# Quantitative analysis of the splice variants expressed by the major hepatitis B virus genotypes

Chun Shen Lim[a], Vitina Sozzi[b], Peter A. Revill[b,c]^#, Chris M. Brown[a]^#

[a]Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand

[b]Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

[c]Department of Microbiology and Immunology, University of Melbourne, Melbourne, Victoria, Australia

**Running Head:** Quantitative analysis of the HBV splice variants

^Equal contributions.

#Address correspondence to Chris M. Brown, chris.brown@otago.ac.nz, or Peter A. Revill, peter.revill@mh.org.au

Word count Abstract: 232 + 149

Word count Text: 3321

## ABSTRACT

Hepatitis B virus (HBV) is a major human pathogen that causes liver diseases. The main HBV RNAs are unspliced transcripts that encode the key viral proteins. Recent studies show that some of the HBV spliced transcript isoforms are predictive of liver cancer, yet the roles of these spliced transcripts remain elusive. Furthermore, a total of 9 major HBV genotypes were isolated from discrete geographical regions of the world, it is likely that these genotypes may express a broad variety of spliced transcript isoforms. To systematically study the HBV splice variants, we transfected the human hepatoma cells Huh7 with 4 HBV genotypes (A2, B2, C2, and D3), followed by deep RNA-sequencing. We found that 12-25% of HBV RNAs were splice variants, which were reproducibly detected across independent biological replicates. This accounted for a total of 6 novel and 6 previously identified splice variants. In particular, 2 highly abundant novel splice variants, in which we called the putative splice variants 1 and 5 (pSP1 and pSP5), were specifically expressed at high levels in genotypes D3 and B2, respectively. In general, the HBV splicing profiles varied across the genotypes except for the known spliced pgRNAs SP1 and SP9, which were present in all 4 major genotypes. Counterintuitively, these singly spliced SP1 and SP9 had a suboptimal 5′ splice site, suggesting that splicing of HBV RNAs is tightly controlled by the viral post-transcriptional regulatory RNA element.

## IMPORTANCE

HBV infection affects over 257 million people worldwide. HBV is a major cause of liver diseases including cancer and there is no cure. Some HBV RNAs are spliced variants and their roles are largely unclear, although some splice variants have been previously found to be associated with liver cancer. HBV exists as 9 genotypes worldwide with marked differences in replicative capacity and disease sequelae. Whether HBV splice variants vary for the different genotypes is yet to be investigated in depth. Here we sequenced RNAs from 4 major HBV genotypes using a cell culture system. We found 6 new and 6 previously known splice variants across these genotypes. Some novel splice variants were present at high levels, suggesting they could be functionally important. Interestingly, although HBV has adapted to human hosts for over 50,000 years, the most frequently spliced location shared little flanking sequence similarity with that of humans.

## INTRODUCTION

Hepatitis B virus is a common human pathogen that is a major cause of liver cirrhosis and liver cancer. The genomic DNA of HBV is approximately 3.2kb, but can be transcribed into the greater than genome length pregenomic RNA (pgRNA) and preC RNA (pcRNA) (1). The pgRNA encodes the core (C) and polymerase (P) proteins, whereas the pcRNA encodes the pre-core (PC) protein that is subsequently processed into the Hepatitis B e Antigen (HBeAg). The HBV genome is also transcribed into several subgenomic transcripts, namely the preS1, preS2, S, and X mRNAs. The preS1, S2, and S mRNAs encode the 3 surface (S) structural proteins of the HBV particles and subviral particles (HBsAg). The X mRNA encodes the HBx protein.

Many strains of HBV have arisen from distinct geographical distributions of the world. This is partly due to the long history of virus-host coevolution (over 50,000 years) and the lack of proofreading function of the viral reverse transcriptase (2–4). These strains were grouped into 9 major genotypes (A to I) and putative J, and about 30 sub-genotypes (4, 5). There are marked differences in replication phenotype and disease natural history across HBV genotypes (6, 7), yet the pathogenicity of different HBV genotypes and their implications for treatment are still not fully understood. For example, it is possible that severe liver injury caused by genotype C is related to its high replication capacity (8), and/or its more frequent mutations at the basal core promoter and pre-core regions (9, 10).

