Alternative splicing of its 5' untranslated region controls CD20 mRNA translation and enables resistance to CD20-directed immunotherapies

Zhiwei Ang[1], Katharina E. Hayer[2], Manuel Torres Diz[1], Carolin Schmidt[1], Sisi Zheng[1], Feng Xu[3], Catherine D. Falkenstein[4], Joseph P. Loftus[4], Scarlet Y. Yang[1], Mukta Asnani[1], Vinodh Pillai[5,6,7], Emeline Chong[8], Marilyn M. Li[3,6,7], Sarah K. Tasian[4,6,9], Yoseph Barash[10], Stephen J. Schuster[8], and Andrei Thomas-Tikhonenko[1,4,6,7,9*]

[1] Division of Cancer Pathobiology, [2] Department of Biomedical and Health Informatics, [3] Division of Genomic Diagnostic, [4] Division of Oncology, [5] Division of Hematopathology, and [6]Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104

[7] Department of Pathology & Laboratory Medicine, [8] Lymphoma Program, Abramson Cancer Center, [9] Department of Pediatrics, [10] Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104

* Correspondence to: andreit@pennmedicine.upenn.edu

Running title: Alternative splicing of CD20 5' UTR in B-cell lymphomas

Keywords: CD20, rituximab, mosunetuzumab, follicular lymphoma, chronic lymphocytic leukemia, alternative splicing, protein synthesis, mRNA translation, untranslated regions

Abstract

CD20-directed immunotherapies, such as rituximab, are mainstays of clinical management of mature B-cell neoplasms, previously known as non-Hodgkin lymphomas (NHL). More recently, the bispecific CD20/CD3 antibody mosunetuzumab was approved in the EU and US for treatment of relapsed or refractory follicular lymphoma (FL). The clinical efficacy of these CD20-directed therapies is limited by instances of intrinsic or acquired resistance, frequently mediated by complete or partial loss of CD20 expression. For example, in most chronic lymphocytic leukemia (CLL) at diagnosis and in a subset of post-mosunetuzumab FL relapses, CD20 protein levels are inexplicably low, despite near normal levels of CD20 mRNA and no detectable CD20 genetic variants. Here, we show that in both normal and malignant B-cells, CD20 down-regulation commonly occurs via alternative splicing in its 5' UTR. This event generates an extended 5' UTR isoform (V1) with a stem-loop structure and upstream open reading frames (uORFs), which cooperatively inhibit translation. This translation-deficient V1 was abundant in most CLL, in a CD20-negative post-mosunetuzumab FL relapse, and in a subset of diffuse large B-cell lymphoma (DLBCL) where it correlated with low CD20 protein levels. In a panel of B-cell lines, knockout of the Sam68/KHDRBS1 splicing factor with multiple putative binding sites in V1 shifted CD20 splicing toward translation-competent variants and increased CD20 levels. The expression of CD20 was further enhanced by modulating its splicing with Morpholino oligomers, which also increased rituximab-mediated cytotoxicity. Thus, this widespread splicing-mediated mechanism of CD20 antigen loss potentially could be targeted to enhance CD20-directed immunotherapies.

Introduction

The plasma membrane protein CD20, encoded in humans by the *MS4A1* gene, is a clinically significant target for multiple monoclonal antibody (mAb) therapies, due to its lineage-specific expression on mature B-cells and the malignancies derived therefrom [1]. In adult patients, CD20-directed therapies are administered alone or in combination with chemotherapy as standards of care for mature B-cell neoplasms such as DLBCL, FL, Burkitt (BL), CLL and high-grade B-cell lymphomas (HGBCL) [2,3]. By far the most widely prescribed anti-CD20 mAb is rituximab, which is on the World Health Organization's List of Essential Medicines [4]. In pediatric patients, rituximab and chemotherapy combinations are approved for previously untreated, advanced stage CD20-positive DLBCL, BL, and HGBCL [5]. In EU and other countries, rituximab is also used to treat children and young adults with CD20-positive B-cell lymphoblastic leukemia (B-ALL) [6]. Newer CD20-directed immunotherapies include mosunetuzumab, a CD20xCD3 bispecific monoclonal antibody that redirects T cells to engage and eliminate malignant B-cells [7,8]. In 2022, mosunetuzumab was granted accelerated approval in the EU and by the FDA for the treatment of relapsed or refractory (r/r) FL as a third-line or later therapy: www.accessdata.fda.gov/drugsatfda_docs/label/2022/761263s000lbl.pdf [9].

While CD20-targeted immunotherapies have increased the median overall survival of patients with B-cell malignancies, immunotherapy resistance due to antigen loss remains a challenge [2,10]. Overall, up to half of all adult patients with DLBCL, one fifth of adult patients with CLL, and the majority of patients with FL are not cured by rituximab-chemotherapy combinations [11,12]. This resistance is commonly associated with the down-modulation of CD20 [13-17]. For example, reduced CD20 expression in DLBCL was reported to correlate with inferior outcomes in patients undergoing R-CHOP chemotherapy (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) [18]. Furthermore, patients with r/r B-cell NHLs (B-NHL) that are low in CD20 protein at baseline (16/293 or 5.5% of cases) failed to respond to mosunetuzumab monotherapy. Beyond this intrinsic resistance, CD20 protein was lost in 7/26 (27%) of patients with B-NHL that relapsed after initial responses to mosunetuzumab [19]. Even at diagnosis, CD20 protein levels are generally lower in CLL relative to other mature B-cell neoplasms [20-23]. This is believed to contribute to the relatively low *in vitro* activity of rituximab monotherapy against CLL [24].

Despite the clear role of CD20 antigen loss in immunotherapy resistance, significant gaps in our understanding of the underlying mechanism(s) still exist. We previously reported that the Myc oncoprotein down-regulates CD20 mRNA expression in human B-cells and renders them partly resistant to rituximab in vitro [25]. What, if any, role this regulation plays in lymphoma patients remains to be determined. Most recently, the apparent loss of CD20 protein in 3 out of 4 mosunetuzumab-resistant B-NHL tumors analyzed using next generation sequencing (NSG) could not be explained by the disappearance of CD20 mRNA or the emergence of *MS4A1* genetic variants [19]. A lack of concordance between CD20 mRNA and protein was also observed in CLL samples, which were found to be low in CD20 protein despite a near normal level of CD20 mRNA relative to normal B-cells [26].

Here, we show that in some post-mosunetuzumab FL relapses as well as in many CLLs at diagnosis, CD20 expression is suppressed due to the prevalence of mRNA isoforms, dubbed V1 and V2 (for variants 1 and 2), with extended 5' UTRs. These extended 5' UTRs contain a prominent stem-loop structure and uORFs, which cooperatively inhibit translation. Knocking out or knocking down the Sam68/KHDRBS1 splicing factor with several putative binding sites in V1 and V2 led to a shift in CD20 splicing towards the translation-competent isoforms. We then designed Morpholino antisense oligonucleotides (ASO) to specifically modulate 5' UTR exon splicing. Treatment with these ASOs led to elevated CD20 expression and enhanced rituximab-mediated cytotoxicity in a panel of lymphoma cells. Of note, our data mining effort revealed that aberrant CD20 5' UTR splicing limits its protein levels in a large DLBCL cohort. Thus, this mechanism appears to be at play in a wide variety of B-cell neoplasms, attesting to its clinical relevance.

Results

CD20 protein loss in a follicular lymphoma following mosunetuzumab treatment coincides with a shift in splicing towards an extended 5' UTR isoform. In 3 out of 4 (75%) mosunetuzumab-resistant B-NHL tumors that were CD20 protein-negative by immunohistochemistry (IHC), whole exome (WES) and RNA sequencing (RNA-Seq) failed to detect any *MS4A1* genetic variants or loss of CD20 mRNA that could explain the apparent loss of CD20 protein [19]. To uncover mechanisms behind this unexplained CD20 antigen loss, we performed RNA-Seq on four additional paired pretreatment/relapse formalin-fixed, paraffin-embedded (FFPE) tumor samples from a similar cohort. In the P195 pair, we failed to detect any CD20 reads either before or after treatment. In the remaining pairs, all pretreatment ("pre") samples had robust levels of CD20 mRNA levels, in excess of 500 transcripts-per-million (TPM) [**Fig. 1a**]. In the P73 pair, we detected an out-of-frame mutation in the relapse ("post") sample, which would account for the loss of protein [**Supplementary Fig. 1**]. In the P100 "post" sample, the loss of CD20 protein coincided with the loss of pan-CD20 RNA-Seq reads. However, in the P29 relapse, 300 TPM of CD20 mRNA were retained, despite complete lack of immunoreactivity for CD20 protein [**Fig. 1a, b**]. These reads contained no deleterious mutations that could explain antigen loss.

A recurrent mechanism that we and others have implicated in resistance to immunotherapies is alternative splicing [reviewed in [27,28]]. To determine if alternative splicing could explain the loss of CD20 in P29, we analyzed the RNA-Seq data using the previously developed MAJIQ package [29]. Specifically, we quantified the change in splicing of the P29-post sample relative to the -pre sample as a delta percent-spliced-in (ΔPSI) value. Within the CD20 transcript, only two alternative splicing events exceeded a ΔPSI value of +/- 0.05. The first event corresponded to a reduction (ΔPSI = -0.2251) in the usage of an alternate donor site which extended exon 5 [**Fig. 1c, top**]. Since the alternate donor site leads to a frameshift and premature stop codon, the reduction in its usage is expected to increase the expression of full-length CD20 protein and would not explain the loss of CD20 expression. The second event, on the other hand, involved exon-exon junctions that mapped to the three annotated 5' UTR variants of CD20, dubbed variants 1-3, or V1, V2 and V3 (NCBI annotation release 109.20211119 for the *MS4A1* gene) [**Fig. 1d**]. Here, the decrease in splicing from exon 1 to 3b (ΔPSI = -0.2622) [**Fig. 1c, bottom**], corresponded to a shift in splicing away from V3, and towards the extended 5' UTR isoform, V1.

To verify the V3-to-V1 shift, we used a custom pipeline to quantify the RNA-Seq reads that mapped to unique exon-exon junctions found in each of the three 5' UTR variants (V1, V2 and V3). Our custom pipeline broadly agreed with MAJIQ, detecting a shift from V3 to V1 in P29-Post relative to the Pre sample. This made V1 the predominant variant in P29-Post. By contrast, V3 predominated in the P73- and P100-Pre samples as well as in the Raji BL, OCI-Ly8 DLBCL and MEC-1 CLL cell lines, making up at least 50% of all reads, followed by V1 and V2 [**Fig. 1e**]. We also detected a fourth, albeit rare variant, with an unannotated permutation of the 5' UTR exons which we dubbed V4 [**Fig. 1d, e**].

To ensure that the distinct 5' UTRs detected by short-read RNA-Seq correspond to functional, cap-to-poly(A) mRNA and not abortive transcripts, we performed long-read Oxford Nanopore direct RNA sequencing on the Raji cell line. Direct RNA sequencing also allowed us to rule out reverse transcription artifacts, common in cDNA-seq approaches [30]. Using direct RNA sequencing, we were able to account for all three annotated variants, V1-3, and confirmed the existence of the unannotated V4 variant. In total, at least two mRNA isoforms existed for each 5' UTR variant due to two potential polyadenylation sites within the CD20 transcript [**Fig. 1f**]. We concluded that in at least in some post-mosunetuzumab FL relapses, CD20 protein loss coincided with a shift in the CD20 mRNA isoform mix towards variants with extended 5' UTR, specifically V1.

The extended 5' UTRs variants (V1 and V2) have increased transcript half-life but reduced translation efficiency, resulting in negligible protein synthesis. While alternative splicing of the exons in the 5' UTR of CD20 does not alter its coding sequence, it may affect CD20 protein levels by altering mRNA turnover and/or translation rates. Investigating these mechanisms necessitates precise quantification of splice isoforms, which is challenging since standard approaches such as Northern blotting or end-point reverse transcription-polymerase chain reaction (RT-PCR) are only semi-quantitative. To solve this problem, we designed separate 5' UTR-specific primers for quantitative real-time PCR (qPCR) that spanned

