

RESEARCH ARTICLE

# Transcriptome-wide analysis of alternative RNA splicing events in Epstein-Barr virus-associated gastric carcinomas

Victoria E. S. Armero<sup>1</sup>, Marie-Pier Tremblay<sup>1</sup>, Andréa Allaire<sup>1</sup>, Simon Boudreault<sup>1</sup>, Camille Martenon-Brodeur<sup>1</sup>, Cyntia Duval<sup>1</sup>, Mathieu Durand<sup>2</sup>, Elvy Lapointe<sup>2</sup>, Philippe Thibault<sup>2</sup>, Maude Tremblay-Létourneau<sup>1</sup>, Jean-Pierre Perreault<sup>1</sup>, Michelle S. Scott<sup>1</sup>, Martin Bisailon<sup>1\*</sup>

**1** Département de biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Quebec, Canada, **2** Plateforme RNomique, Université de Sherbrooke, Sherbrooke, Quebec, Canada

\* [Martin.Bisailon@USherbrooke.ca](mailto:Martin.Bisailon@USherbrooke.ca)



**OPEN ACCESS**

**Citation:** Armero VES, Tremblay M-P, Allaire A, Boudreault S, Martenon-Brodeur C, Duval C, et al. (2017) Transcriptome-wide analysis of alternative RNA splicing events in Epstein-Barr virus-associated gastric carcinomas. *PLoS ONE* 12(5): e0176880. <https://doi.org/10.1371/journal.pone.0176880>

**Editor:** Massimo Caputi, Florida Atlantic University, UNITED STATES

**Received:** January 24, 2017

**Accepted:** April 18, 2017

**Published:** May 11, 2017

**Copyright:** © 2017 Armero et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Primary and processed data are deposited in the National Cancer Institute GDC Data Portal (<https://portal.gdc.cancer.gov/projects/TCGA-STAD>).

**Funding:** This work was supported by grants from the Centre de recherche du Centre hospitalier universitaire de Sherbrooke (M.B. and J.P.P.) and the Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke (M.B., M.S.S. and J.P.P.). J.P.P. holds the Université de Sherbrooke

## Abstract

Multiple human diseases including cancer have been associated with a dysregulation in RNA splicing patterns. In the current study, modifications to the global RNA splicing landscape of cellular genes were investigated in the context of Epstein-Barr virus-associated gastric cancer. Global alterations to the RNA splicing landscape of cellular genes was examined in a large-scale screen from 295 primary gastric adenocarcinomas using high-throughput RNA sequencing data. RT-PCR analysis, mass spectrometry, and co-immunoprecipitation studies were also used to experimentally validate and investigate the differential alternative splicing (AS) events that were observed through RNA-seq studies. Our study identifies alterations in the AS patterns of approximately 900 genes such as tumor suppressor genes, transcription factors, splicing factors, and kinases. These findings allowed the identification of unique gene signatures for which AS is misregulated in both Epstein-Barr virus-associated gastric cancer and EBV-negative gastric cancer. Moreover, we show that the expression of Epstein-Barr nuclear antigen 1 (EBNA1) leads to modifications in the AS profile of cellular genes and that the EBNA1 protein interacts with cellular splicing factors. These findings provide insights into the molecular differences between various types of gastric cancer and suggest a role for the EBNA1 protein in the dysregulation of cellular AS.

## Introduction

Gastric carcinoma (GC) is the second leading cause of cancer-related deaths worldwide [1]. The clinical outcome for patients with GC remains poor, with a 5-year survival rate of only about 20% [2]. Moreover, most patients with GC are diagnosed with advanced stage disease, resulting in a dismal prognosis and highlighting the importance of the identification of diagnostic and prognostic markers [3]. Epstein-Barr virus (EBV) infection is associated with 10% of all GC cases reported worldwide [4]. The molecular characterization of EBV-associated

Research Chair in Structure and Genomic RNA. M. B. is a Chercheur-boursier Sénior from the Fonds de recherche du Québec-Santé (FRQ-S).

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AS, Alternative splicing; ASE, Alternative splicing event; EBV, Epstein-Barr virus; EBVaGC, EBV-associated gastric carcinomas; EBNA1, Epstein-Barr nuclear antigen 1; GC, Gastric carcinoma; NNoV, Normal tissues, no virus; PSI, Percent-spliced-in; TCGA, The Cancer Genome Atlas; TNoV, Virus-free gastric adenocarcinoma tissues; TEBV, EBV-positive gastric adenocarcinoma tissues; TNoV, EBV-negative gastric adenocarcinoma tissues; TPM, Transcript-per-million.

gastric carcinomas (EBVaGC) has shown that EBVaGC exhibit specific clinicopathological features, novel genomic and epigenetic aberrations, and a distinct protein expression profile than that of conventional EBV-negative gastric adenocarcinomas [5]. Moreover, EBVaGC is associated with the expression of specific viral proteins (EBNA1, LMP2A, and secreted BARF1) which have been shown to play important roles in the development of GC [6]. The Cancer Genome Atlas (TCGA) project recently reported a broad molecular classification of GC which resulted in the identification of four subtypes: EBV-positive tumors, microsatellite instable tumors, genomically stable tumors, and tumors with chromosomal instability [7].

Alterations in RNA splicing of cellular genes have been observed in many diseases including GC, and the available data suggest that defects in splicing likely play a role in carcinogenesis [8]. Human cells use alternative splicing (AS) to modify the composition of pre-mRNA transcripts through selection of different exons to be included in mature mRNAs, thereby producing a variability at the proteomic level [9]. These various proteins generated from a single gene can frequently support different or even opposite biological effects. This is clearly demonstrated by the aberrant splicing of the RON gene which encodes for a tyrosine kinase receptor. Exclusion of the RON exon 11 results in a mRNA transcript which encodes a constitutively active receptor,  $\Delta$ Ron, harboring constitutive tyrosine kinase activity and promoting an invasive phenotype [10]. Therefore, it is not surprising that AS is tightly regulated and that variations in splicing patterns have been associated with various human diseases such as cancer [11]. AS alterations can provide selective advantages to tumors, such as angiogenesis, proliferation, cell invasion and avoidance of apoptosis [12]. Recent evidences indicate that some of these splicing alterations can be used as prognostic or diagnostic biomarkers, and the identification of molecules capable of correcting and/or inhibiting pathological splicing events is an imperative issue for future therapeutic approaches [13].

Various studies have shown alterations in the AS patterns of a limited number of specific cellular genes in GC. Examples of aberrantly-spliced genes detected in GC include TACC1 (Transforming, Acidic Coiled-Coil Containing Protein 1) [14], S100A4 [15], HTERT [16] (human telomerase reverse transcriptase), CD44 [17], and MET (or hepatocyte growth factor HGF) [18]. In addition, it has been suggested that differentially expressed isoforms could potentially be exploited as biomarkers for gastric cancer [19]. In the current study, high-throughput RNA sequencing data obtained from TCGA were used to investigate the alterations in the global cellular AS landscape of EBVaGC.

## Materials and methods

### Samples and RNA-seq data analysis

Detailed information on the GC samples can be obtained from the original manuscript describing the comprehensive evaluation of 295 primary gastric adenocarcinomas as part of TCGA project [7]. Essentially, each frozen primary tumour specimen had a companion normal tissue specimen [7]. Adjacent non-tumour gastric tissue was also submitted for a subset of cases [7]. Pathology quality control was performed on each tumour and adjacent normal tissue (if available) [7]. Hematoxylin and eosin (H&E) stained sections from each sample were subjected to pathology analysis to confirm that the tumour specimen was histologically consistent with gastric cancer and the adjacent tissue specimen contained no tumour cells [7].

RNA-Seq samples from TCGA were obtained through the CGHub data portal (<https://cghub.ucsc.edu/>). Since only BAM files were available, a custom script was used to generate valid FASTQ files. The sequence reads were then aligned on the transcriptome reference sequence database UCSCGene Hg19 using Bowtie v2 aligner (default parameters). The associated gene isoforms were quantified in transcript-per-million (TPM) using RSEM for each

sample [20,21]. RSEM utilizes an Expectation-Maximization (EM) algorithm as its statistical model which allows reads mapping to multiple transcripts to be included in the quantification. Alternative splicing events were automatically identified and further quantified using the percent-spliced-in (PSI,  $\Psi$ ) value based on long (L) and short (S) forms of all splicing events present using the equation below:

$$\Psi = \frac{L}{L + S}$$

For each splicing event in one given gene (cassette-exon, mutually exclusive exons, alternative 5' and 3' splice site, etc), a PSI value was computed based on the ratio of the long form on total form (short form and long form) present to determine the inclusion of exon, intron retention, differential splice-site choice, etc. For instance, the long form of a cassette-exon would be its inclusion, and short form would be its exclusion from the mature transcript.

### Statistical analysis

The list of alternative splicing event (ASE) was filtered to retain only data present in at least two replicates for both GC and normal stomach tissue. ASEs with a P-value of less than 0.05 were conserved. To ensure higher stringency, the ASEs were filtered with a cutoff Q-value of less than 0.05. From these ASEs, only those with a difference higher than 10% in PSI were considered biologically relevant. Welch's t-test (Student's t-test with unequal sample sizes and unequal variances) was evaluated through the GNU Scientific Library (GSL) (<http://www.gnu.org/software/gsl/>) integrated to Perl system analysis for both AS and gene expression data. Finally, false discovery rates (fdr) were determined with the Q-value package in R (<https://cran.r-project.org/src/contrib/Archive/qvalue/>) according to Storey and Tibshirani [22]. For all other analyses, Graph Pad Prism version 6.05 was used to run statistical analysis.

### Gene expression analysis

Initially, the list of selected genes was filtered to retain only data present in at least two replicates for both GC and normal gastric tissue. To allow higher reproducibility, only transcripts with expression levels higher than one transcript-per-million in either dataset were retained. Fold changes (in base 2 logarithm) were then determined between STAD samples and normal gastric tissues. Q-values were finally evaluated, and results under 0.05 were determined significant.

