Published online 24 October 2018 *Nucleic Acids Research, 2018, Vol. 46, No. 21* ***11357–11369***
doi: 10.1093/nar/gky946

Aberrant splicing in B-cell acute lymphoblastic leukemia

Kathryn L. Black[1,†], Ammar S. Naqvi[1,2,†], Mukta Asnani[1], Katharina E. Hayer[1,2], Scarlett Y. Yang[1,3], Elisabeth Gillespie[1], Asen Bagashev[1], Vinodh Pillai[1], Sarah K. Tasian[4], Matthew R. Gazzara[5,6], Martin Carroll[7], Deanne Taylor[2,4], Kristen W. Lynch[3,5], Yoseph Barash[6,8] and Andrei Thomas-Tikhonenko[1,3,4,9,*]

[1]Department of Pathology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [2]Department of Biomedical & Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [3]Immunology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA, [4]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [5]Department of Biochemistry & Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [6]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [7]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [8]Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA and [9]Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Received May 30, 2018; Revised September 24, 2018; Editorial Decision October 01, 2018; Accepted October 04, 2018

ABSTRACT

Aberrant splicing is a hallmark of leukemias with mutations in splicing factor (SF)-encoding genes. Here we investigated its prevalence in pediatric B-cell acute lymphoblastic leukemias (B-ALL), where SFs are not mutated. By comparing these samples to normal pro-B cells, we found thousands of aberrant local splice variations (LSVs) per sample, with 279 LSVs in 241 genes present in every comparison. These genes were enriched in RNA processing pathways and encoded ∼100 SFs, e.g. hnRNPA1. HNRNPA1 3′UTR was most pervasively mis-spliced, yielding the transcript subject to nonsense-mediated decay. To mimic this event, we knocked it down in B-lymphoblastoid cells and identified 213 hnRNPA1-regulated exon usage events comprising the hnRNPA1 splicing signature in pediatric leukemia. Some of its elements were LSVs in DICER1 and NT5C2, known cancer drivers. We searched for LSVs in other leukemia and lymphoma drivers and discovered 81 LSVs in 41 additional genes. Seventy-seven LSVs out of 81 were confirmed using two large independent B-ALL RNA-seq datasets, and the twenty most common B-ALL drivers, including NT5C2, showed higher prevalence of aberrant splicing than of somatic mutations. Thus, post-transcriptional deregulation of SF can drive widespread changes in B-ALL splicing and likely contributes to disease pathogenesis.

INTRODUCTION

Despite advances in the treatment of pediatric B-ALL, children with relapsed or refractory disease account for a substantial number of childhood cancer-related deaths. Adults with B-ALL experience even higher relapse rates and long-term event-free survival of <50% (1). Recently, significant gains in the treatment of B-ALL have been achieved through the use of immunotherapies directed against CD19, a protein expressed on the surface of most B-cell neoplasms (2,3). These gains culminated in the recent FDA approval of tisagenlecleucel and axicabtagene ciloleucel, CD19-redirected chimeric antigen receptor (CAR) T-cell immunotherapies, for patients with refractory/relapsed B-cell malignancies. However, relapses occur in 10–20% of patients with B-ALL treated with CD19-directed immunotherapies, often due to epitope loss and/or B-cell dedifferentiation into other lineages (4–7). Other targets for immunotherapy include CD20 and CD22 (8–11). However, neither antigen is uniformly expressed in B-ALL, and factors accounting for this mosaicism are poorly understood (3).

*To whom correspondence should be addressed. Tel: +1 267 426 9699; Fax: +1 267 426 8125; Email: andreit@pennmedicine.upenn.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Elisabeth Gillespie, Inovio Pharmaceuticals, Plymouth Meeting, PA 19462, USA.

We previously reported a new mechanism of pediatric B-ALL resistance to CD19-directed immunotherapy. We discovered that in some cases, resistance to CD19 CAR T cells was generated through alternative splicing of CD19 transcripts. This post-transcriptional event was mediated by a specific splicing factor SRSF3 and generated a CD19 protein isoform invisible to the immunotherapeutic agent via skipping of exon 2 [(12,13), reviewed in (14)].

Our discovery of a resistance mechanism based on alternative splicing prompted us to investigate the extent of this phenomenon in additional B-ALL cases. While driver mutations in splicing factors such as SRSF2, SF3B1 and U2AF1 have recently been discovered in myelodysplastic syndrome/acute myelogenous leukemia (15–17) and chronic lymphocytic leukemia (18,19), SF mutations have not been reported in B-ALL. Nevertheless, our prior work suggested the possibility that SRSF3 (and by inference other SFs) could be deregulated in B-ALL (12), bringing about wide-spread splicing aberrations.

This model would be particularly attractive because B-ALL is a chromosome translocation-driven disease where the prevalence of somatic mutations and copy number variations is relatively low. For example, the commonly mutated *IKZF1* gene (which encodes the Ikaros transcription factor) is affected by missense mutations in just ∼20% of B-ALL cases. Similarly, mutations in the key tumor suppressor gene (TSG) TP53 are found in only ∼7% of B-ALLs (per COSMIC database) (20,21). In addition, both genes are robustly transcribed across individual B-ALLs and thus are not epigenetically silenced. This raises the possibility that they and other TSGs are dysregulated by post-transcriptional events, such as alternative splicing.

MATERIALS AND METHODS

Bone marrow fractionation

Isolated mononuclear cells and whole bone marrow aspirates were obtained, respectively, from the University of Pennsylvania Stem Cell and Xenograft Core facility and CHOP Hematopathology Laboratory. For pediatric bone marrow samples, mononuclear cells were isolated by spinning over Ficoll gradient, as described earlier (22). Residual red blood cells were lysed with ammonium chloride lysis buffer with gentle rocking at room temperature for 10 min. Cells were pelleted by spinning at 250 × g for 10 min at 4°C and washed once with cold PBS/2% FBS. Cells were resuspended in 1 ml PBS/2%FBS and incubated with 500 μl FC Block on ice for 10 min. Cells were stained with 1 ml CD34-PE, 500 μl CD19-APC, and 500 μl IgM-FITC for 30 min on ice. Cells were pelleted at 1300 RPM for 6 min at 4°C and washed twice in cold PBS. Cells were resuspended in 3 ml PBS/2%FBS containing 1 μl/ml of 0.1 mg/ml DAPI. Cells were sorted 4-ways using MoFlo ASTRIOS directly into RLT Lysis buffer (Qiagen) at a ratio of 1:3.

Primary B-ALL samples acquisition

Twenty four primary pediatric B-ALL samples were obtained from the CHOP Center for Childhood Cancer Research leukemia biorepository. Mononuclear cells from fresh bone marrow or peripheral blood specimens were purified via Ficoll gradient separation (22) and cryopreserved for downstream experimental use.

RNA-seq of bone marrow fractions, primary samples and cell lines

RNA was isolated using Qiagen RNeasy Mini Kit. RNA integrity and concentration were determined using Eukaryote Total RNA Nano assay on BioAnalyzer (CHOP NAPCore). RNA-seq was performed on 10 ng–1 μg of total RNA according to the GeneWiz Illumina Hi-seq protocol for poly(A)-selected samples (2 × 150 bp pair-end sequencing, 350M raw reads per lane).

RNA-Seq alignment, quantification and differential expression

Fastq files of RNA-seq obtained from GeneWiz were mapped using STAR aligner (23). STAR was run with the option 'alignSJoverhangMin 8′. We generated STAR genome reference based on the hg19 build. Alignments were then quantified for each mRNA transcript using HTSeq with the Ensembl-based GFF file and with the option '-m intersection-strict'. Normalization of the raw reads was performed using the trimmed mean of *M*-values (TMM). Principal component analysis (PCA) was done on normalized count values of the samples using a correlation matrix and a calculated score for each principal component. Differential expression of wild-type and knock-down or pro-B and B-ALL RNA-Seq datasets were assessed based on a model using the negative binomial distribution, a method employed by the R package DESeq2 (24). Subsequent bar charts were generated using the R package ggplot2. Those differential genes that had a *P*-value of <0.05 were deemed as significantly up or down-regulated.

Splicing analysis

In order to detect LSVs, the MAJIQ tool (version 1.03) was used (25). MAJIQ was run on the Ensembl-based GFF annotations, disallowing *de novo* calls (26). LSVs that had at least a 20% change at a 95% confidence interval between two given conditions were chosen for further analysis. The 95% confidence interval in this context is defined as follows: given observed reads in both experiments there is a 95% posterior probability, according to the MAJIQ model, that there is a change of 20% or more in PSI. In-house customized Perl and Ruby scripts were used to filter for events that corresponded to exon inclusion events only, forcing one event per LSV. For the genes containing differential LSVs in all 18 comparisons, enriched gene ontologies (GO) were identified using the gene functional classification tool DAVID (v. 6.7), reporting the top most significant hits based on *P*-value and false discovery rate. Heatmaps were generated for each ΔPSI value of each LSV that passed the 20% change threshold between Pro-B and B-ALL and additionally filtered for a frequency of $n \geq 2$ in B-ALL samples. The R package gplots was used for this purpose. LSVs detected by MAJIQ were also validated using Leafcutter version 0.2.7 (27). In addition, splicing of HNRNPA1 3′ UTR

was validated by RT-qPCR using a forward primer spanning exons 9–10 and a reverse primer in exon 11. Expression levels were normalized to actin and the PHL156 leukemia.

