M., and Dustin, L.B. (2013). Somatic hypermutations confer rheumatoid factor activity in hepatitis C virus–associated mixed cryoglobulinemia. Arthritis Rheum. *65*, 2430–2440. https://doi.org/10.1002/art.38041.
34. Comarmond, C., Lorin, V., Marques, C., Maciejewski-Duval, A., Joher, N., Planchais, C., Touzot, M., Biard, L., Hieu, T., Quiniou, V., et al. (2019). TLR9 signalling in HCV-associated atypical memory B cells triggers Th1 and rheumatoid factor autoantibody responses. J. Hepatol. *71*, 908–919. https://doi.org/10.1016/j.jhep.2019.06.029.
35. Ferri, C., Cacoub, P., Mazzaro, C., Roccatello, D., Scaini, P., Sebastiani, M., Tavoni, A., Zignego, A.L., and De Vita, S. (2011). Treatment with rituximab in patients with mixed cryoglobulinemia syndrome: results of multicenter cohort study and review of the literature. Autoimmun. Rev. *11*, 48–55. https://doi.org/10.1016/j.autrev.2011.07.005.
36. Monti, G., Pioltelli, P., Saccardo, F., Campanini, M., Candela, M., Cavallero, G., De Vita, S., Ferri, C., Mazzaro, C., Migliaresi, S., et al. (2005). Incidence and characteristics of non-Hodgkin lymphomas in a multicenter case file of patients with hepatitis C virus–related symptomatic mixed cryoglobulinemias. Arch. Intern. Med. *165*, 101–105. https://doi.org/10.1001/archinte.165.1.101.
37. Pozzato, G., Mazzaro, C., Dal Maso, L., Mauro, E., Zorat, F., Moratelli, G., Bulian, P., Serraino, D., and Gattei, V. (2016). Hepatitis C virus and non-Hodgkin's lymphomas: meta-analysis of epidemiology data and therapy options. World J. Hepatol. *8*, 107–116. https://doi.org/10.4254/wjh.v8.i2.107.
38. Ivanovski, M., Silvestri, F., Pozzato, G., Anand, S., Mazzaro, C., Burrone, O.R., and Efremov, D.G. (1998). Somatic hypermutation, clonal diversity, and preferential expression of the VH 51p1/VL kv325 immunoglobulin gene combination in hepatitis C virus–associated immunocytomas. Blood *91*, 2433–2442.
39. De Re, V., De Vita, S., Marzotto, A., Rupolo, M., Gloghini, A., Pivetta, B., Gasparotto, D., Carbone, A., and Boiocchi, M. (2000). Sequence analysis of the immunoglobulin antigen receptor of hepatitis C virus–associated

non-Hodgkin lymphomas suggests that the malignant cells are derived from the rheumatoid factor–producing cells that occur mainly in type II cryoglobulinemia. Blood *96*, 3578–3584.

40. Burnet, F.M. (1972). A reassessment of the forbidden clone hypothesis of autoimmune disease. Aust. J. Exp. Biol. Med. Sci. *50*, 1–9. https://doi.org/10.1038/icb.1972.1.
41. Dameshek, W., and Schwartz, R.S. (1959). Leukemia and auto-immunization—some possible relationships. Blood *14*, 1151–1158. https://doi.org/10.1182/blood.V14.10.1151.1151.
42. Goodnow, C.C. (2007). Multistep pathogenesis of autoimmune disease. Cell *130*, 25–35. https://doi.org/10.1016/j.cell.2007.06.033.
43. Goodnow, C.C., Cyster, J.G., Hartley, S.B., Bell, S.E., Cooke, M.P., Healy, J.I., Akkaraju, S., Rathmell, J.C., Pogue, S.L., and Shokat, K.P. (1995). Self-tolerance checkpoints in B lymphocyte development. Adv. Immunol. *59*, 279–368. https://doi.org/10.1016/s0065-2776(08)60633-1.
44. Young, R.M., Phelan, J.D., Wilson, W.H., and Staudt, L.M. (2019). Pathogenic B-cell receptor signaling in lymphoid malignancies: new insights to improve treatment. Immunol. Rev. *291*, 190–213. https://doi.org/10.1111/imr.12792.
45. Singh, M., Jackson, K.J.L., Wang, J.J., Schofield, P., Field, M.A., Koppstein, D., Peters, T.J., Burnett, D.L., Rizzetto, S., Nevoltris, D., et al. (2020). Lymphoma driver mutations in the pathogenic evolution of an iconic human autoantibody. Cell *180*, 878–894.e19. https://doi.org/10.1016/j.cell.2020.01.029.
46. Potter, K.N., Li, Y., Mageed, R.A., Jefferis, R., and Capra, J.D. (1999). Molecular characterization of the VH1-specific variable region determinants recognized by anti-idiotypic monoclonal antibodies G6 and G8. Scand. J. Immunol. *50*, 14–20. https://doi.org/10.1046/j.1365-3083.1999.00524.x.
47. Singh, M., Al-Eryani, G., Carswell, S., Ferguson, J.M., Blackburn, J., Barton, K., Roden, D., Luciani, F., Giang Phan, T., and Junankar, S. (2019). High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat. Commun. *10*, 3120.
48. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.
49. Kasar, S., Kim, J., Improgo, R., Tiao, G., Polak, P., Haradhvala, N., Lawrence, M.S., Kiezun, A., Fernandes, S.M., Bahl, S., et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat. Commun. *6*, 8866. https://doi.org/10.1038/ncomms9866.
50. Machado, H.E., Mitchell, E., Øbro, N.F., Kübler, K., Davies, M., Leongamornlert, D., Cull, A., Maura, F., Sanders, M.A., Cagan, A.T.J., et al. (2022). Diverse mutational landscapes in human lymphocytes. Nature *608*, 724–732. https://doi.org/10.1038/s41586-022-05072-7.
51. Zhang, L., Dong, X., Lee, M., Maslov, A.Y., Wang, T., and Vijg, J. (2019). Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. Proc. Natl. Acad. Sci. USA *116*, 9014–9019. https://doi.org/10.1073/pnas.1902510116.
52. Piva, R., Deaglio, S., Famà, R., Buonincontri, R., Scarfò, I., Bruscaggin, A., Mereu, E., Serra, S., Spina, V., Brusa, D., et al. (2015). The Krüppel-like factor 2 transcription factor gene is recurrently mutated in splenic marginal zone lymphoma. Leukemia *29*, 503–507. https://doi.org/10.1038/leu.2014.294.
53. Clipson, A., Wang, M., De Leval, L., Ashton-Key, M., Wotherspoon, A., Vassiliou, G., Bolli, N., Grove, C., Moody, S., Escudero-Ibarz, L., et al. (2015). KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. Leukemia *29*, 1177–1185. https://doi.org/10.1038/leu.2014.330.
54. Ren, W., Ye, X., Su, H., Li, W., Liu, D., Pirmoradian, M., Wang, X., Zhang, B., Zhang, Q., and Chen, L. (2018). Genetic landscape of hepatitis B virus–associated diffuse large B-cell lymphoma. Blood *131*, 2670–2681.
55. Jaramillo Oquendo, C., Parker, H., Oscier, D., Ennis, S., Gibson, J., and Strefford, J.C. (2019). Systematic review of somatic mutations in splenic marginal zone lymphoma. Sci. Rep. *9*, 10444. https://doi.org/10.1038/s41598-019-46906-1.
56. Barrio, S., Shanafelt, T.D., Ojha, J., Chaffee, K.G., Secreto, C., Kortüm, K.M., Pathangey, S., Van-Dyke, D.L., Slager, S.L., Fonseca, R., et al. (2017). Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome. Leukemia *31*, 170–176. https://doi.org/10.1038/leu.2016.172.
57. Fabbri, G., Rasi, S., Rossi, D., Trifonov, V., Khiabanian, H., Ma, J., Grunn, A., Fangazio, M., Capello, D., Monti, S., et al. (2011). Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J. Exp. Med. *208*, 1389–1401. https://doi.org/10.1084/jem.20110921.
58. Landau, D.A., Tausch, E., Taylor-Weiner, A.N., Stewart, C., Reiter, J.G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. Nature *526*, 525–530. https://doi.org/10.1038/nature15395.
59. Nadeu, F., Delgado, J., Royo, C., Baumann, T., Stankovic, T., Pinyol, M., Jares, P., Navarro, A., Martín-García, D., and Beà, S. (2016). Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. Blood *127*, 2122–2130.
60. Puente, X.S., Pinyol, M., Quesada, V., Conde, L., Ordóñez, G.R., Villamor, N., Escaramis, G., Jares, P., Beà, S., González-Díaz, M., et al. (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature *475*, 101–105. https://doi.org/10.1038/nature10113.
61. Rosati, E., Sabatini, R., Rampino, G., Tabilio, A., Di Ianni, M., Fettucciari, K., Bartoli, A., Coaccioli, S., Screpanti, I., and Marconi, P. (2009). Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. Blood *113*, 856–865. https://doi.org/10.1182/blood-2008-02-139725.
62. Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., and Aster, J.C. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science *306*, 269–271. https://doi.org/10.1126/science.1102160.
63. Ni, C.Z., Welsh, K., Leo, E., Chiou, C.K., Wu, H., Reed, J.C., and Ely, K.R. (2000). Molecular basis for CD40 signaling mediated by TRAF3. Proc. Natl. Acad. Sci. USA *97*, 10395–10399. https://doi.org/10.1073/pnas.97.19.10395.
64. Liao, G., Zhang, M., Harhaj, E.W., and Sun, S.-C. (2004). Regulation of the NF-κB-inducing kinase by tumor necrosis factor receptor-associated factor 3-induced degradation. J. Biol. Chem. *279*, 26243–26250. https://doi.org/10.1074/jbc.M403286200.
65. He, J.Q., Saha, S.K., Kang, J.R., Zarnegar, B., and Cheng, G. (2007). Specificity of TRAF3 in its negative regulation of the noncanonical NF-κB pathway. J. Biol. Chem. *282*, 3688–3694. https://doi.org/10.1074/jbc.M610271200.
66. Annunziata, C.M., Davis, R.E., Demchenko, Y., Bellamy, W., Gabrea, A., Zhan, F., Lenz, G., Hanamura, I., Wright, G., Xiao, W., et al. (2007). Frequent engagement of the classical and alternative NF-κB pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell *12*, 115–130. https://doi.org/10.1016/j.ccr.2007.07.004.
67. Bushell, K.R., Kim, Y., Chan, F.C., Ben-Neriah, S., Jenks, A., Alcaide, M., Fornika, D., Grande, B.M., Arthur, S., and Gascoyne, R.D. (2015). Genetic inactivation of TRAF3 in canine and human B-cell lymphoma. Blood. J. Am. Soc. Hematol. *125*, 999–1005.
68. Keats, J.J., Fonseca, R., Chesi, M., Schop, R., Baker, A., Chng, W.-J., Van Wier, S., Tiedemann, R., Shi, C.-X., Sebag, M., et al. (2007). Promiscuous mutations activate the noncanonical NF-κB pathway in multiple myeloma. Cancer Cell *12*, 131–144. https://doi.org/10.1016/j.ccr.2007.07.003.