### Gene ontology analysis

Gene ontology analysis was performed with PANTHER using the database for annotation, visualization, and integrated discovery (DAVID) [23].

### Functional ASE prediction

The splicing patterns of selected genes were visualized Using the FAST-DB or EASANA suite. DNA sequences of representative transcripts presenting short and long isoforms were downloaded and translated into proteins using ExPASy translation tool [24] Predicted proteins were then compared using Multalin (truncation and frameshift event) [25], PFAM (loss or appearance of functional domain) [26], and NLS Mapper (loss or gain of nuclear localization signal) [27].

## PCR validation

Gastroesophageal cancer cDNA Arrays from Origene TissueScan plates (Rockville, MD) were assessed for the expression of the various transcripts using the manufacturer's protocol. The plates contained cDNAs from 6 normal and 42 gastroesophageal cancer tissues. All forward and reverse primers were individually resuspended in 20–100  $\mu$ M stock solution in Tris-EDTA buffer and diluted as a primer pair to 1  $\mu$ M in RNase DNase-free water. PCR reactions (10  $\mu$ l) were performed in 96 well plates on a CFX-96 thermocycler (BioRad). For each PCR run, control reactions were performed in the absence of template and carried out for each primer pair. These reactions were consistently negative. The PCR products were finally analyzed using automated chip-based microcapillary electrophoresis on Caliper LC-90 instruments (Caliper LifeSciences). Both amplicon sizing and quantitation were analyzed using the manufacturer's software.

## Knockdown of splicing factors and effect on alternative splicing

The siRNAs used to knockdown the expression of specific splicing factors and the methodology used to measure the effect on alternative splicing are described in [28]. Proteins and RNA were extracted from mock-transfected (lipofectamine 2000, Invitrogen) and small interfering RNA (siRNA)-transfected cells 96 h posttransfection. Control siRNAs included the siRNA All-Stars Negative Control (Qiagen) which has no homology to any known mammalian gene and the KITLG Trilencer-27 Human siRNA (Origene) which contains three unique 27-mer siRNA duplexes targeting the gene encoding the ligand of the tyrosine-kinase receptor.

## Survival analysis for GC patients

The PROGgeneV2 prognostic biomarker identification tool [29] was used to study the implications of splicing factors gene expression on overall survival of GC patients. The preprocessed dataset from TCGA (including RNA-Seq data and clinicopathological features) was used for analysis. The Cox proportional hazard model was used to calculate hazard ratio and p-value of each parameter, and the median gene expression values were used as bifurcation points. A p-value of less than 0.05 was considered significant.

## EBNA1 Expression and Western blot analysis

Stable HEK293T cell lines expressing the EBNA1 protein were generated. HEK293T cells were first transfected with the pMSCV vector (addgene) encoding for the viral EBNA1 protein. Cells with stable EBNA1 protein expression were selected by Puromycin selection (5  $\mu$ g/ml) for 20 days. To maintain the selection, cells were kept in 3  $\mu$ g/ml Puromycin-DMEM complete medium. Cells were grown, pelleted, resuspended and mixed with 50  $\mu$ l of SDS loading buffer and heated for 5 min at 95°C. Samples were resolved by electrophoresis on 10-% SDS-polyacrylamide gels and transferred onto a polyvinylidene difluoride (PVDF) membrane. The presence of the EBNA1 protein was detected using a mouse anti-FLAG antibody (Sigma). A mouse anti-actin antibody (Sigma) was also used as a control. Following incubation with primary antibodies, PVDF membranes were washed and incubated with the horseradish peroxidase-conjugated secondary antibody (Amersham Biosciences). Proteins were revealed using the enhanced chemiluminescence (ECL) kit from Amersham Biosciences.

## Immunofluorescence

HEK293T cells stably expressing the EBNA1-FLAG-HA protein were seeded in 24 well plates at a density of  $5 \times 10^4$  cells/well. The cells were grown for 24 h and washed twice with PBS and

fixed with 4% paraformaldehyde and 4% sucrose for 20 min at 22°C. Cells were permeabilized with 0.15% triton X-100 in PBS for 5 min at 22°C and blocked in 10% normal goat serum (Wisent). The EBNA1 protein was stained with primary mouse anti-HA antibody (Abcam) for 4 h at 22°C. Cells were washed and incubated in the dark for 1 h at 22°C with an Alexa Fluor 488-labelled anti-mouse secondary antibody (Abcam). Nuclei were stained with Hoechst (1 µg/ml) for 15 min at 22°C. Cover glasses were affixed on slides with SlowFade mounting medium (Invitrogen S36937). Epifluorescence microscopy was conducted using a Nikon Eclipse TE2000-E visible/epifluorescence inverted microscope using bandpass filters for Hoechst and Alexa Fluor 488.

## Nuclear extracts preparations and mass spectrometry

Hypotonic lysis of HEK293T cells followed by Dounce homogenization, differential centrifugation, and separation of cytoplasm and nuclei was performed as described before [30]. Nuclear extracts were washed three times with PBS and treated with 20 mg/ml RNase A (Invitrogen) and 200 units DNase RQ1 (Promega) for two hours. Immunoprecipitation of the EBNA1 protein was performed using anti-HA beads (Roche) following the manufacturer's instructions. LC-MS/MS was performed as described previously [31]. Non-specific binders were determined using the CRAPOME database (Contaminant Repository for Affinity Purification and Mass Spectrometry) [32].

## Co-immunoprecipitation

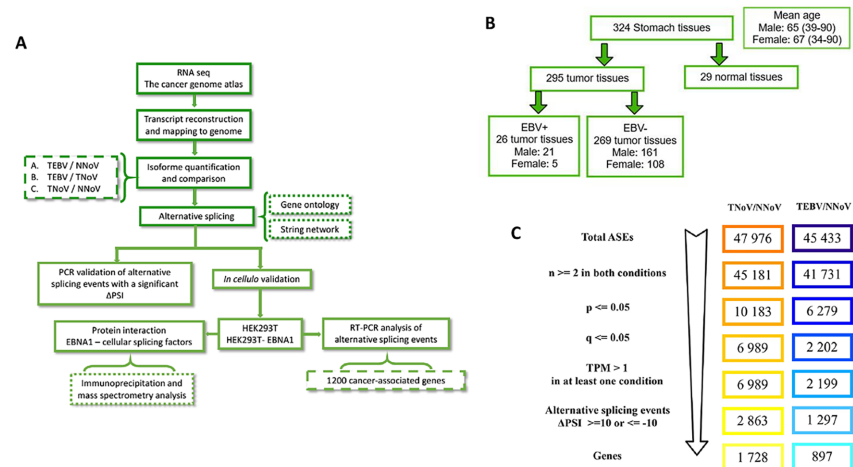
Nuclear extracts were incubated with anti-HA beads (Roche) and the presence of the EBNA1, hnRNP H1, and GAPDH proteins in nuclear extracts was detected using the following specific primary mouse antibodies: anti-FLAG antibody (Sigma), anti-GAPDH (Abcam), and anti-hnRNP H1 (Abcam). Following incubation with primary antibodies, membranes were washed and incubated with the horseradish peroxidase-conjugated secondary antibody (Amersham Biosciences). Proteins were revealed using an enhanced chemiluminescence (ECL) kit (Amersham Biosciences).

## Results

### Alteration of the cellular AS landscape in EBVaGC

Modifications to the global RNA splicing landscape of more than 295 primary gastric adenocarcinoma tissues from The Cancer Genome Atlas (TCGA), including data from 26 EBVaGC tissues, were examined using high-throughput RNA sequencing data. Fig 1A and 1B shows an overview of the analyses performed in the current study. RNA-seq data were mapped to the reference genome, followed by transcript assembly and analysis of the various RNA isoforms abundance. We initially focused our study on EBVaGC (EBVaGC, TEBV, Tumors with EBV), and we compared the splicing patterns with normal tissues (NNoV, Normal tissues, no virus). To identify the cellular AS patterns that are modified in EBVaGC, we evaluated the modifications in the relative splice abundances by quantifying all the primary RNA transcripts that produce two or more isoforms. We detected and quantified the various alternative splicing events (ASEs) using the percent-spliced-in (PSI) parameter based on isoform expression (TPM, transcripts per million) for the long and the short isoforms (see Experimental Procedures). The lists of ASEs (47,976 ASEs) were filtered to keep only the transcripts with at least two replicates for both TEBV and NNoV tissues (Fig 1C). Moreover, only significant ASEs (with a P-value of less than 0.05) that differed between tumor and normal gastric tissues were conserved. To provide higher stringency, the ASEs were additionally filtered with a cutoff Q-value (fdr; false-





**Fig 1. Identification of alternative splicing events in gastric cancer.** (A) Overview of the strategy used to identify the changes in alternative splicing in GC. (B) Classification of the TCGA RNA sequencing data for GC and healthy tissues. (C) The splicing events list for both EBV-negative gastric adenocarcinomas (TNoV, Tumors, no virus) and EBV-associated gastric carcinomas (EBVaGC, TEBV, Tumors with EBV) compared the global splicing patterns with normal tissues (NNoV, Normal tissues, no virus) was filtered to keep only significant ASEs.