Spearman's correlations

Correlations and their significance were computed using the nonparametric Spearman's rank-order correlation implemented in R function *cor.test()*. Correlation analysis was performed between a) RQ values of HNRNPA1 exon 11 inclusion as determined by RT-qPCR and PSI values of HNRNPA1 exon 11 inclusion as determined by MAJIQ; b) RQ values of HNRNPA1 constitutive exons 2–3 and RQ values of HNRNPA1 exon 11 inclusion as determined by RT-qPCR (normalized to actin and PHL156), and c) normalized counts of HNRNPA1 constitutive exon 9 and cassette exon 11, as determined by RNA-seq.

Motif analysis

The motif detection pipeline utilized RBPMap (v. 1.1) with a medium stringency level (P-value ≤ 0.005) and conservation filter (28). The analysis was performed on affected exons and ±200nts into adjacent intronic regions. hnRNPA1 motifs (29) were identified and average z-scores and P-values for all significant hits were calculated.

Actinomycin D mRNA stability assay

2×10^6 Nalm6 B-ALL cells were treated with 5 μg/ml of actinomycin D for various time intervals between 0 and 24 h followed by RNA extraction. RT-qPCR was performed to assess the abundance of HNRNPA1 3′UTR isoforms using forward primer spanning exons 9–10 and reverse primers in exon 11 or spanning exons 12–13. Data was normalized to actin and 0 hr time point. Linear regression was used to plot the decay curves.

Nonsense-mediated decay inhibition via cyclohexamide treatment and UPF1/2 knockdown

2×10^6 Nalm6 B-ALL cells were treated in duplicates with either DMSO or 30 μg/ml cyclohexamide for 6 h followed by RNA extraction. Alternatively, they were electroporated with 133 nM non-targeting control pool of four siRNAs or 750nM ON-TARGET Plus Human UPF1 SMARTpools of four siRNAs + 300–750 nM ON-TARGET Plus Human UPF2 SMARTpool of four siRNAs (all from Dharmacon). To confirm inhibition of NMD, gel-based PCR with primers spanning SRSF3 poison exon 4 to detect accumulation of a canonical NMD target was performed. To determine if either HNRNPA1 3′UTR was an NMD substrate, RT-qPCR with forward primer spanning exons 9–10 and reverse primers in exon 11 or spanning exons 12–13 were performed. Data were normalized to HNRNPA1 constitutive exons 2–3 and DMSO.

siRNA knock-down of hnRNPA1 and SRSF3

Biological replicate experiments were performed on 2×10^6 P493-6 B-lymphoblastoid cells (30) electroporated using Amaxa Program O-006 with either 133 nM non-targeting control pool of four siRNAs, 300nM ON-TARGET Plus Human HNRNPA1 SMARTpool of four siRNAs or 50–300 nM ON-TARGET Plus Human SRSF3 SMARTpool of four siRNAs (all from Dharmacon). Cells were re-plated in 2 ml warm tetracycline-free RPMI for 24 h. Knockdown was validated through RT-qPCR and western blotting with anti-hnRNPA1 antibody (Abcam#ab5832) or as described previously for SRSF3 (12).

LNA-GapmeR-mediated knock-down of HNRNPA1

To knock-down hnRNPA1 by a non-RNAi mechanism, 2×10^6 P493-6 B-lymphoblastoid cells were electroporated with either 100 nM custom non-targeting control pool of four GapmeRs or 100 nM of custom HNRNPA1-ATG-targeting GapmeR using Neon Transfection System (ThermoFisher). All GapmeRs were hybrids of phosphorothioate-bonded DNA and 2′-O-methyl modified RNA. Cells were plated in 2 ml warm tetracycline-free RPMI for 48 h. RNA isolation was performed as described above. Knockdown was validated through RT-qPCR and western blotting with anti-hnRNPA1 antibody (Abcam #ab5832). Alternative splicing of DICER1 (LSV ID ENSG00000100697:95623567-95623607:target) was validated using RT-qPCR with primers amplifying the alternative exon 3–4 junction normalized to primers amplifying the constitutive exon 25. Alternative splicing of NT5C2 (LSV ID ENSG00000076685:104865463-104865558:target) was validated using RT-qPCR with primers amplifying the exon4-exon4a junction and normalized to primers amplifying constitutive exons 5–7.

Radioactive gel-based RT-PCR

To validate the regulation of alternative splicing of TP53 (LSV ID: ENSG00000141510:7573927-7574033:source) 26 cycles of radioactive RT-PCR were performed as described previously (12) using a forward primer in TP53 exon 9 and reverse primer in exon 10.

Oligonucleotide sequences

Sequences of all oligonucleotides are shown in Supplementary Table S4.

Datasets

Cancer gene symbols and annotations were downloaded from the COSMIC database (20,21). Known splice factors were annotated and obtained from published studies or Ensembl-annotated databases. Pediatric B-ALL samples from the TARGET consortium (phs000218.v19.p7) were accessed via NCBI dbGaP (the Database of Genotypes and Phenotypes) Authorized Access system (project #10088). Pediatric B-ALL samples from the St Jude Children's Research Hospital (EGAD00001002704 and EGAD00001002692) were accessed by permission from the Computational Biology Committee though The European Bioinformatics Institute (EMBL-EBI). CHOP B-ALL samples used in this submission were passed to investigators

with a coded identifier and no protected health information (PHI). Basic demographic, treatment, relapse, and survival outcome data for delivered specimens were provided through an online honest broker system. Specimens were obtained via informed consent on institutional research protocols in accordance with the Declaration of Helsinki.

RESULTS

RNA-seq analysis of bone marrow-derived human B-cells

To determine if patterns of splicing dysregulation occur in B-ALL, we first generated normal B-lymphocyte datasets corresponding to potential cells of origin. To this end, we obtained from the University of Pennsylvania Stem Cell and Xenograft Core facility two healthy adult bone marrow biopsies and from the Children's Hospital of Philadelphia (CHOP) Hematopathology Laboratory two bone marrow biopsies from children without leukemia. We enriched for mononuclear cells using Ficoll gradient separation, stained cells for combinations of stage-specific surface markers, and sorted B-cell subsets using flow cytometry. Specifically, we fractionated bone marrow progenitors into early progenitors ($CD34^+/CD19^-/IgM^-$), pro-B ($CD34^+/CD19^+/IgM^-$), pre-B ($CD34^-/CD19^+/IgM^-$), and immature B ($CD34^-/CD19^+/IgM^+$) populations (Figure 1A top and Supplementary Figure S1a). We then extracted RNA, performed RNA-seq, and quantified transcript levels. Concordant with flow cytometric profiles, CD19 mRNA levels increased throughout differentiation stages, CD34 mRNA was confined to early progenitors and pro-B fractions, and IgM transcript was expressed only in the immature fraction (Figure 1A, bottom). Furthermore, we performed principal component analysis (PCA) on all expression datasets and found tight clustering of the four fractions from different donors, suggesting that at the level of mRNA expression B-cell differentiation supersedes individual variations (Figure 1B). To ensure that the clustering of fractions was not driven solely by CD19 and CD34 expression, we repeated the PCA, removing CD19 and CD34 expression contributions and observe very similar clustering of fractions (Supplementary Figure S1B).

RNA-Seq analysis of primary B-ALL samples

We obtained 24 primary pediatric B-ALL samples from the CHOP Center for Childhood Cancer Research leukemia biorepository. Ficoll-purified mononuclear cells from fresh bone marrows or peripheral blood were used for total RNA extraction. Based on the presence of intact RNA (evident by 28S and 18S bands on gels), an RNA integrity score >5.2, and RNA concentration of $>40ng/\mu L$, we successfully extracted high quality RNA for sequencing from 18 out of 24 frozen samples (Supplementary Figure S1C, red asterisks indicate samples that did not pass QC and were not sequenced). These samples were comprised of several different phenotypic and genetic B-ALL subtypes (31) (Supplementary Table S1 and Supplementary Figure S1D). We performed RNA-seq of these leukemia samples and first compared them to adult and pediatric normal bone marrow fractions with respect to CD19 and CD34 expression. We found that B-ALL samples closely resembled the pro-B fractions in that CD19 and CD34 transcripts were readily detectable in the low-to-medium range (Figure 1C). To extend this analysis to the entire transcriptome, we performed PCA on B-ALL samples and four sets of normal bone marrow fractions (two pediatric and two adult). Once again, the 18 B-ALL samples clustered most closely with the pro-B fractions (Figure 1D). This similarity was reflected in shortest Euclidian distances between PC1 coordinates for the centroids of the B-ALL and pro-B cell samples (Supplementary Figure S1E). Therefore, we chose the pro-B fractions as cell-of-origin controls for B-ALL splicing analysis.

Global patterns of aberrant splicing in pediatric B-ALL

To detect patterns of alternative splicing in B-ALL, we utilized the MAJIQ (Modeling Alternative Junction Inclusion Quantification) algorithm (25). MAJIQ offers the ability to detect, quantify, and visualize complex splicing variations, including de-novo variations, while comparing favorably in reproducibility and false discovery rate to common alternatives (25,32). Using MAJIQ, we performed 18 independent pairwise comparisons between the average of pediatric pro-B cell fractions and leukemia samples. To assess heterogeneity in splicing across samples, we measured the number of differential LSVs (minimum of 20% change in splicing, 95% CI—see Materials and Methods) in each sample and compared their identities. We observed that each B-ALL specimen had >3000 LSVs but most LSVs were either unique (first bar in Figure 2A) or shared by only a small number of B-ALL patient samples (subsequent bars in Figure 2A). However, we also found a total of 279 aberrant LSVs in 241 genes that were detected in all of our 18 pairwise comparisons (last bar in Figure 2A). Provocatively, when this 241-gene set was analyzed using DAVID (Database for Annotation, Visualization, and Integrated Discovery) (33), the top gene ontology (GO) categories (34) were related to RNA splicing (Figure 2A, inset). When we investigated individual SF transcripts, we found that of the 167 well-characterized SF genes [Supplementary Table S2, from (35)], 101 were alternatively spliced in B-ALL compared to pro-B cells (Figure 2B, bottom left, proB- versus B-ALL bar chart). Moreover, these changes were highly specific to the malignant phenotype and not to states of normal B-cell differentiation (Figure 2B, upper and middle panels). For example, only six SF transcripts were alternatively spliced during the early progenitors-to-proB transition and only nine during the proB-to-preB transition (Figure 2B, middle).