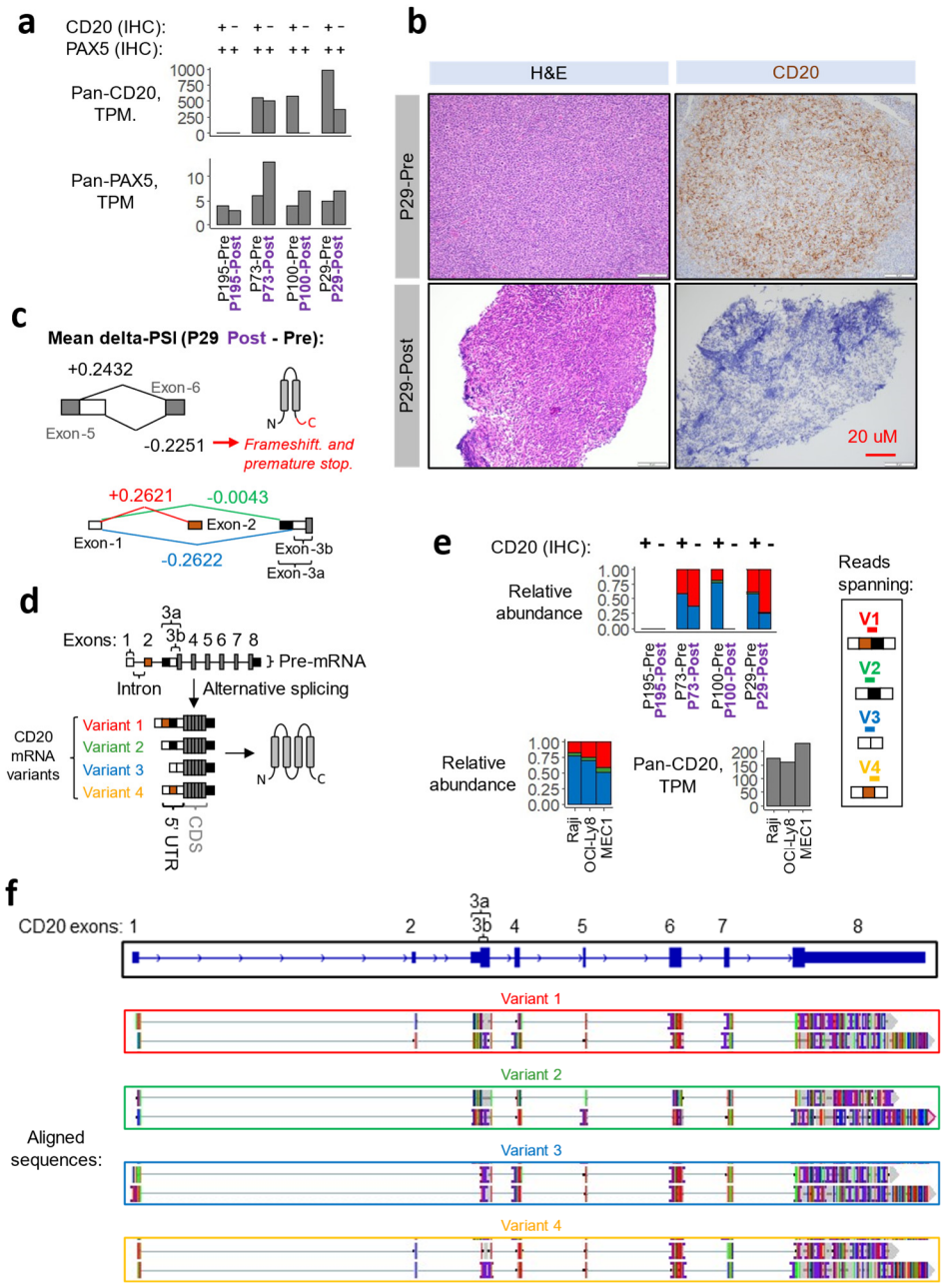


Figure 1. CD20 protein loss coincides with a shift in splicing towards an extended 5' UTR isoform, V1. a, CD20 status per IHC and transcript read abundance per RNA-Seq (expressed as TPM, or transcripts-per-million") in paired pre- and post-mosunetuzumab relapse FL. PAX5 is shown as a tumor marker. **b,** Micrographs of P29 Pre- and Post-mosunetuxumab follicular lymphoma (FL). FFPE sections were IHC stained for CD20 (brown colorimetric detection with DAB) with hematoxylin nuclear counterstain (blue). **c,** Changes in percent spliced-in values (delta-PSI) or the change in ratio between reads including or excluding exons of the P29-post sample relative to the -pre sample. The RNA-Seq data was analyzed by the MAJIQ package for all possible splicing changes in CD20 but only delta-PSI values above 0.05 are shown. **d,** Diagram depicting the CD20 pre-mRNA and the four distinct 5' UTR splice variants, V1, V2, V3 and V4. **e**, CD20 5' UTR isoform abundance in patient samples -post and -pre mosunetuzumab relapse [top], and in Raji, OCI-Ly8 and MEC-1 cells [bottom]. RNA-Seq reads that mapped to the unique exon-exon junctions found in each 5' UTR variant of CD20 were normalized to read depth and displayed as the JPM, or "junctions per million". Pan CD20 mRNA levels are indicated by TPM or "transcripts per million" values. Color-coded isoforms are depicted on the right. **f,** Raji cell mRNA sequences that align to the four distinct 5' UTR splice variants of CD20 as detected by nanopore long-read direct RNA sequencing. Alignments visualized and extracted from IGV viewer.

the unique exon-exon junctions specific for each variant. We also designed a "Pan CD20" primer pair to span a common exon-exon junction. To verify their specificity, we tested these primers against individual and admixed PCR products corresponding to all four CD20 isoforms. As expected, the amplification curves were isoform-specific, displaying linear dynamic ranges that spanned eight orders of magnitude with optimal 0.95 to 1.00 PCR efficiencies ($10^{-1/slope} - 1$) for their cognate CD20 templates [**Supplementary Fig. 2a**]. For example, compared to the cognate V3 template, V1 had to be added in 240-fold excess before the V3-specific primer pair would yield a product at the same cycle threshold (Ct) value. The RT-qPCR assay also accurately measured the relative fractions of each variant within predefined admixtures [**Supplementary Fig. 2b**] and in HEK293T cells transiently transfected with different CD20 isoforms [**Supplementary Fig. 2c**]. In Raji, OCI-Ly8 and MEC-1 cells, the RT-qPCR assay detected V1 and V3 as the two most abundant 5' UTR variants [**Fig. 2a**] which matched our RNA-Seq findings [**Fig. 1e**].

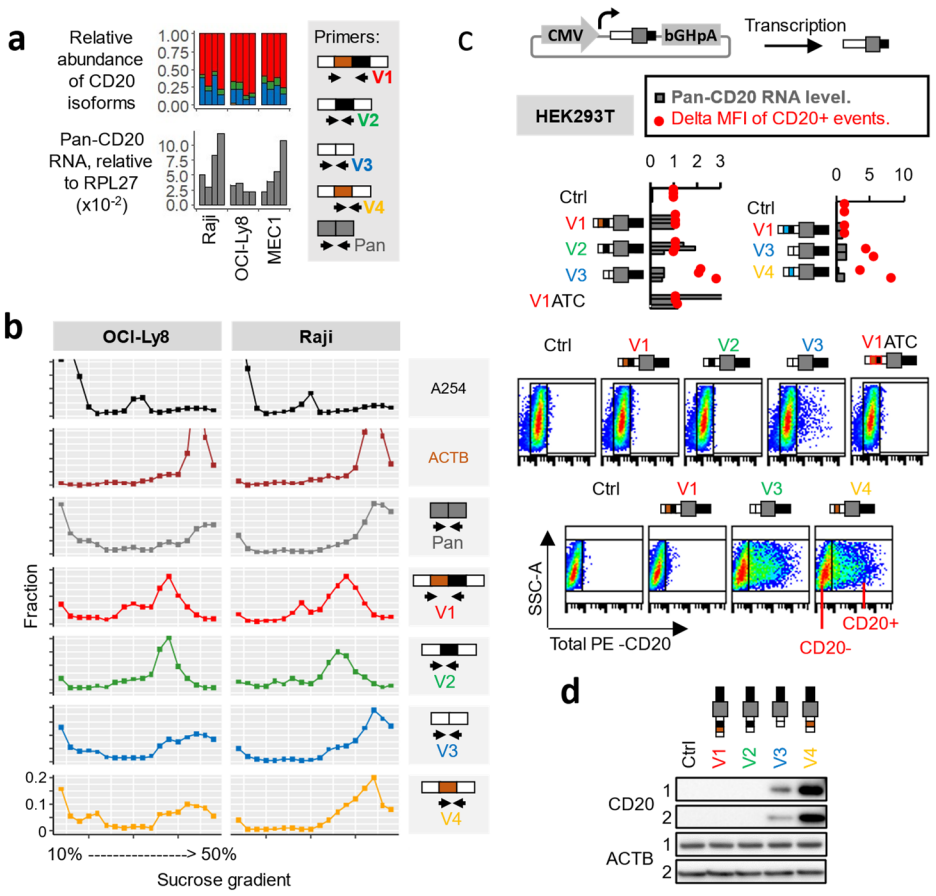


Figure 2. The extended 5' UTRs, V1 and V2, inhibit translation. a and **b**, RT-qPCR-mediated quantification of pan and 5' UTR variant-specific CD20 levels in OCI-Ly8, Raji, and MEC-1 cells (**a**) and in the sucrose gradient fractions of polysome profiling assays of OCI-Ly8 and Raji cells (**b**). In **a**, RNA levels are normalized to the reference gene, RPL27. In **b**, the top panel shows the ribosomal content measured at absorbance 254 nm while the bottom panels show the relative distribution of specific transcripts across the gradient [see Materials and Methods]. **c** and **d,** CD20 protein and RNA levels in HEK293T cells that were transiently transfected to express V1, V2, V3 or V4. Shown at the top of **c** is a diagram of the expression vector used. All ATG start codons found in the 5' UTR of V1 were mutated to ATC to form the V1ATC construct. Pan CD20 RNA levels were quantified with RT-qPCR. Total cellular CD20 protein were measured with flow cytometry and expressed as the delta median fluorescence intensity (see Materials and Methods). All values are further normalized to the V1 sample. Representative scatter plots are shown [bottom half]. In **d,** Western blots of CD20 protein with ACTB as the loading control. Each row is a repeated experiment (N=2). In **a** and **c,** each bar or dot is a repeated experiment (N=4 for **a**; N=3 for **c**).

Armed with this RT-qPCR assay, we measured the levels of each isoform in OCI-Ly8 and Raji cells during an 8-hour treatment with Actinomycin D, which inhibits transcription. Relative to V1 and V2, we detected a higher rate of mRNA decay for V3 and V4 [**Supplementary Fig. 3**]. This suggested that the former have slightly shorter half-lives than the latter. The RT-qPCR assay also allowed us to measure the rate of translation for each 5' UTR isoform in Raji and OCI-Ly8 cells using polysome profiling. Relative to V3 and V4, a smaller fraction of V1 and V2 molecules was detected in the high-density sucrose gradient fractions [**Fig. 2b**]. This pattern is indicative of a lower rate of translation for V1 and V2 relative to V3 and V4.

The reduced rate of translation for V1 and V2 should eventually lead to lower steady-state CD20 protein levels. To confirm this, we transfected HEK293T cells to transiently express all four 5' UTR isoforms (V1-4) of CD20 as full-length mRNAs. These cells had elevated CD20 mRNA levels (as measured by RT-qPCR) relative to the empty vector control (Ctrl), albeit with slightly more mRNA detected for V1 and V2 [**Fig. 2c**]. This difference is consistent with the longer half-life observed for endogenous V1 and V2 [**Supplementary Fig. 3**]. In the V3- and V4-transfected cells, the increase in mRNA was accompanied by increased CD20 protein, as measured by flow cytometry [**Fig. 2c**] and western blotting [**Fig. 2d**], relative to the Ctrl cells. Strikingly, V1- and V2-transfected cells remained CD20 protein-negative and were indistinguishable from the Ctrl cells [**Fig. 2c**]. This 5' UTR-specific difference in CD20 protein levels was independent of the 3' UTR since cells transfected with V3:del-3' (the 3' UTR-deletion mutant version of V3) still had markedly more CD20 protein staining than those transfected with V1:del-3' or V2:del-3', which remained CD20-protein negative [**Supplementary Fig. 4a**]. CD20 proteins levels of the V1:del-3' constructs were also repressed relative to a control with the GCCACC Kozak consensus sequence, kz:del-3' [**Supplementary Fig. 4b**]. Similar results were obtained when the 5' UTR isoforms were stably expressed in HEK293T cells with lentiviral vectors (V1-Puro, V2-Puro etc.) [**Fig. 3a**]. Altogether, our findings reveal that CD20 protein levels are largely dependent on the productive V3 and V4 isoforms since negligible protein amounts are synthesized from V1 and V2 despite their longer half-lives.

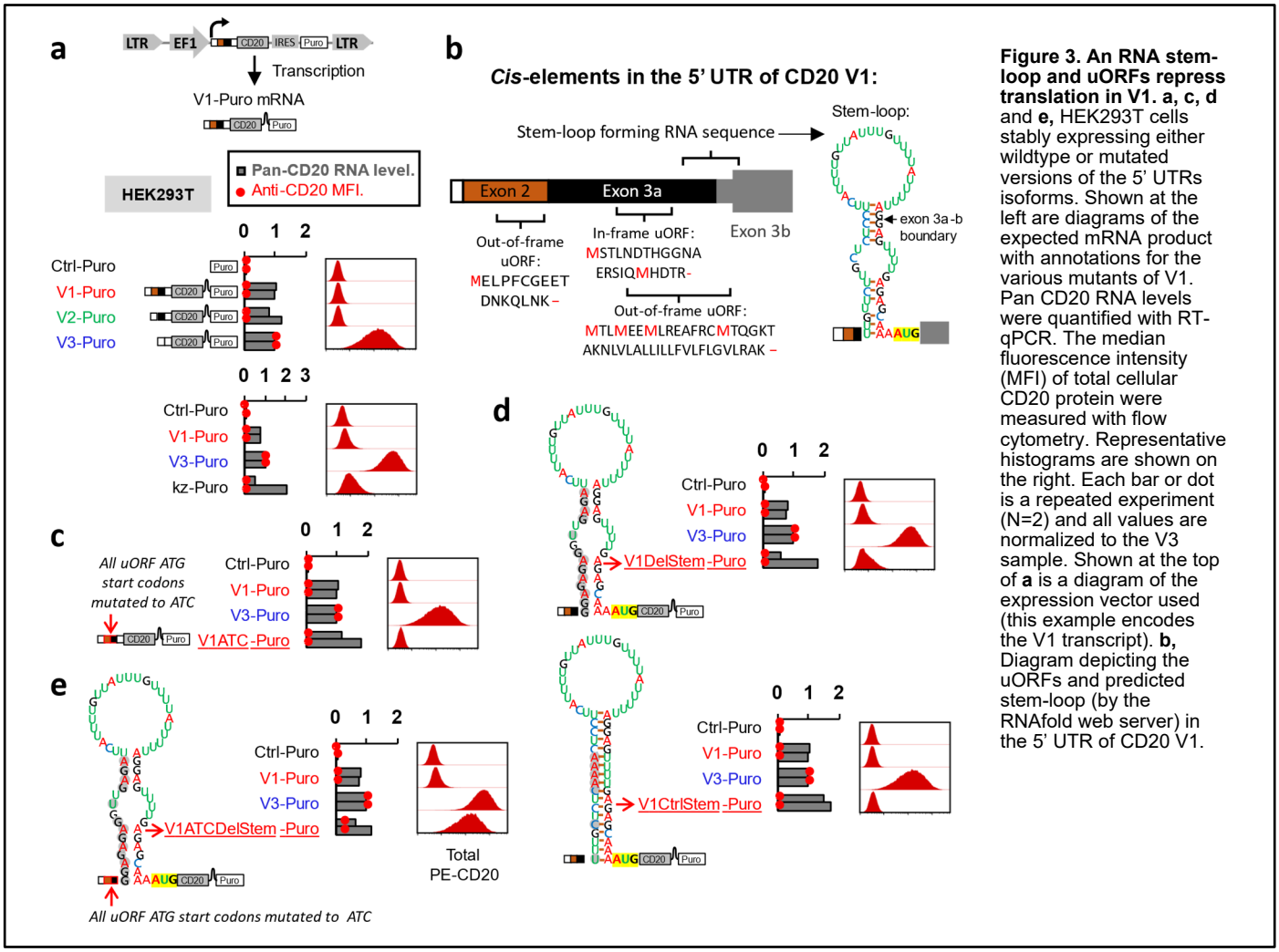


Figure 3. An RNA stem-loop and uORFs repress translation in V1. a, c, d and **e,** HEK293T cells stably expressing either wildtype or mutated versions of the 5' UTRs isoforms. Shown at the left are diagrams of the expected mRNA product with annotations for the various mutants of V1. Pan CD20 RNA levels were quantified with RT-qPCR. The median fluorescence intensity (MFI) of total cellular CD20 protein were measured with flow cytometry. Representative histograms are shown on the right. Each bar or dot is a repeated experiment (N=2) and all values are normalized to the V3 sample. Shown at the top of **a** is a diagram of the expression vector used (this example encodes the V1 transcript). **b,** Diagram depicting the uORFs and predicted stem-loop (by the RNAfold web server) in the 5' UTR of CD20 V1.