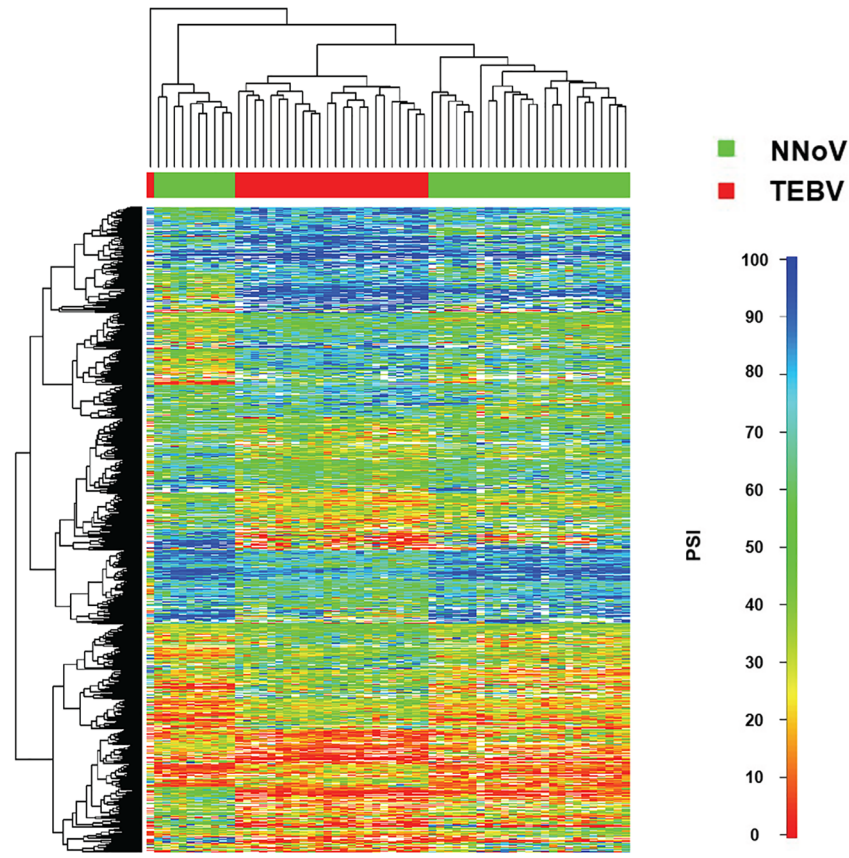
<https://doi.org/10.1371/journal.pone.0176880.g001>

discovery rate) of less than 0.05. From these selected ASEs, only those with a variation higher than 10% in  $|\Delta\text{PSI}|$  were conserved and considered biologically relevant. This strategy allowed the identification of 1297 primary transcripts belonging to 897 genes for which was significantly modified in patients with EBVaGC ( $Q \leq 0.05$ ,  $|\Delta\text{PSI}| \geq 10$ ) (Fig 1C). The comprehensive list of the differential ASEs associated with EBVaGC is provided in S1 Table.

Our analysis revealed that the AS of numerous cellular transcripts was modified in patients with EBVaGC as compared to normal healthy tissues (Fig 2). Hierarchical clustering clearly distinguished between normal and EBVaGC tissues (Fig 2). Moreover, additional analysis revealed the presence of two distinct clusters for EBVaGC tissues. However, multivariate clustering and statistical relationships among the EBVaGC samples could not detect any correlation between the clusters and clinical parameters (such as age, sex, histological grade, etc.). Further analysis revealed that 198 transcripts had  $|\Delta\text{PSI}|$  values higher than 30% (S1 Fig). To detect cellular pathways for the differential ASEs associated with EBVaGC, we performed a statistical analysis to analyze the overrepresentation of biological processes. The analysis revealed that certain processes such as vesicle-mediated transport, cellular component movement/organization, and metabolic process are statistically overrepresented in the differential ASEs associated with gastric adenocarcinoma (S2 Fig). Through manual curation of functional annotations, we also identified several modifications in the AS patterns of genes potentially involved in oncogenesis (Table 1). Indeed, EBVaGC tissues displayed alterations in the AS patterns of 77 tumor suppressors, 62 transcription factors, and 36 kinases. Some of the dysregulated ASEs that were identified include CD44, which encodes a cell-surface protein implicated in cell-cell interactions, migration, and cell adhesion and which has previously been involved in tumor metastasis and GC [17], and BRCA1 which encodes a nuclear protein involved in genomic stability which also functions as a tumor suppressor and is frequently mutated in breast and ovarian cancers.

## Characterization of the ASEs that are altered in EBVaGC

We next examined the functional consequences of altered AS on protein function. Among the transcripts for which AS was modified in EBVaGC, many ASEs that we identified affect



**Fig 2. Global profiling of alternative splicing event modifications in EBVaGC.** Hierarchical cluster and heatmap representation of alternative splicing events modified in EBVaGC tissues. The row at the top shows the clustering information in the form of a dendrogram and the similarity relationships among the alternative splicing events and patients. The column at the left of the heatmap shows the different clusters associated with the PSI values. The heatmap shows PSI values for EBVaGC tissues (TEBV, Tumors with EBV, *in red*) compared with normal tissues (NNoV, Normal tissues, no virus, *in green*).

<https://doi.org/10.1371/journal.pone.0176880.g002>

known protein domains. [S2 Table](#) shows the functional consequences of the differentially spliced transcripts on protein function for 48 transcripts that are differentially spliced in gastric adenocarcinoma. Among the differential ASEs, we noted the loss of predicted nuclear localization signals (NLS) for KDM3B (Lysine-specific demethylase 3B), a histone demethylase with potential tumor suppressor activity. Eleven other ASEs were predicted to disrupt known functionally critical protein domains.

### Comparison between EBV-positive and EBV-negative GC

The important number of GC samples analyzed by high-throughput RNA sequencing also allowed us to acquire data for the modifications to the global RNA splicing landscape of EBV-negative gastric adenocarcinomas (TNoV, Tumors, no virus). Our RNA-Seq analysis performed on 269 EBV-negative gastric adenocarcinomas tissues allowed us to identify 2863 primary transcripts belonging to 1728 genes for which the AS pattern was significantly modified in patients with EBV-negative gastric adenocarcinomas in comparison to normal tissues ([Fig 1C](#)). A complete list of the differential ASEs associated with EBV-negative gastric adenocarcinomas is provided in [S3 Table](#). Unsupervised clustering on data from EBV-negative gastric adenocarcinomas and integration of these results yielded five distinct clusters associated with

**Table 1. Protein families encoded by transcripts that are differentially spliced in GC.**

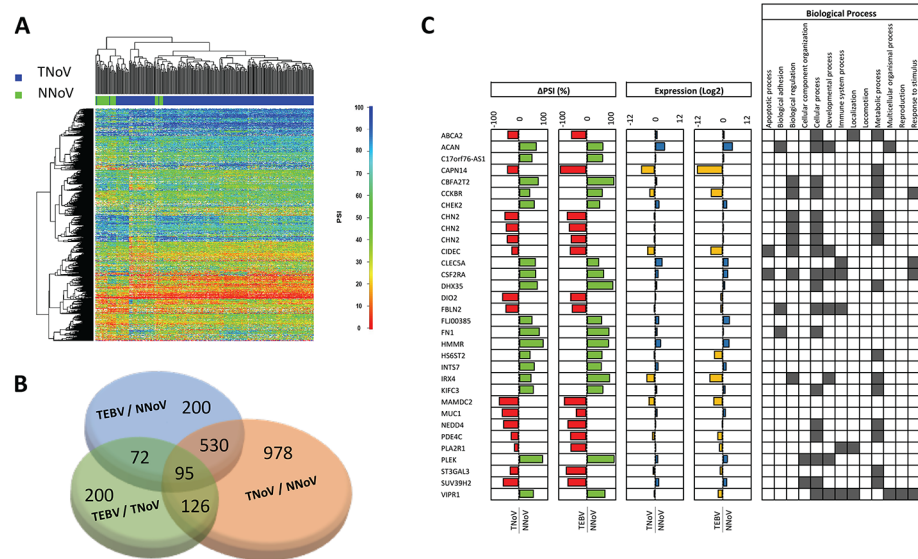
Protein classification	TNoV / NNoV		TEBV / NNoV	
	Genes	Frequency %	Genes	Frequency %
Extracellular matrix protein (PC00102)	35	2.0%	16	1.8%
Protease (PC00190)	63	3.6%	22	2.5%
Cytoskeletal protein (PC00085)	107	6.2%	70	7.8%
Transporter (PC00227)	115	6.7%	50	5.6%
Transmembrane receptor regulatory (PC00226)	7	0.4%	7	0.8%
Transferase (PC00220)	104	6.0%	63	7.0%
Oxidoreductase (PC00176)	51	2.9%	32	3.6%
Lyase (PC00144)	9	0.5%	8	0.9%
Cell adhesion molecule (PC00069)	55	3.2%	22	2.5%
Ligase (PC00142)	51	2.9%	28	3.1%
Nucleic acid binding (PC00171)	192	11.1%	96	10.7%
Signaling molecule (PC00207)	92	5.3%	45	5.0%
Enzyme modulator (PC00095)	140	8.1%	88	9.8%
Calcium-binding protein (PC00060)	36	2.1%	15	1.7%
Defense/immunity protein (PC00090)	59	3.4%	22	2.5%
Hydrolase (PC00121)	162	9.4%	84	9.4%
Transfer/carrier protein (PC00219)	28	1.6%	14	1.6%
Membrane traffic protein (PC00150)	50	2.9%	30	3.3%
Phosphatase (PC00181)	36	2.1%	19	2.1%
Transcription factor (PC00218)	144	8.3%	62	6.9%
Chaperone (PC00072)	14	0.8%	6	0.7%
Cell junction protein (PC00070)	15	0.9%	14	1.6%
Surfactant (PC00212)	5	0.3%	3	0.3%
Structural protein (PC00211)	5	0.3%	4	0.4%
Kinase (PC00137)	57	3.3%	36	4.0%
Storage protein (PC00210)	3	0.2%	2	0.2%
Receptor (PC00197)	142	8.2%	62	6.9%
Isomerase (PC00135)	9	0.5%	2	0.2%
Splicing factors	6	0.3%	12	1.3%
Tumor suppressors	140	8.1%	77	8.6%

<https://doi.org/10.1371/journal.pone.0176880.t001>

modifications in AS (Fig 3A). Since three distinct non-viral subtypes of GC had previously been described following the molecular classification of TCGA (microsatellite instable tumors, genomically stable tumors, and tumors with chromosomal instability), we investigated the possibility that the five clusters associated with EBV-negative gastric adenocarcinomas might be linked to specific molecular subtypes of GC. However, no correlation was found between the non-viral molecular subtypes of GC and the clusters associated with AS, indicating that similar AS changes were occurring in all molecular subtypes. Similarly, we could not detect any correlation between the clusters and clinical parameters (such as age, sex, histological grade, etc.).