69. Rossi, D., Deaglio, S., Dominguez-Sola, D., Rasi, S., Vaisitti, T., Agostinelli, C., Spina, V., Bruscaggin, A., Monti, S., Cerri, M., et al. (2011). Alteration of BIRC3 and multiple other NF-κB pathway genes in splenic marginal zone lymphoma. Blood *118*, 4930–4934. https://doi.org/10.1182/blood-2011-06-359166.

70. Visentini, M., Del Padre, M., Colantuono, S., Yang, B., Minafò, Y.A., Antonini, S., Carnovale, M., De Santis, A., Pulsoni, A., De Sanctis, G.M., et al. (2019). Long-lasting persistence of large B-cell clones in hepatitis C virus-cured patients with complete response of mixed cryoglobulinaemia vasculitis. Liver Int. *39*, 628–632. https://doi.org/10.1111/liv.14053.

71. Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., Döhner, K., Bentz, M., and Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. N. Engl. J. Med. *343*, 1910–1916. https://doi.org/10.1056/NEJM200012283432602.

72. Raponi, S., Del Giudice, I., Ilari, C., Cafforio, L., Messina, M., Cappelli, L.V., Bonina, S., Piciocchi, A., Marinelli, M., and Peragine, N. (2019). Biallelic BIRC3 inactivation in chronic lymphocytic leukaemia patients with 11q deletion identifies a subgroup with very aggressive disease. Br. J. Haematol. *185*, 156–159.

73. Rawstron, A.C., Bennett, F.L., O'Connor, S.J.M., Kwok, M., Fenton, J.A.L., Plummer, M., de Tute, R., Owen, R.G., Richards, S.J., Jack, A.S., et al. (2008). Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N. Engl. J. Med. *359*, 575–583. https://doi.org/10.1056/NEJMoa075290.

74. Haslinger, C., Schweifer, N., Stilgenbauer, S., Döhner, H., Lichter, P., Kraut, N., Stratowa, C., and Abseher, R. (2004). Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. J. Clin. Oncol. *22*, 3937–3949. https://doi.org/10.1200/JCO.2004.12.133.

75. Abruzzo, L.V., Herling, C.D., Calin, G.A., Oakes, C., Barron, L.L., Banks, H.E., Katju, V., Keating, M.J., and Coombes, K.R. (2018). Trisomy 12 chronic lymphocytic leukemia expresses a unique set of activated and targetable pathways. Haematologica *103*, 2069–2078. https://doi.org/10.3324/haematol.2018.190132.

76. Weaver, G.C., Villar, R.F., Kanekiyo, M., Nabel, G.J., Mascola, J.R., and Lingwood, D. (2016). In vitro reconstitution of B cell receptor–antigen interactions to evaluate potential vaccine candidates. Nat. Protoc. *11*, 193–213. https://doi.org/10.1038/nprot.2016.009.

77. Walker, M.R., Li, H., Teutsch, S., Betz-Stablein, B., Luciani, F., Lloyd, A.R., and Bull, R.A. (2016). Incident hepatitis C virus genotype distribution and multiple infection in Australian prisons. J. Clin. Microbiol. *54*, 1855–1861. https://doi.org/10.1128/JCM.00287-16.

78. Martinez-Veracoechea, F.J., and Frenkel, D. (2011). Designing super selectivity in multivalent nano-particle binding. Proc. Natl. Acad. Sci. USA *108*, 10963–10968. https://doi.org/10.1073/pnas.1105351108.

79. Varilly, P., Angioletti-Uberti, S., Mognetti, B.M., and Frenkel, D. (2012). A general theory of DNA-mediated and other valence-limited colloidal interactions. J. Chem. Phys. *137*, 094108. https://doi.org/10.1063/1.4748100.

80. Angioletti-Uberti, S., Varilly, P., Mognetti, B.M., Tkachenko, A.V., and Frenkel, D. (2013). Communication: A simple analytical formula for the free energy of ligand–receptor-mediated interactions. J. Chem. Phys. *138*, 021102. https://doi.org/10.1063/1.4775806.

81. Dubacheva, G.V., Curk, T., Mognetti, B.M., Auzély-Velty, R., Frenkel, D., and Richter, R.P. (2014). Superselective targeting using multivalent polymers. J. Am. Chem. Soc. *136*, 1722–1725. https://doi.org/10.1021/ja411138s.

82. Tian, X., Angioletti-Uberti, S., and Battaglia, G. (2020). On the design of precision nanomedicines. Sci. Adv. *6*, eaat0919. https://doi.org/10.1126/sciadv.aat0919.

83. Shiroishi, M., Ito, Y., Shimokawa, K., Lee, J.M., Kusakabe, T., and Ueda, T. (2018). Structure–function analyses of a stereotypic rheumatoid factor unravel the structural basis for germline-encoded antibody autoreactivity. J. Biol. Chem. *293*, 7008–7016. https://doi.org/10.1074/jbc.M117.814475.

84. DeKosky, B.J., Kojima, T., Rodin, A., Charab, W., Ippolito, G.C., Ellington, A.D., and Georgiou, G. (2015). In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat. Med. *21*, 86–91. https://doi.org/10.1038/nm.3743.

85. Gardam, S., Turner, V.M., Anderton, H., Limaye, S., Basten, A., Koentgen, F., Vaux, D.L., Silke, J., and Brink, R. (2011). Deletion of cIAP1 and cIAP2 in murine B lymphocytes constitutively activates cell survival pathways and inactivates the germinal center response. Blood *117*, 4041–4051. https://doi.org/10.1182/blood-2010-10-312793.

86. Hart, G.T., Wang, X., Hogquist, K.A., and Jameson, S.C. (2011). Krüppel-like factor 2 (KLF2) regulates B-cell reactivity, subset differentiation, and trafficking molecule expression. Proc. Natl. Acad. Sci. USA *108*, 716–721. https://doi.org/10.1073/pnas.1013168108.

87. Winkelmann, R., Sandrock, L., Kirberg, J., Jäck, H.-M., and Schuh, W. (2014). KLF2–a negative regulator of pre-B cell clonal expansion and B cell activation. PLoS One *9*, e97953. https://doi.org/10.1371/journal.pone.0097953.

88. Winkelmann, R., Sandrock, L., Porstner, M., Roth, E., Mathews, M., Hobeika, E., Reth, M., Kahn, M.L., Schuh, W., and Jäck, H.-M. (2011). B cell homeostasis and plasma cell homing controlled by Krüppel-like factor 2. Proc. Natl. Acad. Sci. USA *108*, 710–715. https://doi.org/10.1073/pnas.1012858108.

89. Kohlhof, H., Hampel, F., Hoffmann, R., Burtscher, H., Weidle, U.H., Hölzel, M., Eick, D., Zimber-Strobl, U., and Strobl, L.J. (2009). Notch1, Notch2, and Epstein-Barr virus–encoded nuclear antigen 2 signaling differentially affects proliferation and survival of Epstein-Barr virus–infected B cells. Blood *113*, 5506–5515.

90. Fabbri, G., Holmes, A.B., Viganotti, M., Scuoppo, C., Belver, L., Herranz, D., Yan, X.-J., Kieso, Y., Rossi, D., Gaidano, G., et al. (2017). Common nonmutational NOTCH1 activation in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. USA *114*, E2911–E2919. https://doi.org/10.1073/pnas.1702564114.

91. World Health Organization. (2022). Updated recommendations on Treatment of adolescents and children with chronic HCV infection, and HCV simplified service delivery and diagnostics. https://www.who.int/publications/i/item/9789240052734.

92. Stone, M.J., and Fedak, J.E. (1974). Studies on monoclonal antibodies. II. Immune complex (IgM-IgG) cryoglobulinemia: the mechanism of cryoprecipitation. J. Immunol. *113*, 1377–1385. https://doi.org/10.4049/jimmunol.113.4.1377.