In addition, different HBV genotypes may produce distinct spliced transcript isoforms whose precise roles are largely unknown (11–14). At least 17 spliced transcripts of pgRNA (14–26) and 4 spliced transcripts of preS2/S (27, 28) were identified in various sources including liver, serum, and transfected cells. Interestingly, a recent study showed that HBV RNA splicing is more efficient in human hepatoma cells than other tested cell-types (14). Furthermore, spliced pgRNA SP1 is the most commonly detected (16, 17, 19, 21, 29–33), although SP3 and SP9 have also been commonly observed in some studies (13, 14, 26).

Notably, HBV splice variants can be encapsidated to form defective viral particles, with replication and envelopment requiring polymerase and envelope proteins supplied *in trans* by wild-type HBV (18, 20, 21, 34). The SP1 transcript also encodes the Hepatitis B Spliced Protein (HBSP) (29), as well as a truncated (by 1 amino acid) PC p22 protein that inhibited wild-type HBV replication by interfering with wild-type capsid assembly (35). The HBSP is a fusion product of the first 46 amino acid residues of the P protein and 47 amino acid residues from a distinct reading frame. A recent breakthrough study showed that HBSP could reduce liver inflammation *in vivo* (33). Three other splice variants that have coding potential are SP7, SP10 and SP13. SP7 encodes the Hepatitis B Doubly Spliced Protein (HBDSP), a putative pleiotropic activator (36), whereas SP13 encodes the Polymerase-Surface Fusion Protein (P-S FP), a structural protein that could substitute the large HBV surface protein (37). This fusion protein could inhibit HBV replication and may play a role in persistent infection. Interestingly, SP10 could also act as a functional RNA that reduces wild-type HBV replication through interaction with TATA box binding protein (38).

An increasing number of studies has shown that the HBV splice variants are associated with the development and recurrence of hepatocellular carcinoma (32, 39, 40), and poor response to interferon treatment (24). Therefore, we aimed to utilise RNA-seq on cells that had been transfected with replication competent clones of different HBV genotypes to (i) quantify the composition of splice variants at the RNA level, (ii) investigate the effects of sequence variations on splicing efficiency, (iii) determine the usage of splice sites, and (iv) understand the host response to viral replication across the major HBV genotypes A to D.

## RESULTS AND DISCUSSION

### Six of 12 HBV splice variants detected were novel transcripts

HBV genotypes A to D expressed a large proportion of spliced transcript isoforms, representing 12-25% of the HBV transcriptomes detected (Fig 1A), showing that HBV splicing was indeed widespread across the genotypes. HBV genotype B2 expressed the highest level of HBV transcriptome, followed by A2, C2, and D3 [5830, 5365, 4783, and 4286 TPM (Transcripts Per kilobase Million mapped reads), respectively; see also Supplementary Fig S1 and Table S1 in read counts].

A total of 12 splice variants were consistently detected across independent biological replicates with high confidence (Fig 1B). In particular, we only reported the splice variants with complete, exact match intron chains across independent biological replicates, which were supported by over 10 reads mapped across the splice junctions, indicating spliced reads (Fig 2, and Supplementary Table S2 and S3). Six out of 12 splice variants were novel. In particular, 2 putative splice variants, pSP1 and pSP5 were expressed at high levels in genotypes D3 and B2, respectively. The importance of these 2 highly abundant novel splice variants requires further investigation.

We also identified previously reported splice variants SP1, 5, 6, 7, 9 and 13** across the HBV genotypes (11, 14, 24, 38) (Fig 1B and 2). Notably, SP1 and SP9 were consistently detected in all the genotypes. SP1 was the major spliced transcript detected, ranging from 7 to 16% of the HBV transcriptomes, which is in agreement with previous findings (11, 13, 41). SP9 was the second

most abundant spliced transcript, ranging from 1 to 5% of the HBV transcriptomes. However, the role of SP9 is still unclear. SP13** was previously identified as a novel splice variant, and we were able to detect it across genotypes A2, B2 and D3 at high levels. In addition, SP6 and SP7 were detected in 3 major genotypes (A2, B2 and D3, and A2, C2 and D3, respectively). While any function is yet to be attributed to SP6 or SP13**, SP7 encodes a putative pleiotropic activator (HBDSP) that has been shown to increase replication of wild-type HBV in co-transfection cell culture experiments (36).

