



## OPEN ACCESS

EDITED BY  
Lee Ann Garrett-Sinha,  
University at Buffalo, United States

REVIEWED BY  
Henri Gruffat,  
Institut National de la Santé et de la  
Recherche Médicale (INSERM), France  
Paolo Casali,  
The University of Texas Health Science  
Center at San Antonio, United States

\*CORRESPONDENCE  
Tina M. O'Grady  
cogrady@tulane.edu  
Erik K. Flemington  
erik@tulane.edu

SPECIALTY SECTION  
This article was submitted to  
B Cell Biology,  
a section of the journal  
Frontiers in Immunology

RECEIVED 02 October 2022  
ACCEPTED 28 November 2022  
PUBLISHED 19 December 2022

CITATION  
O'Grady TM, Baddoo M,  
Flemington SA, Ishaq EY,  
Ungerleider NA and Flemington EK  
(2022) Reversal of splicing infidelity is  
a pre-activation step in B  
cell differentiation.  
*Front. Immunol.* 13:1060114.  
doi: 10.3389/fimmu.2022.1060114

COPYRIGHT  
© 2022 O'Grady, Baddoo, Flemington,  
Ishaq, Ungerleider and Flemington. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the  
copyright owner(s) are credited and  
that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is  
permitted which does not comply with  
these terms.

# Reversal of splicing infidelity is a pre-activation step in B cell differentiation

Tina M. O'Grady<sup>1\*</sup>, Melody Baddoo<sup>1</sup>, Samuel A. Flemington<sup>2</sup>,  
Eman Y. Ishaq<sup>1</sup>, Nathan A. Ungerleider<sup>1</sup>  
and Erik K. Flemington<sup>1\*</sup>

<sup>1</sup>Department of Pathology & Laboratory Medicine, Tulane University School of Medicine, New Orleans, LA, United States, <sup>2</sup>Tulane Cancer Center, Tulane University School of Medicine, New Orleans, LA, United States

**Introduction:** B cell activation and differentiation is central to the adaptive immune response. Changes in exon usage can have major impacts on cellular signaling and differentiation but have not been systematically explored in differentiating B cells.

**Methods:** We analyzed exon usage and intron retention in RNA-Seq data from subsets of human B cells at various stages of differentiation, and in an in vitro laboratory model of B cell activation and differentiation (Epstein Barr virus infection).

**Results:** Blood naïve B cells were found to have an unusual splicing profile, with unannotated splicing events in over 30% of expressed genes. Splicing changed substantially upon naïve B cell entry into secondary lymphoid tissue and before activation, involving significant increases in exon commitment and reductions in intron retention. These changes preferentially involved short introns with weak splice sites and were likely mediated by an overall increase in splicing efficiency induced by the lymphoid environment. The majority of transcripts affected by splicing changes showed restoration of encoded conserved protein domains and/or reduced targeting to the nonsense-mediated decay pathway. Affected genes were enriched in functionally important immune cell activation pathways such as antigen-mediated signaling, cell cycle control and mRNA processing and splicing.

**Discussion:** Functional observations from donor B cell subsets in progressive states of differentiation and from timecourse experiments using the in vitro model suggest that these widespread changes in mRNA splicing play a role in preparing naïve B cells for the decisive step of antigen-mediated activation and differentiation.

## KEYWORDS

lymphocyte activation, mRNA, alternative splicing, exon skipping, naïve B cells, lymphoid tissue, nonsense mediated decay, NMD

## 1 Introduction

B-cell recognition of antigens and the production of antibodies lie at the core of the adaptive immune response. B cells originate in the bone marrow *via* a finely tuned pathway that optimizes breadth and specificity of antigen-binding capacity while avoiding autoreactivity. Upon maturation, antigen-naïve B cells (NBCs) with functional B cell receptors (BCRs) emerge from the bone marrow into the bloodstream, and eventually migrate towards secondary lymphoid tissues such as tonsils, adenoids and lymph nodes. The secondary lymphoid tissues are immune system hubs where the initially nonproliferative NBCs may encounter their antigens and cognate T cells in a signaling environment that promotes activation. Upon antigen binding to the BCR combined with T cell stimulation, B cells initiate an elaborate cascade of signaling events. This leads to a rapid and orchestrated increase in the transcription and translation of genes required to induce clonal expansion and the formation of a germinal center (GC). In the GC, proliferating B cells undergo somatic hypermutation to produce BCRs with increased antigen affinity, and class-switch recombination to generate mature antibody subtypes. Successfully mutated and class-switched GC B cells undergo further transcriptomic and phenotypic changes and differentiate into antibody-secreting plasma cells (PCs).

Alternative splicing of mRNA has been found to be linked to many cellular signaling and differentiation processes, with highly specific splicing profiles associated with different tissue types and cell subsets in the immune system and elsewhere (1). Differential abundance or activity of splicing factors can lead to skipping or inclusion of exons, retention of introns, or usage of alternate splice sites. Splicing differences, in particular intron retention variations, have been reported in subsets of neural (2), muscle (3), hematologic (4–6) and immune (7–11) cells.

We used RNA-Seq datasets from multiple different human donors to thoroughly interrogate splicing in B cell subsets across a spectrum of differentiation, including NBCs in venous blood, NBCs in secondary lymphoid tissue, GC B cells and fully differentiated PCs. We identified splicing changes over the course of differentiation that have extraordinary penetrance in the B cell transcriptome, affecting over 30% of expressed genes. Early in the differentiation pathway, blood-derived NBCs have a unique splicing profile, marked by high intron retention and a large number of aberrant exon skipping events that are corrected through the course of differentiation. Surprisingly, most splicing changes occur upon NBC entry into the secondary lymphoid tissue, before antigen-mediated activation induces proliferation and GC formation. These early, pre-proliferative splicing changes are similarly observed in B cell activation *via* a different mechanism (*in vitro* Epstein Barr virus (EBV) infection of blood NBCs) using B cells from different donors,

further supporting a broad increase in splicing fidelity as a prelude to B cell activation.

## 2 Materials and methods

### 2.1 RNA-Seq data acquisition and analysis

RNA-Seq datasets from human B cell subsets from multiple donors were obtained from the European Genome-Phenome Archive (EGA) with permission from the Blueprint Epigenome Consortium. Details of donor B cell subset isolation procedures are available from <https://www.blueprint-epigenome.eu> and Kulis et al.; briefly, peripheral blood NBCs (CD19+, CD27-, IgD+) were obtained from donor buffy coats sorted by FACS, tonsil NBCs (CD19+, CD27-, IgD+), GC B cells (CD20<sup>hi</sup>CD38<sup>med</sup>) and PCs (CD20<sup>med</sup>CD38<sup>hi</sup>) were isolated from donor tonsils using Ficoll-Isopaque, AutoMACS (CD19+ cells or B Cell Isolation Kit II) and FACS (12). Poly(A)+ RNA-Seq datasets from GM12878 cells and infection of primary and immortalized B cells were obtained from the European Nucleotide Archive (ENA). Details of primary B cell isolation are available from the original studies; briefly, CD19+ resting blood lymphocytes for the blood B cell infection timecourse dataset A were negatively selected from donor PBMCs using EasySep (StemCell Technologies) (13), CD19+ blood B cells for the blood B cell infection timecourse B were negatively selected by Precision Bioservices (#8440) (14), and naïve B cells for the lymphoid B cell infection timecourse were isolated from adenoid biopsy tissue using Ficoll Hypaque after T cell rosetting, then sorted by FACS (CD38-, IgD+) (15). Information about the generation of the LCL GM12878 are available from the Coriell Institute at [https://www.coriell.org/0/Sections/Search/Sample\\_Detail.aspx?Ref=GM12878](https://www.coriell.org/0/Sections/Search/Sample_Detail.aspx?Ref=GM12878). Information about the EBV-negative Akata Burkitt Lymphoma cell line is available from Shimizu et al. (16). Poly(A)+ RBP-knockdown datasets from K562 cells were obtained from ENCODE. Details about shRNA knockdown strategy and validation are available from <https://www.encodeproject.org/>. Accession numbers for all datasets are in [Supplemental Table 1](#). Data from donor 2 from blood B cell infection timecourse dataset B (14) were found to be corrupted and were not used. Data from one donor in the lymphoid NBC infection at 8 dpi were reported as not passing quality controls; this dataset was excluded from both the original analysis (15) and the present one.

RNA-Seq datasets were analyzed as follows:

1. Reads were aligned and mapped with STAR v2.5.2a (17) with the following settings, optimized for splicing analysis: `-chimSegmentMin 2 -outFilterMismatchNmax 3`

```
-alignEndsType EndToEnd -outSAMstrandField
intronMotif -alignSJDBoverhangMin 6 -
alignIntronMax 300000.
```

- For analysis of exon inclusion level and statistical assessment of splicing changes (skipped exon, retained intron, alternate 3' splice site, alternate 5' splice site and mutually exclusive exons) between conditions, the BAM alignment files produced by STAR were analyzed with rMATS 4.1.1 or v3.1.0 (18) and splAdder v3.0.3 (19).
- For analysis of intron retention level in each sample, the BAM alignment files produced by STAR were analyzed with IRFinder v1.3.0 (20).
- Gene expression levels were calculated as Transcripts per Million (TPM) using kallisto v0.46.0 (21).
- Differential gene expression was calculated using sleuth (22).
- Gene Set Enrichment Analysis (GSEA) was performed in GSEA 3.0 (23) using a signal-to-noise rank and the TARTE\_PLASMA\_CELL\_VS\_B\_LYMPHOCYTE\_UP gene set.

Alignment, mapping and splicing analysis used a reference genome containing the GRCh38 assembly of the human genome and the EBV genome (Akata strain, NCBI accession number KC207813.1 (24)) and a reference transcriptome annotation containing the Ensembl 90 human transcript annotation and EBV (Akata) transcript annotation (25).

## 2.2 Identification and analysis of unannotated splice junctions

Canonical splice junctions (with a GT-AG splicing pattern) identified by STAR were assigned to genes in the Ensembl 90 human transcriptome annotation using the findOverlaps function in GRanges (26). For each robustly expressed (TPM > 10) gene, the annotated splice junction with the highest number of uniquely mapping reads was identified. Then, unannotated splice junctions with uniquely mapping reads equaling at least 10% that of that gene's highest-depth annotated splice junction were selected for further analysis (Supplemental Table 2).