For each LSV, MAJIQ provides a ΔPSI (percent-spliced-in) value to indicate changes in isoform abundance. We observed that many members of the hnRNP and SRSF families, which play key roles in exon inclusion or skipping (36), exhibited profound changes in ΔPSI values (Figure 2C). These 25 detectable LSVs included increases in the so-called 'poison exons' with in-frame stop codons in several SRSF proteins (37). Of note, some transcripts contained multiple aberrant LSVs. For example, SRSF11 had three distinct LSVs with different start/end UCSC Genome Browser coordinates.

To increase confidence in the detection of splice factor related LSVs in B-ALL compared to normal pediatric pro-B samples we also compared B-ALL to normal adult pro-B

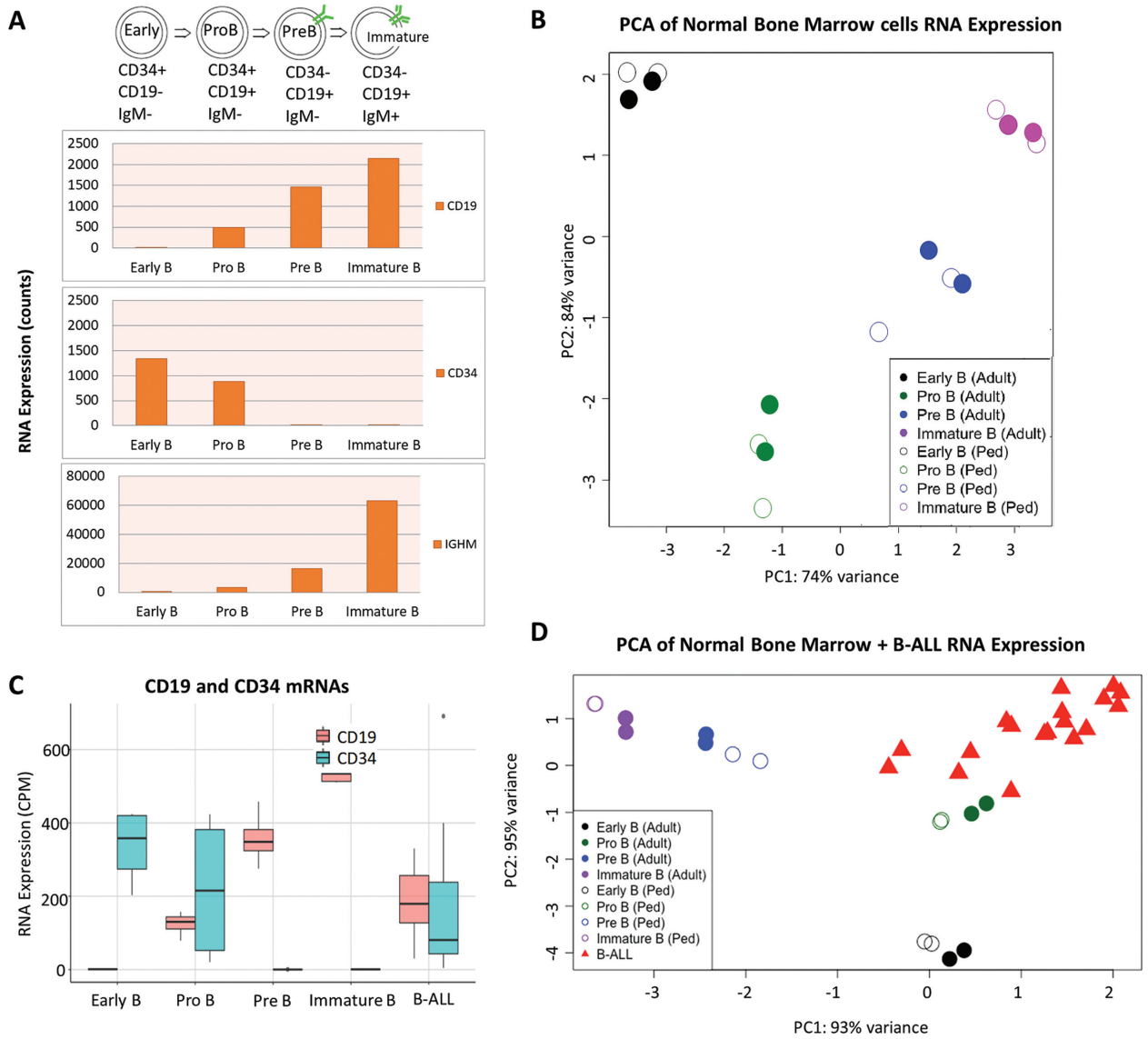

Figure 1. RNA-seq analysis of bone marrow-derived B-cell progenitors and pediatric B-ALL. (**A**) Top: Lymphocytes were isolated from normal whole bone marrow aspirates and fractionated into CD34+/CD19-/IgM-, CD34+/CD19+/IgM-, CD34-/CD19+/ IgM-, and CD34-/CD19+/IgM+ populations. Bottom: Validation of surface marker expression by RNA-Seq in early progenitors, pro-B, pre-B and immature B cells fractions. (**B**) Principal component analysis (PCA) of RNA expression data from bone marrow fractions obtained from 2 adult (solid circles) and two pediatric (open circles) donors. Cumulative variances are shown for each PC. (**C**) Quantification of CD19 and CD34 expression in bone marrow fractions and B-ALL samples. (**D**) PCA on RNA expression data from 18 B-ALL samples and four bone marrow samples.

cells and to publicly available RNA-seq data corresponding to the CD19-positive normal bone marrow cells (BM-GEO) (38). We overlapped these comparisons in a three-way Venn diagram to show that of the 25 LSVs detected in B-ALL versus pediatric pro-B 21 LSVs were also detected in the two other independent and publicly available datasets [B-ALL vs. Adult pro-B and B-ALL vs. BM-GEO (38)] (Figure 2D). Further, we searched for LSVs in the SRSF and hnRNP families in larger, independent TARGET (39) and St Jude Children's Research Hospital (SJCRH) (40) B-ALL datasets using pediatric pro-B cells for comparison. We detected 75 and 78 LSVs in the TARGET and SJCRH datasets, respectively (Supplementary Figure Ss2, left bars, middle and bottom). Of these LSVs, 43 were exclusive to these datasets (Supplementary Figure S2A, third column labeled 43 and 2 connected dots below for TARGET and SJCRH) and 22 were detected in all three datasets (Supplementary Figure S2a, middle, first column labeled 22 and 3 connected dots below for CHOP, TARGET and SJCRH) (see also Supplementary LSV Lists).

To decouple observed changes in splicing from changes in gene expression we performed differential expression analysis on genes with frequently detected LSVs. We found that most genes containing LSVs did not have associated changes in expression, with the notable exception of HNRNPA1, which was observed to be downregulated at least two-fold in approximately half of the B-ALL samples (Figure 2E). Furthermore, after lowering the threshold to 60%