The identification of changes in the AS profiles of both EBVaGC and EBV-negative GC tissues raises the possibility that some of the identified ASEs might be a general feature of cancerous tissues. We therefore compared the cellular AS patterns that are altered in EBVaGC (TEBV) with EBV-negative gastric adenocarcinoma (TNoV), and we evaluated the changes in relative splice abundances. Remarkably, 625 genes for which AS was modified in EBVaGC were found to be common with EBV-negative gastric adenocarcinoma tissues (TNoV), suggesting that these ASEs might be a general feature of both types of GC tissues analyzed in this





**Fig 3. Comparison between EBV-negative and EBV-associated gastric cancer.** (A) Heatmap representation of isoform ratios (PSI values) for EBV-negative gastric adenocarcinomas tissues. EBV-negative gastric adenocarcinomas tissues (TNoV) are shown in blue, and the comparative healthy tissues are shown in green (NNoV). (B) Comparison of the cellular genes with dysregulated ASEs between EBVaGC and EBV-negative GC tissues. (C) The list displays common differentially spliced transcripts for both EBV-negative and EBVaGC with the corresponding Delta PSI values, the associated gene expression (in Log<sub>2</sub>), and the related biological processes.

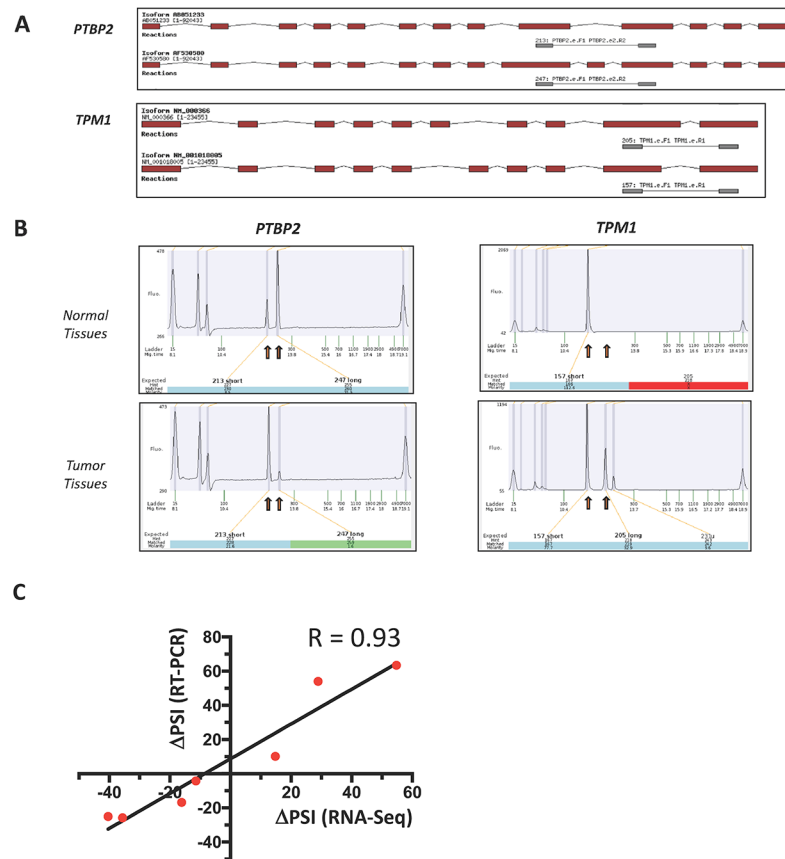
<https://doi.org/10.1371/journal.pone.0176880.g003>

study (Fig 3B, S4 Table). In the case of common misregulated ASEs, the direction of the splicing shift between normal and tumor tissues was frequently the same between the two types of GC (Fig 3C). Using such an approach, we also identified 978 unique genes for which AS is misregulated in EBV-negative gastric adenocarcinoma, while 200 are uniquely misregulated in EBVaGC (Fig 3B, S2 and S3 Tables).

To identify cellular pathways which could be common or different between EBVaGC with EBV-negative gastric adenocarcinoma, we again performed a statistical analysis to analyze the overrepresentation of biological processes. Analysis of the data revealed that certain processes such as vesicle-mediated transport, cellular component movement/organization, and metabolic process are overrepresented for the differential ASEs associated with both EBV-negative GC and EBVaGC (S2 Fig). However, some processes are clearly unique to the differential ASEs associated with EBV-negative GC such as immune system process and intracellular protein transport. The differential ASEs associated with EBV-negative GC also displayed an enrichment for genes involved in heart development although the significance of this finding is not clear.

### Experimental validation on gastric adenocarcinoma tissues

Analysis of high-throughput RNA sequencing data revealed numerous alterations in the global cellular AS landscape of primary gastric adenocarcinomas. To experimentally confirm the results that were acquired from the RNA-Seq data, we next used PCR analysis from cDNA arrays derived from both healthy and cancerous tissues. Specific primers were designed to allow detection of predicted ASEs by PCR. Two examples of differentially spliced transcripts are presented in Fig 4A illustrating the modifications in isoform usage in transcripts encoded by the *PTBP2* and *TPM1* genes. Our PCR results demonstrated that changes in AS levels that



**Fig 4. Validation of ASEs dysregulated in gastric cancer.** (A) Overview of the two isoforms encoded by *PTBP2* and *TPM1* genes. The exons are represented in red and the intervening introns are displayed as thin black lines (not to scale). The primers used to detect the isoforms by RT-PCR are presented in gray and the sizes of the expected amplicons are also presented. (B) The cDNAs acquired from gastric tissues were analyzed by PCR using specific primers to detect both isoforms of the transcripts encoded by the *PTBP2* and *TPM1* genes. Capillary electropherograms of the PCR reactions are presented. The positions and the amplitude of the detected amplicons are emphasized by red boxes. The positions of the internal markers are also indicated. (C) Correlation between PSI values obtained from RNA-Seq and RT-PCR data. The analysis was performed on seven selected ASEs (CAPN14, PLA2G4F, PTBP2, S100A1, SLC52A1, SOGA2, and TPM1). In all cases, the changes in AS levels detected by RT-PCR and the ones revealed through transcriptome sequencing displayed high levels of correlation ( $R = 0.93$ ).

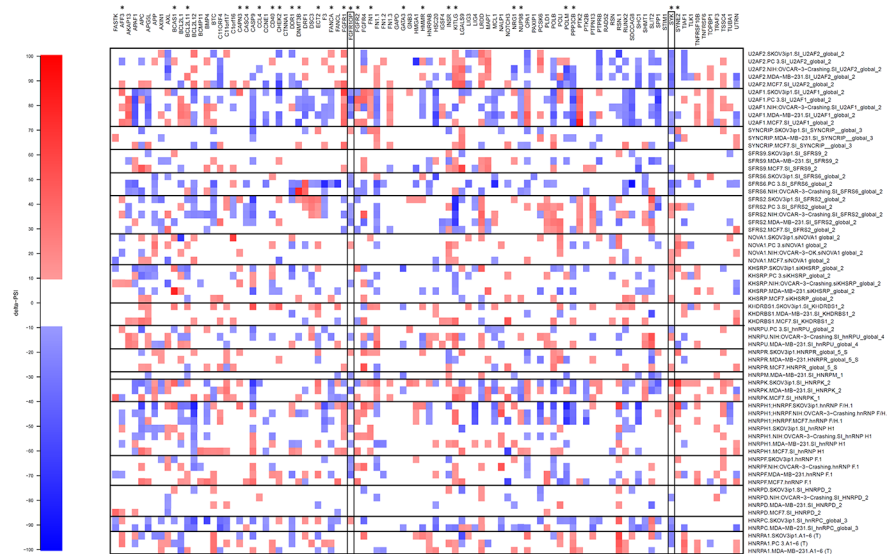
<https://doi.org/10.1371/journal.pone.0176880.g004>

were detected through transcriptome sequencing could also be detected by PCR analysis (Fig 4B). The validation analysis was performed on seven selected ASEs (CAPN14, PLA2G4F, PTBP2, S100A1, SLC52A1, SOGA2, and TPM1) and showed that changes in AS levels detected by RT-PCR were similar to the ones obtained through transcriptome sequencing, and demonstrated high levels of correlation ( $R = 0.93$ ) (Fig 4C, S3 Fig). It should be noted that Sanger sequencing was also performed on several PCR products to validate the predicted ASEs (data not shown).

### Expression levels and AS of RNA splicing factors

The molecular mechanisms which lead to a modification of the AS landscape in GC are currently unknown. However, aberrant splicing is frequently associated with dysregulated expression of individual splicing factors [33,34]. We therefore monitored both the expression levels and the changes in splicing patterns of RNA encoding for splicing factors and spliceosomal





**Fig 6. Modifications to AS of 96 transcripts in response to knockdown of specific splicing factors with siRNAs.** Using specific siRNAs, eighteen splicing factors (U2AF2, U2AF1, SYNCRIP, SFRS9, SFRS6, SFRS2, NOVA1, KHSRP, KHDRBS1, HNRPU, HNRPP, HNRPM, HNRPK, HNRPH1, HNRPF, HNRPD, HNRPC, and HNRPA1) were individually knocked-down in various cell lines to evaluate their implication in splicing of 96 different transcripts. Asterisks (top) indicate transcripts for which AS was altered in GC. Individual knockdowns and ASEs are presented to indicate which knockdowns produced a shift in AS in various cell lines (PC-3, SKOV3, NIH:OVCAR-3, MDA-MB-231, MCF7). Each individual column represents a different knockdown performed with specific siRNAs. The changes in PSI values are indicated. The map displays the changes in PSI values in a color-coded scale. White areas indicate no shifts.