The extended 5' UTRs (V1 and V2) contain uORFs and an RNA stem-loop structure that repress translation. Inefficient translation of V1 and V2 may be due to the presence of repressive elements. We reasoned that such elements would localize in exon 3a based on two observations. Firstly, exon 3a was included in V1 and V2 but not in V3 and V4. Secondly, a truncated mutant construct which retained exon 3a but lacked exons 1 and 2 (i.e., Δex1-2) still retained the ability to repress CD20 protein levels [**Supplementary Fig. 4c**]. Within the exon 3a sequence, we found multiple open reading frames (uORFs). At the exon 3a-b boundary immediately upstream of the CD20 start codon, a stem-loop secondary structure is predicted by the RNAfold algorithm [**Fig. 3b**] [31]. Since translational repression by uORFs and stem-loops has been observed in other genes [32,33], we hypothesized that the same could be happening with CD20. To investigate this possibility, these elements were mutated and expressed in HEK293T cells. We then measured CD20 RNA and protein levels by RT-qPCR and flow cytometry, respectively. We first mutated all the start codons of the uORFs in V1 from ATG to ATC to generate V1ATC [**Fig. 2c**], V1ATC:del-3' [**Supplementary Fig. 4b**] and V1ATC-Puro [**Fig. 3c**], respectively. These mutants remained CD20 protein synthesis-deficient. We also tested V1DelStem-Puro, a V1-Puro mutant with disrupted base pairing in the stem-loop. This led to a minor recovery in CD20 protein levels compared to the V1-Puro construct in [**Fig. 3d**, top]. As a control, enhancing the stem-loop base pairing, as was done in V1Ctrlstem-Puro, led to no such recovery [**Fig. 3d**, bottom]. Unexpectedly, disrupting both the uORFs and the stem-loop led to a near complete recovery in CD20 protein levels [V1ATCDelStem-Puro construct in **Fig. 3e**]. These results suggest that the negligible CD20 protein being produced from V1 and V2 is due to the combined translational repression by the stem-loop and the uORFs.

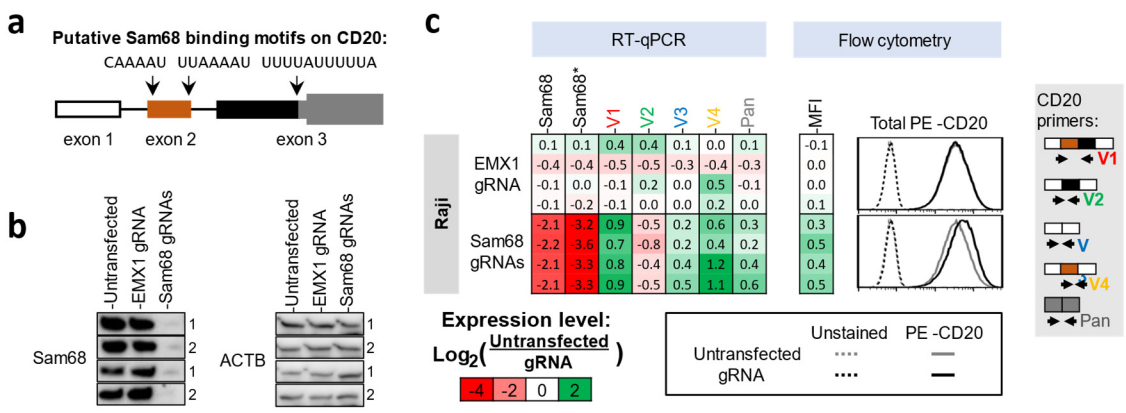


Figure 4. Sam68 knockout redirects CD20 splicing and enhances its expression. a, Diagram depicting the putative Sam68 binding motifs found within exon 2 and exon 3 of V1. **b,** and **c,** Raji cells were electroporated with a mixture of two Cas9-gRNA ribonucleoproteins targeting the CDS of Sam68. In **b,** are western blots of Sam68 with ACTB as the loading control. In **c,** mRNA levels were quantified with RT-qPCR. Two PCR primers pairs measured non-overlapping regions within the Sam68 transcript. The second primer pair is denoted with an asterisk (*). The median fluorescence intensity (MFI) of total cellular CD20 protein were determined with flow cytometry. The $\log_2$ fold change in expression levels relative to the untransfected cells are shown. In **a** and **b,** each row in the western blot image or the heatmap is a repeated experiment (N≥2).

Knockout or knockdown of Sam68 increases V3 and V4 levels and augments CD20 expression. To identify RNA-binding proteins (RBPs) regulating CD20 5' UTR splicing, we searched for RBP motifs in exons 2 and 3 of MS4A1 using the SpliceAid algorithm [34]. Our search identified several putative binding sites for Sam68 (aka KHDRBS1) [**Fig. 4a**], a splicing factor implicated in cancer [35-37] and known to bind to AU-rich sequences [38]. To determine if Sam68 regulates CD20 splicing, we electroporated Raji cells with a mixture of two Cas9-gRNA ribonucleoproteins targeting the coding sequence (CDS) of Sam68. Relative to control gRNAs targeting the irrelevant EMX1 locus and untransfected control cells, the Sam68 gRNAs successfully reduced Sam68 protein [Western blots in **Fig. 4b**] and mRNA [RT-qPCR in **Fig. 4c**, two leftmost columns]. This led to a reduction in V2 mRNA levels, an increase in V1, V3 and V4 mRNA [RT-qPCR in **Fig. 4c**, five rightmost columns], and a modest increase in CD20 protein staining [flow cytometry in **Fig. 4c**, right]. This shift in CD20 mRNA splicing and protein levels was reproduced when we knocked down Sam68 in Raji and OCI-Ly8 cells with two different shRNAs [**Supplementary Fig. 5a, b**]. Taken together, our findings suggest that while disrupting Sam68 activity leads to a modest increase in V3 and V4 by shifting CD20 splicing away from V2, this approach does little to redirect splicing away from V1, which is by far the predominant translation-deficient variant.

Ectopic expression of V3, but not V1, enhances rituximab-mediated cytotoxicity. To robustly over-produce the V1 and V3 isoforms and to study their effects on CD20 protein levels and rituximab sensitivity in OCI-Ly8 cells, we transduced these cells to stably express V1- or V3-Puro. Unlike HEK293T cells, OCI-Ly8 cells express endogenous CD20 which we detected in the empty vector (Ctrl-Puro) transduced cells. Transduction of OCI-Ly8 with V1- or V3-Puro further increased V1 or V3 mRNA relative to Ctrl-Puro cells, which we detected with RT-qPCR [**Fig. 5a**]. However, only cells transduced with the V3-Puro vector displayed elevated CD20 protein levels (as evidenced by flow cytometry) [**Fig. 5a**, right], and had reduced cell viability (via WST assay) in response to rituximab treatment [**Fig. 5b**]. These ectopic overexpression studies suggest that increasing V3 or V4 expression could potentially enhance sensitivity to rituximab-mediated cytotoxicity.

Morpholinos redirect splicing towards V3 and V4 to augment rituximab-mediated cytoxicity. To more directly modulate CD20 splicing, we tested three different phosphorodiamidate Morpholino oligomers (PMO) [39] sequences that target CD20 pre-mRNA against a non-specific control sequence pool (Ctrl-PMO). These PMOs were electroporated into Raji and OCI-Ly8 cells. Among the PMOs tested, the "5ex3-PMO" sequence (targeting the junction between CD20 intron 2 and exon 3a) was the most effective at reducing V1 and V2 mRNA while increasing V3 and V4 mRNA (via RT-qPCR) and CD20 protein levels (via flow cytometry) [**Supplementary Fig. 5c**].

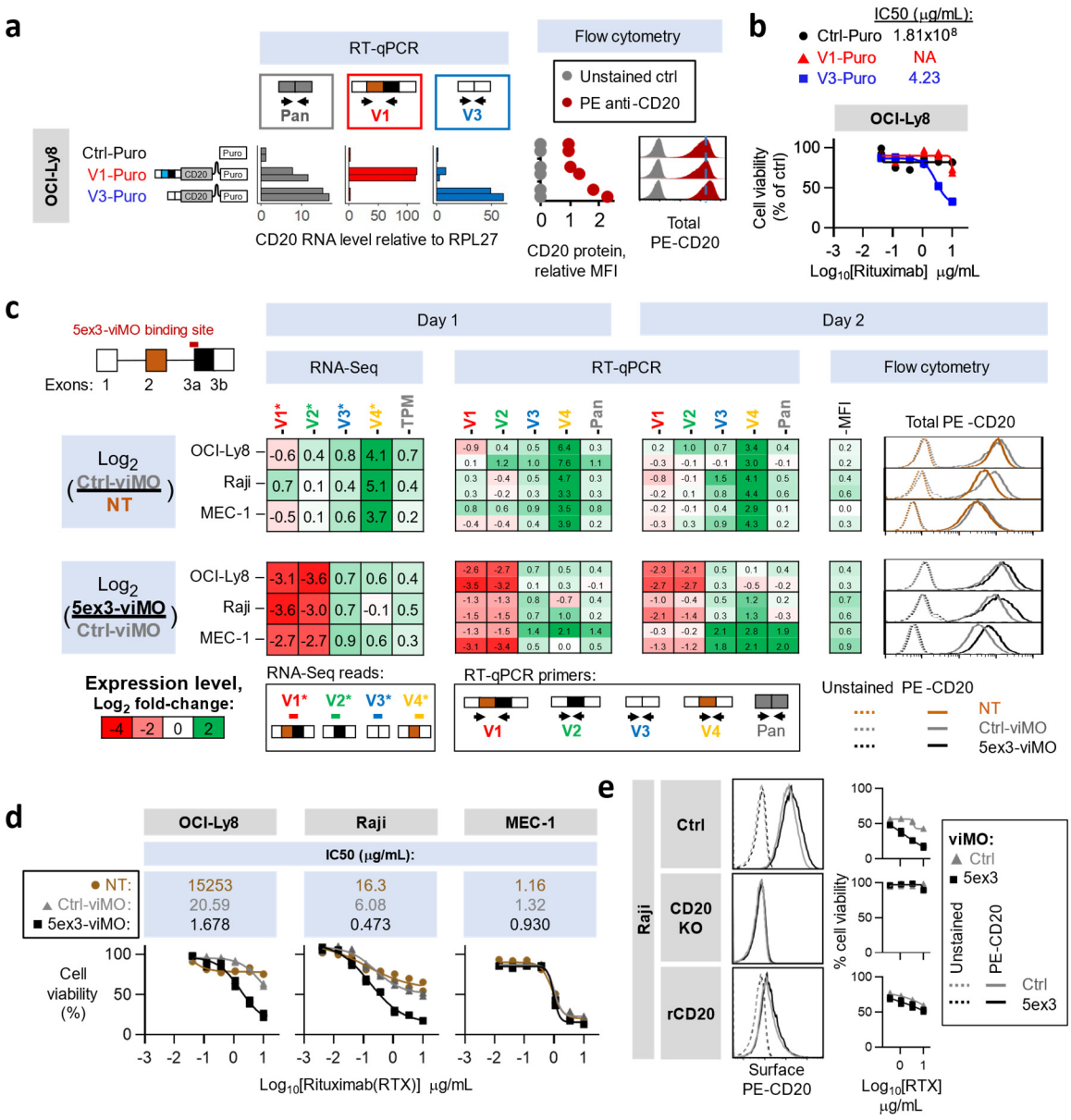


Figure 5. Modulating CD20 5' UTR splicing to increase rituximab-mediated cytotoxicity. a and **b,** OCI-Ly8 cells stably expressing the empty vector (Ctrl-Puro), V1-Puro or V3-Puro. **c** and **d,** OCI-Ly8, Raji and MEC-1 were treated with either the solvent control (NT), Ctrl-viMO or 5ex3-viMO. **e,** The same experiment performed on empty vector Ctrl, CD20 knockout (CD20KO), and CD20-reconstituted (rCD20) Raji cells. In **a** and **c,** mRNA levels were quantified with RT-qPCR. The median fluorescence intensity (MFI) of total cellular CD20 protein were determined with flow cytometry. In **b, d,** and **e,** are rituximab dose response curves. Cell viability was measured with the WST-1 assay. All values are normalized to the no rituximab control. In **c,** the $\log_2$ fold change expression levels in the Ctrl-viMO treatment relative to the no treatment (NT), and in the 5ex3-viMO treatment relative to the Ctrl-viMO, are shown. For the RNA-Seq data, V1*-4* are reads that mapped to the unique exon-exon junctions found in each 5' UTR variant of CD20. The TPM or "transcripts per million" values are based on all reads that map to CD20. In **a-e,** each bar or dot in a graph and row in a heatmap is a repeated experiment (N≥2).