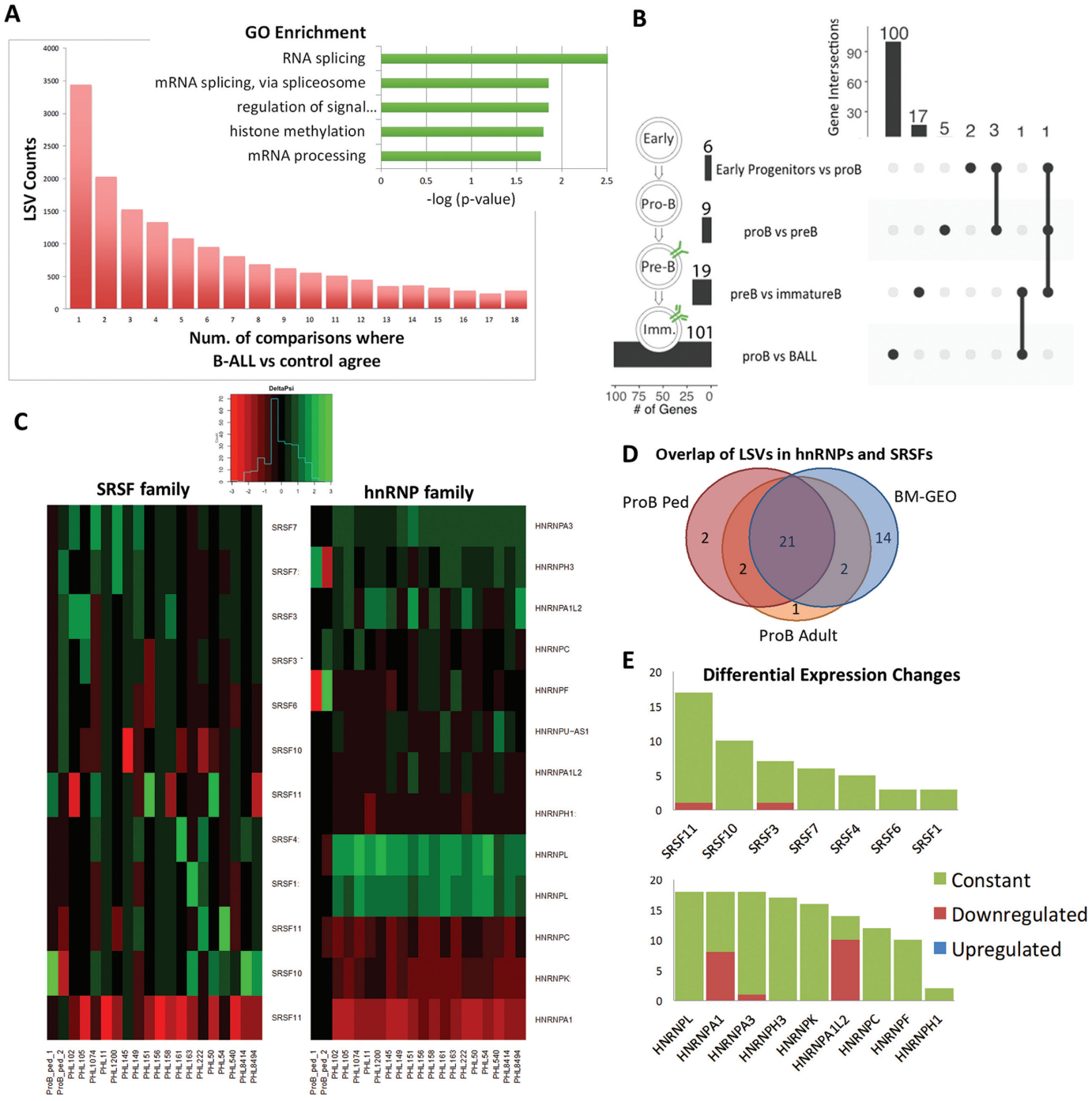

Figure 2. Global patterns of aberrant splicing in pediatric B-ALL. (**A**) Enumeration of local splicing variations (LSVs). The vertical bar chart shows the numbers of LSVs detected by MAJIQ per B-ALL sample compared to normal pro-B cells. The horizontal bar chart (insert) showing enrichment for RNA splicing pathways among genes with most consistent LSVs. (**B**) Overlap of LSVs in SF genes between B-ALL samples and bone marrow fractions. The left panel shows the different sets and number of genes included in them. The middle section visualizes the different combinations of intersection (including each set by itself). The upper panel shows how many genes of the total set are in each intersection. (**C**) Heat maps showing changes in inclusion of exons corresponding to SRSF and hnRNP families members in 18 B-ALL samples compared to pro-B cell fractions. Green color denotes increased inclusion, red color – increased skipping. (**D**) Venn diagram showing overlap of LSVs from hnRNP and SRSF families when B-ALL samples are compared to in-house pediatric pro-B, in-house adult pro-B, and publicly available bone marrow data (GSM1695856 and GSM1695857), as indicated. (**E**) Differential expression analysis of hnRNP and SRSF family members to decouple detection of LSVs with changes in RNA expression. Majority of samples (Y-axis) showed constant expression of indicated genes compared to normal pro-B samples (green). Several samples showed downregulation of a few genes (red), such as HNRNPA1.

change, we found that all 18 samples showed decreased expression relative to proB cells. This interesting expression pattern and the pervasive nature of HNRNPA1 splicing alterations (Figure 2C, right) prompted us to investigate its regulation further.

Dysregulated splicing of HNRNPA1 in B-ALL

According to the Ensembl database (26), there is evidence for HNRNPA1 transcripts with alternatively spliced 3′ UTRs, notably the canonical long proximal exon 11 and two shorter, distal exons (exons 12/13, ENST00000547566) (Figure 3A, top). We observed that the exon 11-containing transcript predominated in normal pro-B cells (Figure 3B, red), but in a typical B-ALL specimen (representative sample PHL50) the predominant event was the skipping of exon 11 to exons 12 and 13 (Figure 3B, blue). To ensure validity of this event, we also ran the LeafCutter splicing algorithm to detect splice events present in the averaged RNA-seq datasets from 18 B-ALL samples and four pro-B samples. Confirming our MAJIQ analysis we were able to detect alternative splicing of HNRNPA1 3′ UTR using this independent analysis (Supplementary Figure S2B). While exon 11 percent-spliced-in (PSI) values varied across leukemia samples, most leukemias had increased skipping of exon 11 compared to pro-B samples (Figure 3C, blue). We also observed intron 10 inclusion in all pro-B and B-ALL samples, although this event was not significantly different between normal and malignant samples (Figure 3C, gray stacks).

We then validated and quantitated exon 11 inclusion by RT-qPCR using a forward primer spanning exons 9–10 and a reverse primer in exon 11 (Figure 3A, D). Using Spearman's *rho* statistic we observed a strong positive association between MAJIQ predictions and RT-qPCR validations (0.61, P-value = 0.00737) (Figure 3E). We next applied the same correlation analysis to HNRNPA1 mRNA expression levels versus exon 11 inclusion and found a positive correlation between the two measurements based on both RNA-seq (0.82, P-value = 0.00002) (Figure 3F) and RT-qPCR (0.6, P-value = 0.00878) (Supplementary Figure S2C). This suggested that preferential splicing from exon 10 to the distal 3′ UTR exons 12/13 (skipping exon 11) may decrease HNRNPA1 RNA steady state levels. To assess the half-lives of HNRNPA1 isoforms we performed actinomycin D time course treatment followed by RT-qPCR. We observed strong differences in mRNA stability between the UTR isoforms. Specifically, the half-life of the exon 11-skipped isoforms was ∼1.5 h (Figure 3G) while the half-life of the exon 11-included isoform was longer than 24 h (Figure 3G and Supplementary Figure S2D).

Based on the exon junction present between exons 12 and 13 downstream from the stop codon we hypothesized that this transcript might be a target of nonsense-mediated decay [NMD; (41,42)]. To test this, we first treated Nalm6 B-ALL cells with the translation inhibitor cyclohexamide to transiently suppress NMD (41,43,44). To show effective suppression of NMD, we performed gel-based PCR for primers spanning SRSF3 exon 4, a canonical target of NMD (45) (Supplementary Figure S2E). We then performed RT-qPCR for HNRNPA1 3′ UTR exons and observed accumulation of the transcript that skips to exon 12/13, but not the transcript including exon 11 (Supplementary Figure S2F)

To further confirm the exons 12/13 transcript as an NMD substrate, we co-knocked down two factors required for NMD, UPF1 and UPF2 (46,47), in Nalm6 B-ALL cells (Supplementary Figure S2g). Despite modest co-knockdown of the factors (∼35% at most), quantification of the canonical NMD substrate SRSF3 exon 4 by gel-based PCR confirmed inhibition of NMD at the highest concentrations of siRNA for UPF1 and UPF2 (750 nM each) (Supplementary Figure S2H). Next, we performed RT-qPCR for HNRNPA1 3′ UTR exons and observed preferential accumulation of the exon 12/13 transcript as compared to the exon 11 transcript (Figure 3H) confirming that only the former is a robust target of degradation by the NMD machinery.

To model downregulation of this splicing factor and to identify affected transcripts, we knocked down HNRNPA1 mRNA with siRNA in the Epstein-Barr Virus (EBV)-transformed human B-lymphoblastoid P493-6 cell line (30). Both the non-targeting scrambled control and the HNRNPA1-targeting reagent were Smart Pools of four siRNAs, designed to target distinct sites within the specific gene of interest. This approach allowed us to reduce the reliance on and the concentration of each individual siRNA and average out their off-target effects, as documented in (48). Knockdown of HNRNPA1 was confirmed by RT-qPCR (Figure 4A) and Western blotting (Figure 4B). Following RNA-Seq and MAJIQ analysis, we identified 213 LSVs in 184 genes associated with knockdown of HNRNPA1 (minimum of 20% ΔPSI, 95% CI). Of these hnRNPA1-dependent LSVs, 74 (or >30%) were present in at least one B-ALL sample (Figure 4C), with more than half of these 74 LSVs altered in 10 or more samples. DAVID analysis of the genes with overlapping LSVs revealed enrichment for macromolecular metabolic pathways (Figure 4D), supporting the well-established role of hnRNP proteins in RNA metabolism (49). We next searched for the presence of the hnRNPA1 binding motif (50) in exons and flanking intronic sequences (±200nt) involved in overlapping LSVs. We were able to identify the UAGGG motif in 68 out of 74 LSVs in question (Figure 4E, top).

Among these genes characterized by overlapping LSVs, involvement in macromolecular metabolism, and the presence of hnRNPA1 binding motifs, was DICER1 (Figure 4E, bottom, purple arrows indicate hnRNPA1 motifs). Interestingly, one of the LSVs in the DICER1 gene was altered in B-ALL samples in the same manner as in the HNRNPA1 KD cells (Figure 4F). We also confirmed the presence of this LSV using LeafCutter algorithm (Supplementary Figure S3A). Notably, this LSV maps to the 5′UTR of the DICER1 transcript, potentially diminishing its translation efficiency. We validated this LSV in Nalm6 B-ALL cells by RT-PCR and Sanger sequencing and noted that this type of deregulation could phenocopy loss-of-function mutations and copy number variations in DICER1 in several types of cancer including leukemias (51).