**Sequence variations surrounding the HBV splice sites affected splicing efficiency**

We next investigated whether the sequence contexts of splice sites contributed to the different types and abundance of splice variants observed across genotypes. MaxEntScan scoring of the HBV splice sites showed that sequence variation affected the strength of the splice sites of the different HBV genotypes (Fig 2 and Supplementary Table S3).

In general, splice sites with weak sequence contexts (negative MaxEntScan scores) were less likely to be used for splicing and vice versa. For example, the splice donor site at position 2087 in genotypes A2 and D3 had poor sequence contexts and spliced reads associated with this site were not detected (see ① in Fig 2 and Supplementary Table S3). In contrast, the same donor position in genotypes B2 and C3 had strong sequence contexts and were supported by over 100 spliced reads. This indicates that the splicing efficiencies of the HBV RNAs are strongly influenced by the HBV sequence variants surrounding the splice sites.

However, we also observed a discordance between splice site sequence contexts and splice read counts. For example, all the genotypes had similar scores at the splice acceptor position 1385 but the splice read counts were markedly different (see ② in Fig 2 and Supplementary Table S3). In particular, the most frequently used 5′ splice site which was used for SP1 and SP9 had a negative MaxEntScan score (see ③ or position 2447 in Fig 2 and Supplementary Table S3). These results suggest that the splicing of this splice junction may be controlled by other *cis*-regulatory elements, such as the HBV post-transcriptional regulatory RNA element (PRE) (1). Indeed, deleting a PRE component called the splicing regulatory element-1 (SRE-1) was previously found to inhibit pgRNA splicing and the production of SP1 (41). Regulation of alternative splicing may play a crucial role in viral-host interactions (42).

**HBV encoded more alternative 3′ splice sites than 5′ splice sites**

A closer examination of the HBV splicing profiles revealed that HBV encoded more alternative 3′ splice sites than 5′ splice sites. (Fig 3A and Supplementary Table S3). Indeed, a trend was observed for more spliced reads mapped across the 5′ splice sites than 3′ splice sites, which reached statistical significance for HBV genotype B2. In contrast, host RNAs had balanced numbers of 5′ and 3′ splice sites (53,009 and 52,998, respectively), as well as the supported read counts (Fig 3, median read counts of 87 for both the 5′ and 3’ splice sites).

To quantify the rates of splicing in HBV versus host cell RNAs, we calculated the completed Splicing Index (coSI) of the 5′ and 3′ splice sites. The 5′ splice sites of HBV showed higher coSI scores than 3′ splice sites (9% versus 5% on average; see also Fig 3B). In contrast, splicing was

87% and 86% completed at the host 5′ and 3′ splice sites, respectively. These results showed that the 5′ splice sites of HBV tend to be more frequently spliced than 3′ splice sites (e.g. see ② in Fig 2), but were much less efficiently utilised than host splice sites.

To identify the key differences between the HBV and human splice sites, we compared their splice site contexts using WebLogo (43), which used the frequencies of the uniquely mapped spliced reads to estimate the most frequently used splice sites. This approach showed that the nucleotide frequency distributions of human splice sites were similar to previous studies (Fig 4) (44). Notably, the most frequently used splice sites differed between the virus and host, e.g. −1 positions of the splice donor sites (Fig 4, left panel, U versus G shaded in gray). The differences between the HBV genotypes were marginal, as the spliced reads were predominantly mapped to SP1 and SP9 (Fig 1B and 2).