<https://doi.org/10.1371/journal.pone.0176880.g006>

these transcripts (AFF3, BCAS1, BCL2L11, CAPN3, CASC4, CCL4, CD40, CHEK2, DDR1, DNMT3B, ECT2, FANCA, FGFR1, FGFR1OP, FGFR2, FN1, GATA3, HMMR, INSR, KITLG, NRG1, NUP98, POLM, PPP3CB, SYK, and SYNE2) were of interest since their AS was altered in GC. The results are depicted in Fig 6 and show that the impact of the individual knockdowns varies considerably. For instance, the knockdown of U2AF1 has a definitive impact on the AS pattern of APAF1, APAPG5L, CASP9, FGFR1, FN1.1, FN1.3, HSC20, MCL1, NUP98, OPA1, POLB, PPP3CB, PTK2, RUNX2, SDCCAG8, SPP1, and TOPBP1 transcripts in all cell lines tested. However, the same knockdown has no impact on the AS profile of numerous other transcripts. Overall, this assay could not specifically correlate the expression of splicing factors which were over- or under-expressed in GC with specific AS modifications observed in GC tissues. However, it demonstrated that the expression of individual splicing factors has an impact on AS profiles, but cannot explain all the changes observed in GC. It should be noted that the use of negative siRNA controls (such as the siRNA AllStars Negative Control and the KITLG Trilencer-27 siRNA) or transfection with lipofectamine alone had no impact on the AS profiles of the 96 transcripts belonging to a subset of apoptotic genes. Not a single modification to the AS profiles of the selected transcripts could be observed (data not shown).

### Expression pattern of DHX15 significantly correlates with patient survival

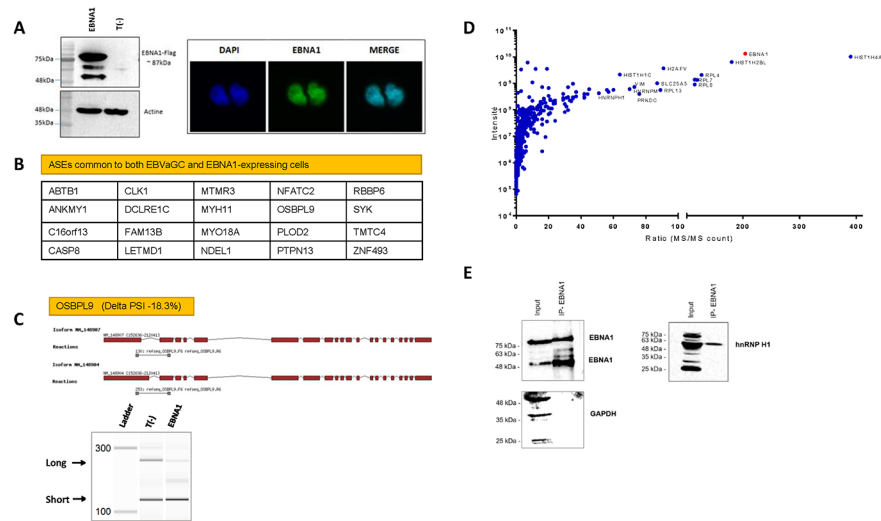
To explore the potential clinical significance of our findings, we next examined the potential link between the mRNA expression levels of protein factors implicated in splicing and patient survival. We used the recently published PROGgeneV2 prognostic biomarker identification tool [29] to study the implications of splicing factors gene expression on overall survival of GC

patients. Kaplan-Meier survival analysis was carried out to determine overall survival for patients as a function of the mRNA expression levels of proteins involved in splicing. Of all the proteins involved in splicing for which the expression level was modulated by more than 2-fold in both EBV-negative GC and EBVaGC, the expression level of DHX15, an ATP-dependent RNA helicase implicated in RNA splicing, showed correlation with patient survival (Fig 5C). A high level of expression of DHX15 was associated with poor patient survival. Using values of median gene expression as bifurcation points, our Cox proportional hazards analyses demonstrated that high expression of this gene was associated with reduced 3-year survival outcomes ( $p < 0.02$ ). Kaplan-Meier plot indicated significant segregation in survival outcomes for patients with high versus low DHX15 expression. We found no significant correlations between high levels of DHX15 expression and clinicopathological features, including age, gender, and tumor stage. The expression levels of other splicing factors did not appear to be correlated with overall survival.

### Expression of EBNA1 alters cellular AS

It is currently not clear whether changes in AS are the cause or the consequence of a malignant phenotype. Moreover, in the case of EBVaGC, it is not clear whether the observed changes in AS result from cellular and/or virus-mediated mechanisms. We therefore investigated the possibility that the presence of a viral protein could contribute to modifications in AS of EBVaGC. We focused on the Epstein-Barr nuclear antigen 1 (EBNA1) protein since EBNA1 is the only viral nuclear protein expressed in EBVaGC, and is also the only viral protein required to maintain viral latency, because of its various roles in the segregation and replication of the EBV episomes [36,37]. There is also increasing evidence that EBNA1 modifies the cellular environment in ways that promote genomic instability, thereby leading to tumorigenesis [38–40]. We relied on a high-throughput reverse transcription-PCR (RT-PCR)-based platform [41] to examine AS in 1200 cancer-associated genes. The stable expression of the EBNA1 protein (harboring a FLAG epitope) was detected by SDS-PAGE in HEK293T cells, a model epithelial cell line extensively used for EBV recombinant molecular genetics [42], followed with anti-FLAG immunoblotting (Fig 7A). HEK293T cells were also chosen because of their high transfection efficiency and their widespread use in RNA splicing studies [43–46]. Nuclear localization of the viral EBNA1 protein was also confirmed by immunofluorescence (Fig 7A). Our results demonstrated that the AS patterns of 112 cellular genes were modified upon EBNA1 expression. Of these, 20 cancer-related genes were also significantly modified in EBVaGC tissues (Fig 7B). Although the presence of these 20 common ASEs between EBNA1-expressing cells and GC tissues is not statistically significant ( $p$ -value of 0.29), we observed that the direction of the splicing shift between EBNA1-expressing cells and tumor tissues was the same between the two types of samples (i.e. the sign of PSI variation was identical for all the 20 cancer-related genes identified between the two sample groups). An example of a modified AS pattern which is modified upon EBNA1 expression is provided in Fig 7C. Of the 20 ASEs that are common between cells expressing EBNA1 and EBVaGC, we identified 11 ASEs which do not occur in EBV-negative GC tissues, thereby demonstrating their EBV specificity. The remaining nine other ASEs that were modified (in EBNA1 expressing cells, in EBVaGC tissues and in EBV-negative GC tissues) likely represent ASEs that can be modified in the presence of EBNA1 but can also be modified in GC tissues in the absence of the virus. It should also be noted that in the case of the 112 differential ASEs associated with EBNA1-expressing cells, we only found a significant enrichment for genes involved in response to stimulus (S2 Fig) which probably reflects the fact that cells expressing EBNA1 are clearly not a perfect reflection of EBVaGC tissues.





**Fig 7. Involvement of EBNA1 in alternative splicing.** (A) Immunoblotting analysis using anti-HA antibody for the detection of EBNA1-HA-FLAG protein in cell lysates from a stable HEK293T cell line expressing EBNA1. Control HEK293T cells (T(-)) were also used in this assay. (B) List of ASEs common to EBVaGC and EBNA1-expressing cells. (C) Example of a modified ASE following the expression of EBNA1. Overview of the two isoforms encoded by *OSBPL9* gene. Exons are represented in red and the intervening introns are displayed as thin black lines (not to scale). The primers used to detect the isoforms by RT-PCR assays are presented in gray and the sizes of the expected amplicons are also specified (top panel). RT-PCR reactions were performed on control cells (T(-)) and cells expressing EBNA1 using specific primers to detect both isoforms of the transcripts encoded by the *OSBPL9* gene. Capillary electrophoresis assays were performed and an image of the detected reaction products is presented (lower panel). The positions of the expected amplicons are shown by arrows. (D) Mass spectrometry analysis of nuclear proteins interacting with EBNA1. The average ratios (MS/MS counts) of the EBNA1 affinity purification-mass spectrometry experiments were plotted versus the total intensities. (E) Validation of the interaction between EBNA1 and splicing factor hnRNP H1. Nuclear extracts were immunoprecipitated with anti-HA. The extracts (input) and immunoprecipitates (IP-EBNA1) were analyzed by immunoblotting and probed with the indicated antibodies.

<https://doi.org/10.1371/journal.pone.0176880.g007>

## EBNA1 interacts with cellular splicing factors

The previous observation that the expression of EBNA1 leads to changes in the AS profile of cellular transcripts led us to examine the capacity of EBNA1 to interact with cellular proteins involved in splicing. We therefore used a quantitative proteomics-based approach using mass spectrometry (LC-MS/MS) to identify cellular proteins that interact with EBNA1. Nuclear extracts of cells stably expressing EBNA1 were prepared and the lysates were treated with DNase and RNase to remove potential non-specific interactions of EBNA1 with splicing factors (see [Materials and Methods](#) for experimental details). Using such an approach, we identified numerous splicing factors which were highly associated with EBNA1 (Fig 7D). These include hnRNP M, hnRNP H1, hnRNP U, hnRNP K, hnRNP C, DHX15, and SFPQ (S4 Table). The physical interaction between EBNA1 and some of these splicing factors was also confirmed by co-immunoprecipitation. An example demonstrating the interaction between EBNA1 and hnRNP H1 is presented in Fig 7E.

## Discussion

The current study provides the first comprehensive portrait of global modifications in the cellular RNA splicing signatures that occur in EBVaGC. Our study showed that the AS patterns of numerous cellular transcripts are modified in EBVaGC. Our data also allowed us to identify unique signatures of genes for which AS is misregulated in EBV-negative and EBVaGC.