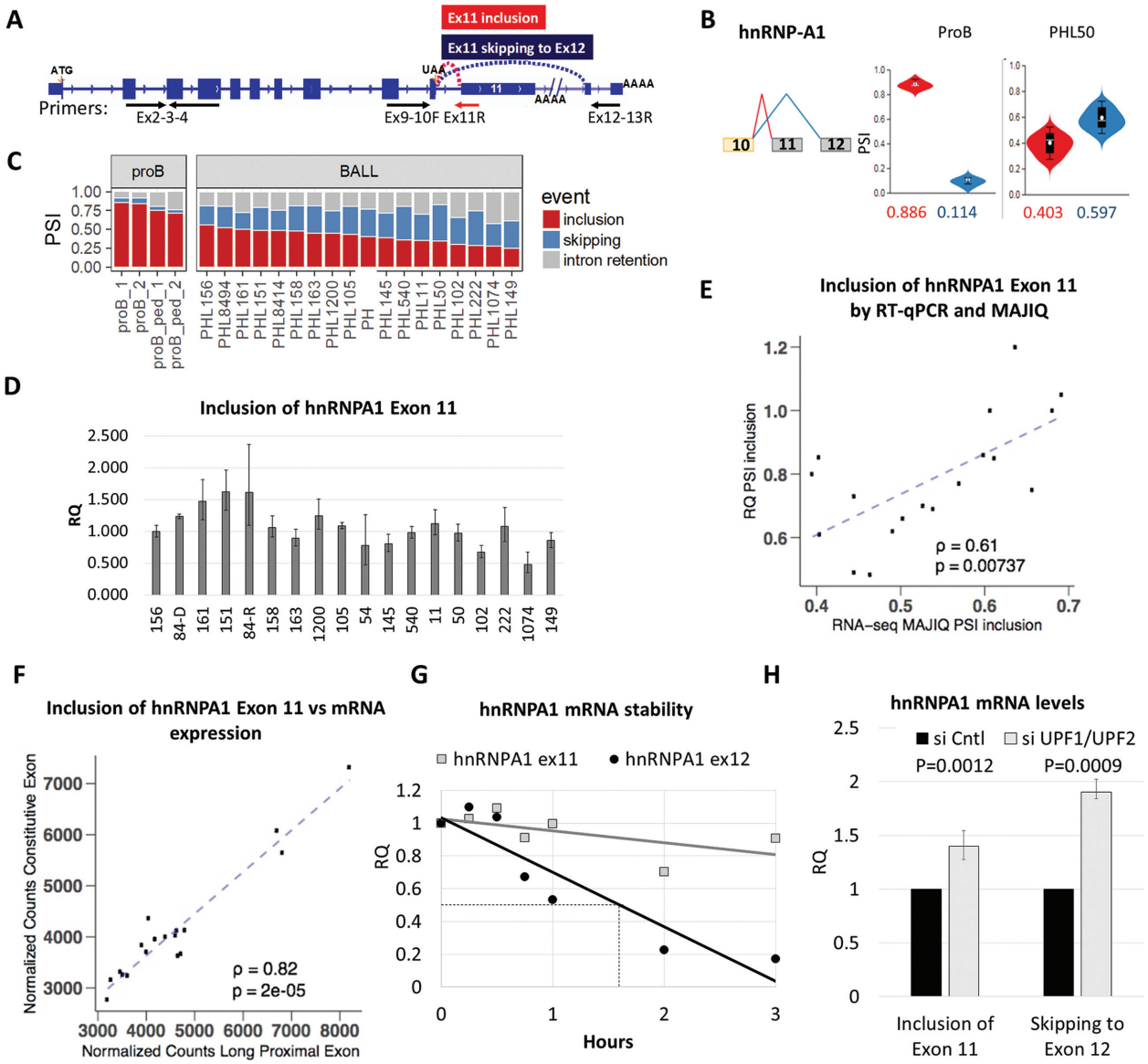

Figure 3. Alternative splicing of HNRNPA1 in pediatric B-ALL. (**A**) Exon–intron structure of the HNRNPA1 transcript and location of RT-qPCR primers. (**B**) Violin plots depicting alternative splicing of the HNRNPA1 transcript. (**C**) Stack plot of PSI values of exon 11 corresponding to pro-B cell fractions and 18 B-ALL samples. (**D**) Analysis of exon 11 inclusion in B-ALL samples by RT-qPCR. (**E**) Spearman's correlation of RT-PCR vs. MAJIQ (rho = 0.06, $P = 0.00737$). (**F**) Spearman's correlations of expression of an HNRNPA1 constitutive exon and alternative exon 11 measured by RNA-Seq (rho = 0.82, $P = 2E{-}05$). (**G**) Measurement of mRNA stability of HNRNPA1 3′ UTR isoforms (proximal exon 11 in gray squares and distal exon 12 in black circles) following actinomycin D treatment at indicated timepoints. The dashed line indicates the half-life of exon 12-included isoform of approximately 1.5 h. (**H**) Relative expression of alternative HNRNPA1 3′UTR exons 11 or 12 in Nalm6 cells after treatment with control siRNA (black) or siRNAs targeting UPF1 and UPF2 (gray) for 48 h.

Since Dicer itself is involved in the RNAi pathway (52), we were concerned that this involvement might affect the results of our siRNA experiments. Thus, we knocked-down HNRNPA1 in a non-RNAi way by using the GapmerR technology (53). To ensure that the knock-down was well-controlled, we custom-designed a pool of four GapmeRs with non-targeting scrambled sequences and compared it to the GapmeR targeting the translation start codon of HNRNPA1. The efficiency of knock-down was validated by RT-qPCR (Figure 4G) and Western blotting (Figure 4H). We were then able to confirm the increased inclusion of DICER1 alternative 5′ UTR exon 4 by RT-qPCR following HNRNPA1 knock-down (Figure 4I).

Aberrant splicing of leukemia drivers in pediatric B-ALL

We next aimed to identify alternative splicing in other genes that contribute to leukemogenesis. We retrieved all genes (141) with known somatic mutations in hematologic malignancies from the COSMIC database (v.82) (20,21) (Supplementary Table S3) and searched for aberrant LSVs affecting these genes in B-ALL samples using the very stringent ΔPSI >50% cutoff for exon inclusion/skipping. We