Donor and acceptor chromosome coordinates of unannotated splice junctions from each replicate were compared to each other in a pairwise fashion to calculate the proportion of shared unannotated splice junctions (the number of common unannotated splice junctions divided by the total number of unannotated splice junctions in the two datasets being compared). A hypergeometric test was used to evaluate the statistical significance of the shared set of unannotated junctions. RNA-Seq read junction depth for selected splice junctions was visualized with SpliceV (27).

## 2.3 Evaluation of exon commitment and intron retention

To determine the exon commitment level for each sample, the exon inclusion level ("IncLevel") of each exon in the rMATS output from genes with TPM > 10 was extracted. In the case of duplicate exons (i.e., the same exon being skipped by multiple alternate splice junctions), the minimum reported inclusion level was used. The exon commitment score for the sample was calculated as:

$$\frac{\text{Number of exons with IncLevel} > 0.99 + \text{Number of exons with IncLevel} < 0.01}{\text{Total nonredundant number of exons}}$$

To determine the global intron retention level, in each sample the "IRratio" was extracted for each intron in IRFinder output for genes with TPM > 10. The number of genes with at least one intron with IRratio > 0.25 was then divided by the total number of genes with TPM > 10.

## 2.4 RT-PCR detection of altered splicing

RNA was harvested from human peripheral blood CD19(+) CD27(-) naïve B cells from 3 different donors (STEMCELL #70032), from DG75 cells (EBV-negative), and from Akata cells (EBV-positive) using the RNeasy Mini Kit (QIAGEN #74106) with on-column DNase digestion. RNA was reverse transcribed to cDNA with the iScript cDNA Synthesis Kit (Bio-Rad #1708891BUN) and PCR was performed using the primers in Supplemental Data Sheet 2. PCR products were run on an agarose gel containing ethidium bromide and imaged using a Bio-Rad ChemiDoc Imaging System.

## 2.5 Cross-dataset splicing alteration comparisons

For pairwise comparisons of splicing profile changes across experiments (B cell differentiation, EBV infection and ENCODE RBP knockdowns), all chromosome coordinates of high-confidence significantly altered splice junctions (FDR < 0.0005) in rMATS output from each comparison were compared to each other in a pairwise fashion to calculate the proportion of splice junctions altered in both conditions. A hypergeometric test was used to evaluate the statistical significance of the shared set of altered junctions.

## 2.6 Splice site strength, intron and exon length, GC content

The sets of exons with significantly higher inclusion (rMATS FDR < 0.05, IncLevelDifference > 0) and introns with lower retention (rMATS FDR < 0.05, IncLevelDifference < 0) in PCs or

EBV infected cells compared to uninfected blood NBCs from the corresponding dataset were examined. Splice site strength was calculated using MaxEntScan (28).

## 2.7 Identification of NMD-targeting exon-skipping events

Alternately spliced exons (rMATS FDR < 0.05) that had nucleotide lengths not divisible by 3 and were in annotated coding sequences were analyzed for NMD-targeting potential. Reading frames for each possible affected isoform of the affected gene were scanned for premature stop codons (PTCs). If a PTC was detected >50 nt upstream of a splice junction, the exon-skipping event was determined to be potentially NMD-targeting.

## 2.8 Examination of conserved protein domain (CD) changes

Alternately spliced exons (rMATS FDR < 0.05) that had nucleotide lengths divisible by 3 and were in annotated coding sequences were analyzed for CD changes. The encoded amino acid sequences of the exons were obtained in the appropriate reading frame using a standard codon table and the amino acid sequences were submitted to the NCBI CD-Search Tool (Conserved Domain database, Expect Value threshold 0.01 with composition-corrected scoring in automatic search mode).

## 2.9 Functional enrichment analysis

To examine genes with unannotated splicing in B cell subsets, genes with identified unannotated splicing (see subsection 2.2, “*Identification and analysis of unannotated splice junctions*”) were compared across replicates. Genes that contained unannotated splice junctions in all replicates of a subset were used for functional enrichment analysis. To examine genes with differentiation-associated alternative splicing, lists of genes with CD restoration, NMD targeting or intron retention were identified as described as above.

Lists of genes were analyzed with Enrichr (29). GO Biological Process 2021 results were downloaded and plots created with R/ggplot2. GO terms with Adjusted p-value < 0.05 were considered to be significantly enriched.

# 3 Results

## 3.1 Pervasive unannotated splicing in blood naïve B cells

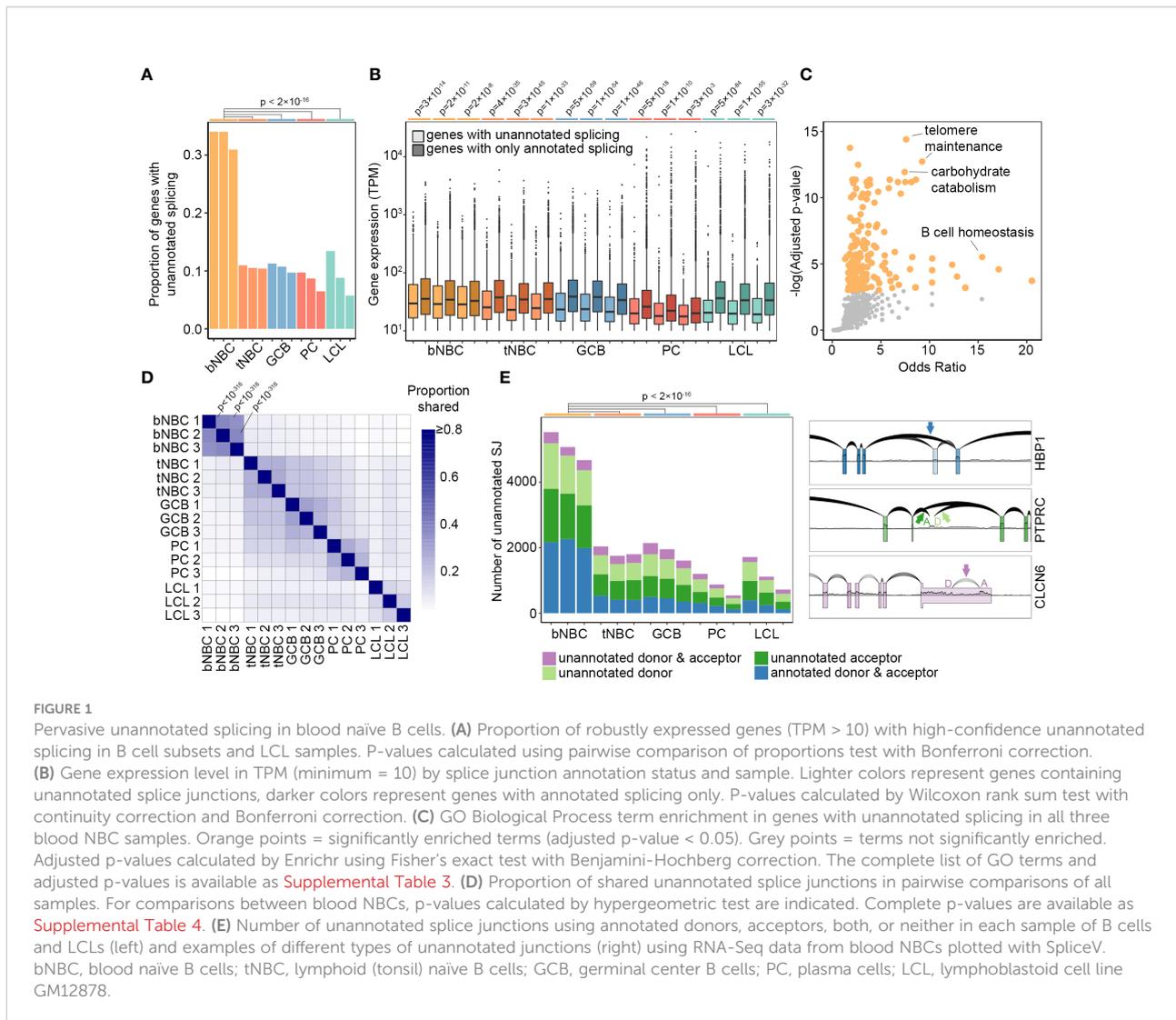
To investigate RNA splicing in human B cell subsets, we obtained high read-count, paired-end RNA-Seq data from

fluorescence activated cell-sorted (FACS) human donor cells *via* the Blueprint Epigenome project, including NBCs isolated from peripheral blood, NBCs isolated from tonsils, GC B cells, and PCs ((30) Supplemental Table 1). Assessing splicing across these datasets, we noticed a remarkably large number of unusual junctions that were not annotated in the Ensembl reference transcriptome, particularly in datasets derived from blood NBCs. To investigate these novel splicing events in more detail, we used stringent criteria to select high confidence unannotated splice junctions. For each robustly expressed gene (transcripts per million (TPM) > 10), we first identified the annotated splice junction with the highest read count. Only unannotated splicing events with a read count of at least 10% of the depth of the gene's highest annotated splice junction were further considered (Supplemental Table 2). Using these criteria, we found that over 30% of genes in blood-derived NBCs from each of the three donors contained at least one unannotated splice junction, noticeably higher than the proportion in lymphoid (tonsil)-derived NBCs, GC B cells, PCs, or the extensively used EBV-immortalized B cell lymphoblastoid cell line (LCL) GM12878 (Figure 1A (31, 32)).

Genes with unannotated splice junctions were on average expressed at slightly lower levels than genes with entirely annotated junctions, but nevertheless included very highly expressed genes (Figure 1B). Notably, however, the most highly expressed genes typically did not contain unannotated splice junctions (Figure 1B). This may indicate an essential requirement for specific isoforms of these genes and commensurately strong *cis* splicing features, and/or it may reflect more comprehensive annotation of alternate isoforms in highly expressed genes.

To evaluate functional implications of aberrant splicing, we identified genes in each subset with unannotated splicing across all replicates and examined these gene sets using Enrichr (29). Blood NBC genes with unannotated splicing were strongly enriched in several tissue-specific Gene Ontology (GO) terms such as *GO:0001782: B cell homeostasis*, possibly indicating the presence of novel stage-specific isoforms of mRNA and protein. GO terms such as *GO:0043470: regulation of carbohydrate catabolic process*, *GO:0043467: regulation of generation of precursor metabolites and energy*, *GO:0010389: regulation of G2/M transition of mitotic cell cycle* and *GO:0032210: regulation of telomere maintenance via telomerase* were also enriched, potentially representing a limited need for canonical forms of gene products involved in energy production and the cell cycle in these resting, unactivated cells. For the remaining B cell subsets, no term or pathway showed statistically significant enrichment (Figure 1C and Supplemental Table 3).