Next, we custom-ordered the best-performing sequence, 5ex3, in the Vivo-Morpholino format [**Fig. 5c**, 5ex3-viMO, top left]. This format can be added directly to cultures for efficient delivery into cells [40]. We tested 5ex3-viMO on OCI-Ly8, Raji and MEC-1 cell cultures and measured the levels of CD20 mRNA (by RT-qPCR and RNA-seq) and total cellular CD20 protein (by flow cytometry). We also tested the reverse-sequence of 5ex3 (Ctrl-viMO) which induced a minor non-specific increase in V4 mRNA and CD20 staining relative to the no-Morpholino treatment (NT) cells [**Fig. 5c**, top rows]. To account for this sequence-independent effect, we normalized the 5ex3-viMO treated samples to the Ctrl-viMO treatment. Even after normalization, 5ex3-viMO treatment induced clear shifts in CD20 splicing, namely reducing V1 and V2 RNA

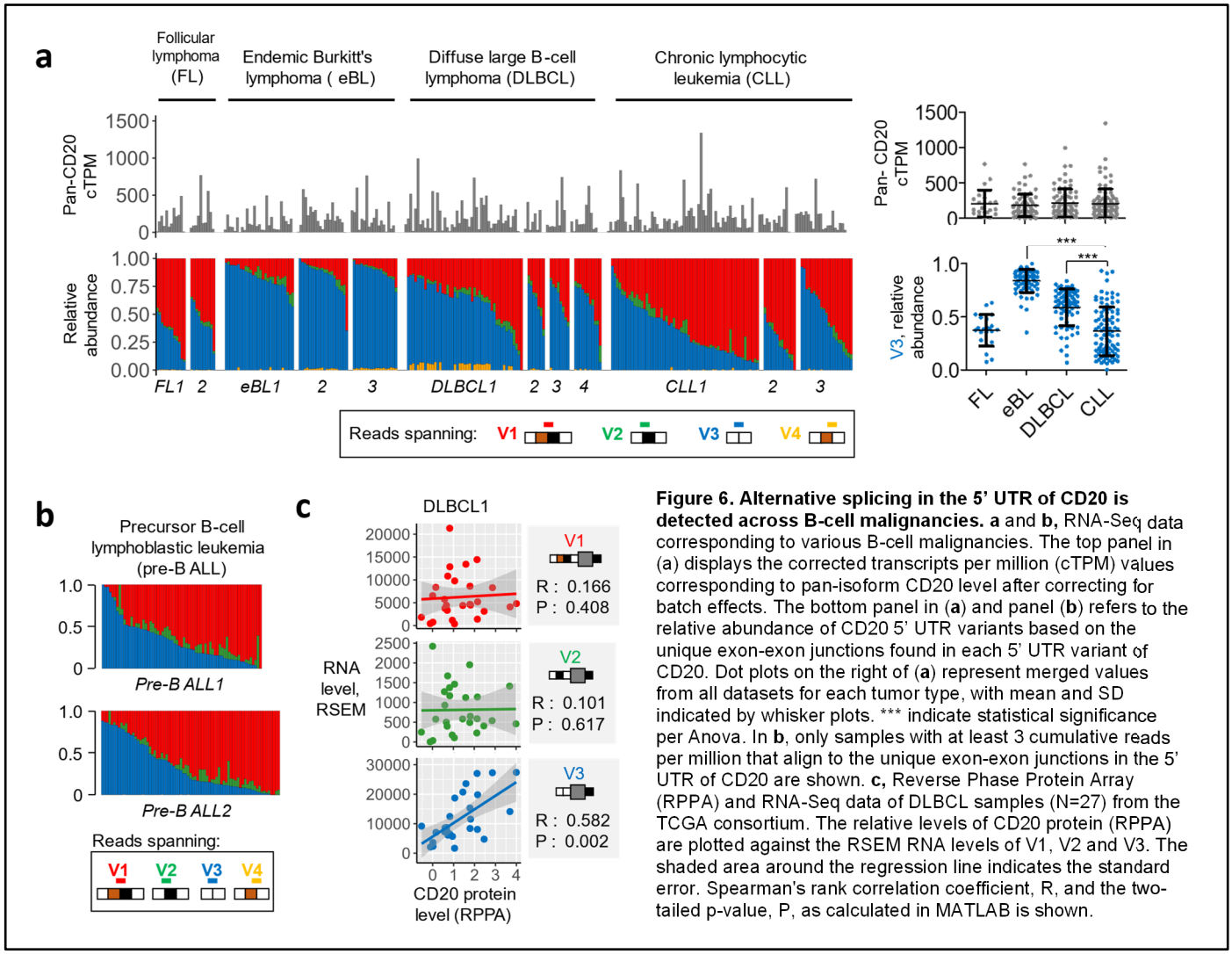


Figure 6. Alternative splicing in the 5' UTR of CD20 is detected across B-cell malignancies. a and **b,** RNA-Seq data corresponding to various B-cell malignancies. The top panel in (a) displays the corrected transcripts per million (cTPM) values corresponding to pan-isoform CD20 level after correcting for batch effects. The bottom panel in (**a**) and panel (**b**) refers to the relative abundance of CD20 5' UTR variants based on the unique exon-exon junctions found in each 5' UTR variant of CD20. Dot plots on the right of (**a**) represent merged values from all datasets for each tumor type, with mean and SD indicated by whisker plots. *** indicate statistical significance per Anova. In **b**, only samples with at least 3 cumulative reads per million that align to the unique exon-exon junctions in the 5' UTR of CD20 are shown. **c,** Reverse Phase Protein Array (RPPA) and RNA-Seq data of DLBCL samples (N=27) from the TCGA consortium. The relative levels of CD20 protein (RPPA) are plotted against the RSEM RNA levels of V1, V2 and V3. The shaded area around the regression line indicates the standard error. Spearman's rank correlation coefficient, R, and the two-tailed p-value, P, as calculated in MATLAB is shown.

levels while increasing V3 and V4 RNA. This coincided with a further increase in CD20 protein staining [**Fig. 5c**, bottom rows]. Most importantly, relative to the Ctrl-viMO treatment, 5ex3-viMO-treated Raji and OCI-Ly8 cells displayed increased sensitivity to rituximab-mediated cell cytotoxicity in WST-1 assays, with a ~1 log difference in IC50 [**Fig. 5d**]. This difference in sensitivity was further verified by propidium iodide (PI) cell viability assays [**Supplementary Fig. 6**].

To confirm that the 5ex3-viMO effect is dependent on CD20 splicing, we generated CD20 knockout Raji cells (CD20KO) that were insensitive to rituximab [**Fig. 5e**, middle row]. We then reconstituted these with V3-Puro (rCD20) which re-sensitized the cells to rituximab [**Fig. 5e**, bottom row]. In the Ctrl Raji cells, treatment with 5ex-viMO enhanced CD20 protein levels and rituximab-mediated cytotoxicity relative to Ctrl-viMO [**Fig. 5e**, top]. Since CD20KO and rCD20 cells lack functional CD20 pre-mRNA which undergoes alternative splicing, the 5ex-viMO had no effect on these cells [**Fig. 5e**, middle and bottom rows]. Altogether, these results suggest that 5ex3 morpholinos could be used to upregulate CD20 and boost responses to CD20-directed immunotherapeutics in a variety of mature B-cell neoplasms.

Inadequate expression of the V3 isoform limits CD20 protein levels in B-cell neoplasms pre-treatment. CD20 downregulation due to alternative 5' UTR exon usage does not have to occur exclusively in response to CD20-directed immunotherapy. For example, many untreated CLL cases have been reported to have lower CD20 protein relative to other mature B-cell neoplasms [20-23]. To determine to what extent splicing controls CD20 protein levels in B-lymphoid malignancies, we profiled pan-isoform CD20 mRNA levels and CD20 5'UTR exon usage across a variety of publicly available datasets, including those corresponding to the most common DLBCL and CLL histotypes. After correcting for batch effects, pan-

isoform CD20 mRNA levels were comparable across the different mature B-cell neoplasm subtypes [**Fig. 6a, top**]. However, when different 5' UTR isoforms were considered, a higher relative abundance of the non-translatable V1 variant was detected in CLL compared to other lymphoma types, except FL [**Fig. 6a, bottom**]. This is consistent with the relatively low CD20 protein levels reported for CLL samples [20-23]. V1 also predominates in many precursor B-ALL samples, which are mostly CD20-negative by flow cytometry [22,23,41] [**Fig. 6b**]. While CD20 protein expression levels were not included in the CLL and B-ALL datasets, we were able to compare RNA-seq and Reverse Phase Protein Array (RPPA) data on primary DLBCL samples from the TCGA consortium. We observed that CD20 protein levels correlate with the abundance of the V3, but not V1 or V2 mRNA isoforms [**Fig. 6c**], suggesting that alternative splicing in the 5' UTR is a major mechanism controlling expression of this important immunotherapy target.

Alternative splicing in the 5' UTR of CD20 is modulated during normal B cell development. It was possible that B-cell malignancies exploit an alternative splicing event that is already used in normal B cells to downregulate CD20 expression. Indeed, evidence of alternative splicing in the 5' UTR of CD20 can be detected in the RNA-Seq datasets corresponding to several normal B cells subsets from the bone marrow and the tonsils reported by us previously [42-44] [**Supplementary Fig. 7**]. Interestingly, the relative abundance of V3 was increased in mature and class-switched B cells, relative to naïve B cells from tonsils [**Supplementary Fig. 7**, middle panel]. This suggests that B cell activation may increase the relative abundance of V3. This effect was replicated by incubating cultured naïve tonsillar B cells with antibodies directed against IgM, resulting in SYK phosphorylation downstream of the B-cell receptor, as described previously [45,46] [**Supplementary Fig. 7**, right panel].

5' UTR isoforms of CD20 emerged in the common ancestor of great apes. Our findings thus far suggested that alternative 5' UTR exon usage governs CD20 expression levels in both normal and malignant B cells. To determine if this mechanism is evolutionarily conserved, we looked for evidence of such alternative splicing in the RNA-Seq datasets for other species. Alternative 5' UTR exon usage was readily detected in chimpanzees. By contrast, in other mammals such as the rhesus macaque, cat and mouse, only exon-exon junctions corresponding to V3 were detected [**Supplementary Fig. 8**].This suggests that translation-competent V3 may be the ancestral 5' UTR isoform of CD20 that is conserved across species and that the translation-deficient V1 and V2 isoforms emerged in the common ancestor of the great apes.

Discussion

Here, we demonstrate that in normal and malignant B-cells, human CD20 mRNA is alternatively spliced to generate up to four distinct 5' UTRs, variants 1 through 4 (V1-4) [**Fig. 1d**]. The resulting mRNA isoforms have identical coding sequences and hence are translated into the same protein product. Among the 5' UTR variants seen in primary lymphoma samples and cell lines, V1 and V3 are the most abundant and together comprise more than 80% of the total CD20 transcript pool [**Fig. 1e** and **6**]. However, ribosome profiling and reporter assays suggest that only V3 and the infrequently observed V4 were efficiently translated. The V1 and similarly structured but much rarer V2 isoforms, on the other hand, generate negligible amounts of CD20 protein due to translational inhibition by a prominent stem-loop and multiple uORFs [**Fig. 2b** and **3**]. This suggests that the vast majority of cellular CD20 protein is produced by V3, and the V3-to-V1 ratio is a key determinant of CD20 positivity in cells with active CD20 transcription.

Our findings further suggest that aberrant splicing of CD20 5' UTR exons play an important role in acquired resistance to CD20-directed immunotherapy. It has been reported previously that in a subset of FL that failed to respond to mosunetuzumab or have relapsed after initial response [19], CD20 antigen loss or downregulation occurs via an unknown mechanism, which is not mediated by the loss of CD20 mRNA or deleterious mutations in *MS4A1*. Here, we show that in these settings, epitope loss can occur via alternative splicing, which has been previously shown by us and others to regulate surface expression of CD19 and CD22 [44,47-53]. However, in the case of CD20, alternative splicing affects not the open reading frame, but rather its 5' UTR. We further show that by shifting 5' UTR exon selection away from translation-competent V3 and towards translation-deficient V1, B-cell neoplasms can dynamically downregulate CD20 protein

expression without resorting to the re-wiring of their transcriptional programs or genomic alterations. Indeed, the V3-to-V1 shift was documented in at least one paired post-mosunetuzumab sample [**Fig. 1e**]. This sample was part of a subset of FLs relapsing with CD20 antigen loss, but without concomitant *MS4A1* mutations or the loss of CD20 RNA [19]. We anticipate that as more paired samples from NCT02500407 and similarly designed clinical trials become available for analysis, the widespread use of this mechanism will become apparent.

Even prior to CD20-directed immunotherapy, a subset of B-cell neoplasms predominantly express V1 and have less (and sometimes very little) translation-competent V3 or V4. The majority of FL tumors clearly fall in that category [**Fig. 6a,** leftmost panels and bar plots on the right]. Other salient examples are CLL and *MS4A1*-transcribing pre-B ALL, which have an especially high frequency of V1-predominant samples [**Fig. 6a, b**]. Our findings resolve, at least partly, a long-standing paradox of low surface expression of CD20 in CLL[20-23,26], despite CD20 mRNA levels being comparable to those observed in DLBCL and BL [**Fig. 6a**, top panel]. Of note, low CD20 levels are known to limit the effectiveness of rituximab against CLL cell *in vitro* [24]. Unlike CLL, where *MS4A1* is universally transcribed, only a more differentiated subset of B-ALL (pre-B ALL) express this gene. In that setting, transcriptional repression and V1-biased splicing may work in concert to keep CD20 protein low [41], limiting the utility of CD20-directed immunotherapies despite the promising results in young adults on the GRAALL-2005 trial [6]. Even in DLBCL which typically have robust CD20 expression [20-23,26], a minority of V1-predominant samples can be detected [**Fig. 6a**, middles panel]. The V3-to-V1 switch also likely mediates at least some of the CD20 downregulation that is frequently observed in patients with DLBCL and is associated with inferior response to R-CHOP treatment [18]. Indeed, in DLBCL from the TCGA consortium, CD20 protein levels correlated with the abundance of the V3, but not V1 or V2 mRNA isoforms [**Fig. 6c**].

One factor possibly complicating patient selection is a high degree of interpatient variability in the relative abundance of the V1 and V3 isoforms [**Fig. 6a**], as compared to a more homogenous distribution in normal B-cell samples from different donors [**Supplementary Fig. 7**]. This observed variability is likely due to the profound dysregulation of mRNA splicing in cancers (reviewed most recently in Ref. [54]). At the same time, it suggests that at least some patients could benefit from carefully designed RNA-based therapeutics (e.g., ASOs) shifting CD20 splicing towards translation-competent V3 or V4 variants with the overall goal of increasing sensitivity to rituximab.

In search for appropriate molecular targets, we first analyzed enrichment for RBP sites in inhibitory exons of CD20 mRNA. We identified KHDRBS1/Sam68 as one such RBP and demonstrated that disrupting its expression with siRNA or shRNA led to increases in V3 and V4 isoform abundance and CD20 protein levels [**Fig. 4** and **Supplementary Fig. 5a, b**]. However, these increases were rather modest and thus unlikely to convert rituximab-resistant cells to rituximab-sensitive. More broadly speaking, there is rather limited literature on direct targeting of RBPs in cancer [55] and even fewer examples of directly interfering with individual RBS [56]. This is likely due to the propensity of individual RBPs (e.g., KHDRBS1/Sam68) to regulate multiple targets as well as to the requirement for multiple RBPs to regulate a single splicing event (e.g., CD20 5' UTR formation). On the other hand, RBP networks are very sensitive to even minor perturbations in signaling pathways, and many targeted therapies exert unintended, but profound effects on splicing of individual transcripts. For example, previous studies have demonstrated that enhanced CD20 expression and rituximab sensitivity could be achieved with protein kinase C (PKC) agonists, histone deacetylase inhibitors (HDACi), or the knockdown of FOXO1 transcription factor [57-59]. It remains to be seen if any of these strategies act at least partly by re-directing CD20 5' UTR splicing towards V3 or V4.