93. Stone, M.J., and Metzger, H. (1968). Binding properties of a Waldenström macroglobulin antibody. J. Biol. Chem. *243*, 5977–5984. https://doi.org/10.1016/S0021-9258(18)94516-3.

94. Brandau, D.T., Trautman, P.A., Steadman, B.L., Lawson, E.Q., and Middaugh, C.R. (1986). The temperature-dependent stoichiometry of mixed cryoimmunoglobulins. J. Biol. Chem. *261*, 16385–16391. https://doi.org/10.1016/S0021-9258(18)66577-9.

95. Hannum, L.G., Ni, D., Haberman, A.M., Weigert, M.G., and Shlomchik, M.J. (1996). A disease-related rheumatoid factor autoantibody is not tolerized in a normal mouse: implications for the origins of autoantibodies in autoimmune disease. J. Exp. Med. *184*, 1269–1278. https://doi.org/10.1084/jem.184.4.1269.

96. Leadbetter, E.A., Rifkin, I.R., Hohlbaum, A.M., Beaudette, B.C., Shlomchik, M.J., and Marshak-Rothstein, A. (2002). Chromatin–IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. Nature *416*, 603–607. https://doi.org/10.1038/416603a.

97. Herlands, R.A., Christensen, S.R., Sweet, R.A., Hershberg, U., and Shlomchik, M.J. (2008). T cell-independent and toll-like receptor-dependent antigen-driven activation of autoreactive B cells. Immunity *29*, 249–260. https://doi.org/10.1016/j.immuni.2008.06.009.

98. Viglianti, G.A., Lau, C.M., Hanley, T.M., Miko, B.A., Shlomchik, M.J., and Marshak-Rothstein, A. (2003). Activation of autoreactive B cells by CpG dsDNA. Immunity *19*, 837–847. https://doi.org/10.1016/s1074-7613(03)00323-6.

99. Hijikata, M., Shimizu, Y.K., Kato, H., Iwamoto, A., Shih, J.W., Alter, H.J., Purcell, R.H., and Yoshikura, H. (1993). Equilibrium centrifugation studies of hepatitis C virus: evidence for circulating immune complexes. J. Virol. *67*, 1953–1958. https://doi.org/10.1128/JVI.67.4.1953-1958.1993.

100. Kaito, M., Gabazza, E.C., Fujita, N., Tanaka, H., Watanabe, S., and Kohara, M. (2006). Immune complex of hepatitis C virus particles detected by immunogold electron microscopy. J. Gastroenterol. *41*, 807–808. https://doi.org/10.1007/s00535-006-1840-z.

101. Riva, E., Maggi, F., Abbruzzese, F., Bellomi, F., Giannelli, G., Picardi, A., Scagnolari, C., Folgori, A., Spada, E., Piccolella, E., et al. (2009). Immune complexed (IC) hepatitis C virus (HCV) in chronically and acutely HCV-infected patients. Med. Microbiol. Immunol. *198*, 13–18. https://doi.org/10.1007/s00430-008-0099-9.

102. Dimitrakopoulos, A.N., Kordossis, T., Hatzakis, A., and Moutsopoulos, H.M. (1999). Mixed cryoglobulinemia in HIV-1 infection: the role of HIV-1. Ann. Intern. Med. *130*, 226–230. https://doi.org/10.7326/0003-4819-130-3-199902020-00027.

103. Ferri, C., Sebastiani, M., Giuggioli, D., Cazzato, M., Longombardo, G., Antonelli, A., Puccini, R., Michelassi, C., and Zignego, A.L. (2004). Mixed cryoglobulinemia: demographic, clinical, and serologic features and survival in 231 patients. Semin. Arthritis Rheum. *33*, 355–374. https://doi.org/10.1016/j.semarthrit.2003.10.001.

104. Mazzaro, C., Dal Maso, L., Mauro, E., Gattei, V., Ghersetti, M., Bulian, P., Moratelli, G., Grassi, G., Zorat, F., and Pozzato, G. (2018). Survival and prognostic factors in mixed cryoglobulinemia: data from 246 cases. Diseases *6*, 35. https://doi.org/10.3390/diseases6020035.

105. Lingala, S., and Ghany, M.G. (2015). Natural history of hepatitis C. Gastroenterol. Clin. North Am. *44*, 717–734. https://doi.org/10.1016/j.gtc.2015.07.003.

106. Simon, V., Ho, D.D., and Abdool Karim, Q.A. (2006). HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet *368*, 489–504. https://doi.org/10.1016/S0140-6736(06)69157-5.

107. Chu, C.M., Chen, Y.C., Tai, D.I., and Liaw, Y.F. (2010). Level of hepatitis B virus DNA in inactive carriers with persistently normal levels of alanine aminotransferase. Clin. Gastroenterol. Hepatol. *8*, 535–540. https://doi.org/10.1016/j.cgh.2010.03.006.

108. Wu, B.-R., Eltahla, A.A., Keoshkerian, E., Walker, M.R., Underwood, A., Brasher, N.A., Agapiou, D., Lloyd, A.R., and Bull, R.A. (2019). A method for detecting hepatitis C envelope specific memory B cells from multiple genotypes using cocktail E2 tetramers. J. Immunol. Methods *472*, 65–74. https://doi.org/10.1016/j.jim.2019.06.016.

109. Nielsen, S.C.A., Roskin, K.M., Jackson, K.J.L., Joshi, S.A., Nejad, P., Lee, J.-Y., Wagar, L.E., Pham, T.D., Hoh, R.A., Nguyen, K.D., et al. (2019). Shaping of infant B cell receptor repertoires by environmental factors and infectious disease. Sci. Transl. Med. *11*, eaat2004. https://doi.org/10.1126/scitranslmed.aat2004.

110. Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381. https://doi.org/10.1038/nmeth.3364.

111. Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. *41*, W34–W40. https://doi.org/10.1093/nar/gkt382.

112. Lefranc, M.-P. (2011). IMGT, the international ImMunoGeneTics information system. Cold Spring Harbor Protoc. *2011*, 595–603. https://doi.org/10.1101/pdb.top115.

113. Rizzetto, S., Koppstein, D.N.P., Samir, J., Singh, M., Reed, J.H., Cai, C.H., Lloyd, A.R., Eltahla, A.A., Goodnow, C.C., and Luciani, F. (2018). B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle. Bioinformatics *34*, 2846–2847. https://doi.org/10.1093/bioinformatics/bty203.

114. Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018). SAVER: gene expression recovery for single-cell RNA sequencing. Nat. Methods *15*, 539–542. https://doi.org/10.1038/s41592-018-0033-z.

115. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

116. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

117. Gori, K., and Baez-Ortega, A. (2018). sigfit: flexible Bayesian inference of mutational signatures. Preprint at bioRxiv.

118. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

119. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

120. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

121. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594. https://doi.org/10.1038/s41592-018-0051-x.

122. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

123. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

124. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

125. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339. https://doi.org/10.1093/bioinformatics/bts378.

126. Mageed, R.A., Dearlove, M., Goodall, D.M., and Jefferis, R. (1986). Immunogenic and antigenic epitopes of immunoglobulins XVII-Monoclonal antibodies reactive with common and restricted idiotopes to the heavy chain of human rheumatoid factors. Rheumatol. Int. *6*, 179–183. https://doi.org/10.1007/BF00541285.

127. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

128. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat. Methods *12*, 519–522. https://doi.org/10.1038/nmeth.3370.

129. R Core Team (2013). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

130. Field, M.A. (2021). Detecting pathogenic variants in autoimmune diseases using high-throughput sequencing. Immunol. Cell Biol. *99*, 146–156. https://doi.org/10.1111/imcb.12372.

131. Field, M.A., Cho, V., Andrews, T.D., and Goodnow, C.C. (2015). Reliably detecting clinically important variants requires both combined variant calls and optimized filtering strategies. PLoS One *10*, e0143199. https://doi.org/10.1371/journal.pone.0143199.

132. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation

from sequencing data. Bioinformatics *27*, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

133. Waardenberg, A.J., and Field, M.A. (2019). consensusDE: an R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. PeerJ *7*, e8206. https://doi.org/10.7717/peerj.8206.

134. Kalovidouris, A.E., and Johnson, R.L. (1978). Rapid cryoglobulin screening: an aid to the clinician. Ann. Rheum. Dis. *37*, 444–448. https://doi.org/10.1136/ard.37.5.444.

135. Wang, J.J., Colella, A.D., Beroukas, D., Chataway, T.K., and Gordon, T.P. (2018). Precipitating anti-dsDNA peptide repertoires in lupus. Clin. Exp. Immunol. *194*, 273–282. https://doi.org/10.1111/cei.13197.

136. Arentz, G., Thurgood, L.A., Lindop, R., Chataway, T.K., and Gordon, T.P. (2012). Secreted human Ro52 autoantibody proteomes express a restricted set of public clonotypes. J. Autoimmun. *39*, 466–470. https://doi.org/10.1016/j.jaut.2012.07.003.

137. Newkirk, M.M., Mageed, R.A., Jefferis, R., Chen, P.P., and Capra, J.D. (1987). Complete amino acid sequences of variable regions of two human IgM rheumatoid factors, BOR and KAS of the Wa idiotypic family, reveal restricted use of heavy and light chain variable and joining region gene segments. J. Exp. Med. *166*, 550–564. https://doi.org/10.1084/jem.166.2.550.