**HBV replication had little effect on the host gene expression**

To understand the impact of HBV replication on the host, we carried out differential gene expression analysis using DESeq2 (45). We found that only one and 12 genes were significantly differentially expressed in A2 and B2 treated samples, respectively (Fig 5, red points; Fig 6 and Supplementary Table S6, FDR-adjusted $p$-value <0.05). Interestingly, both the A2 and B2 genotypes also showed higher levels of HBV transcriptomes (Fig S1; see also Supplementary Table S1 and S2). The accumulation of HBV transcripts may induce a stress response as the stress-related genes INHBE, FAM129A, SESN2, ASNS, and CHAC1 were all upregulated (Fig 7; see also the functional annotation results in Supplementary Table S7). Indeed, previous studies

have also shown that HBV infection could lead to endoplasmic reticulum stress (46–48), including upregulation of INHBE (49). Interestingly, 3 significantly upregulated genes (ADM2, AKNA, and SH3BP2) were previously shown to correlate with the Ishak fibrosis stage (50).

**Concluding remarks**

Our study has used RNA-seq to identify the type and proportion of splice variants produced for HBV genotypes A2, B2, C2 and D3, with a higher degree of sensitivity than 454 sequencing and conventional reverse-transcription PCR (13, 14, 21, 24), all of which require an intervening step to generate cDNA, followed by PCR. We identified a number of novel splice variants across HBV genotypes as well as previously identified variants of yet unknown functions. The role of the SP9 variant in particular needs to be further explored. We acknowledge that this study is limited to one clone for each of the genotypes, and needs to be expanded to include additional HBV genotypes. Nonetheless, this study demonstrates that HBV has a large capacity for alternative splicing likely controlled by *cis*-acting elements such as the PRE (1), which results in high-levels of SP1 and SP9 mRNAs, despite the suboptimal context of the 5′ splice site. With up to 25% of all HBV mRNAs being of splice origin, the importance of these molecules in the HBV "life-cycle" and pathogenesis requires further investigation.

## MATERIALS AND METHODS

### Cell culture

Cell culture and transfection experiments were carried out as previously described with the following modifications (6). Huh7 cells were seeded in 6-well plates at partial confluence. After overnight incubation, the cells were transiently transfected with pUC57 constructs harboring 1.3-mer HBV genomes (genotypes A2, B2, C2, and D3) using FuGene 6 transfection reagent according to the manufacturer's instructions (Promega, WI, USA). The generation of plasmids has been previously described and this transient expression system relies on the endogenous promoters of HBV for transcription (6). The empty pUC57 vector was used as a control. Two independent biological replicates were performed which included 2 technical replicates for each treatment.

### RNA-seq

Total RNA samples were purified using RNeasy kit (Qiagen, Hilden, Germany) and submitted to the Otago Genomics and Bioinformatics Facility at the University of Otago (Dunedin, New Zealand) under contract for library construction and sequencing. The libraries were prepared using TruSeq stranded Total RNA sample preparation kit with Ribo-Zero according to the manufacturer's protocol and sequenced using HiSeq 2500 (Illumina, CA, USA), generating 125-bp paired-end reads (see the RNA-seq analysis workflow in Fig 7).

### Quality control (QC) of RNA-seq

The fastq files were examined using FastQC v0.11.5 (51). Most files passed most of the analysis modules except 'per base sequence content', 'sequence duplication levels', and 'k-mer content', which are common warnings for Illumina TruSeq reads (FastQC documentation). However, some fastq files failed at per base sequence quality and per base N content due to diminish of the quality score over position 100. Some fastq files also failed at per tile sequence quality due to loss of quality at random positions and cycles, which is likely due to the overloading of flowcell. Both of these issues should have minimum impact on downstream analysis because the regions of poor based calling were soft-clipped during alignment. In addition, only uniquely mapped reads were used for gene counting and transcript assembly.

As a post-alignment QC, the mapping statistics of the non-redundant RNA-seq reads were examined. About 60% of the reads were uniquely mapped reads to the human genomes (Supplementary Table S4). The distribution of aligned reads was then analyzed using the CollectRnaSeqMetrics mode of Picard 2.10.2 (http://broadinstitute.github.io/picard; Broad Institute, Cambridge, MA, USA). Over 55% and 27% of the bases of these reads were mapped the coding sequences (CDS) and untranslated regions (UTRs), respectively. Only 10% or lower of the bases of the sequencing reads were aligned to intronic or intergenic regions (Supplementary Table S5). These metrics are comparable with previous findings (52), indicating that our RNA-seq libraries are reliable.