Analysis of the AS landscape revealed numerous GC-specific markers, which considerably increases the number of potential biomarkers that can presently be recognized using standard expression profiling. The currently available GC markers, such as CEA, CA 19-9 and CA 50, primarily detect advanced GC, for which only palliative treatment is available [3,47]. Clearly, future developments in molecular genetics and proteomic analysis are needed to identify other useful GC-specific markers. The current identification of unique signatures for genes in which AS is misregulated in the different types of GC clearly constitutes a step toward that goal.

The expression of numerous cellular splicing factors has been shown to be dysregulated in various human diseases. For example, the expression of the Serine/Arginine-Rich Splicing Factor 1 (SRSF1) is increased in several tumor types [48], and fibroblasts overexpressing SRSF1 can cause tumors when injected into mice [49]. Similarly, an elevated expression of the splicing factor hnRNP H has been demonstrated to drive splicing switches of oncogenic target genes in gliomas [50]. More recently, RBM4, an RNA-binding factor involved in AS, was also shown to control cancer-related splicing events, thereby affecting various cellular processes such as proliferation, apoptosis, and migration [51]. In the present study, we detected many changes in the expression levels and/or splicing patterns of cellular splicing factors in GC tissue samples. We also observed modifications to the AS profiles of many cellular splicing factors including RBFOX2 and MBNL1. How the altered expression/splicing dynamics of these splicing regulators contribute to AS homeostasis during GC remains to be investigated. Interestingly, recent data points to a key role for RBFOX2 and MBNL1 who seem to account for numerous splicing alterations in various cancers including breast, lung, and prostate cancer [52].

Very limited studies have been conducted to find prognosis markers for GC. The most critical prognostic factor influencing overall survival of patients with GC remains the extent of disease as assessed by tumor stage [3]. Indeed, 80% of patients who undergo gastrectomy with stage I disease are alive after 5 years, but only 7% of patients with stage IV disease which has spread to other organs reach the 5-year survival mark. Lymph nodes ratio (ratio of involved to the total resected lymph nodes) has also been shown to have prognostic significance [53]. Moreover, the histological type of the tumor is frequently considered as an essential prognostic factor in GC [3]. In recent years, numerous expression profiles based on the abundance of mRNAs have been put forward as predictors of prognosis in various cancers [54]. For instance, the mRNAs levels of hnRNP K, an important member of the hnRNP family of proteins, was previously shown to be increased in various cancers [55,56]. It has also been reported that high levels of hnRNP K are associated with poorer overall survival of patients with nasopharyngeal carcinoma (NPC) and prostate cancer [55–56]. High levels of expression of another member of the family, namely hnRNP D, is also correlated with reduced survival in oral cancer [57]. In our study, high levels of DHX15 mRNAs were also associated with reduced patient survival in GC. The exact mechanism by which elevated levels of DHX15 mRNAs might result in poor patient survival is currently unknown. However, a previous study on cerium oxide nanoparticles (CNPs) provides a potential link between DHX15 and poor patient survival [58]. CNPs can interfere with cancer development by inducing both cytotoxicity and oxidative stress in cancer cells. The authors of the study demonstrated that treatment of gastric cells with CNPs resulted in an increased expression of DHX15 [58]. This increased expression of DHX15 was shown to activate p38 MAPK signal pathway, leading to the inhibition of proliferation and metastasis in GC both *in vitro* and *in vivo*. It was suggested that CNPs may be a promising approach to suppress malignant activity of GC by increasing the expression of DHX15 [58]. Clinical validation of potential tumor markers such as DHX15 will clearly continue to be a major focus of GC research in the next few years. Indeed, improvements in early diagnosis are still needed since only a minority of patients will be cured of gastric cancer with surgery alone. The identification of valid biomarkers and/or prognosis markers for GC is critical since

patients for whom curative resection is not possible generally develop deadly symptomatic metastatic disease from unresected microscopical tumor remnants [2].

Many viruses perturb host cell nuclear functions, presumably to promote an ideal environment for viral replication. These modifications on host cell nuclear functions can be achieved by a variety of mechanisms such as alteration of nuclear architecture [59–62], disruption of nucleocytoplasmic transport pathways [63–66], and induction of nuclear herniations [67]. The exact mechanism by which viral infection leads to alterations of the cellular AS landscape is currently unknown. Two scenarios can be envisaged. First, the cells could respond to viral infection by modifying various pathways which ultimately lead to changes in the AS landscape of cellular RNA. Second, viral factors could potentially disrupt cellular pathways which could lead to a modification of the AS profiles of cellular genes. Recent data indicate that important changes can be observed in the AS patterns of two cellular pre-mRNAs (i.e. PCBP2 and DST) following sequestration of the cellular HuR protein by Sindbis virus [68]. In addition, the Poliovirus protease 2A (2Apro) also induces a nucleo-cytoplasm translocation of numerous RNA binding proteins, including splicing factors, which might result in changes in AS [69]. Moreover, previous studies reported the inhibition of pre-mRNA splicing by the Herpes simplex virus 1 (HSV-1) ICP27 protein [70,71]. In fact, ICP27 has been shown to interact with components of the splicing apparatus, thereby causing a redistribution of splicing factors [72,73]. However, a more recent transcriptomic found no evidence for a generalized inhibition of cellular splicing during HSV-1 infection [74]. The current study demonstrated that the AS patterns of 112 cellular genes (including 20 cancer-related genes which were modified in EBVaGC) were significantly modified upon EBNA1 expression. Interestingly, a previous study demonstrated that the EBV noncoding RNA (EBER1) can interact with splicing factor AUF1/hnRNP D, which leads to variations in cellular AS patterns [75]. These findings suggest that numerous viral factors, in addition to cellular components, are likely involved in the observed alterations in the global cellular AS landscape of EBVaGC. Additional studies will clearly be required to investigate the potential role(s) of additional EBV gene products, such as LMP2A, EBERs, and BARF1, in the modifications of cellular AS during viral infection.

Recent advances in large-scale studies have revealed the extent of modifications that occur in AS in various types of cancer [11]. Although the precise molecular mechanisms which lead to a modification of the AS landscape in GC are currently unknown, our study showed that the expression of the viral EBNA1 protein can lead to alterations in the AS profiles of cellular genes. Our results demonstrated that the AS patterns of 20 cancer-related genes which were modified in EBVaGC were also significantly modified upon EBNA1 expression, thereby suggesting a potential mechanism by which the presence of a virus can modify cellular AS profiles. Numerous direct and/or indirect molecular mechanisms by which the expression of a viral protein such as EBNA1 can lead to a modification of the cellular AS landscape can be envisaged. However, the demonstration that EBNA1 can interact with various splicing factors, as demonstrated by our mass spectrometry studies, clearly constitutes a step towards identifying the mechanisms leading to the extensive AS modifications observed in GC. In addition, the interaction of EBNA1 with splicing factors, many of which have additional roles unrelated to splicing, most likely results in alterations of additional biological processes.

Although changes in splicing patterns are characteristics of cancer development and progression, it is currently not clear whether these modifications are the cause or the result of a malignant phenotype. We are still at an early stage of characterizing the full repertoire of cancer-associated alternatively spliced isoforms in various cancers. In the case of our study on GC tissues, numerous additional functional studies will be required to demonstrate the role, if any, of the identified AS events in carcinogenesis. Moreover, additional layers of complexity are most likely involved. For instance, the AS of tenascin-C is triggered by low pH in normal cells,

and this is an effect of nearby tumor growth [76]. Therefore, tumor cells may use this heterotypic signaling with their environment to promote invasion and metastasis. As a result, AS appears intimately bound with cancer cells both as a cause and a consequence [76].

## Supporting information

**S1 Table. Bioinformatical prediction of functional changes caused by some identified ASEs.** Negative delta-PSI indicates that the short isoform is favored and positive delta-PSI indicates that the long isoform is favored.

(PDF)

**S2 Table. List of genes for which AS is uniquely dysregulated in EBVaGC.**

(PDF)

**S3 Table. List of genes for which AS is uniquely dysregulated in EBV-negative GC (Tumor, no virus).**

(PDF)

**S4 Table. List of splicing factors that co-purify with EBNA1.**

(PDF)

**S1 Fig. Characterization of splicing dysregulation in gastric cancer.** Distribution of Delta PSI values in GC.

(PDF)

**S2 Fig. Statistical analysis for the overrepresentation of biological processes.** Overrepresentation of biological processes was analyzed for the differential ASEs associated with EBV-negative GC, EBVaGC, and cells expressing the EBNA1 protein. The first column highlights the name of the biological process, the second column displays the total corresponding number of genes in the database (PANTHER), the third column highlights the number of genes from the searched sample, the fourth column displays the expected number of genes in a normal cell, and the fifth column displays the fold-enrichment. The p-value associated with each overrepresentation is also displayed (last column).

(PDF)

**S3 Fig. Validation of ASEs dysregulated in gastric cancer.** The cDNAs acquired from gastric tissues were analyzed by PCR using specific primers to detect both isoforms of the transcripts encoded by the *CARN14*, *PLA2G4F*, *S100A1*, *SLC52A1*, and *SOGA2* genes. Capillary electrophoregrams of the PCR reactions are presented. The positions and the amplitude of the detected amplicons are indicated. The positions of the internal markers are also shown.

(PDF)

**S4 Fig. RNA splicing factors in EBVaGC.** (A) Iris Graph representing the expression profile of splicing factors for EBVaGC. Differences in gene expression levels are shown on a logarithmic color scale (Log2), from red (negative changes in expression) to blue (increase in gene expression). The expression of proteins involved in splicing modulated by more than 2-fold is indicated by an asterisk. (B) Misregulation of splicing factors alternative splicing in GC (TEBV, Tumors with EBV). The delta PSI values are represented in red (negative delta PSI values) and in blue (positive delta PSI values).

(PDF)

**S5 Fig. Variations in the expression of RNA splicing factors in gastric cancer.** Differences in gene expression levels of proteins involved in splicing are shown. The figure only shows the splicing factors for which the expression level varied by more than 2-fold. A logarithmic color

scale (Log<sub>2</sub>) is used. Red indicates negative changes in expression and blue indicates increase in gene expression.