Our second strategy was to use Morpholino splice-switching ASOs to more specifically modulate CD20 splicing. We identified a morpholino sequence, 5ex3-viMO, that enhanced CD20 expression and rituximab-mediated cytotoxicity *in vitro* [**Supplementary Fig. 5c** and **Fig. 5c, d**]. The 5ex3-viMO did not increase CD20 expression and rituximab-mediated sensitivity in CD20 knockout Raji cells that were reconstituted with an intron-less CD20 transcript, attesting to the specificity of this approach [**Fig. 5e**]. However, a potential limiting factor for the widespread use of ASO therapies, such as 5ex3-viMO, has been the difficulty in achieving efficient oligonucleotide delivery in vivo, outside of hepatic tissues [60]. Newer approaches such as chemical modification, bioconjugation and the use of nanocarriers, may enhance the effectiveness of the

5ex3 and other ASO *in vivo*, resulting in new therapeutic modalities. While previous studies have used ASOs to increase translation by blocking inhibitory elements within the 5' UTR [61,62], our results demonstrate that ASOs could also enhance translation by modulating alternative splicing in the 5' UTR. This novel approach expands upon the pool of targets genes that can potentially be upregulated with ASOs. Besides CD20, studies have identified at least a few dozen other genes with mRNA isoforms of varying translation efficiencies due to alternative splicing in their 5' UTR sequences [63]. More broadly speaking, uORF-containing 5' UTRs are now known to regulate translation of as much as 50% of human genes in both diseased and healthy tissues [64].

Consistent with the latter finding, the bi-directional V1-to-V3 switch is not limited to transformed B cells and is also observed in normal B cell subsets from the bone marrow and tonsils [**Supplementary Fig. 7**]. This is unsurprising as B-cell malignancies have been known to exploit alternative splicing mechanisms that already occur in normal B cells to downregulate cell surface expression of CD19 and CD22 [28]. While examples of the complete loss of V3 were rare, the V3 isoform predominates in the earliest bone marrow progenitors, pro-B cells, but is later supplanted by V1 in bone marrow differentiation. Then in the secondary lymphoid organs, V3 levels start to rise again, as B cells progress through naïve, mature, and class-switched stages [**Supplementary Fig. 7,** top panel]. This pattern appears to inversely correlate with pan-isoform CD20 transcript levels, which was measured as CD20 TPM [**Supplementary Fig. 7,** bottom panel]. Thus, by boosting CD20 translation when CD20 transcript levels are low and downregulating translation when CD20 mRNA levels are high, alternative splicing in the 5' UTR of CD20 can modulate its protein levels in B-cells to guard against exaggerated immune responses and/or autoimmunity. While the complete loss of CD20 in both mice and humans has been associated with reduced B-cell signaling and deficient T cell-independent humoral responses [65,66], other studies have suggested that CD20 is an important gatekeeper for the resting state of human B-cells [67].

Interestingly, V1 and V2 are only detectable in great apes, including humans and chimpanzees, while in other mammals such as the rhesus macaque, cat and mouse, only exon-exon junctions [**Supplementary Fig. 8a**] and exon sequences [**Supplementary Fig. 8b**] corresponding to the human V3 5' UTR isoform were detected. Thus, evolutionarily speaking, the translation-competent V3 variant is the ancestral 5' UTR isoform of CD20, with alternatively spliced 5' UTR exons emerging only in the common ancestor of great apes. What advantage, if any, this novel pattern of regulation confers upon their complex immune systems [68] is an intriguing area for further studies.

Materials and Methods

Cell lines and cell culture. OCI-Ly8, Raji and HEK293T cells were a kind gifts from the laboratories of Raju Chaganti (Memorial Sloan Kettering Cancer Center, New York, NY), Paul M. Lieberman (The Wistar Institute, Philadelphia, PA) and Peter Choi (Children's Hospital of Philadelphia, Philadelphia, PA), respectively. MEC-1 cells were obtained from the ATCC (American Type Culture Collection). OCI-Ly8 and Raji cells were authenticated with short tandem repeat profiling. All cell lines were routinely tested to be Mycoplasma free via EZ-PCR Mycoplasma Detection Kit (Sartorius). OCI-Ly8, Raji, and MEC-1 cells were maintained in RPMI-1640 medium supplemented with 10% FBS, 2 mmol/L l-glutamine at 37°C and 5% CO_2. The HEK293T cells were maintained in DMEM medium supplemented with 10% FBS, 2 mmol/L l-glutamine at 37°C and 5% CO_2.

Sample preparation and RNA-Seq. In the Phase I/II trial of mosunetuzumab (NCT02500407),[19] patients with relapsed or refractory B-NHL received mosunetuzumab in 3-week cycles, for 8 to 17 cycles depending on tumor response. For our studies, clinical tumor biopsies that had been collected prior to study enrollment (before mosunetuzumab [-Pre]), and after study discontinuation for disease progression (after mostunetuzumab [-Post]) were utilized. By immunohistochemistry (IHC), these FFPE samples were dual-stained with anti-CD20 (clone L26, VENTANA) and anti-PAX5 (clone DAK-PAX5, DAKO) antibodies. Some patients were indicated by pathologist reports to have been CD20 protein positive pre- and negative post-mosunetuzumab. We obtained FFPE samples paired pre and post mosunetuzumab relapse FL tumor samples from four such patients (P29, P73, P100 and P195). RNA was isolated using the Agencourt

FormaPure Kit (Beckman Coulter LifeSciences, Indianapolis, IN). Sequencing libraries were prepared using 100ng of RNA and the TruSeq RNA Exome kit (Illumina, Inc., San Diego, CA). Libraries were sequenced on a HiSeq2500 (Illumina) to generate 2x150 bp strand-specific paired-end reads. All samples were sequenced simultaneously, in the same sequencing run. For the cell lines, total cellular RNA was isolated using the Maxwell® RSC simplyRNA Cells Kit (Promega), and 500 ng was submitted to Genewiz LLC, for paired-end RNA-Seq (Illumina, 2x150bp) after Poly(A) selection. All control and treatment samples were sequenced simultaneously, in the same sequencing run. For the tonsillar B-cells, normal human tonsils from CHOP [43] were manually homogenized with Precision Large Tissue Grinder (Covidien 3500SA) and passed through a 70 µm strainer. Cells were washed with PBS, blocked with Human TruStain (Biolegend), and stained with anti-CD19 APC (clone HiB19, Biolegend), anti-IgM FITC (Clone G20-127, BD Biosciences), anti-IgD PE (Clone IA6-2, Biolegend). Cells were washed twice in PBS and resuspended in 2% FBS and 0.1 µg/mL DAPI in PBS. CD19+IgM+IgD+, CD19+IgM+, CD19+IgM-, CD19+IgM+IgD-, and CD19+IgM-IgD- cells were sorted on the MoFlo ASTRIOS into 10% FBS RPMI-1640 media with 25 mM HEPES. For BCR stimulations, naïve tonsillar B cells (CD19+IgM+IgD+) from 3 pediatric donors were FACS-sorted and treated with isotype control (Goat F(ab')2 IgG-UNLB, Southern Biotech, Cat.#. 0110-01) or 10ug/mL anti-IgM F(ab')2 (Goat F(ab')2 Anti-Human IgM-UNLB, Southern Biotech, Cat. #. 2022-01) overnight. Viability and proximal BCR signaling were confirmed in cells 20 min after stimulation with various F(ab')2 via phospho-SYK immunoblotting (anti-pSYK, Cell Signaling, Cat. #2710S). From the tonsillar B cell samples, RNA was isolated using Qiagen RNeasy Kit (Micro). RNA integrity and concentration were determined using Eukaryote Total RNA Nano assay on BioAnalyzer. RNA-seq was performed on 10 ng–1µg of total RNA according to the Genewiz Illumina HiSeq protocol for poly(A)-selected samples (2 × 100 or 150 bp, ~350M raw paired-end reads per lane).

Nanopore long-read direct RNA sequencing. From whole Raji cells, total RNA was isolated using the Maxwell® RSC simplyRNA Cells Kit (Promega). 500ng of mRNA was isolated from total RNA using Dyna-beads mRNA DIRECT kit (Invitrogen) and used for direct RNA (SQK-RNA002, ONT) library preparation. Subsequently, each library was loaded into a Spot-ON Flow Cell R9 version (FLO-MIN106D, ONT) and sequenced in a MinION Mk1B device (ONT) for 48 hours. Raw Fast5 files were converted to fastq with guppy (v3.4.5), followed by alignment to the gencode version of hg38 (v30) using minimap2 (v2.18); the resulting bam file was visualized using the Integrative Genomics Viewer (v2.11.0).

Dataset usage. The eBL2 (PRJNA292327), eBL3(PRJNA374464) [Ref [69]], DLBCL2(PRJNA531552), DLBCL3 (PRJNA752102), DLBCL4 (PRJNA373954) [Ref [70]], FL1(PRJNA596663) [Ref [71]], FL2(PRJNA263567) [Ref [72]], CLL2(PRJNA376727) [Ref [73]], CLL3(PRJEB4498), CLL4(PRJNA792609) [Ref [74]], CLL5(PRJNA450999) [Ref [75]] and bone marrow subsets (PRJNA475684) [Ref [42-44]] datasets were from the BioProject databases of the National Center for Biotechnology Information (NCBI). The eBL1(phs001282.V2.p1) and CLL1(phs000767.V1.p1) [Ref [76]] datasets were from the dbGaP database of the NCBI. For the TCGA DLBCL samples, CD20 isoform RNA level RSEM data and protein RPPA data were obtained from the TSVdb (http://www.tsvdb.com/) [77] and cBioPortal (https://www.cbioportal.org/) [78] websites respectively.

RNA-Seq analysis. RNA-Seq reads were first trimmed to remove adapters (BBTools v38.96) and were then aligned using STAR v2.7.9a to the hg38 reference genome while providing known gene isoforms through the GENCODE annotation V32. In addition, we used STAR flags "–quantMode GeneCounts" and "–alignSJoverhangMin 8" to quantify genes and ensure spliced reads had an overhang of at least eight bases. Junction spanning reads were obtained from the STAR "*_SJ.out.tab" result files, and each entry was normalized by dividing by the total number of junction-spanning reads and multiplying by a factor of 1 million to obtain the "junctions per million" or JPM. Visualization and downstream analyses were conducted in R using the ggplot2 and tidyverse packages. TPMs for all samples were calculated using TPMCalculator [79] version 0.0.3 followed by ComBat-seq [80] batch correction as implemented in the R Bioconductor package sva (version 3.42.0).

MAJIQ analysis. For RNA splicing quantification, Majiq v2.4.dev+g85d07819 was used. The bam files from STAR and the gencode.v37 annotation file were processed with the majiq build functionality to generate the splicegraph and majiq files. Next, for each patient sample, the files from the corresponding pre- and post-relapse samples were compared with the majiq deltapsi tool to estimate changes in splicing (deltaPSI). Results were exported with voila tsv into 2 different files: one including all LSVs (flag –show-all) and another file with the flag "–changing-between-group-dpsi 0.2" to select for those LSVs (Local Splice Variations) with a change in inclusion of at least 20% between pre- and post-relapse samples.

RT-qPCR. From whole cells, total RNA was isolated using the Maxwell® RSC simplyRNA Cells Kit (Promega). To generate cDNA, 500 ng of RNA (via NanoDrop™ 2000) was primed with a mixture of Oligod$(T)_{20}$ and random hexamer primers and reverse transcribed with GoScript™ Reverse Transcriptase (Promega) with a 1-hour extension step at 42 ^{o}C. Quantitative PCR was performed using PowerSYBR Green PCR Master Mix (Life Technologies) on an Applied Biosystems Viia7 machine. According to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [81], all primers used had efficiencies exceeding 95% and amplified single PCR products from cDNA mixtures, which was then validated via Sanger sequencing. Reference gene primers used include RPL27 (123 bp amplicon on NM_000988; forward primer:ATCGCCAAGAGATCAAAGATAA; reverse primer:TCTGAAGACATCCTTATTGACG), GAPDH (227 bp amplicon on NM_002046; forward primer:GAAGGTGAAGGTCGGAGTC; reverse primer: GGAAGATGGTGATGGGATTTC) and ACTB (121 bp amplicon on NM_001101.5; forward primer: GTCATTCCAAATATGAGATGCGT; reverse primer: GCTATCACCTCCCCTGTGTG) [82,83]. RPL27 has been previously validated to be stably expressed across a multitude of different cell types and experimental conditions [82]. Other primers used include MS4A1 V1 (147 bp amplicon on NM_152866; forward primer: TGGAGACTCAGATCCTGAACAA; reverse primer: CCTCTTCCGAGTGACCTTGG), MS4A1 V2 (64 bp amplicon on NM_152867; forward primer: AGGCCTTGGAGACTCAGAACTC; reverse primer: CCTCTTCCGAGTGACCTTGG), MS4A1 V3 (126 bp amplicon on NM_021950; forward primer: GCCTGGACTACACCACTCAC; reverse primer: GCTCTCAAAACTCCTGAGTCTCC), MS4A1 V4 (121 bp amplicon on hypothesized V4 of MS4A1; forward primer: TGGAGACTCAGATCCTGAACAA; reverse primer: GTCATTTTGCTCTCAAAACTCCTTAT), pan MS4A1 (107 bp amplicon on NM_152866, NM_152867, NM_021950 and V4; forward primer: AAAGAACGTGCTCCAGACCC; reverse primer: TGTTTCAGTTAGCCCAACCACT) and pan KHDRBS1 (aka Sam68) (95 bp amplicon on NM_006559 and NM_001271878; forward primer: GGATACCTTTGCCTCCACCT; reverse primer: GCCTTCGTAGCCTTCGTAACT). The expression levels relative to the control was calculated using the 2-ΔΔCT method. To control for the possibility that Sam68 knockout or knockdown affects reference gene mRNA levels, the data in Fig. 6c and e were normalized to the mean CT of RPL27, GAPDH and ACTB. All other RT-qPCR data were normalized to RPL27. DNA contamination was routinely assessed via qPCR of negative control reverse transcription reactions which lacked reverse transcriptase (cycle thresholds above 35).

Polysome profiling. Polysome profiling was performed as described previously [50,83]. Briefly, cells were preincubated with cycloheximide (100 µg/mL, Sigma) for 15 min, and cytoplasmic lysates were prepared and fractionated by ultracentrifugation (at 246000 x g for 3 hours on a SW 41 Ti swinging bucket rotor) through 10%–50% linear sucrose gradients; 18 fractions were collected, and their absorbance at 254 nm (A254) were measured with a NanoDrop™ 2000 Spectrophotometer (Thermo Scientific). Total RNA was isolated in the presence of 4 µg glycogen (Ambion) using TRIzol™ LS Reagent (Ambion). To obtain a DNA-free sample, an additional set of TRIzol reaction tubes were prepared in parallel with no cells added. After phase separation, the upper aqueous phase of the cell-free tubes was removed and replaced with the upper aqueous phase from the cell sample tubes. The new mixture was then subjected to another round of shaking and phase separation to remove any residual DNA. To generate cDNA, the entire RNA pellet was primed with a mixture of Oligod$(T)_{20}$ and random hexamer primers and reverse transcribed with the SuperScript IV First-Strand Synthesis System (Invitrogen) with a 50 min extension step at 55 ^{o}C. Purified RNAs were stored at -80 ^{o}C for less than a week before reverse transcription. The cDNA from each fraction was used for quantitative real-time PCR analysis.