138. Frenkel, D., and Smit, B. (2023). Understanding Molecular Simulation: from Algorithms to Applications (Elsevier).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD21 PeCy7 (clone B-Ly4)	BD Biosciences	Cat# 561374, RRID:AB_10681717
CD11c PE-CF594 (clone B-Ly6)	BD Biosciences	Cat# 562393, RRID:AB_11153662
IgD BB700 (clone IA6-2)	BD Biosciences	Cat# 566539, RRID:AB_2744486
CD14 APCCy7 (clone MFP9)	BD Biosciences	Cat# 557831, RRID:AB_396889
CD3 APCCy7 (clone SK7)	BD Biosciences	Cat# 561800, RRID:AB_10895381
CD20 AF700 (clone 2H7)	BD Biosciences	Cat# 560631, RRID:AB_1727447
CD19 BV650 (clone SJ25C1)	BD Biosciences	Cat# 563226, RRID:AB_2744313
CD38 biotin (clone HIT2)	BioLegend	Cat# 303518, RRID:AB_2259861
CD27 BV786 (clone L128)	BD Biosciences	Cat# 563327, RRID:AB_2744353
Anti-Kappa BV510 (clone G20-193)	BD Biosciences	Cat# 752958, RRID:AB_2917913
Anti-Kappa FITC (clone MHK-49)	Biolegend	Cat# 316506, RRID:AB_493611
Anti-IgM BV421 (clone MHM-88)	Biolegend	Cat# 314516, RRID:AB_2561443
Anti-IgM PE (clone MHM-88)	Biolegend	Cat# 314508, RRID:AB_493005
IGHV1-69 F allele (clone G6)	Singh et al.[45]	N/A
Goat Anti-Human IgM, Fc$_{5\mu}$ fragment specific	Jackson ImmunoResearch	Cat# 109-005-129, RRID:AB_2337543
ChromPure™ Human IgM (myeloma), whole molecule	Jackson ImmunoResearch	Cat# 009-000-012, RRID:AB_2337048
Alkaline Phosphatase Anti-Human IgM (μ-chain specific, clone MB-11)	Sigma-Aldrich	Cat# A2189, RRID:AB_257983
ChromPure™ Human IgG, whole molecule	Jackson ImmunoResearch	Cat# 009-000-003, RRID:AB_2337043
Anti-E2 HCV IgG1 (AR3C)	Genscript, Flyak et al.[22]	PDB: 6MEF
Bacterial and virus strains		
One Shot™ TOP10 Chemically Competent *E. coli*	Invitrogen	Cat# C404003
Biological samples		
Human PBMC	St Vincent's Hospital, Sydney, Australia	N/A
IgG from human blood	Sigma	Cat# G4386
Chemicals, peptides, and recombinant proteins		
Fluorescently conjugated dextran (Klickmer®-PE)	Immudex	Cat# DX01K
Recombinant biotinylated E2 (H77 isolate)	Produced in house, Wu et al.[108]	Genbank: AF011751
Recombinant biotinylated E2 (UNK3a.13.6 isolate)	Produced in house, Wu et al.[108]	Genbank: AY894683
Biotin-SP (long spacer) ChromPure™ Human IgG, whole molecule	Jackson ImmunoResearch	Cat# 009-060-003, RRID:AB_2337060
Human recombinant TNFα	R&D Systems	Cat# 210-TA
Fixable Viability Dye eFluor™ 780	ThermoFisher Scientific	Cat# 65-0865
Streptavidin BUV395	BD	Cat# 564176
Critical commercial assays		
Alexa Fluor 488 Antibody Labelling kit	Invitrogen	Cat# A20181
RNeasy Mini Extraction Kit	Qiagen	Cat# 74104
iScript cDNA Synthesis Kit	Bio-Rad	Cat# 1708891
Nextera XT Index Kit v2 Set A-D	Illumina	Cat# FC-131-(2001-2004)
SuperScript II Reverse Transcriptase	Invitrogen	Cat#18064-071
Nextera XT DNA Sample Preparation Kit	Illumina	Cat# FC-131-1096
REPLI-g Single Cell Kit (96)	Qiagen	Cat# 150345

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
QIAamp DNA Micro Kit	Qiagen	Cat# 56304
KAPA Hyper Plus Kit	Roche	Cat# KK8514
SeqCap Adaptor Kit A	Roche	Cat#ROC-07141530001
Chromium Next GEM Single Cell 3′ Library & Gel Bead Kit v2	10X Genomics	Cat# 120267
Chromium™ Single Cell A Chip Kit	10X Genomics	Cat# 1000009
Q5 High-Fidelity Kit	NEB	Cat# E0555S
ExpiFectamine™ 293 Transfection Kit	Gibco	Cat# A14524
Luciferase Assay System	Promega	Cat# E1500
Galacto-*Star* β-Galactosidase Reporter Assay System	Invitrogen	Cat# T1012
Biotin Protein Ligase Kit	GeneCopoeia	Cat# BI001
Deposited data		
Antibody nucleotide sequences	This paper	See Table S5
Whole genome sequencing (WGS)	This paper	NCBI BioProject: PRJNA1125918
RNA-seq (10X Genomics)	This paper	GEO: GSE269333
PCAWG tumour/normal pairs	ICGC Controlled Data; Alexandrov et al.[48]	https://docs.icgc-argo.org/docs/data-access/icgc-25k-data
Paired IGHV1-69/IGKV3-20 sequences from healthy donors	DeKosky et al.[84]	SRA: SRP047462
IGHV1-69 sequences from healthy donors	Nielsen et al.[109]	NCBI BioProject: PRJNA491287
Experimental models: Cell lines		
Expi293F™ Cells	Gibco	Cat# A14527, RRID: CVCL_D615
HEK293T	ATCC	Cat#CRL-3216; RRID: CVCL_0063
Oligonucleotides		
Bulk immunoglobulin sequencing primers	Singh et al.[45]	N/A
KLF2_F CTGGAGGCCAAGCCAAAG; KLF2_R^ GCTGGGAATCCGACCTCT	Clipson et al.[53]	N/A
11q_del_F^ TGACCCAACCCAGACAAACC; 11q_del_R CCCACCTCCAACGTTGTCTA; SEC23IP_11q_ctrl_F CTTGGCTGAGGCACTCCGTA; SEC23IP_11q_ctrl_R ACGTATGAGCACGGGCAAAC; NOTCH1_F^ CTGGCGGTGCACACTATTCTG; NOTCH1_R GCGCGCCGTTTACTTGAAG; TRAF3_F TGTCAGAGTGTGTCAATGCCC; TRAF3_R^ CACCAACTGGAAAGCGAGCA; ^sequencing oligo, F=forward, R=reverse	This paper	N/A
Recombinant DNA		
Custom pcDNA3.1$^+$ plasmids	Genscript	NA
Luciferase NF-kB Reporter Vector	Promega	Cat# N1111
β-Galactosidase Control Vector	Gift, Shane Grey	NA
Software and algorithms		
MiXCR	Bolotin et al.[110]	https://github.com/milaboratory/mixcr.git
RAGE-Seq pipeline	Singh et al.[45]	https://github.com/KCCG/rageseq
Albacore	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads
IgBLAST	Ye et al.[111]	https://www.ncbi.nlm.nih.gov/igblast/
IMGT (V-Quest, HighV-Quest, JunctionAnalysis)	LeFranc[112]	https://www.imgt.org
VDJPuzzle	Rizzetto et al.[113]	https://bitbucket.org/kirbyvisp/vdjpuzzle2
bcl2fastq	Illumina	https://sapac.illumina.com/
Cell Ranger	10X Genomics	https://www.10xgenomics.com/
R	R Development Core Team	https://www.r-project.org/
SAVER	Huang et al.[114]	https://github.com/mohuangx/SAVER

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
limma	Ritchie et al.[115]	https://bioconductor.org/packages/release/bioc/html/limma.html
Dynamic Read Analysis for Genomics (DRAGEN) Somatic Pipeline	Illumina	https://sapac.illumina.com/
BEDTools	Quinlan and Hall[116]	https://bedtools.readthedocs.io/en/latest/
sigfit	Gori and Baez-Ortega[117]	https://github.com/kgori/sigfit
SAMtools	Li et al.[118]	https://samtools.sourceforge.net/
BWA	Li and Durbin.[119]	https://bio-bwa.sourceforge.net/
GATK-MuTect2	McKenna et al.[120]	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Strelka2	Kim et al.[121]	https://github.com/Illumina/strelka
Ensembl Variant Effect Predictor (VEP)	McLaren et al.[122]	https://www.ensembl.org/vep
Integrative Genomics Viewer (IGV)	Robinson et al.[123]	https://igv.org/
Manta	Chen et al.[124]	https://github.com/Illumina/manta
Delly	Rausch et al.[125]	https://github.com/dellytools/delly
XPro	Bioinformatics Solution	https://www.bioinfor.com/peaks-xpro/
FlowJo	Tree Star	https://www.flowjo.com/
Prism	GraphPad	https://www.graphpad.com/features
FluoresenceFit (FluorescenceFit-main)	This paper	https://github.com/sangioletti/fluorescence_fit
Other		
Expi293™ Expression Medium	Gibco	Cat# A1435101

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient samples

Peripheral blood was taken from patients with chronic Hepatitis C virus (HCV) infection preparing to undergo direct acting antiviral (DAA) treatment recruited and consented under the Surveillance for Antiviral Resistant Variants in Chronic Hepatitis C (SEARCH-C) study, St Vincent's Hospital (SVH), Sydney, Australia. The SEARCH-C was approved by the St Vincent's Hospital Human Research Ethics Committee (HREC/11/SVH/197). Peripheral blood was collected and processed by Ficoll-Paque centrifugation for peripheral blood mononuclear cells (PBMC) collection. Blood samples were collected before the start of DAA therapy (baseline) and again between 25-50 weeks later. Because this was a retrospective study, there was no way to determine if all blood samples had been collected and processed at 37̊C to preserve cryoglobulin RFs in plasma. At the baseline timepoint, the patients commenced 12 weeks of DAA therapy. All patients achieved HCV clearance which was confirmed by absence of HCV RNA in serum. Four HCV patients from the SEARCH-C study were retrospectively identified as having been diagnosed with HCV-associated cryoglobulinemic vasculitis before recruitment into the SEARCH-C study based on their clinical features, laboratory and biopsy results. Patient age, sex and clinical features are indicated in Table S1. Patient ethnicity and socioeconomic status were not recorded for the study. Patient P3 suffered from glomerulonephritis and was treated with rituximab 7 weeks before the baseline sample (HCV$^+$) and commencement of DAA therapy (Table S1). Patient P4 experienced a more severe neuropathy and was treated with methotrexate, mycophenolate and prednisolone both before and after DAA (Table S1). All patients were consented and their samples were transferred to and analysed under the SOMAD study/HOPE Research Program at the Garvan Institute of Medical Research, Sydney, Australia approved by the Western Sydney Local Health District Human Research Ethics Committee (HREC17/5449).