To assess whether the extensive unannotated splicing in blood NBCs is a programmed regulatory feature of B cell differentiation, we examined the consistency of unannotated splicing events across multiple donors and B cell subsets. We performed pairwise comparisons of high-confidence



unannotated splice junctions in all samples and calculated the proportion of unannotated junctions shared between the samples (Figure 1D and Supplemental Table 4). Blood NBCs had broad similarity among the three donors and formed a distinct group, separate from the other samples. This high level of consistency indicated that the unique splicing pattern in blood NBCs is an innate, regulated feature of this B cell subtype.

Assessing individual unannotated splicing events in blood NBCs, we noted that many novel splicing events represented exon skipping of one or more constitutive exon(s), with both the splice donor and acceptor sites being annotated (Figure 1E, HBP1 example). Also prevalent were hemi-annotated events in which either the donor or the acceptor sites were unannotated (Figure 1E, PTPRC example). Lastly, a small percentage of unannotated junctions were completely novel with both the donor and acceptor sites unannotated (Figure 1E, CLCN6 example). The varied nature of unannotated splicing events is consistent with compromised activity of multiple splicing factors

in blood NBCs, with the skipping observed in HBP1 (Figure 1E) potentially arising from diminished U2 complex activity at the acceptor site and/or reduced binding of splicing enhancers to the skipped exon, and splicing into and out of intronic sequences (PTPRC – Figure 1E) possibly arising from decreased intronic splicing silencer factor activity. Overall, the diversity of unannotated splicing and the robust numbers of unannotated splicing events in blood NBCs (greater than 4000) are consistent with the hypothesis that splicing fidelity is depressed in blood NBCs due to a diverse restriction of splicing factor function.

### 3.2 Splicing changes are early events in B cell differentiation

Our finding of an unusually high level of novel splicing in blood NBCs compared to other B cell subsets, including even NBCs derived from tonsils, suggests that this splicing pattern is

largely reversed early in differentiation, upon NBC entry into lymphoid tissue. To investigate whether these splicing changes would also occur using an alternative method of B cell activation we turned to the *in vitro* EBV infection model. This model mimics activation, proliferation and differentiation of B cells in a controlled tissue culture environment by inducing a signaling cascade that shares some similarities with, but is distinct from, that induced by B cell antigen exposure and T cell stimulation (15, 33). Our analysis of the fully EBV-immortalized LCL GM12878 indicated that unannotated splicing in this cell line was comparable to that of activated, differentiated B cells *in vivo* (Figure 1). To investigate the stage at which the unique splicing pattern of blood NBCs is reversed, we obtained RNA-Seq datasets from timecourse experiments of EBV infection leading to LCL establishment (13–15, 34). Consistent with previously observed similarities between *in vitro* EBV-associated activation and *in vivo* antigen-mediated activation (15, 33), fully EBV-immortalized LCLs from two different laboratories showed strong enrichment in the expression of genes associated with PCs relative to their parental blood B cells (Figure 2A). Assessing the EBV infection timecourses, we found substantial decreases in unannotated splicing after infection (Figure 2B). Strikingly, the splicing changes occurred predominantly by 2 days post-infection (dpi), early in the timecourses (Figure 2B, Blood A and Blood B) and well before the previously reported onset of infection-associated cell cycle entry and proliferation (35). Notably, NBCs harvested in another laboratory from lymphoid tissue (adenoids) had less unannotated splicing than blood B cells, consistent with our observations of tonsil NBCs (Figure 1) and there were only modest further decreases upon EBV infection (Figure 2B, Lymphoid). Infection of a differentiated EBV-negative Burkitt lymphoma (BL) cell line, Akata-EBV-N, had little impact on the already low level of unannotated splicing in these cells (Figure 2B, BL). Together, these *in vitro* B cell infection experiments support the findings from our analyses of *in vivo* B cell subtypes, indicating that the unusual splicing profile observed in blood NBCs can be reversed by multiple independent mechanisms of B cell activation. Further, the reversal of the blood NBC splicing program is an early event, occurring before NBCs enter the cell cycle and likely upon their entry into secondary lymphoid tissue.

### 3.3 Low exon commitment and high intron retention in blood naïve B cells

In general, unannotated splicing in blood NBCs was robust but not fully penetrant, i.e. a portion of transcripts from each affected gene was aberrantly spliced with the remaining transcripts

displaying conventional splicing (Figure 1E and Figure S1). To globally investigate alternative splicing penetrance we used rMATS, a frequently used software that performs well in independent assessments (36, 37), to assess the inclusion level of exons across the transcriptome. For each sample we calculated an “exon commitment” score, which we define as the proportion of exons that were either completely spliced-in (i.e. included in >99% of transcripts from the gene) or completely spliced out (i.e., included in <1% of transcripts from the gene). In blood NBCs exon commitment scores were low, illustrating a lack of commitment to specific isoforms. Upon differentiation exon commitment increased, with a majority of exons in PCs being committed to either inclusion or exclusion (Figures 2C, S2A). In the EBV infection timecourses, uninfected blood B cells showed even lower exon commitment, which was substantially reversed by EBV infection. In contrast, no change in exon commitment was observed after infection of lymphoid NBCs or differentiated BL cells (Figures 2D, S2B–D). While exon commitment scoring takes into consideration both commitment to the spliced-in and the spliced-out isoforms, it is notable that for most exons, commitment entailed a strong bias toward splicing-in upon differentiation (Figures 2E, S2A–D). The increase in exon inclusion upon differentiation was also observed using a second splicing analysis algorithm, splAdder, whose results correlated well with the results obtained from rMATS (Figure S3). Exon usage differences between blood naïve B cells and more differentiated B cells, both EBV-negative (DG75 cells) and EBV-positive (Akata cells) were also observed by RT-PCR (Figures 2F, S4).

NBCs have previously been reported to have high intron retention in both humans and mice (7). RNA-Seq datasets from the Blueprint Epigenome project are not poly(A)-selected and include nascent RNAs with incomplete splicing, making discerning true intron retention events in mature mRNA less precise. Nevertheless, we observed the previously reported decreases in intron retention in differentiated cells relative to blood NBCs in these datasets (Figures 2G, S5A (7)). The poly (A)-selected RNA-Seq data from the EBV infection timecourse experiments, which reflect only mature RNA transcripts, showed even more striking decreases in intron retention after infection of blood B cells (Figures 2H, I, S5B–D). The changes in lymphoid NBCs after EBV infection were substantially less than those in blood B cells, and BL cells did not show appreciable intron retention changes upon infection. Overall, the aberrant splicing program in blood NBCs is marked by low exon commitment and high intron retention, affecting a broad swath of genes. While canonical transcripts are produced for most expressed genes, the extensive splicing infidelity, in the form of mature transcripts with retained introns or skipped exons, likely dampens output of functional proteins until NBCs encounter the lymphoid environment and prepare for activation.