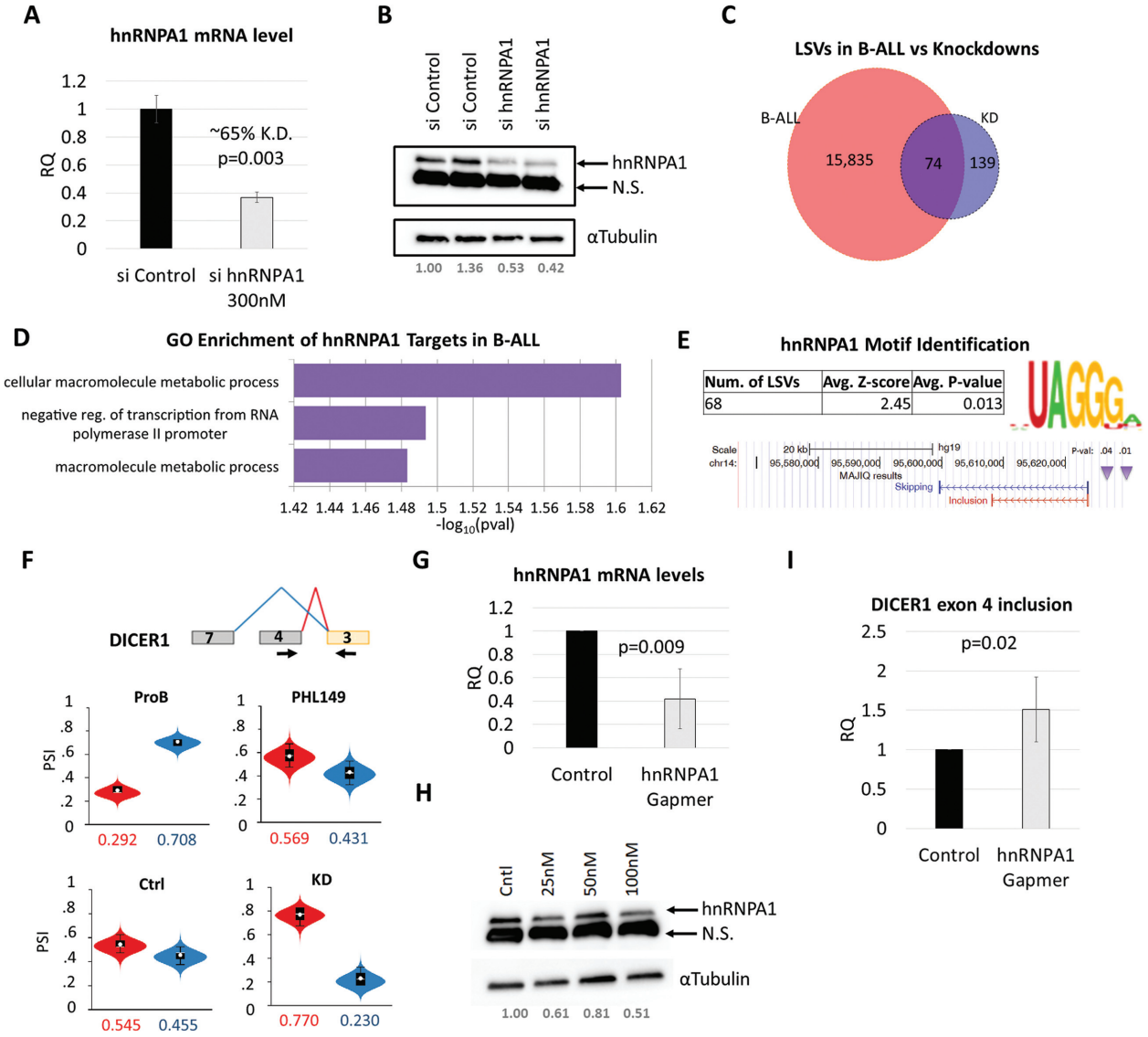

Figure 4. hnRNPA1-regulated splicing of DICER1. (**A**) Expression of HNRNPA1 RNA in P493-6 cells as determined by RT-qPCR (ex2–3–4 primers). A pool of four non-targeting control (black) and four HNRNPA1-targeting (grey) siRNAs were used in duplicate experiments. (**B**) hnRNPA1 protein expression determined by immunoblotting with anti-hnRNPA1 and anti-tubulin antibodies in cells from panel A. (**C**) Overlap in LSVs detected in B-ALL versus pro-B and HNRNPA1 knockdown versus control cells. (**D**) The horizontal bar chart showing enrichment for macromolecular metabolic pathways among genes with LSVs detected in both HNRNPA1 knockdown and B-ALL. (**E**) hnRNPA1 motif analysis of exons and adjacent intronic sequences (±200nt) surrounding LSVs detected in HNRNPA1 knockdown and B-ALL samples (top). Significant hnRNPA1 motifs found ∼200 nts upstream of DICER1 exon 1 (bottom). (**F**) Violin plots depicting alternative splicing of DICER1 in Pro-B vs. PHL149 B-ALL (top) and HNRNPA1 control and knockdown P493-6 cells (bottom). (**G**) Expression of HNRNPA1 RNA in P493-6 cells as determined by RT-qPCR (ex2-3-4 primers). A pool of four non-targeting control (black) GapmeRs and HNRNPA1-ATG-targeting GapmeR (grey) were used in triplicate experiments. (**H**) hnRNPA1 protein expression determined by immunoblotting with anti-hnRNPA1 and anti-tubulin antibodies in cells from panel G. (**I**) DICER1 alternative exon 4 inclusion measured by RT-qPCR following transfection with control or HNRNPA1-targeting GapmeRs.

identified 81 aberrant LSVs in 41 genes that were present in at least two B-ALL samples. They accurately separated leukemia samples from the non-transformed pro-B cell counterparts following hierarchical clustering using Euclidian distances (Figure 5A). Of note, these LSVs affected roughly a third of B-ALL driver genes. Some (in genes such as FBXW7) were present in almost all B-ALL samples, attesting to their potential significance in leukemia pathogenesis. Again, to decouple changes in splicing from changes in expression we performed differential expression analysis on genes containing LSVs in the COSMIC genes. Consistent with SF data analysis, the majority of genes in the majority of samples did not have changes in expression associated with changes in splicing (Supplementary Figure S3A).

To confirm our findings in independent datasets, we searched for aberrant LSVs in the COSMIC genes in the TARGET (39) and SJCRH (40) B-ALL datasets using pediatric pro-B cells for comparison. We found 229 LSVs in 80 genes in the TARGET data and 362 LSVs in 95 genes in the SJCRH data (Figure 5B). Importantly, 64 of

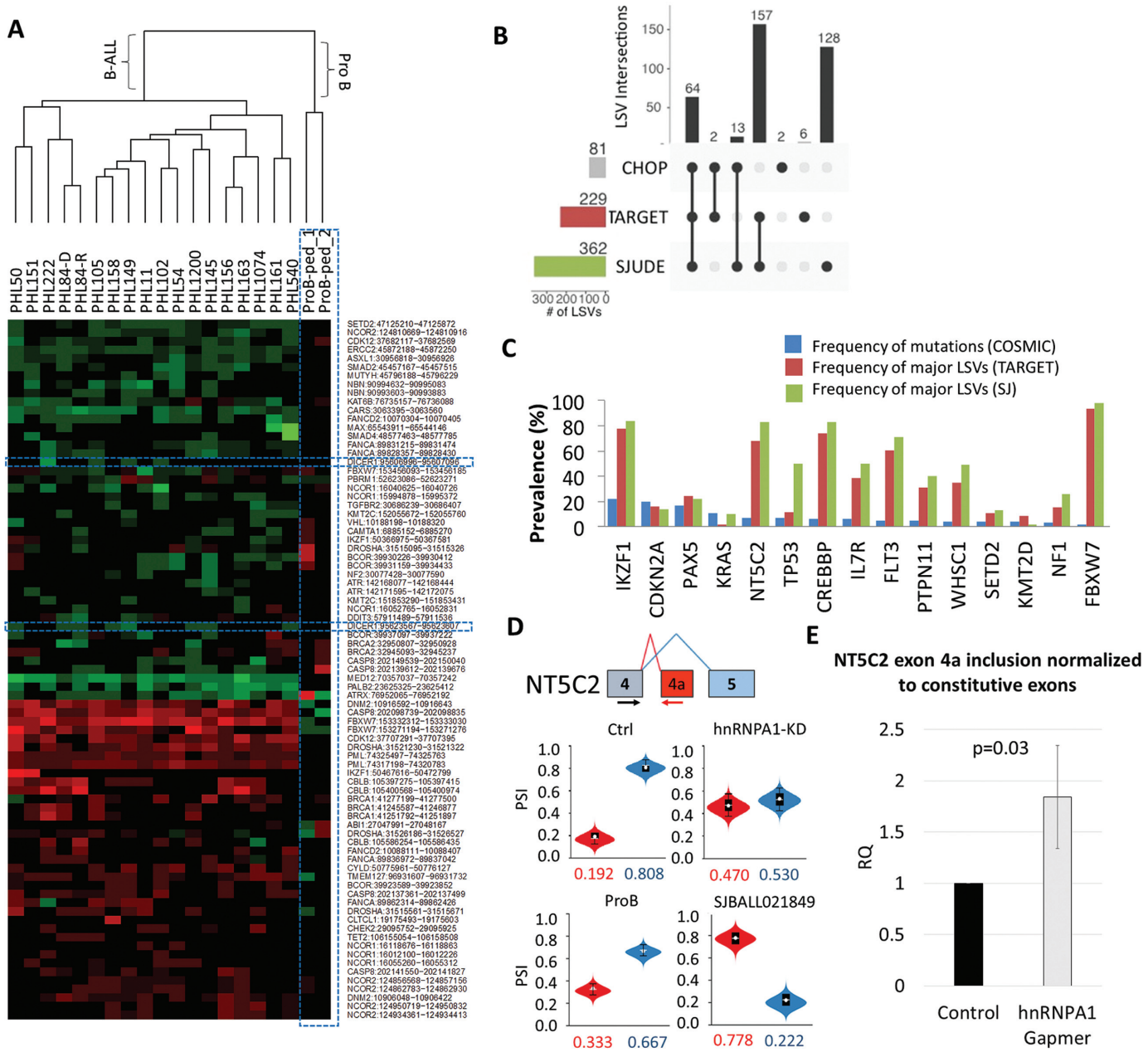

Figure 5. Alternative splicing of leukemia drivers in pediatric B-ALL. (**A**) Heat maps showing changes in splicing in B-ALL of genes commonly mutated in hematologic malignancies. The dendrogram on top of the heatmap represents results of hierarchical clustering. The dotted blue line denotes LSVs in the DICER1 transcript and also demarcates the pro-B samples. (**B**) Overlap between the LSVs from panel (A) and those detected in TARGET and SJCRH B-ALL datasets. (**C**) Bar graph representing frequencies of mutations (COSMIC dataset) and major LSVs as defined by MAJIQ in the indicated datasets. (**D**) Violin plots depicting alternative splicing of NT5C2 exon 4a (red) in control siRNA- vs. HNRNPA1 siRNA-electroporated P493-6 cells (top). The same analysis was applied to pro-B cells vs. the representative B-ALL sample SJBALL021849. (**E**) NT5C2 alternative exon 4a inclusion as measured by RT-qPCR following transfection with control (black) or HNRNPA1-targeting (grey) GapmeRs.

81 LSVs identified in the CHOP dataset significantly overlapped with TARGET (Fisher's Exact Test, $P = 7e{-}124$) and SJCRH datasets (Fisher's Exact Test, $P = 1e{-}139$), 13 LSVs were present only in CHOP and SJCRH datasets and another two LSVs were present only in CHOP and TARGET datasets (Figure 5B). We then narrowed down our analysis to LSVs in top 20 B-ALL tumor suppressors and oncogenes as defined in the COSMIC database. In both TARGET and SJCRH datasets, we discovered largely overlapping LSVs corresponding to 15 out of the top 20 mutated genes in B-ALL. Of note, LSV frequencies were much greater than frequencies of somatic mutations (Figure 5C), suggesting that in B-ALL these driver genes are preferentially affected by post-transcriptional mechanisms, such as alternative splicing. For example, we detected increased inclusion of NT5C2 alternative exon 4a in ∼65% of SJCRH samples analyzed (Figure 5D, top, red trace and violin plot). Apparently, this event is also under the regulation of HN-

RNPA1, as evidenced by increased inclusion of exon 4a in both HNRNPA1 knock-down experiments (siRNA in Figure 5D, GapmeR in Figure 5E). Additionally, we detected increased inclusion of TP53 exon 9i in B-ALL compared to pro-B (Supplementary Figure S3B), which would promote the expression of the SRSF3-regulated p53β isoform (Supplementary Figure S3C) involved in cell cycle and cell death (54–57).