METHOD DETAILS

Flow cytometry

PBMCs were thawed and washed in 2% FCS/PBS and incubated with Fc block for 20 min on ice. PBMCs were stained with eF780 Fixable Viability Dye (ThermoFisher Scientific) and a cocktail of antibodies (see key resources table) for 30 mins on ice. Anti-idiotypic monoclonal antibody G6, specific for F-allele IGHV1-69 heavy chains, was fluorescently labelled using the Alexa Fluor 488 Antibody Labelling kit (Life Technologies).[126] After staining, cells were washed and fixed in 2% formaldehyde (5% formalin, Sigma) for 15 min on ice.

Fluorescence activated cell sorting (FACS)

Cells were either bulk sorted for WGS or single cell RNA-sequencing (10X Genomics) or single cell sorted for downstream single cell RNA-sequencing (Smart-seq2), either done alone or in parallel with single cell DNA sequencing (G&T). For both bulk and single cell sorting, PBMCs were stained using the same buffers and steps as described for flow cytometry (above), with the exception of the anti-idiotypic monoclonal antibody G6, which was only used to stain the PBMCs of patients P1 and P4. For bulk sorting for whole genome sequencing, two populations from the baseline timepoint were sorted: IgM$^+$ B cell clone and polyclonal B cells, according to the gating strategy shown (Figure 2B). For single cell sorting for immunoglobulin and DNA analysis, total B cells (CD19$^+$CD20$^+$) or memory B cells (CD27$^+$) from the baseline (HCV$^+$) timepoint were sorted into 96 well Lo bind plates containing buffers as described below. For single cell RNA-sequencing using the 10X Genomics platform, total B cells (CD19$^+$CD20$^+$) from patient P1 PBMC were sorted at the clearance (HCV$^-$) timepoint.

Bulk immunoglobulin sequencing

Total RNA from roughly 1-3 million thawed patient PBMCs was extracted using a RNeasy Mini Extraction kit (Qiagen) and reverse transcribed to cDNA using oligo-DT primers and the iScript cDNA Synthesis Kit (Bio-Rad). Immunoglobulin heavy and light chain cDNA was amplified using the Q5 High-Fidelity 2X Master mix (NEB) and pooled forward primers that bind the leader sequences of the variable regions and the reverse primers that bind the first exon of the mu and kappa constant regions, as previously described.[45] The primers incorporate 5' sequences compatible with sequencing on the Illumina platform. Two separate PCR reactions were performed for each antibody isotype i.e. mu and kappa. PCR products were indexed (Illumina), pooled and sequenced on an MiSeq machine (Illumina) capturing 300bp paired end reads to a depth of 1 million reads per sample. B cell clones were identified using the MiXCR software tool (v3.0.12),[110] with the top clones based on total number of reads sharing >90% CDR3 amino acid identity. The top five clones were displayed as proportion of all sequenced reads.

Single cell RNA sequencing

Libraries for single cell RNA sequencing on the patient baseline samples (HCV$^+$) were prepared using Smart-seq2 protocol.[127] If genomic DNA was also being extracted simultaneously, cells were processed as described for the Genome & Transcriptome (G&T) protocol.[128] The Smart-seq2 protocol was modified to halve the volume of lysis buffer added to each well of the plate (from 4 μL to 2μL), ensuring the lysis buffer/oligo-DT/dNTP was maintained at a ratio of 2:1:1. cDNA was generated as described,[127] but with reactions performed at half the volumes. cDNA was amplified as described[127] but with the ISPCR primer final concentration reduced to 50nM. For each cell, 1ng of DNA was added to a new 96 well plate for library generation. Amplified DNA was tagmented and barcoded using the Nextera XT DNA Library Preparation Kit (Illumina). The same quantity of each barcoded DNA libraries (range 4-10ng) were pooled together for sequencing. DNA libraries were sequenced on an NextSeq550 machine (Illumina) at a median depth of 1 million reads per cell.

Single cell V(D)J and RNA-sequencing at the clearance (HCV$^-$) timepoint for blood of patient P1 was performed using Repertoire and Gene Expression by Sequencing (RAGE-Seq) using droplet based capture of approximately 10,000 cells using 10X Genomics Chromium 3' system as described.[47] In brief, following reverse transcription and before fragmentation, two extra PCR cycles were performed and full length cDNA equally spilt for short-read gene expression library generation (Illumina) and targeted V(D)J capture followed by long-read library generation (Oxford Nanopore Technologies, ONT). Short-read gene expression libraries were sequenced on a NovaSeq 500 (Illumina) at 50,000 mean reads per cell. Long-read V(D)J libraries were loaded onto R9.5.1 (FLO-MIN107) flowcells (ONT) with base calling performed using the Albacore software pipeline (v2.2.7) (ONT).

Single cell immunoglobulin analysis and clonal tree generation

The immunoglobulin B cell receptor (BCR) heavy and light chain nucleotide sequences for each cell of patient P1 clearance (HCV$^-$) sample were generated by targeted capture and long-read sequencing using the RAGE-Seq protocol, were generated as described.[47] In brief, long-read sequencing data was demultiplexed by cell barcode, subject to *de novo* assembly, aligned and "polished" to generate fasta files containing consensus transcript contigs. The immunoglobulin sequences were determined by aligning the contigs to IgBLAST[111] and BLASTN to determine V(D)J and constant region gene usage respectively. Immunoglobulin sequences that were non-productive, out-of-frame or contained stop codons were removed. Each cell was annotated as polyclonal or clonal, where clones were defined as having >90% amino acid identity with the heavy chain and light chain CDR3 of P1 B cell clone (Figure 1E).

The immunoglobulin B cell receptor (BCR) heavy and light chain nucleotide sequences for each cell generated by the Smart-seq2 protocol from the patient baseline (HCV$^+$) samples was reconstructed from the single cell RNA-seq data using the VDJPuzzle software package.[113] The individual heavy and light chain nucleotide sequences obtained by VDJPuzzle were uploaded to IMGT HighVQuest (v1.9.1). Using the somatic hypermutation annotations, a clonal tree was constructed. The diameter of each circle is proportion to the number of individual B cells sharing an identical immunoglobulin sequence, with the exception of the unmutated ancestor circle (grey). The joining lines between the circles present number of somatic hypermutations between the immunoglobulin sequences, where the length of the lines is proportional to the number of somatic hypermutations (linear scale, 1 unit of distance = 1 mutation). Somatic V(D)J mutations over the whole V(D)J gene loci were counted i.e. including CDR3 and FR4.

Single cell RNA-seq gene expression analysis

Raw sequencing files binary base call (bcl) files were demultiplexed and converted to Fastq using bcl2fastq (v2.19.0.316) (Illumina). Alignment and barcode counting were performed using cell ranger (v6.0.2) (10X Genomics). Gene expression reads were aligned to human genome reference GRCh38 (hg38). Gene expression count matrices were exported by cell barcode linked to a polyclonal or clonal annotation (as determined above) using R.[129] Single cell gene expression counts were normalized using SAVER with default values.[114] Single cells were excluded if the library size or number of expressed genes fell below 2 median absolute deviations, or if mitochondrial reads accounted for more than 20% of the reads. Polyclonal naïve and B cell memory clusters were determined using landmark genes *IGHM*, *IGHG1*, *IGHD*, *TCL1A*. Differentially expressed genes were determined using limma on log normalised values.[115] Bonferroni correction was applied to each set of p-values, to determine the family-wise error rate (FWER). Statistically significant differentially expressed genes were defined as FWER <0.05.

Whole genome sequencing

For the two patients with an IGHV1-69 B cell clone (P1 and P4), the G6 antibody was used to sort the expanded B cell clone ($G6^+$ IgK^+ B cells) and the control polyclonal B cell population ($G6^-$ B cells) (Figure 2B). For the other two patient B cell clones encoded by heavy chains for which no anti-idiotypic antibody exists (P2; IGHV4-59 and P3; IGHV3-7), the B cell clone was identified and sorted from the IgM^+ kappa light chain restricted memory B cell population ($CD27^+$ IgM^+ IgK^+ B cells) and the control polyclonal B cell population consisted of all remaining B cells that did not co-express IgM and IgK (Figure 2B). Polyclonal B cells were not fractionated into naïve and memory B cells based on the assumption that the additional burden of AID off-target somatic mutations in memory B cells would be different in each cell and comprise a low percentage of reads, falling below the threshold for variant calling. The gating strategy for the expanded clone were predicted to result in purity >90% (>95% for the IGHV1-69 B cell clones).