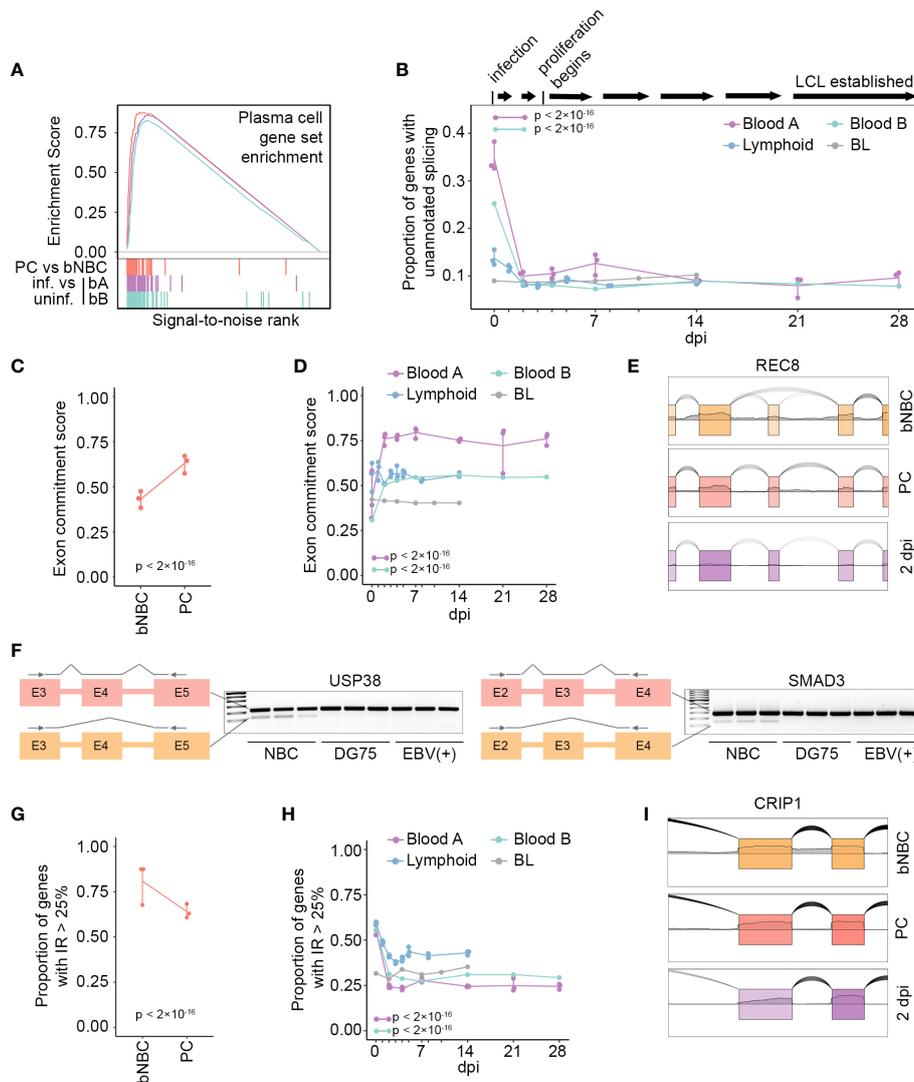


FIGURE 2

*In vivo or in vitro* activation of B cells restores the splicing program. **(A)** Enrichment scores from GSEA of gene expression level changes in PCs compared to blood NBCs (PC vs bNBC, red, maximum enrichment score = 0.880,  $p = 0$ ), EBV-infected blood B cells from timecourse A at 28 dpi compared to uninfected cells (bA: purple, maximum enrichment score = 0.863,  $p = 0$ ) and EBV-infected blood B cells from timecourse B at 28 dpi compared to uninfected cells (bB: green, maximum enrichment score = 0.829,  $p = 0$ ). Gene set: TARTE\_PLASMA\_CELL\_VS\_B\_LYMPHOCYTE\_UP from MSigDB. Colored vertical lines indicate genes in the gene set. **(B)** Proportion of genes (TPM > 10) with unannotated splicing after EBV infection. A timeline of EBV infection events is at top. P-values calculated using a pairwise comparison of proportions test with Bonferroni correction are indicated for Blood A and Blood B, 2dpi compared to 0 dpi. **(C)** Exon commitment score for blood NBC (bNBC) and PC samples. Exon commitment score = proportion of exons either fully included (rMATS IncLevel > 0.99) or fully excluded (rMATS IncLevel < 0.01) in genes with TPM > 10. P-value calculated using 2-sample test for equality of proportions with continuity correction. **(D)** Exon commitment scores after EBV infection. P-values calculated using a pairwise comparison of proportions test with Bonferroni correction are indicated for Blood A and Blood B timecourses, 2dpi compared to 0 dpi. **(E)** Example of uncommitted exon in blood NBCs that is committed by increased inclusion upon differentiation to PC or EBV infection (2 dpi, Blood A), plotted with SpliceV. **(F)** Transcript diagrams and RT-PCR gels showing skipped exons in transcripts of USP38 (left) or SMAD3 (right) in naive B cells from 3 donors but not in differentiated B cell lines DG75 (EBV-negative) or Akata (EBV-positive). Arrows in diagrams indicate primer positions. **(G)** Proportion of genes containing introns with high retention (IRFinder IRratio > 0.25) in blood NBC (bNBC) and PC samples. P-value calculated using 2-sample test for equality of proportions with continuity correction. **(H)** Proportion of genes containing introns with high retention levels at multiple timepoints after EBV infection. P-values calculated using a pairwise comparison of proportions test with Bonferroni correction are indicated for Blood A and Blood B, 2dpi compared to 0 dpi. **(I)** Example of intron retention in blood NBCs reduced upon differentiation to PC or EBV infection (2 dpi, Blood A), plotted with SpliceV. Blood A = blood infection timecourse A, 3 replicates, Blood B = blood infection timecourse B, 1 replicate, Lymphoid = adenoid infection timecourse, 3 replicates. BL = Akata-EBV-N Burkitt lymphoma cell infection timecourse, 1 replicate.

### 3.4 Involvement of splicing regulatory factors in differentiating B cells

To investigate the basis for the observed splicing infidelity in blood NBCs, we compared differences in exon skipping events between blood NBCs and lymphoid NBCs, GC B cells and PCs to changes in exon skipping observed after knockdown of each of 182 RNA-binding proteins (RBPs), including known and potential splicing regulators (38). RNA-seq datasets for each pair of control or RBP-shRNA knockdowns in the lymphoblast cell line K562 were downloaded from ENCODE (39, 40) and statistically significant differences in exon skipping were analyzed using rMATS. Strikingly, substantially greater numbers of differential exon skipping events were observed when comparing blood NBCs to lymphoid NBCs, GC B cells, PCs or EBV-infected cells than after the knockdown of any splicing factor tested, including the core acceptor factors U2AF1 and U2AF2 (Figure S6A). We then performed pairwise comparisons of altered splicing events in each RBP knockdown experiment to altered splicing events in blood NBCs vs each B cell subset and in uninfected blood B cells vs each EBV infection timepoint. Knockdown of several RBPs led to statistically significant overlap in differential exon skipping events with blood NBCs as compared to lymphoid NBCs, PCs, or EBV-infected cells. These included U2AF1, U2AF2, and other U2 acceptor factors (e.g. SF3B4 and SF3B1), the splicing enhancer SRSF1, and splicing suppressors such as HNRNPC and HNRNPA2B1 (Figure 3A and Supplemental Table 5). The greatest overlap was observed with knockdown of AQR, a component of the spliceosome that is involved in small nucleolar ribonucleoprotein (snoRNP) assembly. As a negative control, we compared overlap in changed exon skipping events after artificially reversing the direction of change for blood NBCs vs lymphoid NBCs, GC B cells, and PCs and observed minimal overlap with changes detected in RBP knockdown experiments (Figure S6B). Our findings that the knockdown of any tested splicing factor is insufficient to recapitulate the number or type of changes in exon skipping upon blood NBC differentiation is consistent with a coordinated set of RBP changes leading to a broad, general increase in splicing fidelity as an early step in B cell differentiation.

Notably, some RBPs showing functional relationships with exon skipping differences in blood NBCs vs differentiating cells showed modest increases in mRNA abundance during differentiation (Figure 3A). Others showed changes in exon inclusion or decreases in intron retention that likely lead to more efficient production of functional splicing factors (Figure 3A and Supplemental Table 5). A parallel analysis of overlap in intron retention events in RBP knockdowns and blood NBCs vs differentiating cells also implicated a group of possible RBP effectors of intron retention changes, with many

of these showing increased abundance, less exon skipping or less intron retention (Figure S6C and Supplemental Tables 5, 6). Altogether, the splicing changes upon RBP knockdowns combined with the abundance and splicing changes of RBPs themselves is consistent with a general increase in splicing efficiency occurring alongside the known increase in transcription (41) as B cells progress from resting, unactivated cells to activated, proliferating and/or antibody-secreting cells.

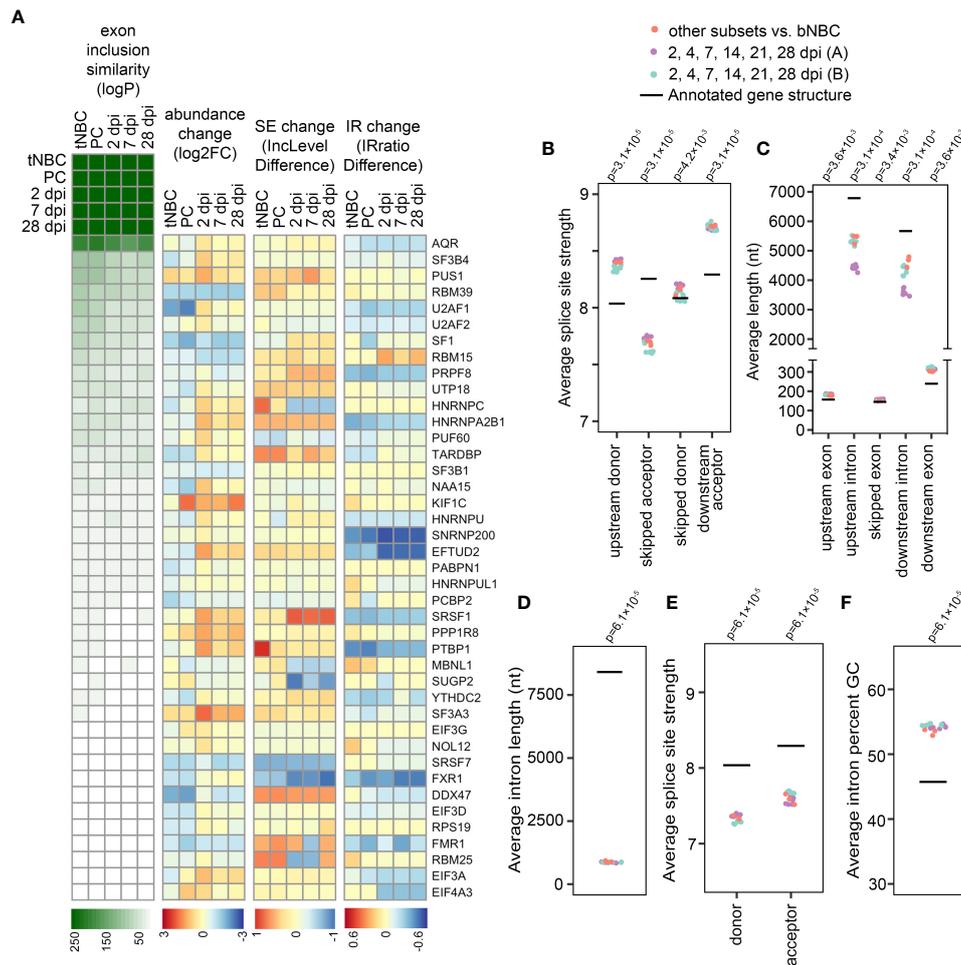
### 3.5 Weaker splice sites and shorter introns are associated with missplicing in blood naïve B cells

Exon-skipping junctions specific to blood NBCs most often skipped a single exon, though some junctions spanned multiple exons (Figure S6D). The Maximum Entropy Model (28) splice site strengths for skipped exons had a distinctive pattern, with an unusually weak acceptor site for the skipped exon flanked by unusually strong upstream exon donor and downstream exon acceptor sites (Figure 3B). This is consistent with limited availability of active acceptor recognition factors in blood NBCs causing reduced spliceosome engagement with the weak skipped acceptor but minimal impacts at the strong upstream donor and downstream acceptor sites. Interestingly, blood NBC skipped exons also tended to reside between introns that were relatively short (Figure 3C), a characteristic that has been linked to suboptimal transcription rates (42), a property previously reported in naïve T cells (43) that may also be a feature of blood NBCs.

Introns with decreased retention in differentiating B cells also had distinct characteristics. They were substantially shorter than average, with weaker donor and acceptor splice sites, and higher GC content (Figures 3D–F). These features are all known to be associated with intron retention both in basal conditions and during splicing impairment (44–46), and have previously been observed in retained introns in mouse NBCs (7). Overall, these observations suggest an inefficient splicing program in blood NBCs with decreased recognition of sub-optimal splice sites.