DISCUSSION

Personalized cancer diagnostics employ carefully selected next-generation sequencing panels, which can identify mutations in specific genes known or suspected to be drivers in human malignancies. Hematologic malignancy panels typically include dominant oncogenes (e.g. FLT3 and IL7R), recessive tumor suppressors (e.g. TP53 and FBXW7), and haploinsufficient DNA/RNA caretakers (e.g. splicing factor SRSF2). Ostensibly, genomic profiling of diagnostic cancer specimens should be able to identify all prognostic mutations and inform treatment selection for patients. However, our data demonstrate that in B-ALL deregulation of cancer genes occurs primarily at the level of splicing, rather than via genetic mutations. For example, MAJIQ analyses of B-ALL transcriptomes revealed that several SRSF genes controlling exon inclusion (36), while unmutated, show widespread variations in their own splicing patterns, some of which are bound to decrease protein levels (37). Consequently, we observed aberrant splicing of some of their known target transcripts such as TP53 (57), which encodes the key tumor suppressor gene. Of note, the majority of aberrant LSVs were highly concordant across different datasets. This reproducibility validates our conclusions and alleviates the potential concern that sample preparation conditions (e.g. storage at ambient temperature) could be affecting RNA surveillance and thus impacting analysis of alternative splicing (58).

One of the most consistent changes in exon usage was the non-canonical selection of 3′ UTR exons of HNRNPA1, a splice factor implicated in cancer progression (59,60). While selection of alternative poly(A) sites is emerging as an important mechanism of gene expression in leukemia (61), that mechanism is distinct from alternative splicing involved in HNRNPA1 regulation. Indeed, as transcription and splicing occur on similar timescales (62), with splicing estimated to be half-complete when Pol II is less than 50nt downstream of introns (63), the decision to utilize 3′ splice site in exon 11 or 12 of HNRNPA1 has to be made prior to RNA polymerase II approaching either of the alternative poly(A) sites. Regardless of the underlying mechanism, while HNRNPA1 is overexpressed in some cancers, in the B-ALL model HNRNPA1 LSV correlated with a decrease in HNRNPA1 mRNA abundance. This is in agreement with our data showing that this multi-exon 3′UTR triggers nonsense mediated decay of the transcript. It is also possible that new 3′UTR sequences could create additional sites for targeting by microRNAs, which are known to play a key role in HNRNPA1 downregulation in chemotherapy-resistant ovarian cancer cells (64). Interestingly, knockdown of HNRNPA1 in an Epstein-Barr Virus (EBV)-transformed human B-cell lymphoblastoid cell line resulted in aberrant splicing of Dicer, a key enzyme in microRNA biogenesis [reviewed in (52)]. Moreover, four non-hnRNPA1-dependent LSVs present in B-ALL samples (Figure 5A) affected Drosha, another key enzyme in the microRNA pathway (52), suggesting a strong link between splicing and miRNA machineries.

Of even greater importance is the fact that B-ALL-specific LSVs affect 15 out of 20 top leukemia driver genes (including the aforementioned NT5C2 and *TP53*), with frequencies far exceeding those of somatic mutations. This discovery could explain why the prevalence of somatic mutations and copy number variations in B-ALL is low compared to other human cancers. It remains to be determined whether splicing alterations in oncogenes and tumor suppressor genes are functionally equivalent to gain-of-function and loss-of-function mutations, respectively. If this proves to be the case, interfering with splicing using RNA-based therapeutics and/or available small molecule inhibitors could be used to inhibit oncogenes such as *FLT3* and to activate dormant tumor suppressor gene such as *TP53*. Such strategies could yield tangible therapeutic benefits across a broad spectrum of childhood B-ALL subtypes.

DATA AVAILABILITY

The original RNA-Seq data sets have been deposited in the NCBI GEO database under accession number GSE115656.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all members of our laboratories and all participants in the RNA Society-sponsored RNP Discussion Group at the University of Pennsylvania for many helpful discussions.

FUNDING

National Institutes of Health (NIH) [T32 HL007439 to K.L.B., T32 CA 115299 to A.S.N., S.Y.Y., E.G., K08 CA184418 to S.K.T., R01 AG046544 to Y.B. and U01 CA232563 to A.T.T., Y.B.]; William Lawrence and Blanche Hughes Foundation [2016 Research Grant to A.T.T.]; St. Baldrick's Foundation [RG 527717 to A.T.T.]; Alex's Lemonade Stand Foundation [2017 Innovation Award to A.T.T.]; CURE Childhood Cancer [Basic Research Award to K.L.B.]; St. Baldrick's-Stand Up To Cancer Pediatric Dream Team Translational Research Grant [SU2C-AACR-DT-27-17]. Stand Up to Cancer is a division of the Entertainment Industry Foundation. Its research grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. Funding for open access charge: St. Baldrick's-Stand Up To Cancer Pediatric Dream Team Translational Research Grant [SU2C-AACR-DT-27-17].

Conflict of interest statement. None declared.

REFERENCES

1. Roberts,K.G. and Mullighan,C.G. (2015) Genomics in acute lymphoblastic leukaemia: insights and treatment implications. *Nat. Rev. Clin. Oncol.*, **12**, 344–357.
2. Scheuermann,R.H. and Racila,E. (1995) CD19 antigen in leukemia and lymphoma diagnosis and immunotherapy. *Leuk. Lymphoma*, **18**, 385–397.
3. Sikaria,S., Aldoss,I. and Akhtari,M. (2016) Monoclonal antibodies and immune therapies for adult precursor B-acute lymphoblastic leukemia. *Immunol. Lett.*, **172**, 113–123.
4. Topp,M.S., Gokbuget,N., Stein,A.S., Zugmaier,G., O'Brien,S., Bargou,R.C., Dombret,H., Fielding,A.K., Heffner,L., Larson,R.A. *et al.* (2015) Safety and activity of blinatumomab for adult patients with relapsed or refractory B-precursor acute lymphoblastic leukaemia: a multicentre, single-arm, phase 2 study. *Lancet Oncol.*, **16**, 57–66.
5. Maude,S.L., Frey,N., Shaw,P.A., Aplenc,R., Barrett,D.M., Bunin,N.J., Chew,A., Gonzalez,V.E., Zheng,Z., Lacey,S.F. *et al.* (2014) Chimeric antigen receptor T cells for sustained remissions in leukemia. *N. Engl. J. Med.*, **371**, 1507–1517.
6. Gardner,R., Wu,D., Cherian,S., Fang,M., Hanafi,L.A., Finney,O., Smithers,H., Jensen,M.C., Riddell,S.R., Maloney,D.G. *et al.* (2016) Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy. *Blood*, **127**, 2406–2410.
7. Jacoby,E., Nguyen,S.M., Fountaine,T.J., Welp,K., Gryder,B., Qin,H., Yang,Y., Chien,C.D., Seif,A.E., Lei,H. *et al.* (2016) CD19 CAR immune pressure induces B-precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic plasticity. *Nat. Commun.*, **7**, 12320.
8. Maino,E., Bonifacio,M., Scattolin,A.M. and Bassan,R. (2016) Immunotherapy approaches to treat adult acute lymphoblastic leukemia. *Expert. Rev. Hematol.*, **9**, 563–577.
9. Haso,W., Lee,D.W., Shah,N.N., Stetler-Stevenson,M., Yuan,C.M., Pastan,I.H., Dimitrov,D.S., Morgan,R.A., FitzGerald,D.J., Barrett,D.M. *et al.* (2013) Anti-CD22-chimeric antigen receptors targeting B-cell precursor acute lymphoblastic leukemia. *Blood*, **121**, 1165–1174.
10. Raetz,E.A., Cairo,M.S., Borowitz,M.J., Blaney,S.M., Krailo,M.D., Leil,T.A., Reid,J.M., Goldenberg,D.M., Wegener,W.A., Carroll,W.L. *et al.* (2008) Chemoimmunotherapy reinduction with epratuzumab in children with acute lymphoblastic leukemia in marrow relapse: a Children's Oncology Group Pilot Study. *J. Clin. Oncol.*, **26**, 3756–3762.
11. Fry,T.J., Shah,N.N., Orentas,R.J., Stetler-Stevenson,M., Yuan,C.M., Ramakrishna,S., Wolters,P., Martin,S., Delbrook,C., Yates,B. *et al.* (2018) CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. *Nat. Med.*, **24**, 20–28.
12. Sotillo,E., Barrett,D.M., Black,K.L., Bagashev,A., Oldridge,D., Wu,G., Sussman,R., Lanauze,C., Ruella,M., Gazzara,M.R. *et al.* (2015) Convergence of acquired mutations and alternative splicing of CD19 enables resistance to CART-19 immunotherapy. *Cancer Discov.*, **5**, 1282–1295.
13. Bagashev,A., Sotillo,E., Tang,C.A., Black,K.L., Perazzelli,J., Seeholzer,S.H., Argon,Y., Barrett,D.M., Grupp,S.A., Hu,C.A. *et al.* (2018) CD19 alterations emerging after CD19-directed immunotherapy cause retention of the misfolded protein in the endoplasmic reticulum. *Mol. Cell. Biol.*, **38**, e00383-18.
14. Behjati,S. (2015) Hiding from the enemy. *Sci.Transl. Med.*, **7**, 313ec193
15. Yoshida,K., Sanada,M., Shiraishi,Y., Nowak,D., Nagata,Y., Yamamoto,R., Sato,Y., Sato-Otsubo,A., Kon,A., Nagasaki,M. *et al.* (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, **478**, 64–69.
16. Papaemmanuil,E., Cazzola,M., Boultwood,J., Malcovati,L., Vyas,P., Bowen,D., Pellagatti,A., Wainscoat,J.S., Hellstrom-Lindberg,E., Gambacorti-Passerini,C. *et al.* (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.*, **365**, 1384–1395.
17. Graubert,T.A., Shen,D., Ding,L., Okeyo-Owuor,T., Lunn,C.L., Shao,J., Krysiak,K., Harris,C.C., Koboldt,D.C., Larson,D.E. *et al.* (2012) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.*, **44**, U53–U77.
18. Quesada,V., Conde,L., Villamor,N., Ordonez,G.R., Jares,P., Bassaganyas,L., Ramsay,A.J., Bea,S., Pinyol,M., Martinez-Trillos,A. *et al.* (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.*, **44**, 47–52.
19. Wang,L.L., Lawrence,M.S., Wan,Y.Z., Stojanov,P., Sougnez,C., Stevenson,K., Werner,L., Sivachenko,A., DeLuca,D.S., Zhang,L. *et al.* (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **365**, 2497–2506.
20. Forbes,S.A., Beare,D., Gunasekaran,P., Leung,K., Bindal,N., Boutselakis,H., Ding,M., Bamford,S., Cole,C., Ward,S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
21. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
22. Tasian,S.K., Doral,M.Y., Borowitz,M.J., Wood,B.L., Chen,I.M., Harvey,R.C., Gastier-Foster,J.M., Willman,C.L., Hunger,S.P., Mullighan,C.G. *et al.* (2012) Aberrant STAT5 and PI3K/mTOR pathway signaling occurs in human CRLF2-rearranged B-precursor acute lymphoblastic leukemia. *Blood*, **120**, 833–842.
23. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
24. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
25. Vaquero-Garcia,J., Barrera,A., Gazzara,M.R., Gonzalez-Vallinas,J., Lahens,N.F., Hogenesch,J.B., Lynch,K.W. and Barash,Y. (2016) A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife*, **5**, e11752.
26. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Giron,C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
27. Li,Y.I., Knowles,D.A., Humphrey,J., Barbeira,A.N., Dickinson,S.P., Im,H.K. and Pritchard,J.K. (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.*, **50**, 151–158.
28. Paz,I., Kosti,I., Ares,M. Jr, Cline,M. and Mandel-Gutfreund,Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*, **42**, W361–W367.
29. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
30. Pajic,A., Staege,M.S., Dudziak,D., Schuhmacher,M., Spitkovsky,D., Eissner,G., Brielmeier,M., Polack,A. and Bornkamm,G.W. (2001) Antagonistic effects of c-myc and Epstein-Barr virus latent genes on the phenotype of human B cells. *Int. J. Cancer*, **93**, 810–816.
31. Mullighan,C.G., Goorha,S., Radtke,I., Miller,C.B., Coustan-Smith,E., Dalton,J.D., Girtman,K., Mathew,S., Ma,J., Pounds,S.B. *et al.* (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, **446**, 758–764.
32. Norton,S., Vaquero-Garcia,J., Lahens,N.F., Grant,G.R. and Barash,Y. (2018) Outlier detection for improved differential splicing quantification from RNA-Seq experiments with replicates. *Bioinformatics*, **34**, 1488–1497.
33. Dennis,G. Jr, Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, 3.
34. Blake,J.A., Dolan,M., Drabkin,H., Hill,D.P., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T., McCarthy,F. *et al.* (2013) Gene ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
35. Sebestyen,E., Singh,B., Minana,B., Pages,A., Mateo,F., Pujana,M.A., Valcarcel,J. and Eyras,E. (2016) Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.*, **26**, 732–744.
36. Busch,A. and Hertel,K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA*, **3**, 1–12.