Approximately 30,000 cells (minimum 8,000) of either the B cell clone or polyclonal B cells of the patients were directly FACS sorted into 100μL ALT lysis buffer (Qiagen MicroKit). gDNA was extracted as per manufacturer's instructions (Qiagen MicroKit). Approximately 2/3 of gDNA was taken to generate DNA libraries (minimum 10ng), with remaining gDNA to be stored for variant confirmation. DNA libraries for whole genome sequencing were generated using the Kapa Hyper Plus Kit (Roche) for low input DNA. DNA libraries were generated according to manufacturer instructions, with the following modifications: fragmentation for 17 minutes, adapter ligation for 1 hour and PCR amplification with 3-5 cycles depending on starting DNA amount (5 cycles for ~10ng DNA). DNA libraries were sequenced on a HiSeqX instrument (Illumina) with 150bp paired end reads at 30X coverage for the polyclonal population and at 60X coverage for the B cell clone. WGS sequencing quality was verified by calculating the coverage (unique reads) per nucleotide base over ~180 genes at exonic locations totalling ~1,430kb, using the BAM files generated from somatic variant analysis and BEDtools.[116]

SNV mutational signatures

To detect small somatic variants (SNV), the Dynamic Read Analysis for Genomics (DRAGEN) Somatic Pipeline (v3.3.7) (Illumina) for tumour/normal pair was used, with the expanded B cell clone assigned as "tumour" and polyclonal B cells assigned as "normal." Reads were aligned to GRCh37 (hg19) reference genome. The COSMIC (v3) single base substitution (SBS) mutational signatures of somatic SNV calls of the HCV-cryovas clones were fit to mutational signatures relevant to lymphocytes (SBSBlood, SBS1, SBS7a, SBS8, SBS9, SBS18) as described[50] using the R package sigfit (v2.0).[117] The SBS signatures were plotted as a relative proportion for each clone.

Small (<50bp) somatic variant (SNV and indel) analysis

Four PCAWG tumour/normal sample pairs from the CLL-mutated ("Lymph-CLL") and BNHL ("Lymph-BNHL) patient cohort were identified[48] and extracted from ICGC Controlled Data. Variants were called using a workflow previously described.[130,131] In brief, the PCAWG BAM files were downloaded and HCV B cell clone tumour/normal BAM files generated using the Dynamic Read Analysis for Genomics (DRAGEN) Somatic Pipeline (v3.3.7) (Illumina) (see above) were used to extract reads using SAMtools[132] bam2fq. Read were aligned to GRCh38.p14 (hg38) using BWA[119] and prepared for variant calling using GATK best practices.[120] For detecting somatic SNV and indel calls, default filtering was applied for both GATK-MuTect2[120] and Strekla2[121] and for each sample a consensus of SNVs and indels were generated using an approach previously described[133] (Table S2). Somatic variant output file was annotated using the Ensembl Variant Effect Predictor (VEP) algorithm.[122] Somatic variants were filtered selecting for coding variants resulting a frameshift, in-frame deletion or insertion, missense, stop gain or loss or synonymous variant that occurred within a gene of a curated list of 178 genes frequently mutated in B cell lymphomas.[45] Somatic variants were verified using the Integrative Genomics Viewer (IGV) (v2.5.3)[123] and used to calculate the proportion of reads carrying the variant (i.e. variant allele frequency, VAF).

Large (>50bp) structural somatic variant analysis

To detect somatic SVs in the patient HCV-cryovas clones, the CLL-mutated ("Lymph-CLL") and BNHL ("Lymph-BNHL) samples (see above),[48] Manta[124] and Delly[125] were run in somatic tumour/normal mode with default filtering. Manta and Delly identify and score possible SVs based on two types of evidence; discordant pair end reads and split read alignments. For each sample, a consensus of SVs were generated using an approach previously described[133] (Table S3). The consensus SV list was manually annotated for variants occurring at the heavy, kappa or lambda immunoglobulin loci using the hg38 UCSC Genome Browser.

Somatic ploidy analysis

To detect the presence of any ploidy event such as a chromosome gain, the average whole genome sequencing coverage was calculated for the whole genome for chromosomes 1 to 22 using the Samtools depth command.[61] An average across every base for the entire genome for each of the B cell clone and the polyclonal B cells was first calculated, and this ratio utilised as a baseline. The same was done for each chromosome and this ratio compared to the whole genome baseline ratio (Table S4). If a whole chromosome gain was present, we would expect ~1.5X more reads in the chromosome ratio than the baseline genomic ratio. Any chromosome loss (monoploidy) was defined as a ratio ≤ 0.6, any single whole chromosome gain (triploidy) was defined as a ratio ≥ 1.4 but ≤ 1.8 and whole duplication of both chromosomes (tetraploidy) was defined as a ratio ≥ 1.9.

Somatic variant validation

PCR with or without sanger sequencing was performed on either remaining bulk sorted gDNA or from single cell amplified gDNA performed in parallel with mRNA extraction for single cell immunoglobulin sequencing using the Genome & Transcriptome (G&T) protocol.[128] Following separation of single cell mRNA and gDNA, gDNA was amplified via multiple displacement amplification (MDA) using the REPLI-g Single Cell Kit (Qiagen) according to manufacturer's instructions. MDA rapidly amplifies DNA at a constant temperature, by using a combination of random hexamer primers and a high-fidelity bacteriophage DNA polymerase (error rate 1 in 10^6-10^7 bases). PCR primers and PCR conditions used on bulk or single cell derived gDNA are listed in the key resources table. All PCR primers were designed in house with the exception of the KLF2 primers.[53] All PCR reactions used a standard Taq polymerase (Invitrogen), with the exception of KLF2 which was performed using the Q5 High-Fidelity kit (NEB). For the 11q deletion variant (P3), primers for a control gene (*SEC23IP*) giving rise to a PCR product half the size was also included in the same PCR reaction. Purified PCR products (with the exception of the *SEC23IP* control gene) were submitted for Sanger sequencing.

NF-κB luciferase assay

Adherent HEK293T cells were co-transfected with 0.3μg NF-kB luciferase reporter (pcDNA3.1-luc) (Promega), 0.2μg of β-galactosidase plasmid (CMV.β-galactosidase, gifted by Shane Grey) and either 0.5μg pcDNA3.1 vectors encoding wildtype or KLF2 variants or empty pcDNA3.1 vector (total 1μg DNA) (Genscript). Transfection was performed using Lipofectamine 3000 (ThermoFisher) in serum-free conditions. Two hours before harvest, cells were stimulated with human recombinant TNFα (R&D Systems) at 300IU/mL. Luciferase activity was measured in cell lysates collected 8 hours after transfection. To correct for transfection efficiency, luciferase results were normalised to β-galactosidase activity.

Soluble IgM antibody expression

Secretory IgM antibodies were produced in house by transient transfection of non-adherent Expi293F (Gibco) cells using the Expi293 Expression System Kit (Gibco). The heavy immunoglobulin variable region sequences (Table S5) were inserted into the human heavy chain mu constant region secretory sequence (Genbank: BC073758.1) and synthesised into mammalian vector pcDNA3.1 (Genscript). Similarly, the light chain immunoglobulin variable region sequences were inserted into the human kappa constant region sequence (Genbank: OM584289.1) and synthesised into mammalian vector pcDNA3.1 (Genscript). Leader sequences inserted upstream of the heavy and light chain variable region were obtained from IMGT. Human J chain (Genbank: AK312014.1) was also synthesised into mammalian vector pcDNA3.1. Broadly neutralising anti-E2 HCV AR3C heavy and light chain nucleotide sequences were derived from the immunoglobulin amino acid sequences (PBD: 6MEF).[22] Expi293F cells were transfected according to manufacturer's instructions with a 10μg of heavy chain plasmid, 10μg of light chain plasmid and 10μg of J chain plasmid (ratio 1:1:1) for total of 30μg DNA per flask at a density of ~3 x 10^6 cells/mL. On Day 7 post-transfection, supernatant for each flask was collected, and spun for 10 min at 3000g and passed through a 0.22μM filter. Soluble IgM was purified by concentrating the supernatant using centrifugal filter units (Amicon, Merek) with a 30K membrane cut-off spun at 4000g for 40-60 mins at room temperature until <2mL concentrate remained.

Soluble IgM quantification by ELISA

A 384-well plate (Corning) was coated with 0.5μg/mL purified goat anti-human IgM (Jackson ImmunoResearch) diluted in 1X PBS for 1 hour at room temperature and blocked with 3% BSA/1XPBS for 1 hour at 37°C. An IgM standard was prepared using human whole IgM derived from serum (Jackson ImmunoResearch) starting at 10μg/mL concentration followed by a 1:2 serial dilution. Concentrated supernatants containing soluble IgM were diluted 1:100 followed by a 1:2 serial dilution performed in duplicates and incubated for 1 hour at room temperature. Presence of IgM was detected using anti-human monoclonal IgM-alkaline phosphatase (Sigma Aldrich) (1:1000) incubated for 1 hour at room temperature before addition of para-nitrophenylphosphate substrate (Sigma Aldrich). ELISA plate was read an optical density (OD) at 405nm (ClarioStar instrument, BMG Labtech). Concentration of IgM (μg/mL) in each supernatant was interpolated using the standard curve.