### 3.6 B cell exon commitment restores conserved protein domains and evades nonsense-mediated mRNA decay

Alternative splicing events can have drastically different downstream effects in different contexts. Skipped coding exons with lengths divisible by 3 (in-frame skipped exons) remove sequence without disrupting the reading frame. The resulting protein isoforms may serve distinct functions in the cell, may be nonfunctional, or may even act as dominant negative isoforms if



**FIGURE 3**  
 Increased splicing program efficiency alters the splicing profile. **(A)** Heatmaps indicating exon inclusion similarity (logP = -log(p-value) by hypergeometric test), mRNA abundance change (log2FC = log2(fold change) by sleuth), skipped exon (SE) change (IncLevelDifference = inclusion level difference by rMATS), and intron retention (IR) change (IRratio difference = intron retention difference by IRFinder) for RBPs with exon inclusion changes most similar to those of B cell differentiation/EBV infection. For RBP genes with multiple splice junctions the SE and IR events with the highest absolute inclusion/retention difference measurable in all compared replicates are displayed. PC = PCs compared to blood NBCs; tNBC = lymphoid (tonsil) NBCs compared to blood NBCs; 2, 7 and 28 dpi = EBV-infected cells at specified timepoints compared to uninfected cells in Blood timecourse **(A, B)** Average splice site strength calculated by MaxEntScan for exons showing increased inclusion in differentiating or EBV-infected B cells (FDR < 0.05 by rMATS). **(C)** Average length in nucleotides (nt) of exons showing increased inclusion in differentiating or EBV-infected B cells (FDR < 0.05 by rMATS), and their flanking introns and exons. P-values compared to annotated averages calculated by Wilcoxon signed rank exact test with Bonferroni correction. **(D)** Average length in nucleotides (nt) of introns showing decreased retention in differentiating or EBV-infected B cells (FDR < 0.05 by rMATS). **(E)** Average splice site strength calculated by MaxEntScan for introns showing decreased retention in differentiating or EBV-infected B cells (FDR < 0.05 by rMATS). **(F)** Average GC content of introns showing decreased retention in differentiating or EBV-infected B cells (FDR < 0.05 by rMATS). Blood infection timecourse A (purple) includes 6 timepoint comparisons with 3 replicates each, Blood infection timecourse B (green) includes 6 timepoint comparisons with 1 replicate each, B cell subset comparisons include 3 subsets (tonsil NBCs, GC B cells, and PCs) compared to blood NBCs. Each B cell subset includes 3 replicates **(B-F)**, p-values compared to annotated averages were calculated by Wilcoxon signed rank exact test with Bonferroni correction.

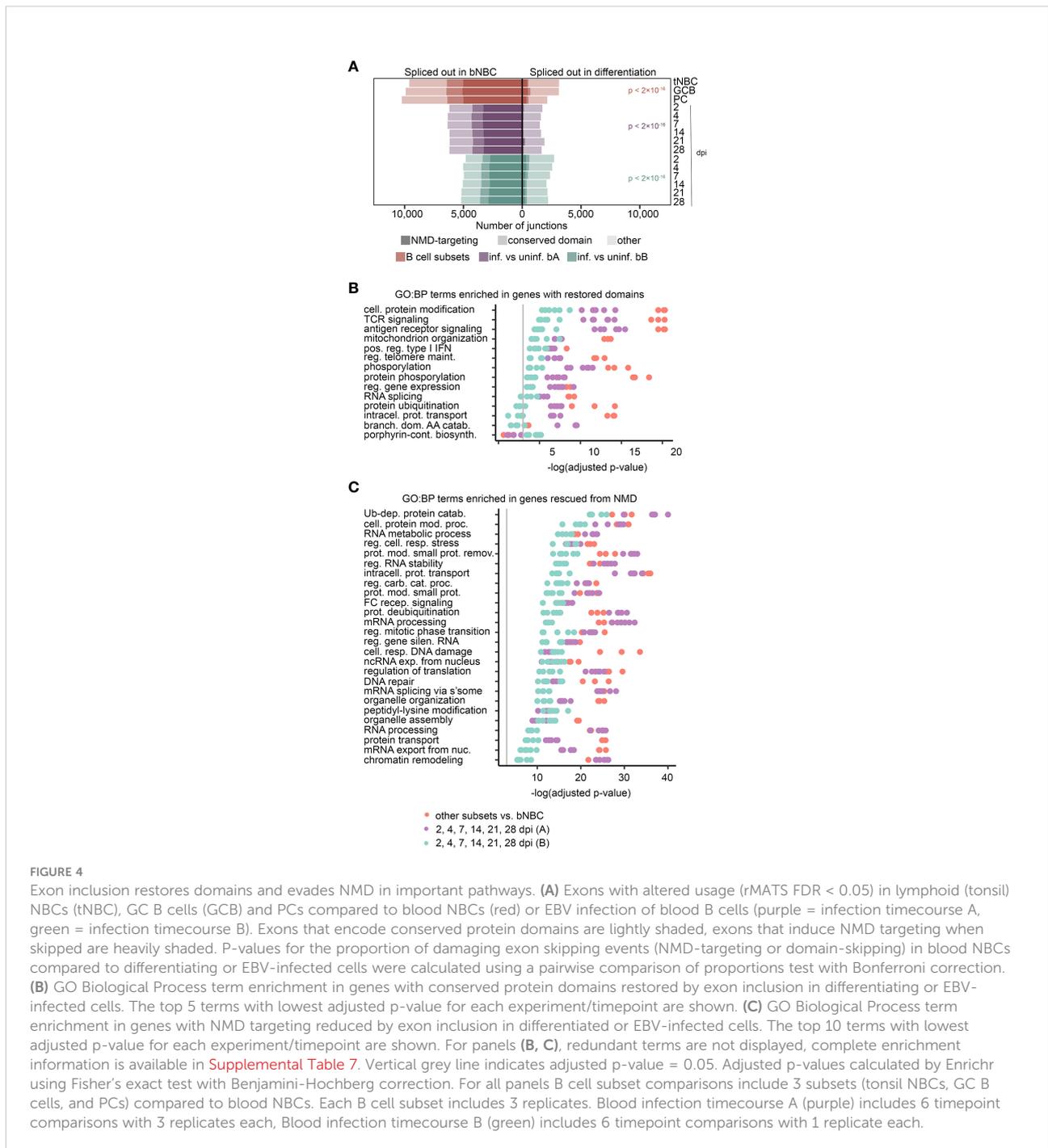
a regulatory domain is spliced out. Out-of-frame skipped exons disrupt open reading frames, potentially introducing a premature stop codon (PTC) and targeting the transcript for degradation *via* the nonsense-mediated decay (NMD) pathway (47).

We investigated the impact of exon skipping events by first determining if the skipped exon was in-frame or out-of-frame. For

out-of-frame exons, we scanned the new reading frame downstream of the alternatively spliced exon for PTCs. Transcripts with PTCs more than 50 nt upstream of splice junctions were identified as likely NMD targets, according to previously identified NMD-targeting characteristics (47). A large proportion of skipped exons in blood NBCs appeared to target their transcripts to NMD, while a very small proportion of skipped

exons in more differentiated cells were NMD-targeting (Figure 4A). For in-frame skipped exons, we determined their encoded amino acid sequences and investigated whether they included conserved protein domains (CD) using the NCBI's Batch CD-Search tool (48). The proportion of skipped exons in blood NBCs that encoded a conserved domain was much higher than that in differentiated cells. Interestingly, while splicing changes after EBV infection of lymphoid (adenoid) B cells are

less pronounced than after infection of blood B cells (Figure 2D), in this case others have similarly observed a large number of infection-induced exon changes that either restore a transcript's open reading frame or avoid an NMD-targeting PTC (49). The high number of NMD-targeting and domain-deleted transcripts in blood NBCs suggests that an early splicing program that produces nonfunctional, potentially deleterious transcripts is repaired during B cell differentiation, leading to restored inclusion of critical exons.



### 3.7 Functional pathways impacted by restored transcripts in activated B cells

To predict cellular functions impacted by restoration of splicing fidelity, we separately assessed GO term enrichment in genes with altered mRNA splicing in each B cell subset compared to blood NBCs, and in EBV-infected cells at each infection timepoint compared to parental uninfected B cells. Results from these different experiments were remarkably consistent. Genes with transcripts whose restored exons encoded conserved domains were enriched for GO terms related to B cell activation and differentiation, such as *GO:0050851: antigen receptor-mediated signaling pathway* and *GO:0032481: positive regulation of type I interferon production*; and terms related to cellular signaling, such as *GO:0006468: protein phosphorylation*. Consistent with previous reports that splicing factors themselves are frequently alternatively spliced (50, 51), *GO:0008380: RNA splicing* and related terms were also enriched (Figure 4B and Supplemental Table 7). Analysis of genes with transcripts rescued from NMD by exon inclusion in differentiating B cells revealed a set of enriched GO terms that partially overlapped those of genes with restored protein domains. Many of these genes are involved in *GO:0043470: regulation of carbohydrate catabolic process* and *GO:1901990: regulation of mitotic cell cycle phase transition*, suggesting a role for alternative splicing in supporting the cell growth and proliferation induced by both antigen-mediated activation and EBV infection. Splicing-related GO terms, such as *GO:0000398: mRNA splicing, via spliceosome*, were also enriched in genes with reduced mRNA NMD targeting (Figure 4C and Supplemental Table 7). Genes with reduced mRNA intron retention in differentiating B cells, especially at the lymphoid NBC and GC B cell stages, were also enriched in terms related to mRNA splicing and processing (e.g. *GO:0000398: mRNA splicing, via spliceosome* and *GO:0006397: mRNA processing*; Figure S7 and Supplemental Table 7), consistent with previous findings in mice (7). The enrichment of activation-related GO terms in genes with altered mRNA splicing points to a multifaceted role for splicing fidelity enhancement in B cell differentiation. Furthermore, enrichment in splicing-related GO terms suggests a positive feedback mechanism whereby production of splicing factors themselves is improved, leading to further increases in splicing factor availability and function.

### 3.8 Pre-activation splicing changes prepare naïve B cells for antigen encounter

Exon usage in blood B cells changes rapidly after EBV infection, with a majority of splicing alterations established by 2 dpi and very few further changes at later timepoints (Figures 2, 3, S2). In contrast, mRNA abundance levels change dramatically both

before and after 2 dpi (13–15), and major physiological changes occur later, in particular a phase of hyperproliferation that begins 3–4 dpi (35). Overall, fewer splicing changes occurred upon EBV infection of NBCs derived from lymphoid tissue than from blood.