37. Jumaa,H. and Nielsen,P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.*, **16**, 5077–5085.
38. Casero,D., Sandoval,S., Seet,C.S., Scholes,J., Zhu,Y., Ha,V.L., Luong,A., Parekh,C. and Crooks,G.M. (2015) Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat. Immunol.*, **16**, 1282–1291.
39. Hunger,S.P., Loh,M.L., Whitlock,J.A., Winick,N.J., Carroll,W.L., Devidas,M., Raetz,E.A. and Committee,C.O.G.A.L.L. (2013) Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia. *Pediatr. Blood Cancer*, **60**, 957–963.
40. Gu,Z., Churchman,M., Roberts,K., Li,Y., Liu,Y., Harvey,R.C., McCastlain,K., Reshmi,S.C., Payne-Turner,D., Iacobucci,I. *et al.* (2016) Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat. Commun.*, **7**, 13331.
41. Amrani,N., Ganesan,R., Kervestin,S., Mangus,D.A., Ghosh,S. and Jacobson,A. (2004) A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature*, **432**, 112–118.
42. Maquat,L.E. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.*, **5**, 89–99.
43. Carter,M.S., Li,S. and Wilkinson,M.F. (1996) A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.*, **15**, 5965–5975.
44. Huang,L., Lou,C.H., Chan,W., Shum,E.Y., Shao,A., Stone,E., Karam,R., Song,H.W. and Wilkinson,M.F. (2011) RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell*, **43**, 950–961.
45. Anko,M.L., Muller-McNicoll,M., Brandl,H., Curk,T., Gorup,C., Henry,I., Ule,J. and Neugebauer,K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.
46. Gupta,P. and Li,Y.R. (2018) Upf proteins: highly conserved factors involved in nonsense mRNA mediated decay. *Mol. Biol. Rep.*, **45**, 39–55.
47. Hug,N., Longman,D. and Caceres,J.F. (2016) Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.*, **44**, 1483–1495.
48. Parsons,B.D., Schindler,A., Evans,D.H. and Foley,E. (2009) A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *PLoS One*, **4**, e8471.
49. Weighardt,F., Biamonti,G. and Riva,S. (1996) The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *Bioessays*, **18**, 747–756.
50. Burd,C.G. and Dreyfuss,G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.*, **13**, 1197–1204.
51. Foulkes,W.D., Priest,J.R. and Duchaine,T.F. (2014) DICER1: mutations, microRNAs and mechanisms. *Nat. Rev. Cancer*, **14**, 662–672.
52. Sotillo,E. and Thomas-Tikhonenko,A. (2011) Shielding the messenger (RNA): microRNA-based anticancer therapies. *Pharmacol. Ther.*, **131**, 18–32.
53. Kurreck,J., Wyszko,E., Gillen,C. and Erdmann,V.A. (2002) Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.*, **30**, 1911–1918.
54. Fujita,K., Mondal,A.M., Horikawa,I., Nguyen,G.H., Kumamoto,K., Sohn,J.J., Bowman,E.D., Mathe,E.A., Schetter,A.J., Pine,S.R. *et al.* (2009) p53 isoforms D133p53 and p53b are endogenous regulators of replicative cellular senescence. *Nat. Cell Biol.*, **11**, 1135–1142.
55. Bourdon,J.C., Fernandes,K., Murray-Zmijewski,F., Liu,G., Diot,A., Xirodimas,D.P., Saville,M.K. and Lane,D.P. (2005) p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.*, **19**, 2122–2137.
56. Marcel,V., Fernandes,K., Terrier,O., Lane,D.P. and Bourdon,J.C. (2014) Modulation of p53b and p53g expression by regulating the alternative splicing of *TP53* gene modifies cellular response. *Cell Death Differ.*, **21**, 1377–1387.
57. Tang,Y., Horikawa,I., Ajiro,M., Robles,A.I., Fujita,K., Mondal,A.M., Stauffer,J.K., Zheng,Z.M. and Harris,C.C. (2013) Downregulation of splicing factor SRSF3 induces p53beta, an alternatively spliced isoform of p53 that promotes cellular senescence. *Oncogene*, **32**, 2792–2798.
58. Dvinge,H., Ries,R.E., Ilagan,J.O., Stirewalt,D.L., Meshinchi,S. and Bradley,R.K. (2014) Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 16802–16807.
59. David,C.J., Chen,M., Assanah,M., Canoll,P. and Manley,J.L. (2010) HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature*, **463**, 364–368.
60. Michlewski,G. and Caceres,J.F. (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.*, **17**, 1011–1018.
61. Lee,S.H., Singh,I., Tisdale,S., Abdel-Wahab,O., Leslie,C.S. and Mayr,C. (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature*, **561**, 127–131.
62. Herzel,L., Ottoz,D.S.M., Alpert,T. and Neugebauer,K.M. (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell. Biol.*, **18**, 637–650.
63. Oesterreich,F.C., Herzel,L., Straube,K., Hujer,K., Howard,J. and Neugebauer,K.M. (2016) Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell*, **165**, 372–381.
64. Rodriguez-Aguayo,C., Monroig,P.D.C., Redis,R.S., Bayraktar,E., Almeida,M.I., Ivan,C., Fuentes-Mattei,E., Rashed,M.H., Chavez-Reyes,A., Ozpolat,B. *et al.* (2017) Regulation of hnRNPA1 by microRNAs controls the miR-18a-K-RAS axis in chemotherapy-resistant ovarian cancer. *Cell Discov.*, **3**, 17029.