(PDF)

## Acknowledgments

The authors would like to thank François-Michel Boisvert and Dominique Lévesque for mass spectrometry experiments and analysis.

## Author Contributions

**Conceptualization:** VESA MTL JPP MSS MB.

**Data curation:** VESA MPT AA SB CD.

**Formal analysis:** VESA MPT AA SB CD CMB.

**Funding acquisition:** MB JPP MSS.

**Investigation:** VESA CD EL MD PT.

**Methodology:** PT JPP MSS MB VESA.

**Project administration:** MB MSS JPP.

**Resources:** MB JPP MSS EL MD PT.

**Software:** PT MSS.

**Supervision:** MB JPP MSS.

**Validation:** EL MD PT.

**Visualization:** SB AA VESA MTL CMB PT.

**Writing – original draft:** VESA MB JPP MSS.

**Writing – review & editing:** VESA MB JPP MSS.

## References

1. Bozzetti F, Marubini E, Bonfanti G, Miceli R, Piano C, Gennari L. Subtotal versus total gastrectomy for gastric cancer: five-year survival rates in a multicenter randomized Italian trial. Italian Gastrointestinal Tumor Study Group. *Ann Surg.* 1999; 230:170–178. PMID: [10450730](https://pubmed.ncbi.nlm.nih.gov/10450730/)
2. Dicken BJ, Bigam DL, Cass C, Mackey JR, Joy AA, Hamilton SM. Gastric adenocarcinoma: review and considerations for future directions. *Ann Surg.* 2005; 241:27–39 <https://doi.org/10.1097/01.sla.0000149300.28588.23> PMID: [15621988](https://pubmed.ncbi.nlm.nih.gov/15621988/)
3. Sturgeon CM, Diamandis E. Tumor Markers in Gastric Cancer. In *Use of Tumor Markers in Liver, Bladder, Cervical, and Gastric Cancers*. American Association for Clinical Chemistry. 2010.
4. Uozaki H, Fukayama M. Epstein-Barr Virus and Gastric Carcinoma—Viral Carcinogenesis through Epigenetic Mechanisms. *Int J Clin Exp Pathol.* 2008; 1:198–216. PMID: [18784828](https://pubmed.ncbi.nlm.nih.gov/18784828/)
5. Shinozaki-Ushiku A, Kunita A, Fukayama M. Update on Epstein-Barr virus and gastric cancer (Review). *Int J Oncol.* 2015; 46:1421–1434. <https://doi.org/10.3892/ijo.2015.2856> PMID: [25633561](https://pubmed.ncbi.nlm.nih.gov/25633561/)
6. Luo B, Wang Y, Wang XF, Liang H, Yan LP, Huang BH, Zhao P. Expression of Epstein-Barr virus genes in EBV-associated gastric carcinomas. *World J Gastroenterol.* 2005; 11:629–33. <https://doi.org/10.3748/wjg.v11.i5.629> PMID: [15655811](https://pubmed.ncbi.nlm.nih.gov/15655811/)
7. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014; 513, 202–9 <https://doi.org/10.1038/nature13480> PMID: [25079317](https://pubmed.ncbi.nlm.nih.gov/25079317/)



8. Bonomi S, Gallo S, Catillo M, Pignataro D, Biamonti G, Ghina C. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. *Int J Cell Biol*. 2013;962038. <https://doi.org/10.1155/2013/962038> PMID: 24285959
9. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*. 2003; 72:291–336. <https://doi.org/10.1146/annurev.biochem.72.121801.161720> PMID: 12626338
10. Ghigna C, Giordano S, Shen H, Benvenuto F, Castiglioni PM, Green MR, et al. Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol Cell*. 2005; 20:881–890 <https://doi.org/10.1016/j.molcel.2005.10.026> PMID: 16364913
11. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev*. 2003; 17,419–437. <https://doi.org/10.1101/gad.1048803> PMID: 12600935
12. Bechara EG, Sebestyén E, Bernardis I, Eyraes E, Valcarcel J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell*. 2013; 52:720–733. <https://doi.org/10.1016/j.molcel.2013.11.010> PMID: 24332178
13. Brosseau JP, Lucier JF, Lamarche AA, Shkreta L, Gendron D, Lapointe E, et al. Redirecting splicing with bifunctional oligonucleotides. *Nucleic Acids Res*. 2014; 42:e40. <https://doi.org/10.1093/nar/gkt1287> PMID: 24375754
14. Line A, Slucka Z, Stengrevics A, Li G, Rees RC. Altered splicing pattern of TACC1 mRNA in gastric cancer. *Cancer Genet Cytogenet*. 2002; 139:78–83. PMID: 12547166
15. Li Y, Zhang KL, Sun Y, Yang Y, Chen XY, Kong QY, et al. Frequent S100A4 Expression with Unique Splicing Pattern in Gastric Cancers: A Hypomethylation Event Paralleled with E-cadherin Reduction and Wnt Activation1. *Transl Oncol*. 2008; 1:165–176. PMID: 19043527
16. Xu JH, Wang YC, Geng X, Zhang WM. Changes of alternative splicing variants of human telomerase reverse transcriptase during gastric carcinogenesis. *Ai Zheng*. 2008; 27:1271–6. PMID: 19079992
17. Branco da Cunha C, Oliveira C, Wen X, Gomes B, Sousa S, Suriano G, et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest*. 2010; 90:1604–14. <https://doi.org/10.1038/labinvest.2010.155> PMID: 20856229
18. Smyth EC, Sclafani F, Cunningham D. Emerging molecular targets in oncology: clinical potential of MET/hepatocyte growth-factor inhibitors. *Onco Targets Ther*. 2014; 7:1001–14. <https://doi.org/10.2147/OTT.S44941> PMID: 24959087
19. Wu H-H, Lin W, Tsai K-W. Advances in molecular biomarkers for gastric cancer: miRNAs as emerging novel cancer markers. *Exp Rev Mol Med*. 2014; 16:e1.
20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9(4):357–9. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
21. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323. <https://doi.org/10.1186/1471-2105-12-323> PMID: 21816040
22. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003; 100:9440–45. <https://doi.org/10.1073/pnas.1530509100> PMID: 12883005
23. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*. 2012; 28:1805–1806. <https://doi.org/10.1093/bioinformatics/bts251> PMID: 22543366
24. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012; 40:W597–W603. <https://doi.org/10.1093/nar/gks400> PMID: 22661580
25. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*. 1988; 16:10881–10890. PMID: 2849754
26. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res Database Issue*. 2014; 42:D222–D230.
27. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci USA*. 2009; 106:10171–10176. <https://doi.org/10.1073/pnas.0900604106> PMID: 19520826
28. Venables JP, Koh CS, Froehlich U, Lapointe E, Couture S, Inkel L, et al. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol*. 2008; 28:6033–43. <https://doi.org/10.1128/MCB.00726-08> PMID: 18644864
29. Goswami CP, Nakshatri H. PROGgeneV2: enhancements on the existing database. *BMC Cancer*. 2014; 14:970. <https://doi.org/10.1186/1471-2407-14-970> PMID: 25518851
30. Boisvert FM, Lam YW, Lamont D, Lamond AI. A Quantitative Proteomics Analysis of Subcellular Proteome Localization and Changes Induced by DNA Damage. *Mol Cell Proteomics*. 2010; 9:457–70. <https://doi.org/10.1074/mcp.M900429-MCP200> PMID: 20026476