IgG (RF) ELISA

384-well plates (Corning) were coated with 5μg/mL whole IgG obtained from human serum (Jackson ImmunoResearch) in NPP coating buffer overnight. ELISA plates were blocked with 3% BSA/PBS for 1 hour at 37°C. Samples were added in duplicate or triplicate to the plate at a starting concentration of 100μg/ml, followed by 1:2 serial dilution and incubated for 1 hour at 37°C. Anti-human monoclonal IgM-alkaline phosphatase (Sigma Aldrich) (1:1000) was used as secondary detection antibody incubated at 45 mins at

37°C before addition of para-nitrophenylphosphate substrate (Sigma Aldrich). Plates were read at an optical density (OD) at 405nm (ClarioStar instrument, BMG Labtech).

Cryoglobulin assay

Cryoglobulin activity was analysed via a temperature-controlled spectrophotometer (Cary Series UV-Vis spectrometer, Agilent). Antibody cryoaggregation causes changes in light scattering.[134] IgM supernatants stored at 4°C were allowed to warm to 37°C on a heat-block for 20 min (cryoglobulins redissolve upon warming). An equal volume of IgM antibody at a concentration of 150-200μg/mL was mixed with equal volume of human IgG obtained from blood (Sigma-Aldrich) at a concentration of 1mg/mL (ratio ~1:5) in 1X PBS. The mixture of IgM and IgG was incubated for another 20 min at 37°C on a heat-block before immediately transferring to the spectrometer set to 37°C. Approximately 200 optical density (OD) measurements were taken at 500nm as the spectrophotometer reduced in temperature by 0.1°C/min from 37°C to 4°C.

RF peptide sequencing and data analysis

IgM RF antibody was purified from patient P3 plasma by precipitating with heat-aggregated IgG followed by reduced SDS-PAGE[135] and non-reduced SDS-PAGE,[45] respectively. The IgM RF gel band of the heavy and light chain was excised and digested with Pierce trypsin protease (ThermoFisher Scientific) and chymotrypsin (Promega) to generate peptides for LC-MS/MS using a Thermo Scientific Orbitrap Exploris 480 mass spectrometer coupled to an Ultimate 3000 UHPLC (Dionex). Peptide sequence analysis was performed by de novo sequencing and the IMGT database matching using Peaks studio XPro software (Bioinformatics Solution Inc). Parameters for database searches, data refinement and immunoglobulin variable region subfamily assignments were described previously.[135,136] An average local confidence score threshold of 80% was applied to select high quality de novo peptides. A false discovery rate (FDR) threshold of 1.0% was applied at the peptide level to each data set. The immunoglobulin variable region subfamily is assigned from the presence of a unique peptide corresponding to the subfamily.

Recombinant HCV E2 (including biotinylated)

The E2 protein sequence was derived from HCV Genotype 1a (isolate H77, GenBank accession number AF011751) and HCV Genotype 3a (isolate UNK3a.13.6, GenBank accession number AY894683). Recombinant HCV E2 (rE2) antigen was produced as described.[108] In brief, the E2 construct was incorporated the HCV polyprotein amino acid residues 384 to 661 encoding the E2 ectodomain (no stem or transmembrane domain) into a pcDNA3.1 vector with a N-terminus secretion signal, a C-terminal Avitag™ and a six-histidine tag (6×His) and transiently transfected into HEK293F cells. Media was harvested 96 hours later and rE2 was purified by passing through a 1 ml HiTrap chelating HP chromatography column. Biotinylation of rE2 was performed at the C-terminal Avitag™ using *E.coli* biotin ligase (BirA).

Surface IgM expression and antigen staining

Methods for transient transfection for expression of membrane-bound IgM on ExpiHEK293F cells were adapted from the published protocol.[76] Desired heavy immunoglobulin variable region sequences (Table S5) were inserted into the human heavy chain mu constant region membrane sequence that includes the transmembrane region (Genbank accession no. BC009851.2) and synthesised into mammalian vector pcDNA3.1 (Genscript). Leader sequences upstream of the variable region were obtained from IMGT. Immunoglobulin light chain constructs were the same as used for secretory IgM expression. Broadly neutralising anti-E2 HCV AR3A heavy and light chain nucleotide sequences were derived from the immunoglobulin amino acid sequence (PBD: 6BKB).[19] Expi293F cells were transfected according to manufacturer's instructions with 10μg of heavy chain plasmid and10μg of light chain plasmid (ratio 1:1) for total of 20μg DNA per flask at a density of approximately 1 x 10^6 cells/mL. On Day 2 post-transfection, Expi293 cells were harvested and 1 x 10^6 cells were transferred to a 96 well U bottom plate (Corning). Cells were stained with 10uL of biotinylated IgG (2mg/mL, Jackson Immunoresearch) or 10uL of commercial fluorescently labelled dextran molecule (Klickmer®, Immudex) loaded with biotinylated IgG, biotinylated rE2 or biotinylated rE2 bound with anti-E2 IgG1 monoclonal antibody (clone AR3C, PDB: 6MEF)[22] (Genscript), for 10 min at room temperature. After staining with the antigen loaded Klickmer dextran, anti-human IgM BV421 antibody (clone MHM-88) was added to the cells and incubated for a further 25 min on ice. Cells were washed with 1%BSA/PBS/0.02% sodium azide buffer and run immediately on a LSRII Fortessa or Symphony (BD) instrument.

Klickmer® molecules carry 20-25 biotin acceptor sites (~5-6 streptavidin molecules) attached along a dextran backbone conjugated to phycoerythrin (PE). Biotinylated E2 (described above) or biotin-SP (long spacer) ChromPure human IgG (Jackson Immunoresearch) were mixed with PE-Klickmers according to manufacturer's instructions (HCV rE2; 60kDa, whole IgG; 150kDa) at molar ratios to achieve full loading of acceptor sites. Antigen was incubated with Klickmer® for 30 mins in the dark at room temperature. Where required, Klickmer® molecules loaded with biotinylated rE2 were incubated with 5uL purified anti-E2 AR3C monoclonal IgG1 antibody (1mg/mL, Genscript) for 20 minutes in the dark. Antigen-Klickmer® molecules were diluted with 1%BSA/PBS/0.02% sodium azide buffer to 16nM (1:10 dilution), ready for staining. Antigen-Klickmer® mix was stored in the dark at 4°C for up to 4 weeks.

IGHV1-69/IGKV3-20 anti-HCV E2 antibodies

Antibodies B04 and H04 were identified by single cell mRNA sequencing of single cell sorted HCV E2 (cocktail of Genotype 1a and 3a) tetramer-binding memory B cells from a patient chronically infected with HCV (Genotype 3a)[108] (R.A. Bull, unpublished data).

Antibodies were confirmed to be HCV E2 specific by ELISA. Mutations in heavy and light chains were reverted to the unmutated ancestor using IMGT[112] (Table S5). Non-synonymous mutations in the heavy chain and light chain CDR3 were also reverted if the mutation resided within the V and J gene. In the case of H04 and B04, non-synonymous mutations were also reverted in the D genes (deemed to reside at positions clear of N addition sites).

IGHV1-69 CDR3 amino acid lengths

Seventeen IGHV1-69/IGKV3-20 RF antibody sequences confirmed to bind IgG were acquired from this project (P1,P4) and published data.[33,34,45,83,137] Ten IGHV1-69/IGKV3-20 anti-HCV E2 antibody sequences were derived from the same HCV patient as described above (R.A. Bull, unpublished data). IGHV1-69/IGKV3-20 antibody sequences with unknown specificity were obtained from single memory B cells from three healthy donors.[84] Unmutated (<2% SHM) IGHV1-69 IgM (“naïve”) and mutated (≥ 2% SHM) IGHV1-69 IgG (“memory”) antibody sequences were from bulk repertoire sequencing of healthy donor PBMCs filtered for least 50,000 clones (61 healthy adult donors).[109] CDR3 amino acid lengths were defined as according to IMGT, with anchor residues removed.

Estimating K_D using the IgM onset of binding

For a full description of how the IgM threshold or onset of binding is linked to the affinity constant (K_D), see methods below. In brief, if the K_D value for a reference antibody *A* is known, we can extrapolate the K_D value of any other antibody to the same antigen by comparing the shift in IgM fluorescence threshold at which antigen-dextran binding is first detectable on a logarithmic scale, using the equation shown below (Equation 1). To obtain the IgM fluorescence threshold at which antigen binding can be detected, two methods were used. The first method relies on subjective visual inspection of the flow cytometry plots. The geometric mean fluorescence on a narrow gate of ~200 events was generated at the threshold IgM showing onset of IgG-binding as identified by visual inspection using FlowJo™ software (v.10). In cases where binding was visually ambiguous, binding was considered to be occurring if the binding signal of cells with the highest density of membrane IgM was statistically significant ($p<0.0001$) compared to cells with intermediate levels of membrane IgM. The IgG-dextran fluorescence intensity for individual cells with “medium” and “high” surface IgM density was exported using Flowjo™ software. The second method relied on objective fitting of the antigen-dextran binding curves. The flow cytometry fluorescent scale values for the IgM (x-axis) and IgG-Dextran (y-axis) parameters for each Expi293F cell expressing IgM were exported and fit to an equation using a basin-hopping Monte Carlo procedure using original code generated for this paper (“FluorescenceFit-main”, see key resources table). For a full description of the fitting procedure see methods below. Given the fitted curve, we define the IgM threshold of binding (“calculated onset” or “onset for best solution at x”) as the intercept along the x-axis between the intensity baseline and a straight line tangent to the fitting curve at the mid-point between the minimum and maximum intensity value (Figure S6C). The fitting procedure was run three times for the reference antibody (e.g. TJ4). The average onset of the reference antibody was used to estimate the K_D for the unknown antibodies using Equation 1. Antibodies with an unknown K_D were only run through the fitting procedure once.