While the Blueprint Epigenome samples represent snapshots of B cell subsets from different donors rather than a timecourse experiment, they can be arranged to reflect the progression of NBCs from blood to secondary lymphoid tissue, followed by differentiation into GC B cells and finally into PCs. Differentiation-associated splicing changes also occurred remarkably early, with blood NBCs displaying a singular splicing profile that was lost upon entry into secondary lymphoid tissue (tonsil). Indeed, the lymphoid NBC splicing profile more closely resembled that of activated B cells than that of blood NBCs (Figures 1A, D, E, S2A). To investigate more closely the relationships of the B cell subsets we used rMATS to perform pairwise comparisons of blood NBCs, lymphoid NBCs and GC B cells to fully differentiated PCs. By far the largest number of exon usage differences was observed when comparing blood NBCs to PCs. Both lymphoid NBCs and GC B cells had relatively few exon usage differences compared to PCs (Figure 5A). The reduction in intron retention occurred more slowly, with a progressive stepwise reduction from blood NBCs to lymphoid NBCs and GC B cells relative to PCs (Figure 5B). Transcript abundance changes also occurred between blood and lymphoid NBCs, with the transition to a fully differentiated PC mRNA expression program that had begun but was by no means complete in lymphoid NBCs (Figure 5C). Because NBCs, unlike GC B cells and PCs, have not yet encountered their cognate antigen, this suggests that some transcript abundance changes and a majority of exon splicing changes occur upon NBC migration to secondary lymphoid tissue in preparation for - not in response to - antigen encounter and BCR signaling.

To further assess the similarity of lymphoid NBC and PC splicing profiles we examined GO term enrichment in genes with altered mRNA splicing relative to blood NBCs. GO term enrichment in these comparisons was strikingly similar, both when examining genes with protein domains restored and genes with mRNA NMD targeting reduced by B cell (pre-)activation (Figures 5D, E). One of the GO terms with the strongest enrichment in genes with restored protein domains was *GO:0050851: antigen receptor-mediated signaling pathway* (Supplemental Table 8), further supporting a potential role for splicing fidelity enhancement in preparing cells for antigen encounter and BCR signaling. An in-depth examination of mRNA splicing changes in the BCR signaling pathway (52) revealed that an extraordinary number of genes in this pathway showed a change in exon usage as NBCs entered the secondary lymphoid tissue (Figure 5F and Supplemental Table 9). Few further exon usage changes occurred upon differentiation into GC B cells or PCs (Supplemental Table 9), suggesting that changes at the mRNA splicing level contribute to preparing the BCR signaling pathway for activation at the protein level.

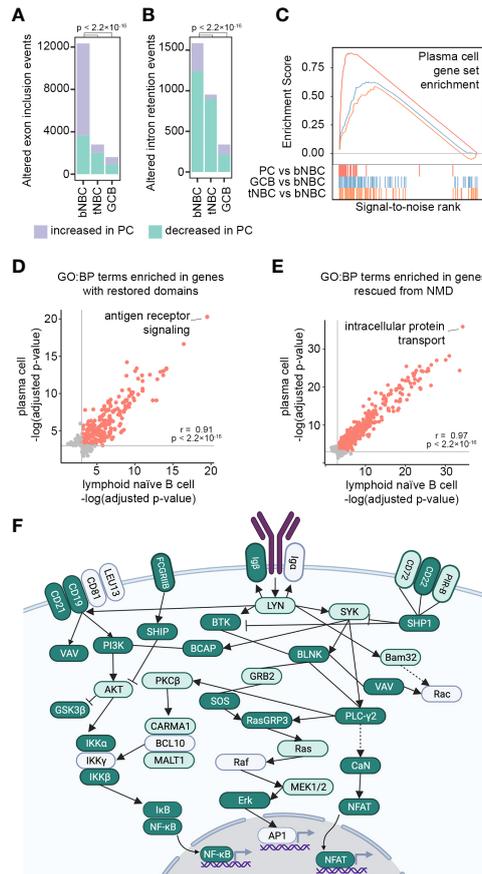


FIGURE 5

Lymphoid tissue-associated splicing repair prepares naïve B cells for antigen encounter. (A) Exon usage changes (rMATS FDR < 0.05, |IncLevelDifference| ≥ 0.1) in blood NBCs (bNBC), lymphoid (tonsil) NBCs (tNBC) and GC B cells (GCB) compared to PCs. P-values calculated by Pearson's Chi-squared test with Yates' continuity correction. (B) Intron retention changes (rMATS FDR < 0.05, |IncLevelDifference| ≥ 0.1) in blood NBCs (bNBC), tonsil NBCs (tNBC) and GC B cells (GCB) compared to PCs. P-values calculated by Pearson's Chi-squared test with Yates' continuity correction. (C) Enrichment scores from GSEA of RNA abundance changes in PCs (red, maximum enrichment score = 0.880, p = 0), GC B cells (GCB, blue, maximum enrichment score = 0.627, p = 0.001) or tonsil NBCs (tNBC, orange, maximum enrichment score = 0.588, p = 0.021) compared to blood NBCs. Gene set: TARTE\_PLASMA\_CELL\_VS\_B\_LYMPHOCYTE\_UP from MSigDB. Colored vertical lines indicate genes in the gene set. (D) GO Biological Process term enrichment in genes with conserved protein domains restored by exon inclusion in lymphoid (tonsil) NBCs or in PCs compared to blood NBCs. All significantly enriched terms (adjusted p-value < 0.05) are shown. Grey lines indicate adjusted p-value = 0.05. Red points: significant enrichment in both PCs and lymphoid NBCs. Grey points: significant enrichment in lymphoid NBCs or PCs only. Correlation (r) and p-value calculated using Pearson's product-moment correlation. Full GO term information is shown in Supplemental Table 8. (E) GO Biological Process term enrichment in genes with NMD targeting reduced by exon inclusion in lymphoid (tonsil) NBCs or in PCs compared to blood NBCs. All significantly enriched terms (adjusted p-value < 0.05) are shown. Grey lines indicate adjusted p-value = 0.05. Red points: significant enrichment in both PCs and lymphoid NBCs. Grey points: significant enrichment in lymphoid NBCs or PCs only. Correlation (r) and p-value calculated using Pearson's product-moment correlation. Full GO term information is shown in Supplemental Table 8. (F) BCR signaling pathway (KEGG). Genes in dark green show altered exon usage (rMATS FDR < 0.05) in lymphoid (tonsil) NBCs compared to blood NBCs and in EBV-infected cells compared to parental blood NBCs (both experiments, all timepoints). Genes in light green show altered exon usage in lymphoid NBCs compared to blood NBCs and in at least one EBV infection experiment timepoint. Image created with Biorender.com. Full information is shown in Supplemental Table 9.

## 4 Discussion

We used RNA-seq datasets from four different human B cell subtypes, each with multiple donors, to examine stage-specific splicing profiles during B cell differentiation. Blood-derived NBCs had a remarkable number of unannotated splicing events with 2127 high-confidence novel splice junctions

detected consistently across all donors and over 30% of expressed genes containing at least one novel splicing event (Figure 1A). While statistically significant splicing changes were observed at each stage of B cell differentiation, the preponderance of changes occurred, surprisingly, in the blood NBC to lymphoid NBC transition. Assessing the functional significance of alternative splicing in blood NBCs, we found

that a high percentage of exon skipping events specific to blood NBCs are predicted to undergo nonsense mediated RNA decay or to exclude a known conserved domain (Figure 4A), suggesting that the transition to the lymphoid microenvironment increases splicing fidelity in preparation for activation. Using EBV infection as another method of B cell activation, in blood B cells we similarly found that exon skipping and intron retention were substantially reduced (i.e. splicing fidelity increased) prior to B cell proliferation, and in lymphoid NBCs the remaining lower-level splicing infidelity was further corrected, again prior to EBV-induced proliferation (Figure 2).

While the major splicing changes occurred in naïve B cells upon entry into secondary lymphoid tissue, further splicing changes occurred in GC B cells and PCs (Figure 1D, 5A, B). These additional changes involved fewer events, were not biased toward increased exon inclusion, and reflected splicing programs that almost entirely used annotated splice junctions (Figures 5A, B, 1A). This suggests controlled alterations of specific splicing events leading to defined biological impacts, as opposed to the wholesale change in splicing fidelity of naïve B cells. For example, in PCs relative to GC B cells we observed decreased usage of exon 5a in IKZF3 (aka Aiolos), a transcription factor involved in B cell differentiation. This exon has been reported to affect the usual nuclear localization of Aiolos, leading to increased cytoplasmic distribution (53) and so possibly fine-tuning its transcriptional control function in GC B cells and PCs. In conjunction with changes in transcript abundance and protein signaling, targeted splicing alterations such as that in IKZF3 likely function to closely calibrate gene expression and functionality in each B cell subset (54).

The striking changes in exon usage between blood NBCs and lymphoid NBCs, with exon usage in lymphoid NBCs more closely resembling that in antigen-experienced, fully differentiated PCs, were an unexpected finding. These changes precede the major physiological changes of activation and are accompanied by the earliest changes in gene expression levels (Figure 5C (12, 55)) The previous observation that, unlike the other B cell subsets, the two sets of NBCs do not significantly differ in DNA methylation (12) suggests that the lymphoid NBC samples in the Blueprint Epigenome project are unlikely to be significantly contaminated by activated B cells and do in fact represent a distinct population of antigen-naïve cells. B cell “pre-activation” at the RNA level is supported by our observations that adenoid NBCs show fewer splicing changes upon EBV infection than blood B cells, despite the fact that the blood B cells used in both EBV infection timecourses were not explicitly enriched for naïve cells, although they likely consisted largely of NBCs (56). Further support comes from the observations of others that EBV infection-induced proliferation proceeds much more quickly in adenoid B cells than in blood B cells (57). These parallel observations of NBCs from tonsils and from adenoids suggest that exon usage changes are a common feature of NBCs

entering secondary lymphoid tissues throughout the body, though splicing changes associated with lymph nodes, spleen or gut lymphoid tissues remain to be resolved.