31. Drissi M, Dubois ML, Douziech M, Boisvert FM. Quantitative Proteomics Reveals Dynamic Interactions of the Minichromosome Maintenance Complex (MCM) in the Cellular Response to Etoposide Induced DNA Damage. *Mol Cell Proteomics*. 2015; 14:2002–13. <https://doi.org/10.1074/mcp.M115.048991> PMID: 25963833
32. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*. 2013; 10:730–6. <https://doi.org/10.1038/nmeth.2557> PMID: 23921808
33. Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer. *EMBO Rep*. 2008; 9:1087–1093. <https://doi.org/10.1038/embor.2008.189> PMID: 18846105
34. Fackenthal JD, Godley LA. Aberrant RNA splicing and its functional consequences in cancer cells. *Dis Model Mech*. 2008; 1:37–42. <https://doi.org/10.1242/dmm.000331> PMID: 19048051
35. Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell*. 2005; 19:1–13. <https://doi.org/10.1016/j.molcel.2005.05.026> PMID: 15989960
36. Kieff E, Rickinson AB. Epstein–Barr virus and its replication. In: Knipe Dm, Howley Pm (eds). *Fields Virology*. Lippincott Williams & Wilkins: Philadelphia. 2001.
37. Sivachandran N, Thawe NN, Frappier L. Epstein-Barr Virus Nuclear Antigen 1 Replication and Segregation Functions in Nasopharyngeal Carcinoma Cell Lines. *J Virol*. 2011; 85:10425–10430. <https://doi.org/10.1128/JVI.05293-11> PMID: 21795327
38. Gruhne B, Sompallae R, Marescotti D, Kamranvar SA, Gastaldello S, Masucci MG. The Epstein-Barr virus nuclear antigen-1 promotes genomic instability via induction of reactive oxygen species. *Proc Natl Acad Sci USA*. 2009; 106:2313–2318 <https://doi.org/10.1073/pnas.0810619106> PMID: 19139406
39. Lu J, Murakami M, Verma SC, Cai Q, Haldar S, Kaul R, et al. Epstein-Barr virus nuclear antigen 1 (EBNA1) confers resistance to apoptosis in EBV-positive B-lymphoma cells through up-regulation of survivin. *Virology*. 2011; 410:64–75 <https://doi.org/10.1016/j.virol.2010.10.029> PMID: 21093004
40. Wilson JB, Bell JL, Levine AJ. Expression of Epstein-Barr virus nuclear antigen-1 induces B cell neoplasia in transgenic mice. *EMBO J*. 1996; 15:3117–3126 PMID: 8670812
41. Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Gervais-Bird J, Madden R, et al. Multiple Alternative Splicing Markers for Ovarian Cancer. *Cancer Res*. 2008; 68:657–663. <https://doi.org/10.1158/0008-5472.CAN-07-2580> PMID: 18245464
42. Geiser V, Cahir-McFarland E, Kieff E. Latent Membrane Protein 1 Is Dispensable for Epstein-Barr Virus Replication in Human Embryonic Kidney 293 cells. *PLoS ONE*. 2011; 6:e22929. <https://doi.org/10.1371/journal.pone.0022929> PMID: 21853056
43. Stamm S, Smith C, Luhrmann R (eds.). In: *Alternative Pre-mRNA Splicing: Theory and Protocols*. Wiley-Blackwell, Germany, eScholarID:158349. 2012.
44. Pawellek A, McElroy S, Samatov T, Mitchell L, Woodland A, Ryder U, et al. Identification of small molecule inhibitors of pre-mRNA splicing. *J Biol Chem*. 2014; 289:34683–98. <https://doi.org/10.1074/jbc.M114.590976> PMID: 25281741
45. Convertini P, Shen M, Potter PM, Palacios G, Lagisetti C, de la Grange P, et al. Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic Acids Research*. 2014; 42:4947–61. <https://doi.org/10.1093/nar/gku151> PMID: 24623796
46. Liu X, Biswas S, Berg MG, Antapli CM, Xie F, Wang Q, Tang MC, et al. Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from *Burkholderia thailandensis* MSMB43. *J Nat Prod*. 2013; 76:685–93. <https://doi.org/10.1021/np300913h> PMID: 23517093
47. Cidón EU, Bustamante R. Gastric cancer: tumor markers as predictive factors for preoperative staging. *J Gastrointest Cancer*. 2011; 42:127–30. <https://doi.org/10.1007/s12029-010-9161-0> PMID: 20499209
48. Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol*. 2012; 19:220–228. <https://doi.org/10.1038/nsmb.2207> PMID: 22245967
49. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol*. 2007; 14:185–193. <https://doi.org/10.1038/nsmb1209> PMID: 17310252
50. LeFave CV, Squatrito M, Vorlova S, Rocco GL, Brennan CW, Holland EC, et al. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J*. 2011; 30:4084–4097. <https://doi.org/10.1038/emboj.2011.259> PMID: 21915099
51. Wang Y, Chen D, Qian H, Tsai YS, Shao S, Liu Q, et al. The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. *Cancer Cell*. 2014; 26:374–89. <https://doi.org/10.1016/j.ccr.2014.07.010> PMID: 25203323

52. Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res.* 2015; 43:5130–5144. <https://doi.org/10.1093/nar/gkv210> PMID: 25908786
53. Siewert JR, Bottcher K, Stein HJ, Roder JD. Relevant prognostic factors in gastric cancer. Ten-year results of the German Gastric Cancer Study. *Ann Surg.* 1998; 228:449–461. PMID: 9790335
54. Mehta S, Shelling A, Muthukaruppan A, Lasham A, Blenkiron C, Laking G, Print C. Predictive and prognostic molecular markers for cancer medicine. *Ther Adv Med Oncol.* 2010; 2:125–48. <https://doi.org/10.1177/1758834009360519> PMID: 21789130
55. Barboro P, Repaci E, Rubagotti A, Salvi S, Boccardo S, Spina B, et al. Heterogeneous nuclear ribonucleoprotein K: altered pattern of expression associated with diagnosis and prognosis of prostate cancer. *Br J Cancer.* 2009; 100:1608–1616. <https://doi.org/10.1038/sj.bjc.6605057> PMID: 19401687
56. Chen LC, Hsueh C, Tsang NM, Liang Y, Chang KP, Hao SP, et al. Heterogeneous Ribonucleoprotein K and Thymidine Phosphorylase Are Independent Prognostic and Therapeutic Markers for Nasopharyngeal Carcinoma. *Clin Cancer Res.* 2008; 14:3807–13. <https://doi.org/10.1158/1078-0432.CCR-08-0155> PMID: 18559600
57. Kumar M, Matta A, Masui O, Srivastava G, Kaur J, Thakar A, et al. Nuclear heterogeneous nuclear ribonucleoprotein D is associated with poor prognosis and interactome analysis reveals its novel binding partners in oral cancer. *J Transl Med.* 2015; 13:1–15.
58. Xiao YF, Li JM, Wang SM, Yong X, Tang B, Jie MM, et al. Yang S.-M. Cerium oxide nanoparticles inhibit the migration and proliferation of gastric cancer by increasing DHX15 expression. *Int J Nanomedicine.* 2016; 11:3023–34. <https://doi.org/10.2147/IJN.S103648> PMID: 27486320
59. Gustin KE. Effects of poliovirus infection on nucleocytoplasmic trafficking and nuclear pore complex composition. *EMBO J.* 2001; 20:240–9. <https://doi.org/10.1093/emboj/20.1.240> PMID: 11226174
60. Gustin KE, Sarnow P. Inhibition of nuclear import and alteration of nuclear pore complex composition by rhinovirus. *J Virol.* 2002; 76:8787–96. <https://doi.org/10.1128/JVI.76.17.8787-8796.2002> PMID: 12163599
61. Muranyi W, Haas J, Wagner M, Krohne G, Koszinowski UH. Cytomegalovirus recruitment of cellular kinases to dissolve the nuclear lamina. *Science.* 2002; 297:854–7. <https://doi.org/10.1126/science.1071506> PMID: 12161659
62. Scott ES O'Hare P. Fate of the inner nuclear membrane protein lamin B receptor and nuclear lamins in herpes simplex virus type 1 infection. *J Virol.* 2001; 75:8818–30.
63. Enninga J, Levy DE, Blobel G, Fontoura BMA. Role of nucleoporin induction in releasing an mRNA nuclear export block. *Science.* 2002; 295:1523–5. <https://doi.org/10.1126/science.1067861> PMID: 11809937
64. Fortes P, Beloso A, Ortín J. Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *EMBO J.* 1994; 13:704–12. PMID: 8313914
65. Petersen JM, Her L-S, Varvel V, Lund E, Dahlberg JE. The matrix protein of vesicular stomatitis virus inhibits nucleocytoplasmic transport when it is in the nucleus and associated with nuclear pore complexes. *Mol Cell Biol.* 2000; 20:8590–601. PMID: 11046154
66. Petersen JM, Her L-S, Dahlberg JE. Multiple vesiculoviral matrix proteins inhibit both nuclear export and import. *Proc Natl Acad Sci.* 2001; 98:8590–5. <https://doi.org/10.1073/pnas.151240998> PMID: 11447272
67. Noronha CMC de Sherman MP, Lin HW Cavrois MV, Moir RD Goldman RD, et al. Dynamic disruptions in nuclear envelope architecture and integrity induced by HIV-1 Vpr. *Science.* 2001; 294:1105–8. <https://doi.org/10.1126/science.1063957> PMID: 11691994
68. Barnhart MD, Moon SL, Emch AW, Wilusz CJ, Wilusz J. Changes in cellular mRNA stability, splicing and polyadenylation through HuR protein sequestration by a cytoplasmic RNA virus. *Cell Rep.* 2013; 5:909–17. <https://doi.org/10.1016/j.celrep.2013.10.012> PMID: 24210824
69. Álvarez E, Castelló A, Carrasco L, Izquierdo JM. Poliovirus 2A protease triggers a selective nucleocytoplasmic redistribution of splicing factors to regulate alternative pre-mRNA splicing. *PLoS ONE.* 2013; 8:e73723. <https://doi.org/10.1371/journal.pone.0073723> PMID: 24066065
70. Lindberg A, Kreivi J-P. Splicing inhibition at the level of spliceosome assembly in the presence of herpes simplex virus protein ICP27. *Virology.* 2002; 294:189–98. <https://doi.org/10.1006/viro.2001.1301> PMID: 11886277
71. Sciabica KS. ICP27 interacts with SRPK1 to mediate HSV splicing inhibition by altering SR protein phosphorylation. *EMBO J.* 2003; 22:1608–19. <https://doi.org/10.1093/emboj/cdg166> PMID: 12660167
72. Bryant HE, Wadd SE, Lamond AI, Silverstein SJ, Clements JB. Herpes simplex virus IE63 (ICP27) protein interacts with spliceosome-associated protein 145 and inhibits splicing prior to the first catalytic step. *J Virol.* 2001; 75:4376–85. <https://doi.org/10.1128/JVI.75.9.4376-4385.2001> PMID: 11287586

73. Sandri-Goldin RM, Hibbard MK, Hardwicke MA. The C-terminal repressor region of herpes simplex virus type 1 ICP27 is required for the redistribution of small nuclear ribonucleoprotein particles and splicing factor SC35; however, these alterations are not sufficient to inhibit host cell splicing. *J Virol.* 1995; 69:6063–76. PMID: [7666511](#)
74. Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, et al. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun.* 2015; 6:7126. <https://doi.org/10.1038/ncomms8126> PMID: [25989971](#)
75. Pimenta G, Fok V, Haslip M, Nagy M, Takyar S, Steitz JA. Proteomics and transcriptomics of BJAB cells expressing the Epstein-Barr virus noncoding RNAs EBER1 and EBER2. *PLoS One.* 2015; 10: e0124638. <https://doi.org/10.1371/journal.pone.0124638> PMID: [26121143](#)
76. Borsi L, Allemanni G, Gaggero B, Zardi L. Extracellular pH controls pre-mRNA alternative splicing of tenascin-C in normal, but not in malignantly transformed, cells. *Int J Cancer.* 1996; 66:632–5 [https://doi.org/10.1002/\(SICI\)1097-0215\(19960529\)66:5<632::AID-IJC9>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0215(19960529)66:5<632::AID-IJC9>3.0.CO;2-U) PMID: [8647625](#)