Using the IgM binding onset as calculated by either method (i.e. visual inspection or fitting algorithm) including for the reference antibody with a known K_D, the IgM binding thresholds were used to calculate the K_D of the unknown antibody using the following equation:

$$\log K_{D,A} = \log K_{D,B} + \log \sigma_{R,A} - \log \sigma_{R,B} \qquad \text{(Equation 1)}$$

An example calculation is included below. Here, we are estimating the affinity of HCV B cell clone of patient P1 to IgG-dextran using TJ4 as a reference with known IgG affinity constant ($K_{D,A} = 11.2 \times 10^{-6}$ M). Visual inspection of flow cytometry plots indicated TJ4 has an IgM IgG binding onset $(\sigma_{R,A}) = 3000$, and P1 has an IgM IgG binding onset $(\sigma_{R,B}) = 3600$ (Figure 6B). To estimate the affinity constant ($K_{D,B}$) for the HCV B cell clone of patient P1:

$$\log K_{D,A} - \log K_{D,B} = \log \sigma_{R,A} - \log \sigma_{R,B} \qquad \text{(Equation 1), rearranged}$$

$$\log\left(11.2 \times 10^{-6}\right) - \log\left(K_{D,P1}\right) = \log(3000) - \log(3600)$$

$$\log\left(11.2 \times 10^{-6} \big/ K_{D,P1}\right) = \log(3000 \,/\, 3600)$$

$$11.2 \times 10^{-6} / K_{D,P1} = 3000/3600$$

$$K_{D,P1} = 11.2 \times 10^{-6} \times 3600/3000$$

$$K_{D,P1} = 13.4 \times 10^{-6}\ \mathrm{M}$$

$$= 13\mu\mathrm{M}$$

Modelling of multivalent antigen binding to membrane IgM

Our mathematical model for surface IgM expression binding to IgG-dextran arises from adapting previous results on the interaction between multivalent ligand-functionalised nanoparticles and surfaces bearing cognate receptors.[78–82] Equilibrium binding probabilities in multivalent systems can be accurately described within the context of statistical mechanics. If a receptor coated surface (i.e. surface IgM expressed on Expi293F cells) is in contact with a solution containing ligand-functionalised constructs (i.e. IgG-dextran),[78] the number of constructs bound is given by the Langmuir formula:

$$N_{ads} = M\,\frac{z\,Q}{(1+z\,Q)} \qquad \text{(Equation 2)}$$

M is the maximum number of adsorption sites on the surface and $Q = Q(N_L, \sigma_R, \chi)$, the binding partition function, depends on the number of ligands N_L, the surface density of receptors σ_R and χ, a parameter related to the strength of the ligand-receptor bond. Furthermore, z is the so-called activity in solution[138] of the multivalent construct, which for all practical cases of interest is given by $z = v_0 \rho_B$, where ρ_B is the molar concentration of the multivalent construct in solution and v_0 the adsorption volume. Within the Langmuir description used here, $\theta = N_{ads}/M$ can be interpreted as the binding probability of a single multivalent particle binding to the surface.[78] In Equation 2, in contrast to the more general definition,[78] we assumed that each binding site contains exactly the same number of receptors.

In order to calculate Q, we assume that each of the N_L ligands in the multivalent construct can independently form a bond with each of the $N_R = \sigma_R A$ receptors, where A is the area of a binding site on the surface. In this case, the total bound partition function is given by:

$$Q = (1.0 + N_R\,\chi)^{N_L} \qquad \text{(Equation 3)}$$

Where χ is related to the ligand-receptor dissociation constant K_D by Varilly et al.[79]:

$$\chi = \exp(-\Delta G_{bond}/k_B T)K\_D = (K_D/\rho_0)\chi_{cnf} \qquad \text{(Equation 4)}$$

where ΔG_{bond} is the effective bond energy, $\rho_0 = 1M$ is the standard molar concentration and $0 < \chi_{cnf} < 1$ a (bond-weakening) correction that arises from the fact that the ligands are not free in solution, but attached to a scaffold (i.e. dextran) that limits their translational and rotational freedom. Although the exact value of χ_{cnf} depends on details of how the ligands are grafted to the particle,[79] in our experiments it is the same across all systems, since the mixture of ligand (IgG) and the scaffold (dextran) is kept consistent within the same experiment.

N_{ads} as described by Equation 2 has the form of a Hill-equation with an exponent equal to N_L. Qualitatively, on a logarithmic scale the adsorption curve thus resembles a smoothened step function, with $N_{ads} \approx 0$ until the receptor density reaches a certain critical value, σ_R^*, after which N_{ads} increases very rapidly (at least until it reaches its maximum value, M). This is the behaviour we observed in our experiments (Figures 5C and 6A).

To estimate the value of the bond dissociation constant (K_D), we note that in the partition function Q, and thus also in the expression for the number of adsorbed particles N_{ads}, N_R (or σ_R) and χ only appear as a product, not alone. Although the expression for Q is derived assuming independent ligands, more general treatments that do not make these assumptions[80,82] still lead to expressions where only the $N_R\chi$ product appears, making the following conclusions quite robust. Because these two quantities only appear as a product, mathematically, this means that when plotting log N_{ads} vs log σ_R, multiplying K_D by an arbitrary (positive) factor γ, simply rigidly shifts all points of the curve by exactly -log γ. This happens for every single point on the curve, and thus also for the critical receptor density σ_R^* at which bound construct (IgG-dextran) begins to rapidly rise. In other words, if we plot N_{ads} vs σ_R for two systems A and B that differ only in their ligand-receptor pairs, in our case, different IgM antibody sequences, and we compare two points for these systems with the same value of N_{ads}, we have:

$$N_{ads,A} = N_{ads,B} \rightarrow \chi_A \sigma_{R,A} = \chi_B \sigma_{R,B} \qquad \text{(Equation 5)}$$

Or, using Equation 4 to connect with K_D and rearranging the terms in Equation 1.

Notably, in Equation 1 the dependence on the configurational contribution χ_{cnf} to the bond strength, which would be otherwise difficult to estimate, drops out because it is the same for two systems that only differ by the ligand-receptor pair used. Thus, if the K_D value for a reference system A is known, Equation 1 allows us to calculate the K_D value of any other ligand-receptor pair by simply comparing the shift in the adsorption (i.e. IgM fluorescence threshold at which IgG-dextran binding occurs) on a logarithmic scale.

Fitting procedure for calculating the IgM binding threshold

First, we make the natural assumption that the measured fluorescence signal increases linearly with the amount of dextran bound, i.e.:

$$S = A + I\theta \tag{Equation 6}$$

Which allows us to interpolate the intensity with the functional form:

$$S = A + B\frac{C\,(\,1+D\,x\,)^{E}}{1+C(1+D\,x)^{E}} \tag{Equation 7}$$

where x is the number of IgM receptors per cell and A,B,C,D and E are fitting parameters. In principle, one can try to relate the values of these parameters to other physical quantities characterising the system and, if known, they could be directly substituted inside Equation 7. In this regard, A is related to the baseline value of the fluorescence signal, while B is the proportionality constant between the amount of dextran adsorbed and the fluorescence provided, $C = \rho\, v_B$ is the activity of dextran in solution, where ρ is the number concentration of dextran and v_B the so-called binding volume.[82] $D = K_D\, \sigma$ is the product of the dissociation constant of a single bond, expressed in M, and σ the area occupied by an adsorbing dextran molecule. Finally E is the number of ligands on a dextran molecule that can form bonds with the receptors on cell's surface. Whereas this microscopic interpretation could be useful in extracting additional information, compensation of errors can make the exact values obtained for the fitting parameters quite sensitive, and thus their exact interpretation should be done with care. Nevertheless, the general shape of the adsorption curve described by Equation 6 fits the data well and together with a geometric construction, can be reliably used to extract the IgM threshold or onset of binding, using the following procedure.

First, we find the optimal value for the fitting parameters by performing a numerical fit of the flow cytometry fluorescence scale values for the IgM (x-axis) and IgG-dextran (y-axis) parameters for IgM expressing Expi293F cells. We do this by using a basin-hopping Monte Carlo procedure to find the global optimum, as implemented in the Scipy python library, with an effective temperature of T=3 and 2000 iterations. The code used for fitting and for building the geometrical construction necessary is located in the key resources table. Second, given the fitted curve, we define the IgM onset of binding as the intercept along the x-axis between the intensity baseline (i.e., the line described by the equation $y = A$) and a straight line tangent to the fitting curve at the mid-point between the minimum and maximum intensity value (see Figure S6C). The IgM onset of binding can be used to estimate the K_D using Equation 1, and as described in the methods section above. While performing the fitting procedure, we use an F-test based on the variance measured using our binding model versus that obtained by fitting the data with a constant value (reported p-values). In other words, we test our hypothesis ("there is binding with measurable fluorescence signal"), against the null hypothesis ("no binding is measured"). The threshold of the p-value for statistical relevance between the difference in the obtained variance calculated by the two models is set to 1%. As such, when the presence of a binding signal is not statistically significant, we cannot calculate the binding onset, and thus cannot calculate the value of the dissociation constant K_D. In this scenario, if any binding does occur, we assume the onset of binding occurs at an IgM fluorescence intensity X equal or greater than the maximum IgM value recorded in the relevant experiment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were conducted using Prism Graphpad software, R[129] and original code. Statistical details can be found in Figure legends and in relevant STAR Methods sections.