Transient expression of CD20 mRNA isoforms. The full-length sequence of MS4A1 V1 (NM_152866), V2 (NM_152867), V3 (NM_021950) and V4 were cloned into the MXS_CMV::PuroR-bGHpA plasmid, replacing the PuroR ORF in the process, to generate the V1, V2, V3 or V4 expression plasmids. MXS_CMV::PuroR-bGHpA was a gift from Pierre Neveu (RRID:Addgene_62439) [84]. Subsequent deletions and mutations were also generated via Gibson assembly. Plasmids were transfected into HEK293T cells using the ViaFect™ Transfection Reagent (Promega) according to manufacturer instructions [85].

Lentiviral transduction and antibiotic selection. Lentiviral particles were produced from HEK293T via co-transfection of the transfer plasmids with the psPAX2 and pMD2.G packaging plasmids. The psPAX2 (Addgene plasmid # 12260) and pMD2.G (Addgene plasmid # 12259) plasmids were gifts from Didier Trono. Two days after transduction with these viral particles, Raji and OCI-Ly8 cells were cultured in media containing 1 mg/mL Puromycin or Blasticidin (Invitrogen) for at least two weeks before samples were collected. HEK293T cells were cultured in media containing 10 mg/mL Puromycin.

Stable expression of CD20. The T2A sequence within the pCDH-EF1α-MCS-T2A-Puro plasmid (System Biosciences) was replaced with the IRES sequence from the pMXs-IRES-Puro (Cell Biolabs) to generate the pCDH-MCS-EF1α-IRES-Puro lentiviral transfer plasmid. The 5' UTR and CDS of MS4A1 V1, V2 and V3 were cloned into the pCDH-MCS-EF1α-IRES-Puro vector to generate the V1-, V2- and V3-Puro plasmids. Subsequent deletions and mutations were also generated via Gibson assembly. These plasmids were stably expressed in HEK293T cells

Sam68 knockout and knockdown. Alt-R™ CRISPR-Cas9 sgRNAs (Integrated DNA Technologies or IDT) targeting the TGAACCCTCTCGTGGACGTG and CAAGTACCTGCCCGAACTCA sequences within the Sam68 CDS, were complexed with Alt-R® S.p. Cas9 Nuclease V3 (IDT) and electroporated in OCI-Ly8 cells with the Neon™ Transfection System 10 µL Kit (ThermoFisher) following manufacturer recommendations. Samples were collected between 1 to 3 weeks after electroporation. The pLKO.1-puro lentiviral vectors with predesigned shRNAs which were dubbed sh-Sam68-1 (TRC clone ID: TRCN0000000046; target sequence: GACGGCAGAAATTGAGAAGAT) and sh-Sam68-2 (TRC clone ID: TRCN0000000048; target sequence: GATGAGGAGAATTACTTGGAT), were obtained from Sigma-Aldrich. We stably expressed these plasmids in Raji and OCI-Ly8 cells to knockdown Sam68.

CD20 knockout and rescue. A gRNA sequence targeting the GGATCATCAGAAGACCCCCC sequence within the CD20 CDS, and a non-specific sequence targeting GTTCCGCGTTACATAACTTA, was cloned into the lentiCRISPR v2 lentiviral vector for co-expression with Cas9. The lentiCRISPR v2 plasmid backbone was a gift from Feng Zhang (Addgene plasmid # 52961) [86]. We stably expressed these plasmids in Raji cells and used fluorescence activated cell sorting (FACS) to enrich for cells that stained negative with PE anti-human CD20 antibodies (clone 2H7, Biolegend). This sorted CD20-negative cell population were the Raji CD20 KO cells. The rV3-BSD lentiviral transfer plasmid was generated by mutating the CCTGGGGGGTCTTCTGATGATCC sequence, found within the CD20 CDS of the V3-Puro plasmid, to GTTGGGCGGACTACTTATGATTC; and replacing the puromycin resistance gene with a blasticidin resistance gene. This plasmid was stably expressed in Raji CD20 KO cells to generate the Raji CD20 rescue cells.

Morpholino treatments. The Ex2-1, Ex2-2, and 5ex3 morpholino sequences are AGTAGAGATTTTGTTCTCTCTTGTT, ATTGTCAGTCTCTTCCCCACAGAAT and CTGCTGAGTTCTGAGAAAGGAGATG respectively. Standard morpholinos (Gene Tools, LLC) were electroporated into Raji and OCI-Ly8 cells using the Neon™ Transfection System 10 µL Kit (ThermoFisher). Briefly, for every 10 µL tip, half a million live cells were electroporated in buffer R containing 10 mM of the morpholino at 1350 V and 30 ms pulse width. The Random Control 25-N morpholino (Gene Tools) was electroporated as the Ctrl-PMO. Raji, OCI-Ly8 and MEC-1 cells were incubated in cell culture media containing 10 mM Vivo Morpholinos (viMO) for 3 h before the media was replaced with fresh media. The reverse sequence of 5ex3, GTAGAGGAAAGAGTCTTGAGTCGTC, was used as the control in Ctrl-viMO. Cells were assayed for rituximab sensitivity at day 2 post morpholino treatment.

Flow cytometry. Cells were fixed and permeabilized with the BD Cytofix/Cytoperm Fixation/Permeabilization Solution Kit, and stained with PE anti-human CD20 antibodies (clone 2H7, Biolegend) before acquisition on a BD Accuri™ C6 Cytometer (BD Biosciences). Manual gating was performed on the FlowJo Software v.10 (BD Biosciences).

Western blotting. Total cell lysates were prepared in Laemmli Sample Buffer (Bio-Rad) with 50 mM DTT, resolved on a 4–15% precast polyacrylamide gel (Mini-PROTEAN® TGX™, Bio-Rad) in Tris/glycine/SDS running buffer (Bio-Rad), and then transferred onto a PVDF membrane (Millipore Sigma, Immobilon-P, 0.45 µm) via the Mini Trans-Blot® Cell system (Bio-Rad). These membranes were then probed with antibodies against CD20 (clone E7B7T, Cell Signaling Technology), ACTB (clone 8H10D10, Cell Signaling Technology), and Sam68 (clone 7-1, Santa Cruz Biotechnology, Inc.); followed by anti-rabbit or -mouse horseradish peroxidase–linked secondary antibodies (Cell Signaling Technology). Western blot chemiluminescent signals were captured with an G:Box Chemi XX6 Gel Doc (Syngene).

Evaluation of rituximab sensitivity. To cell cultures containing a million live cells per mL, rituximab (Roche) was added. After 15 min, normal human serum (Complement Technology) was then added to a 20% (v/v) concentration as a source of complement, for complement-dependent-cytotoxicity. After another 2 h incubation, cell viability was evaluated by PI staining before acquisition on a BD Accuri™ C6 Cytometer (BD Biosciences), or by the WST-1 cell proliferation reagent according to the manufacturer's protocol (Sigma-Aldrich). The WST-1 absorbance signal (A450nm – A690nm) was measured using the BioTek Synergy 2 instrument. The blank control well contains the same volume of culture medium and WST-1 but no cells. The % cell viability was calculated as (absorbance at X concentration of rituximab − absorbance of blank control well)/(absorbance with no rituximab – absorbance of blank control well). Dose response curves and IC50 values were calculated using the nonlinear regression curve fit function in GraphPad Prism 5 (v5.01) with the top and bottom of the curve constrained to 100 and 0% respectively.

Acknowledgements

We are grateful to all members of the ATT laboratory, in particular Priyanka Sehgal, for many helpful discussions. This work was supported by grants from the National Institutes of Health (U01 CA232563 and U01 CA232563-S3 to ATT and YB, U01 CA232486 and U01 CA243072 to SKT, and T32 CA009615 to SZ), United States Department of Defense (CA180683P1 to SKT), The V Foundation for Cancer Research (T2018-014 to ATT), The Emerson Collective (886246066 to ATT), and Mildred L. Roeckle (ATT) and Joshua Kahan (SKT) Endowed Chairs at Children's Hospital of Philadelphia. .

References

1. Pavlasova, G. & Mraz, M. The regulation and function of CD20: an "enigma" of B-cell biology and targeted therapy. *Haematologica* **105**, 1494-1506 (2020).
2. Pierpont, T.M., Limper, C.B. & Richards, K.L. Past, Present, and Future of Rituximab-The World's First Oncology Monoclonal Antibody Therapy. *Front Oncol* **8**, 163 (2018).
3. Grillo-Lopez, A.J., Hedrick, E., Rashford, M. & Benyunes, M. Rituximab: Ongoing and future clinical development. *Semin Oncol* **29**, 105-112 (2002).
4. Marks, V.A., Latham, S.R. & Kishore, S.P. On Essentiality and the World Health Organization's Model List of Essential Medicines. *Ann Glob Health* **83**, 637-640 (2017).
5. Minard-Colin, V. *et al.* Rituximab for High-Risk, Mature B-Cell Non-Hodgkin's Lymphoma in Children. *N Engl J Med* **382**, 2207-2219 (2020).
6. Maury, S. *et al.* Rituximab in B-Lineage Adult Acute Lymphoblastic Leukemia. *N Engl J Med* **375**, 1044-53 (2016).
7. Sun, L.L. *et al.* Anti-CD20/CD3 T cell-dependent bispecific antibody for the treatment of B cell malignancies. *Sci Transl Med* **7**, 287ra70 (2015).

8. Budde, L.E. *et al.* Safety and efficacy of mosunetuzumab, a bispecific antibody, in patients with relapsed or refractory follicular lymphoma: a single-arm, multicentre, phase 2 study. *Lancet Oncol* **23**, 1055-1065 (2022).
9. Kang, C. Mosunetuzumab: First Approval. *Drugs* **82**, 1229-1234 (2022).
10. Glennie, M.J., French, R.R., Cragg, M.S. & Taylor, R.P. Mechanisms of killing by anti-CD20 monoclonal antibodies. *Mol Immunol* **44**, 3823-37 (2007).
11. Nowakowski, G.S. *et al.* Beyond RCHOP: A Blueprint for Diffuse Large B Cell Lymphoma Research. *J Natl Cancer Inst* **108**, djw257 (2016).
12. Brown, J.R. *et al.* The Role of Rituximab in Chronic Lymphocytic Leukemia Treatment and the Potential Utility of Biosimilars. *Oncologist* **23**, 288-296 (2018).
13. Haidar, J.H. *et al.* Loss of CD20 expression in relapsed lymphomas after rituximab therapy. *Eur J Haematol* **70**, 330-2 (2003).
14. Hiraga, J. *et al.* Down-regulation of CD20 expression in B-cell lymphoma cells after treatment with rituximab-containing combination chemotherapies: its prevalence and clinical significance. *Blood* **113**, 4885-93 (2009).
15. Miyoshi, H. *et al.* Comparison of CD20 expression in B-cell lymphoma between newly diagnosed, untreated cases and those after rituximab treatment. *Cancer Sci* **103**, 1567-73 (2012).
16. D'Auria, F. *et al.* Modulation of CD20 antigen expression after rituximab treatment: a retrospective study in patients with chronic lymphocytic leukemia. *Clin Ther* **32**, 1911-6 (2010).
17. Yu, H. *et al.* Repeated loss of target surface antigen after immunotherapy in primary mediastinal large B cell lymphoma. *Am J Hematol* **92**, E11-E13 (2017).
18. Johnson, N.A. *et al.* Diffuse large B-cell lymphoma: reduced CD20 expression is associated with an inferior survival. *Blood* **113**, 3773-80 (2009).
19. Schuster, S.J. *et al.* Characterization of CD20 expression loss as a mechanism of resistance to mosunetuzumab in patients with relapsed/refractory B-cell non-Hodgkin lymphomas. *Journal of Clinical Oncology* **40**, 7526-7526 (2022).
20. Almasri, N.M., Duque, R.E., Iturraspe, J., Everett, E. & Braylan, R.C. Reduced expression of CD20 antigen as a characteristic marker for chronic lymphocytic leukemia. *Am J Hematol* **40**, 259-63 (1992).
21. Ginaldi, L. *et al.* Levels of expression of CD19 and CD20 in chronic B cell leukaemias. *J Clin Pathol* **51**, 364-9 (1998).
22. Perz, J., Topaly, J., Fruehauf, S., Hensel, M. & Ho, A.D. Level of CD 20-expression and efficacy of rituximab treatment in patients with resistant or relapsing B-cell prolymphocytic leukemia and B-cell chronic lymphocytic leukemia. *Leuk Lymphoma* **43**, 149-51 (2002).
23. Olejniczak, S.H., Stewart, C.C., Donohue, K. & Czuczman, M.S. A quantitative exploration of surface antigen expression in common B-cell malignancies using flow cytometry. *Immunol Invest* **35**, 93-114 (2006).
24. Golay, J. *et al.* CD20 levels determine the in vitro susceptibility to rituximab and complement of B-cell chronic lymphocytic leukemia: further regulation by CD55 and CD59. *Blood* **98**, 3383-9 (2001).
25. Yu, D., Dews, M., Park, A., Tobias, J.W. & Thomas-Tikhonenko, A. Inactivation of Myc in murine two-hit B lymphomas causes dormancy with elevated levels of interleukin 10 receptor and CD20: implications for adjuvant therapies. *Cancer Res* **65**, 5454-61 (2005).
26. Sarro, S.M. *et al.* Quantification of CD20 mRNA and protein levels in chronic lymphocytic leukemia suggests a post-transcriptional defect. *Leuk Res* **34**, 1670-3 (2010).
27. Zheng, S., Asnani, M. & Thomas-Tikhonenko, A. Escape From ALL-CARTaz: Leukemia Immunoediting in the Age of Chimeric Antigen Receptors. *Cancer J* **25**, 217-222 (2019).
28. Bourcier, J. & Abdel-Wahab, O. Splicing-Mediated Antigen Escape from Immunotherapy for B-cell Malignancies. *Blood Cancer Discov* **3**, 87-89 (2022).
29. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752 (2016).
30. Schulz, L. *et al.* Direct long-read RNA sequencing identifies a subset of questionable exitrons likely arising from reverse transcription artifacts. *Genome Biol* **22**, 190 (2021).
31. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
32. Morris, D.R. & Geballe, A.P. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**, 8635-42 (2000).