**Sequence alignment**

Adapter sequences were trimmed from the RNA-seq reads using Skewer v0.2.2 (53). To detect novel splice junctions, alignment was performed using STAR 2.5.2a in 2-pass mode (54). Duplicated reads were removed and uniquely mapped reads were retained using SAMtools (55).

**Transcriptome assembly**

Transcriptome assembly of HBV splice variants was carried out using StringTie v1.3.3b (56). This tool assembles and quantifies spliced transcript isoforms using a network flow algorithm. Annotations of the spliced transcript isoforms were merged by biological replicates using gtfmerge (https://github.com/Kingsford-Group/rnaseqtools) and GffCompare (57). Only the assembled spliced transcripts that were found in both biological replicates were reported (intersection of complete, exact match intron chain). After merging the BAM files by biological replicates using SAMtools, a spliced graph of HBV transcripts was plotted using Gviz and GenomicFeatures (58, 59).

**Splice site analysis**

Splice site sequence contexts were scored using MaxEntScan (60). This tool is a key plugin of the Ensembl Variant Effect Predictor (61) and performed the best in a recent benchmark (62). Splice site mapping frequencies were parsed from the SJ.out.tab file from STAR. IPSA was used to calculate the coSI score of 5′ and 3′ splice sites (https://github.com/pervouchine/ipsa) (63). WebLogo 3.5.0 was used to plot the nucleotide frequencies surrounding the splice sites.

**Differential gene expression analysis**

306 To examine the reproducibility of the biological replicates, the uniquely mapped reads were first
307 counted and summarized at the gene level using mmquant v1.3 (64). The correlation of samples
308 was analyzed. The Spearman's correlations between the biological replicates were >0.9,
309 suggesting a good reproducibility (Supplementary Fig S2). However, the Spearman's
310 correlations between biological replicates were smaller than those within the same batch (e.g.
311 A2_rep1 versus A2_rep2 is 0.938, whereas A2_rep1 versus B2_rep1 is 0.996). These results
312 suggest the presence of batch effects, which is likely due to the second biological replicate being
313 performed a year after. This was further examined using principal component analysis (PCA).
314 Indeed, the samples were clustered by batches (Supplementary Fig S3).
315
316 To resolve the issue of batch effects, read counts were transformed using the vst
317 (variance-stabilizing transformation) function of DESeq2 (45). Transformed read counts were
318 examined using the plotPCA function of DESeq2, before and after correction using the
319 removeBatchEffect function of limma (65). To take batch effects into account, differential
320 expression analysis was carried out using batch as a linear term in the DESeqDataSetFromMatrix
321 function. Differentially expressed genes were examined using DAVID functional annotation
322 tools v6.8 (66, 67).
323
324 **Statistical analysis**
325 Welch two-sample t-test and permutation test were performed using the exactRankTests R
326 package (68, 69). Plotting was carried out using ggplot2 unless otherwise stated (70).
327

## Code and data availability

The raw RNA-seq libraries are available on Gene Expression Omnibus (71) (GSE155983). Scripts and data for the analysis can be found at https://github.com/lcscs12345/HBV_splicing_paper_2020

## ACKNOWLEDGMENTS

CMB and CSL were funded by the University of Otago. CSL is a recipient of a Dr. Sulaiman Daud 125th Jubilee Postgraduate Scholarship and the Marjorie McCallum travel award. PR and VS were funded by the NHMRC grant APP1145977.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## REFERENCES

1. Lim CS, Brown CM. 2016. Hepatitis B virus nuclear export elements: RNA stem-loop α and β, key parts of the HBV post-transcriptional regulatory element. RNA Biol 13:743–747.
2. Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, Rinne C, Kornell S-C, Yepes D, Franzenburg S, Heyne HO, Meier T, Lösch S, Meller H, Friederich S, Nicklisch N, Alt KW, Schreiber S, Tholey A, Herbig A, Nebel A, Krause J. 2018. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife 7.
3. Yuen LKW, Littlejohn M, Duchêne S, Edwards R, Bukulatjpi S, Binks P, Jackson K, Davies J, Davis JS, Tong SYC, Locarnini S. 2019. Tracing Ancient Human Migrations into Sahul

Using Hepatitis B Virus Genomes. Mol Biol Evol 36:942–954.