The progression from blood NBC to antigen-activated GC B cell and differentiated PC constitutes a major coordinated overhaul of the cell, from DNA methylation (12) and chromatin structure (58) through gene expression (59) and up to the physiological level. A global increase in splicing efficiency appears to begin surprisingly early and be well-integrated into this complex process, with a decrease in the transcriptomic “noise” of aberrantly-spliced, nonfunctional transcripts ushering in the known overall increase in transcription. This is likely supported by a suite of biological processes. Increased splicing factor availability, particularly factors that recognize weak splice acceptor sites (see Figure 3B) could potentially increase splicing fidelity (7). As shown in Figure 3A, several splicing factors increase in mRNA abundance during B cell differentiation, and/or undergo mRNA splicing modifications that could lead to better translation efficiency and higher protein abundance. Alterations in the subcellular localization of splicing factors could also lead to changes in splicing efficiency (60). In addition to splicing factor proteins, snRNA and snoRNA abundance, availability and post-transcriptional modifications can impact splicing. U1 and U6 snRNAs in particular have been reported to increase in abundance during lymphocyte activation (61). The RNA polymerase II elongation rate, which has been shown to vary during T cell development, has also been linked to splicing changes (42, 43). Finally, there is the possibility that the NMD-targeting splicing events we observed in NBCs arose not from altered splicing mechanics but from accumulation of aberrantly-spliced isoforms that are more efficiently degraded in differentiated cells; indeed, an increase in NMD efficiency has been reported in mouse activated B cells (62).

The pervasive similarity of splicing changes between *in vitro* EBV infection and *in vivo* B cell differentiation suggests a common mechanism, with the virus hijacking a cellular system rather than directly intervening in the splicing process. Importantly, splicing changes occur early in infection and are complete by 2 dpi, before most viral genes reach their full expression level (13–15). Among the viral genes with high early expression is Epstein Barr virus nuclear antigen 2 (EBNA2). EBNA2 interacts with cellular splicing factors and has the capacity to impact splicing (63). However, infection of blood B cells with recombinant EBV lacking EBNA2 induced a large-scale shift in cellular exon splicing very similar to that induced by wild-type EBV (Figure S8), despite very different impacts on host mRNA levels (14) and a complete abrogation of infection-associated proliferation and immortalization (35). Similarly, the viral EBNA-LP gene is expressed early after infection and has in fact been demonstrated to control a subset of cellular splicing changes, but the majority of infection-associated splicing changes are not dependent on

EBNA-LP expression (49). Interestingly, EBNA2 and EBNA-LP expression, together with physical binding of the viral protein gp350 to the cellular complement receptor CR2 (also known as CD21) are sufficient to drive proliferation in blood NBCs (64). While the splicing profile of these gp350/EBNA2/EBNA-LP-stimulated cells is not known, it raises the possibility that infection-induced splicing changes are triggered not by intracellular viral gene expression but by the virion's physical engagement of CR2 and the resultant cellular signaling cascade.

CR2 signaling is also known to enhance antigen-mediated B cell activation (65). While the secondary lymphoid tissues are rich in antigens and immune cells, these tissues are also enriched in complement proteins. Although BCR affinity for antigens is highly specific and the chances of an NBC encountering its antigen at any given time are relatively low, CR2 has generic affinity for the abundant complement protein fragment C3dg, suggesting that NBCs may encounter and bind complement proteins in lymphoid tissue before or in the absence of their cognate antigens. In addition to C3dg, CR2 can be stimulated by CD23 (66) and IFN- $\alpha$  (67), both present in the secondary lymphoid tissues (68). While it is not yet known whether complement signaling impacts splicing, this pathway represents an intriguing candidate for the initiation of the pre-proliferative splicing fidelity enhancement that is common to both B cell differentiation and EBV infection.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Datasets from the BLUEPRINT consortium are accessible to researchers with permission from the BLUEPRINT Data Access Committee. Requests to access these datasets should be directed to <https://www.blueprint-epigenome.eu/>. All other datasets analyzed are publicly available. Sources and accession numbers for all datasets used are listed in [Supplemental Table 1](#).

## Author contributions

TO'G and EF conceived the study, analyzed the data and wrote the manuscript. EI performed experiments. MB, SF and NU analyzed the data. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the NIH grants R01CA262090, R01CA243793, and P01CA214091 (EKF). NU was supported by a fellowship from the Leukemia & Lymphoma Society.

## Acknowledgments

This study makes use of data generated by the Blueprint Consortium. A full list of the investigators who contributed to the generation of the data is available from [www.blueprint-epigenome.eu](http://www.blueprint-epigenome.eu). Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 – BLUEPRINT. This study also makes use of data generated by the ENCODE consortium, in particular by the Brenton Graveley laboratory at the University of Connecticut. This research was supported in part using high performance computing (HPC) resources and services provided by Technology Services at Tulane University, New Orleans, LA. We also acknowledge the use of computational resources and expertise from the Tulane Cancer Crusaders Next Generation Sequence Analysis Core at the Tulane Cancer Center.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.1060114/full#supplementary-material>

**SUPPLEMENTARY DATA SHEET 2**  
PCR primers used in this study.

**SUPPLEMENTARY TABLE 1**  
Datasets used in this study.

**SUPPLEMENTARY TABLE 2**  
Unannotated splice junctions detected in B cell subset RNA-Seq samples.

**SUPPLEMENTARY TABLE 3**  
Gene Ontology Biological Process term enrichment (from Enrichr) genes with unannotated splicing in all three blood NBC samples.

**SUPPLEMENTARY TABLE 4**  
Hypergeometric test p-values for pairwise comparisons of unannotated splice junctions in B cell subset samples.

## SUPPLEMENTARY TABLE 5

Hypergeometric test p-values for pairwise comparisons of exon inclusion changes in B cell differentiation, EBV infection of blood B cells, and RBP knockdowns. Log2(FoldChange) by sleuth of mRNA abundance, IncLevelDifference (SE\_max\_dPSI) by rMATS for exon usage and intron retention change (IR\_max\_dPSI) by IRFinder per gene in differentiated B cells compared to blood NBCs and in EBV-infected cells compared to parental blood NBCs. For genes with multiple splice junctions the SE and IR events with the highest absolute inclusion/retention difference measurable in all compared replicates are displayed.

## SUPPLEMENTARY TABLE 6

Hypergeometric test p-values for pairwise comparisons of intron retention changes in B cell differentiation, EBV infection of blood B cells, and RBP knockdowns.

## SUPPLEMENTARY TABLE 7

Gene Ontology Biological Process term enrichment (from Enrichr) in genes with protein domains restored or NMD targeting reduced by exon inclusion in differentiated or EBV-infected cells. The top 10 terms with lowest adjusted p-value for each experiment/timepoint are shown.

## SUPPLEMENTARY TABLE 8

Gene Ontology Biological Process term enrichment (from Enrichr) in genes with restored protein domains or reduced NMD targeting in tonsil NBCs (tNBCs) compared to blood NBCs (bNBCs) and in PCs compared to bNBCs.

## SUPPLEMENTARY TABLE 9

Adjusted p-values from rMATS for altered exon usage in genes in the KEGG BCR signaling pathway in differentiated B cells compared to blood NBCs and in EBV-infected cells compared to parental blood NBCs.

## References

- Rodriguez JM, Pozo F, di Domenico T, Vazquez J, Tress ML. An analysis of tissue-specific alternative splicing at the protein level. *PLoS Comput Biol* (2020) 16(10):e1008287. doi: 10.1371/journal.pcbi.1008287
- Furlanis E, Scheiffle P. Regulation of neuronal differentiation, function, and plasticity by alternative splicing. *Annu Rev Cell Dev Biol* (2018) 34:451–69. doi: 10.1146/annurev-cellbio-100617-062826
- Yue L, Wan R, Luan S, Zeng W, Cheung TH. Dek modulates global intron retention during muscle stem cells quiescence exit. *Dev Cell* (2020) 53(6):661–76.e6. doi: 10.1016/j.devcel.2020.05.006
- Edwards CR, Ritchie W, Wong JJ, Schmitz U, Middleton R, An X, et al. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood* (2016) 127(17):e24–34. doi: 10.1182/blood-2016-01-692764
- Pimentel H, Parra M, Gee SL, Mohandas N, Pachter L, Conboy JG. A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res* (2016) 44(2):838–51. doi: 10.1093/nar/gkv1168
- Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science* (2014) 345(6204):1251033. doi: 10.1126/science.1251033
- Ullrich S, Guigo R. Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during b-cell development. *Nucleic Acids Res* (2020) 48(3):1327–40. doi: 10.1093/nar/gkz1180
- Pai AA, Baharian G, Page Sabourin A, Brinkworth JF, Nedelec Y, Foley JW, et al. Widespread shortening of 3' untranslated regions and increased exon inclusion are evolutionarily conserved features of innate immune responses to infection. *PLoS Genet* (2016) 12(9):e1006338. doi: 10.1371/journal.pgen.1006338
- Ni T, Yang W, Han M, Zhang Y, Shen T, Nie H, et al. Global intron retention mediated gene regulation during CD4+ T cell activation. *Nucleic Acids Res* (2016) 44(14):6817–29. doi: 10.1093/nar/gkw591
- Wong JJ, Ritchie W, Ebner OA, Selbach M, Wong JW, Huang Y, et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* (2013) 154(3):583–95. doi: 10.1016/j.cell.2013.06.052
- Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* (2016) 167(5):1398–414.e24. doi: 10.1016/j.cell.2016.10.026
- Kulis M, Merkel A, Heath S, Queiros AC, Schuyler RP, Castellano G, et al. Whole-genome fingerprint of the DNA methylome during human b cell differentiation. *Nat Genet* (2015) 47(7):746–56. doi: 10.1038/ng.3291
- Wang C, Li D, Zhang L, Jiang S, Liang J, Narita Y, et al. RNA Sequencing analyses of gene expression during Epstein-Barr virus infection of primary B lymphocytes. *J Virol* (2019) 93(13):e00226–19. doi: 10.1128/JVI.00226-19
- Yanagi Y, Okuno Y, Narita Y, Masud H, Watanabe T, Sato Y, et al. RNAseq analysis identifies involvement of EBNA2 in PD-L1 induction during Epstein-Barr virus infection of primary B cells. *Virology* (2021) 557:44–54. doi: 10.1016/j.virol.2021.02.004
- Mrozek-Gorska P, Buschle A, Pich D, Schwarzmayr T, Fechtner R, Scialdone A, et al. Epstein-Barr Virus reprograms human B lymphocytes immediately in the prelatent phase of infection. *Proc Natl Acad Sci U S A* (2019) 116(32):16046–55. doi: 10.1073/pnas.1901314116
- Shimizu N, Tanabe-Tochikura A, Kuroiwa Y, Takada K. Isolation of Epstein-Barr virus (EBV)-negative cell clones from the EBV-positive burkitt's lymphoma (BL) line akata: Malignant phenotypes of BL cells are dependent on EBV. *J Virol* (1994) 68(9):6069–73. doi: 10.1128/JVI.68.9.6069-6073.1994
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29(1):15–21. doi: 10.1093/bioinformatics/bts635
- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc Natl Acad Sci U S A* (2014) 111(51):E5593–601. doi: 10.1073/pnas.1419161111
- Kahles A, Ong CS, Zhong Y, Ratsch G. SplAdder: Identification, quantification and testing of alternative splicing events from RNA-seq data. *Bioinformatics* (2016) 32(12):1840–7. doi: 10.1093/bioinformatics/btw076
- Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ, et al. IRFinder: Assessing the impact of intron retention on mammalian gene expression. *Genome Biol* (2017) 18(1):51. doi: 10.1186/s13059-017-1184-4
- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34(5):525–7. doi: 10.1038/nbt.3519
- Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* (2017) 14(7):687–90. doi: 10.1038/nmeth.4324
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
- Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, et al. Whole-genome sequencing of the akata and mutu Epstein-Barr virus strains. *J Virol* (2013) 87(2):1172–82. doi: 10.1128/JVI.02517-12
- O'Grady T, Wang X, Honer Zu Bentrup K, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res* (2016) 44(18):e145. doi: 10.1093/nar/gkw629
- Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol* (2013) 9(8):e1003118. doi: 10.1371/journal.pcbi.1003118
- Ungerleider N, Flemington E. SpliceV: Analysis and publication quality printing of linear and circular RNA splicing, expression and regulation. *BMC Bioinf* (2019) 20(1):231. doi: 10.1186/s12859-019-2865-7
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* (2004) 11(2-3):377–94. doi: 10.1089/106652704110418
- Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. *Curr Protoc* (2021) 1(3):e90. doi: 10.1002/cpz1.90
- Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* (2012) 30(3):224–6. doi: 10.1038/nbt.2153
- Cho H, Davis J, Li X, Smith KS, Battle A, Montgomery SB. High-resolution transcriptome analysis with long-read RNA sequencing. *PLoS One* (2014) 9(9):e108095. doi: 10.1371/journal.pone.0108095

32. Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci U S A* (2014) 111(27):9869–74. doi: 10.1073/pnas.1400447111
33. Thorley-Lawson DA. EBV persistence—introducing the virus. *Curr Top Microbiol Immunol* (2015) 390(Pt 1):151–209. doi: 10.1007/978-3-319-22822-8\_8
34. Inagaki T, Sato Y, Ito J, Takaki M, Okuno Y, Yaguchi M, et al. Direct evidence of abortive lytic infection-mediated establishment of Epstein-Barr virus latency during b-cell infection. *Front Microbiol* (2020) 11:575255. doi: 10.3389/fmicb.2020.575255
35. Pich D, Mrozek-Gorska P, Bouvet M, Sugimoto A, Akidil E, Grundhoff A, et al. First days in the life of naive human B lymphocytes infected with Epstein-Barr virus. *mBio* (2019) 10(5):e01723–19. doi: 10.1128/mBio.01723-19
36. Mehmood A, Laiho A, Venalainen MS, McGlinchey AJ, Wang N, Elo LL. Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief Bioinform* (2020) 21(6):2052–65. doi: 10.1093/bib/bbz126
37. Muller IB, Meijers S, Kampstra P, van Dijk S, van Elswijk M, Lin M, et al. Computational comparison of common event-based differential splicing tools: Practical considerations for laboratory researchers. *BMC Bioinform* (2021) 22(1):347. doi: 10.1186/s12859-021-04263-9
38. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A Large-scale binding and functional map of human RNA-binding proteins. *Nature* (2020) 583(7818):711–9. doi: 10.1038/s41586-020-2077-3
39. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. doi: 10.1038/nature11247
40. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res* (2018) 46(D1):D794–801. doi: 10.1093/nar/gkx1081
41. Kouzine F, Wojtowicz D, Yamane A, Resch W, Kieffer-Kwon KR, Bandle R, et al. Global regulation of promoter melting in naive lymphocytes. *Cell* (2013) 153(5):988–99. doi: 10.1016/j.cell.2013.04.033
42. Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* (2014) 28(23):2663–76. doi: 10.1101/gad.252106.114
43. Zorca CE, Kim LK, Kim YJ, Krause MR, Zenklusen D, Spilianakis CG, et al. Myosin VI regulates gene pairing and transcriptional pause release in T cells. *Proc Natl Acad Sci U.S.A.* (2015) 112(13):E1587–93. doi: 10.1073/pnas.1502461112
44. Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* (2014) 24(11):1774–86. doi: 10.1101/gr.177790.114
45. Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. *RNA* (2004) 10(5):757–65. doi: 10.1261/rna.5123504
46. Sakabe NJ, de Souza SJ. Sequence features responsible for intron retention in human. *BMC Genomics* (2007) 8:59. doi: 10.1186/1471-2164-8-59
47. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci* (1998) 23(6):198–9. doi: 10.1016/s0968-0004(98)01208-0
48. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res* (2020) 48(D1):D265–D8. doi: 10.1093/nar/gkz991
49. Manet E, Polveche H, Mure F, Mrozek-Gorska P, Roisne-Hamelin F, Hammerschmidt W, et al. Modulation of alternative splicing during early infection of human primary B lymphocytes with Epstein-Barr virus (EBV): A novel function for the viral EBNA-LP protein. *Nucleic Acids Res* (2021) 49(18):10657–76. doi: 10.1093/nar/gkab787
50. Lareau LF, Brenner SE. Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* (2015) 32(4):1072–9. doi: 10.1093/molbev/msv002
51. Homa NJ, Salinas R, Forte E, Robinson TJ, Garcia-Blanco MA, Luftig MA. Epstein-Barr Virus induces global changes in cellular mRNA isoform usage that are important for the maintenance of latency. *J Virol* (2013) 87(22):12291–301. doi: 10.1128/JVI.02464-13
52. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27
53. Caballero R, Setien F, Lopez-Serra L, Boix-Chornet M, Fraga MF, Ropero S, et al. Combinatorial effects of splice variants modulate function of aiolos. *J Cell Sci* (2007) 120(Pt 15):2619–30. doi: 10.1242/jcs.007344
54. Chang X, Li B, Rao A. RNA-Binding protein hnRNPL regulates mRNA splicing and stability during b-cell to plasma-cell differentiation. *Proc Natl Acad Sci U.S.A.* (2015) 112(15):E1888–97. doi: 10.1073/pnas.1422490112
55. Chokeshai-u-saha K, Lepoivre C, Grieco L, Nguyen C, Ruxrungtham K. Comparison of immunological characteristics of peripheral, splenic and tonsillar naive B cells by differential gene expression meta-analyses. *Asian Pac J Allergy Immunol* (2012) 30(4):326–30.
56. Morbach H, Eichhorn EM, Liese JG, Girschick HJ. Reference values for b cell subpopulations from infancy to adulthood. *Clin Exp Immunol* (2010) 162(2):271–9. doi: 10.1111/j.1365-2249.2010.04206.x
57. Zeidler R, Meissner P, Eissner G, Lazis S, Hammerschmidt W. Rapid proliferation of B cells from adenoids in response to Epstein-Barr virus infection. *Cancer Res* (1996) 56(24):5610–4.
58. Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, et al. Myc regulates chromatin decompaction and nuclear architecture during b cell activation. *Mol Cell* (2017) 67(4):566–78.e10. doi: 10.1016/j.molcel.2017.07.013
59. Tesi A, de Pretis S, Furlan M, Filipuzzi M, Morelli MJ, Andronache A, et al. An early myc-dependent transcriptional program orchestrates cell growth during b-cell activation. *EMBO Rep* (2019) 20(9):e47987. doi: 10.15252/embr.201947987
60. Huang S, Spector DL. Dynamic organization of pre-mRNA splicing factors. *J Cell Biochem* (1996) 62(2):191–7. doi: 10.1002/(sici)1097-4644(199608)62:2<191::aid-jcb7>3.0.co;2-n
61. Haaland RE, Herrmann CH, Rice AP. Increased association of 7SK snRNA with tat cofactor p-TEFb following activation of peripheral blood lymphocytes. *AIDS* (2003) 17(17):2429–36. doi: 10.1097/00002030-200311210-00004
62. Tinguely A, Chemin G, Peron S, Sirac C, Reynaud S, Cogne M, et al. Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during b cell development. *Mol Cell Biol* (2012) 32(1):107–17. doi: 10.1128/MCB.06138-11
63. Peng Q, Wang L, Wang J, Liu C, Zheng X, Zhang X, et al. Epstein-Barr Virus EBNA2 phase separation regulates cancer-associated alternative RNA splicing patterns. *Clin Transl Med* (2021) 11(8):e504. doi: 10.1002/ctm2.504
64. Sinclair AJ, Palmero I, Peters G, Farrell PJ. EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus. *EMBO J* (1994) 13(14):3321–8. doi: 10.1002/j.1460-2075.1994.tb06634.x
65. Barrington RA, Zhang M, Zhong X, Jonsson H, Holodick N, Cherukuri A, et al. CD21/CD19 coreceptor signaling promotes b cell survival during primary immune responses. *J Immunol* (2005) 175(5):2859–67. doi: 10.4049/jimmunol.175.5.2859
66. Aubry JP, Pochon S, Graber P, Jansen KU, Bonnefoy JY. CD21 is a ligand for CD23 and regulates IgE production. *Nature* (1992) 358(6386):505–7. doi: 10.1038/358505a0
67. Delcayre AX, Salas F, Mathur S, Kovats K, Lotz M, Lernhardt W. Epstein Barr Virus/Complement C3d receptor is an interferon alpha receptor. *EMBO J* (1991) 10(4):919–26. doi: 10.1002/j.1460-2075.1991.tb08025.x
68. Sandberg K, Eloranta ML, Campbell IL. Expression of Alpha/Beta interferons (IFN-Alpha/Beta) and their relationship to IFN-Alpha/Beta-Induced genes in lymphocytic choriomeningitis. *J Virol* (1994) 68(11):7358–66. doi: 10.1128/JVI.68.11.7358-7366.1994