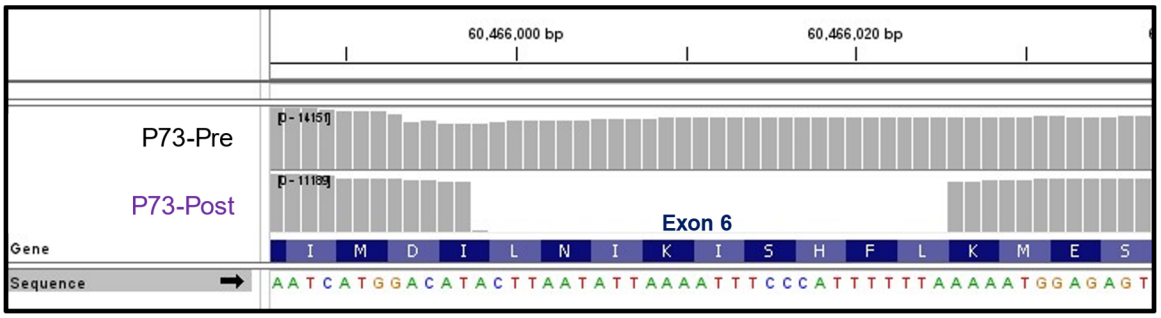
33. Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* **9**, 5134-42 (1989).
34. Piva, F., Giulietti, M., Nocchi, L. & Principato, G. SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* **25**, 1211-3 (2009).
35. Lin, Q., Taylor, S.J. & Shalloway, D. Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. *J Biol Chem* **272**, 27274-80 (1997).
36. Pedrotti, S. *et al.* The splicing regulator Sam68 binds to a novel exonic splicing silencer and functions in SMN2 alternative splicing in spinal muscular atrophy. *EMBO J* **29**, 1235-47 (2010).
37. Bielli, P., Busa, R., Paronetto, M.P. & Sette, C. The RNA-binding protein Sam68 is a multifunctional player in human cancer. *Endocr Relat Cancer* **18**, R91-R102 (2011).
38. Itoh, M., Haga, I., Li, Q.H. & Fujisawa, J. Identification of cellular mRNA targets for RNA-binding protein Sam68. *Nucleic Acids Res* **30**, 5452-64 (2002).
39. Morcos, P.A. Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. *Biochem Biophys Res Commun* **358**, 521-7 (2007).
40. Morcos, P.A., Li, Y. & Jiang, S. Vivo-Morpholinos: a non-peptide transporter delivers Morpholinos into a wide array of mouse tissues. *Biotechniques* **45**, 613-4, 616, 618 passim (2008).
41. Jeha, S. *et al.* Prognostic significance of CD20 expression in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* **108**, 3302-4 (2006).
42. Black, K.L. *et al.* Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Res* **46**, 11357-11369 (2018).
43. Yang, S.Y. *et al.* FBXW7beta isoform drives transcriptional activation of the proinflammatory TNF cluster in human pro-B cells. *Blood Adv* (2022).
44. Zheng, S. *et al.* Modulation of CD22 Protein Expression in Childhood Leukemia by Pervasive Splicing Aberrations: Implications for CD22-Directed Immunotherapies. *Blood Cancer Discov* **3**, 103-115 (2022).
45. Chung, E.Y. *et al.* CD19 is a major B cell receptor-independent activator of MYC-driven B-lymphomagenesis. *J Clin Invest* **122**, 2257-66 (2012).
46. Psathas, J.N. *et al.* The Myc-miR-17-92 axis amplifies B-cell receptor signaling via inhibition of ITIM proteins: a novel lymphomagenic feed-forward loop. *Blood* **122**, 4220-9 (2013).
47. Sotillo, E. *et al.* Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. *Cancer Discov* **5**, 1282-95 (2015).
48. Fischer, J. *et al.* CD19 Isoforms Enabling Resistance to CART-19 Immunotherapy Are Expressed in B-ALL Patients at Initial Diagnosis. *J Immunother* **40**, 187-195 (2017).
49. Bagashev, A. *et al.* CD19 Alterations Emerging after CD19-Directed Immunotherapy Cause Retention of the Misfolded Protein in the Endoplasmic Reticulum. *Mol Cell Biol* **38**, e00383-18 (2018).
50. Asnani, M. *et al.* Retention of CD19 intron 2 contributes to CART-19 resistance in leukemias with subclonal frameshift mutations in CD19. *Leukemia* **34**, 1202-1207 (2020).
51. Rabilloud, T. *et al.* Single-cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse after CAR-T therapy. *Nat Commun* **12**, 865 (2021).
52. Henry, C. *et al.* Identification of an alternative CD20 transcript variant in B-cell malignancies coding for a novel protein associated to rituximab resistance. *Blood* **115**, 2420-9 (2010).
53. Gamonet, C. *et al.* New CD20 alternative splice variants: molecular identification and differential expression within hematological B cell malignancies. *Exp Hematol Oncol* **5**, 7 (2015).
54. Bradley, R.K. & Anczukow, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* (2023).
55. Lee, S.C. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat Med* **22**, 976-86 (2016).
56. Wang, E. *et al.* Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell* **35**, 369-384 e7 (2019).
57. Pyrzynska, B. *et al.* FOXO1 promotes resistance of non-Hodgkin lymphomas to anti-CD20-based therapy. *Oncoimmunology* **7**, e1423183 (2018).
58. Bobrowicz, M., Dwojak, M., Bojarczuk, K., Winiarska, M. & Golab, J. HDAC6 Inhibition Increases CD20 Level and Improves The Efficacy Of Anti-CD20 Monoclonal Antibodies. *Blood* **122**, 4406-4406 (2013).

59. Wojciechowski, W., Li, H., Marshall, S., Dell'Agnola, C. & Espinoza-Delgado, I. Enhanced expression of CD20 in human tumor B cells is controlled through ERK-dependent mechanisms. *J Immunol* **174**, 7859-68 (2005).
60. Roberts, T.C., Langer, R. & Wood, M.J.A. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov* **19**, 673-694 (2020).
61. Liang, X.H. *et al.* Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels. *Nucleic Acids Res* **45**, 9528-9546 (2017).
62. Liang, X.H. *et al.* Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat Biotechnol* **34**, 875-80 (2016).
63. Floor, S.N. & Doudna, J.A. Tunable protein synthesis by transcript isoforms in human cells. *Elife* **5**(2016).
64. Lee, D.S.M. *et al.* Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun* **12**, 1515 (2021).
65. Uchida, J. *et al.* Mouse CD20 expression and function. *Int Immunol* **16**, 119-29 (2004).
66. Kuijpers, T.W. *et al.* CD20 deficiency in humans results in impaired T cell-independent antibody responses. *J Clin Invest* **120**, 214-22 (2010).
67. Klasener, K. *et al.* CD20 as a gatekeeper of the resting state of human B cells. *Proc Natl Acad Sci U S A* **118**, e2021342118 (2021).
68. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
69. Abate, F. *et al.* Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma. *PLoS Pathog* **11**, e1005158 (2015).
70. Park, H.Y. *et al.* Whole-exome and transcriptome sequencing of refractory diffuse large B-cell lymphoma. *Oncotarget* **7**, 86433-86445 (2016).
71. Parsa, S. *et al.* The serine hydroxymethyltransferase-2 (SHMT2) initiates lymphoma development through epigenetic tumor suppressor silencing. *Nat Cancer* **1**, 653-664 (2020).
72. Koues, O.I. *et al.* Enhancer sequence variants and transcription-factor deregulation synergize to construct pathogenic regulatory circuits in B-cell lymphoma. *Immunity* **42**, 186-98 (2015).
73. Agrawal, A.A. *et al.* Novel SF3B1 in-frame deletions result in aberrant RNA splicing in CLL patients. *Blood Adv* **1**, 995-1000 (2017).
74. Ghia, E.M. *et al.* High expression level of ROR1 and ROR1-signaling associates with venetoclax resistance in chronic lymphocytic leukemia. *Leukemia* **36**, 1609-1618 (2022).
75. Wernig-Zorc, S. *et al.* Global distribution of DNA hydroxymethylation and DNA methylation in chronic lymphocytic leukemia. *Epigenetics Chromatin* **12**, 4 (2019).
76. Kaymaz, Y. *et al.* Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences. *Mol Cancer Res* **15**, 563-576 (2017).
77. Sun, W. *et al.* TSVdb: a web-tool for TCGA splicing variants analysis. *BMC Genomics* **19**, 405 (2018).
78. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
79. Vera Alvarez, R., Pongor, L.S., Marino-Ramirez, L. & Landsman, D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960-1962 (2019).
80. Zhang, Y., Parmigiani, G. & Johnson, W.E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).
81. Bustin, S.A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611-22 (2009).
82. de Jonge, H.J. *et al.* Evidence based selection of housekeeping genes. *PLoS One* **2**, e898 (2007).
83. Ang, Z. *et al.* Novel AU-rich proximal UTR sequences (APS) enhance CXCL8 synthesis upon the induction of rpS6 phosphorylation. *PLoS Genet* **15**, e1008077 (2019).
84. Sladitschek, H.L. & Neveu, P.A. MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. *PLoS One* **10**, e0124958 (2015).
85. Bodbin, S.E., Denning, C. & Mosqueira, D. Transfection of hPSC-Cardiomyocytes Using Viafect Transfection Reagent. *Methods Protoc* **3**, 57 (2020).
86. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784 (2014).

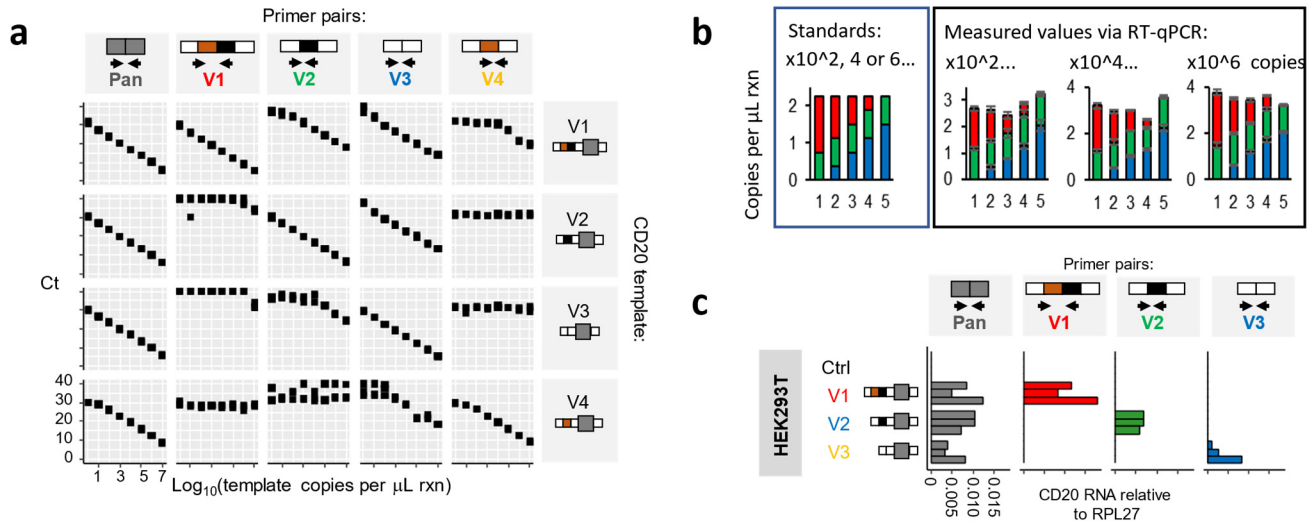
Supplementary data for Ang et al

Alternative splicing of its 5' untranslated region controls CD20 mRNA translation and enables resistance to CD20-directed immunotherapies

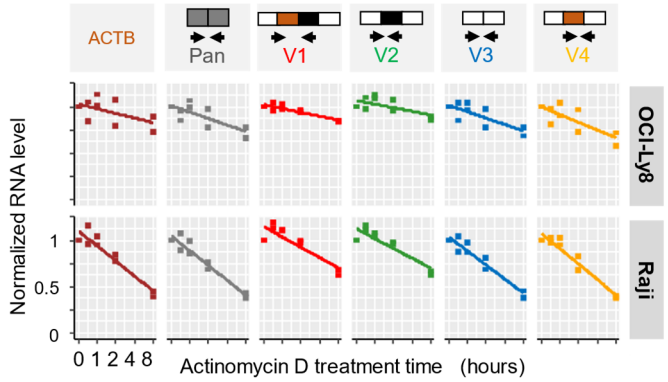
Figures S1 - S8



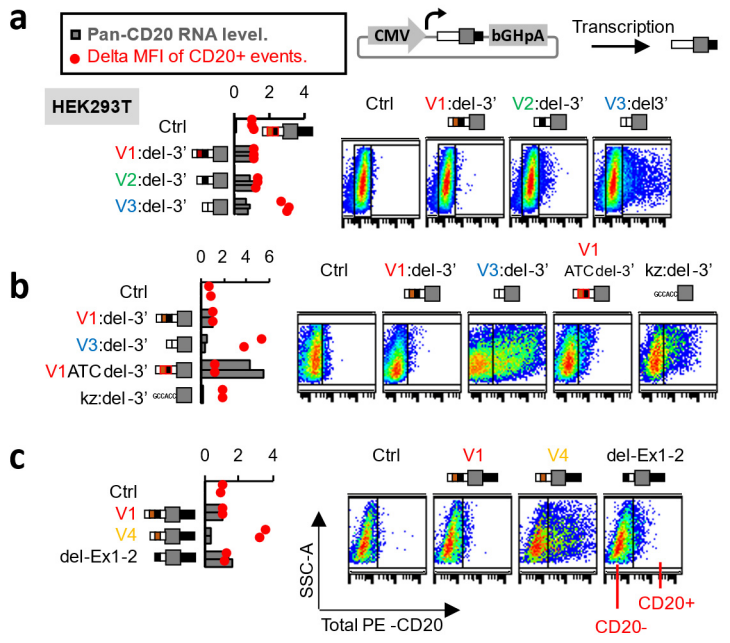
Supplementary Fig. 1: Twenty-eight nucleotide deletion in exon 6 of the CD20 transcript in the P73-Post sample. RNA-Seq read alignments were visualized in and extracted from the IGV viewer.



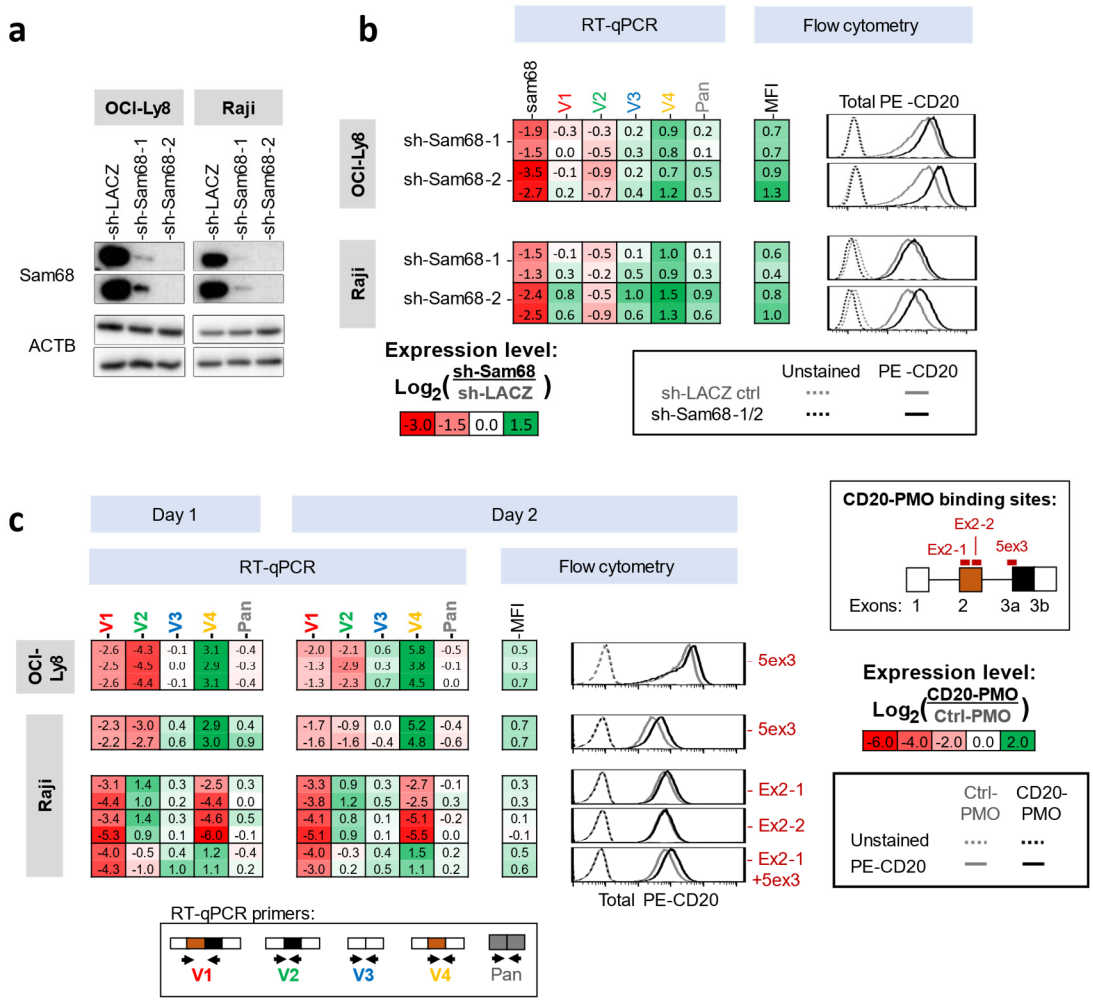
Supplementary Fig. 2: The 5' UTR-specific primer pairs amplify their intended targets efficiently and specifically. a and **b,** Real-time PCR quantification of single (**a**) and mixed template standards (**b**) using 5' UTR variant-specific (V1, V2, V3 and V4) and pan-CD20 primer pairs. **c**, RT-qPCR-mediated quantification of pan and 5' UTR variant-specific CD20 levels in HEK293T cells after transient transfection with either the MXS-CMV-V1, -V2 or -V3 plasmids. In **b**, error bars represent the standard deviation of 3 technical replicates. In **c**, RNA levels are normalized to the reference gene, RPL27. Each bar or dot is a repeated experiment (N=2 for **a**; N=3 for **c**).



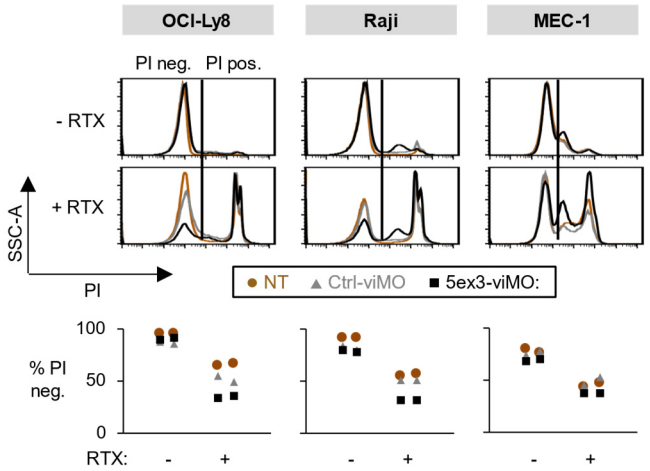
Supplementary Fig. 3: The extended 5' UTRs, V1 and V2, have longer half-lives. RT-qPCR-mediated quantification of pan- and 5' UTR variant-specific CD20 levels in 10 mg/mL actinomycin D-treated OCI-Ly8 and Raji cells. RNA levels are normalized to the 0-hour time-point. Each dot is a repeated experiment (N=2).



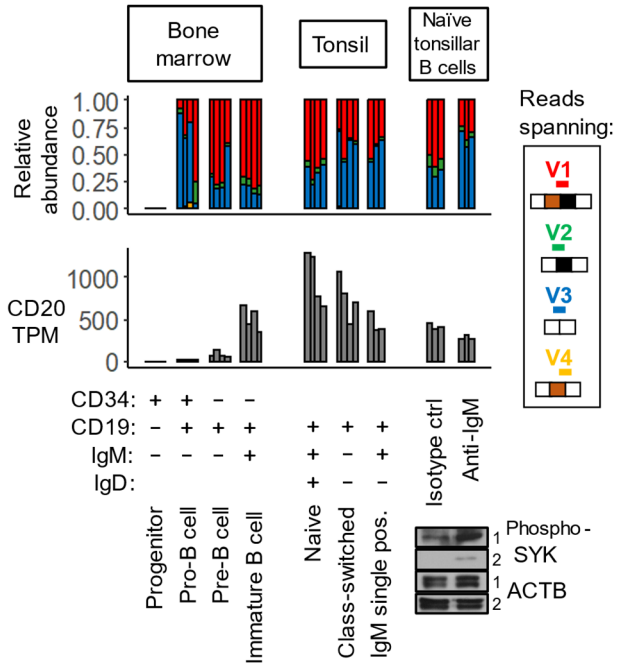
Supplementary Fig. 4: Elements within exon 3a inhibit translation. a, b and **c,** CD20 protein and RNA levels in HEK293T cells that were transiently transfected to express CD20 5' UTR isoforms. Shown at the top right corner of **a** is a diagram of the expression vector used. Constructs denoted with the "del-3'" annotation lack the CD20 3' UTR sequence. V1ATC denotes a V1 mutant where all ATG start codons found in the 5' UTR were mutated to ATC. Exons 1 and 2 were deleted from V1 to form del-Ex1-2. Pan CD20 RNA levels were quantified with RT-qPCR. Total cellular CD20 protein were measured with flow cytometry and expressed as the delta median fluorescence intensity (delta MFI = MFI of CD20+ events – MFI of CD20- events). Representative scatter plots are shown on the right. All values are normalized to the V1 sample. Each bar or dot in a graph is a repeated experiment (N=3 for **a**; N=2 for **b** and **c**).



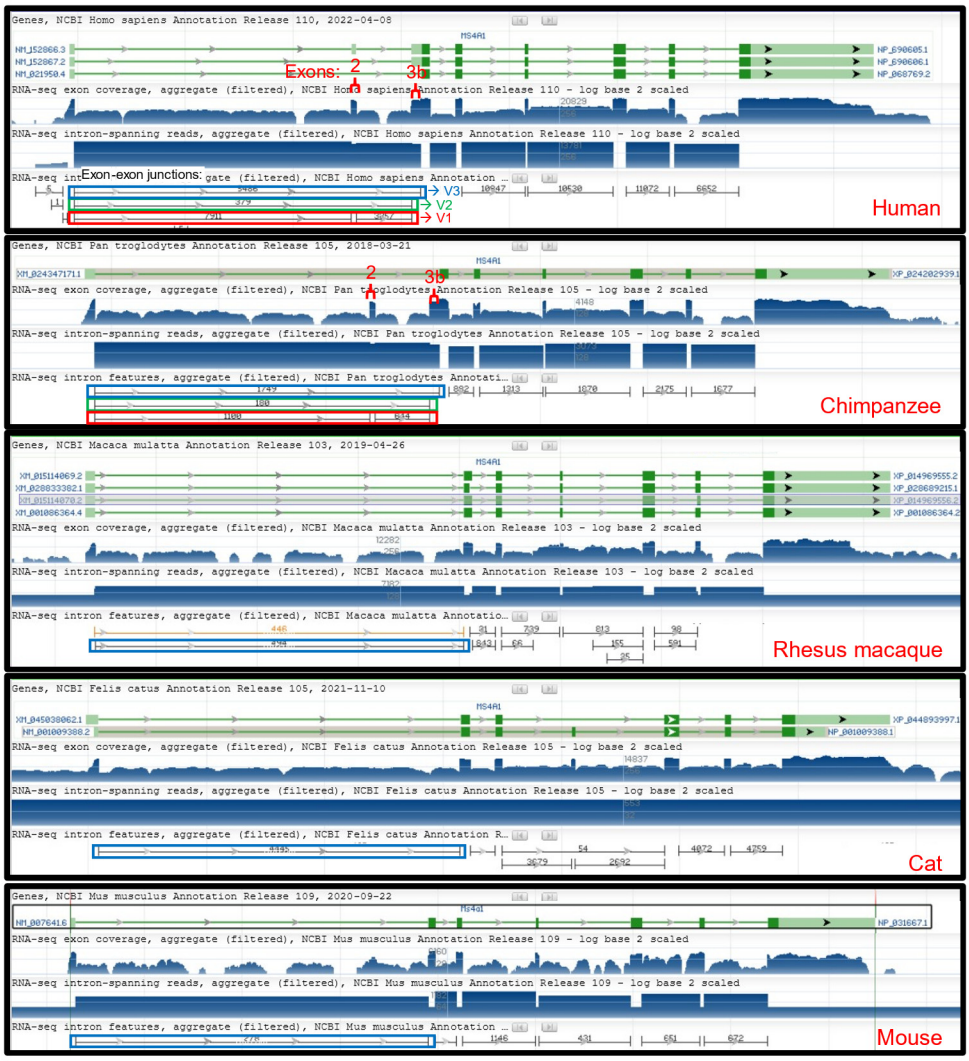
Supplementary Fig. 5: Modulating CD20 splicing to enhance its expression. a, and **b,** OCI-Ly8 and Raji cells were transduced to stably express either sh-Sam68-1 or sh-Sam68-2 or the sh-LACZ negative control. **c,** OCI-Ly8 and Raji cells were electroporated with various CD20-targeted PMOs (i.e.: Ex2-1, Ex2-2 and 5ex3). In **a,** are western blots of Sam68 with ACTB as the loading control. Each row is a repeated experiment (N≥2). In **b** and **c,** mRNA levels were quantified with RT-qPCR. The median fluorescence intensity (MFI) of total cellular CD20 protein were determined with flow cytometry. Each heatmap row is a repeated experiment (N≥2). Representative histograms are shown. In **b**, the $\log_2$ fold change in expression levels relative to cells expressing the sh-LACZ control are shown. In **c,** the $\log_2$ fold change expression levels in the in the 5ex3-PMO treatment relative to the Ctrl-PMO, are shown.

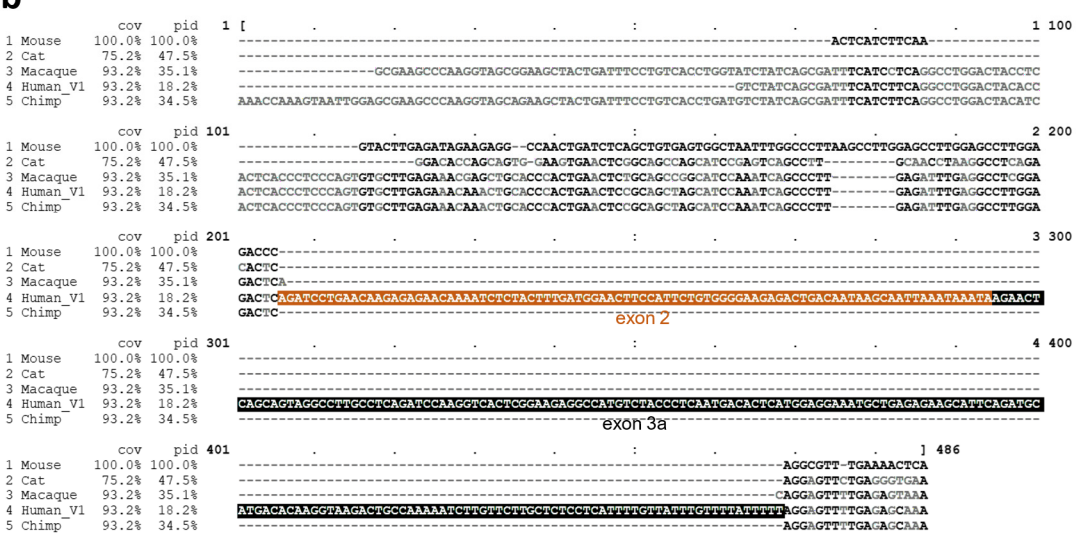


Supplementary Fig. 6: Treatment with 5ex3-viMO enhances rituximab-mediated cytotoxicity. OCI-Ly8, Raji and MEC-1 were treated with either solvent control (NT), Ctrl-viMO or 5ex3-viMO; followed by 10 mg/mL rituximab. PI negative, viable cells were analyzed with flow cytometry. Representative histograms are shown with the Y-axis scaled as a percentage of the maximum frequency. Each dot is a repeated experiment (N=2)



Supplementary Fig. 7: Alternative splicing events in the 5' UTR of CD20 in normal B cells RNA-Seq data of corresponding to normal B cell subsets that were FACS-enriched from the human bone marrow and tonsils. Data for each sample is presented as a stacked bar and indicates the relative abundance of RNA-Seq reads that map to the unique exon-exon junctions found in each 5' UTR variant of CD20 (as shown on the right). Naïve tonsillar B cells (CD19+IgM+IgD+) from the lymph node were treated with anti-IgM or an isotype control. Western blots of phosphorylated SYK is shown at the bottom right corner as a marker of B cell activation with actin as a loading control.





Supplementary Fig. 8: Sequence conservation of CD20 exons 2 and 3a. a, RNA-Seq reads that align to CD20 from human, chimpanzee, rhesus macaque, cat and mouse samples. Data adapted from the NCBI "Gene" database (https://www.ncbi.nlm.nih.gov/gene/; retrieved on 17 Jan 2023). **b,** The 5' UTR sequences from the orthologs of CD20 (From the top: NM_007641.6, NM_001009388.2, XM_001086364.4, NM_152866.3 and XM_024347171.1) are aligned using the MUSCLE multiple sequence alignment tool.