4. Revill PA, Tu T, Netter HJ, Yuen LKW, Locarnini SA, Littlejohn M. 2020. The evolution and clinical impact of hepatitis B virus genome diversity. Nat Rev Gastroenterol Hepatol.

5. McNaughton AL, Revill PA, Littlejohn M, Matthews PC, Ansari MA. 2020. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. J Gen Virol 101:271–283.

6. Sozzi V, Walsh R, Littlejohn M, Colledge D, Jackson K, Warner N, Yuen L, Locarnini SA, Revill PA. 2016. In Vitro Studies Show that Sequence Variability Contributes to Marked Variation in Hepatitis B Virus Replication, Protein Expression, and Function Observed across Genotypes. J Virol 90:10054–10064.

7. Kramvis A. 2014. Genotypes and genetic variability of hepatitis B virus. Intervirology 57:141–150.

8. Kao J-H. 2011. Molecular epidemiology of hepatitis B virus. Korean J Intern Med 26:255–261.

9. Kramvis A, Kostaki E-G, Hatzakis A, Paraskevis D. 2018. Immunomodulatory Function of HBeAg Related to Short-Sighted Evolution, Transmissibility, and Clinical Manifestation of Hepatitis B Virus. Front Microbiol 9:2521.

10. Chotiyaputta W, Lok ASF. 2009. Hepatitis B virus variants. Nat Rev Gastroenterol Hepatol 6:453–462.

11. Lee GH, Wasser S, Lim SG. 2008. Hepatitis B pregenomic RNA splicing--the products, the regulatory mechanisms and its biological significance. Virus Res 136:1–7.

12. Candotti D, Allain J-P. 2016. Biological and clinical significance of hepatitis B virus RNA

splicing: an update. Ann Blood 2:6–6.

13. Huang C-C, Kuo T-M, Yeh C-T, Hu C-P, Chen Y-L, Tsai Y-L, Chen M-L, Chou Y-C, Chang C. 2013. One single nucleotide difference alters the differential expression of spliced RNAs between HBV genotypes A and D. Virus Res 174:18–26.

14. Ito N, Nakashima K, Sun S, Ito M, Suzuki T. 2019. Cell Type Diversity in Hepatitis B Virus RNA Splicing and Its Regulation. Front Microbiol 10:207.

15. Sommer G, van Bömmel F, Will H. 2000. Genotype-specific synthesis and secretion of spliced hepatitis B virus genomes in hepatoma cells. Virology 271:371–381.

16. Chen PJ, Chen CR, Sung JL, Chen DS. 1989. Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. J Virol 63:4165–4171.

17. Su TS, Lai CJ, Huang JL, Lin LH, Yauk YK, Chang CM, Lo SJ, Han SH. 1989. Hepatitis B virus transcript produced by RNA splicing. J Virol 63:4011–4018.

18. Terré S, Petit MA, Bréchot C. 1991. Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs in vivo. J Virol 65:5539–5543.

19. Wu HL, Chen PJ, Tu SJ, Lin MH, Lai MY, Chen DS. 1991. Characterization and genetic analysis of alternatively spliced transcripts of hepatitis B virus in infected human liver tissues and transfected HepG2 cells. J Virol 65:1680–1686.

20. Rosmorduc O, Petit MA, Pol S, Capel F, Bortolotti F, Berthelot P, Brechot C, Kremsdorf D. 1995. In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology 22:10–19.

21. Günther S, Sommer G, Iwanska A, Will H. 1997. Heterogeneity and common features of

394 defective hepatitis B virus genomes derived from spliced pregenomic RNA. Virology
395 238:363–371.
396 22. Abraham TM, Lewellyn EB, Haines KM, Loeb DD. 2008. Characterization of the
397 contribution of spliced RNAs of hepatitis B virus to DNA synthesis in transfected cultures
398 of Huh7 and HepG2 cells. Virology 379:30–37.
399 23. El Chaar M, El Jisr T, Allain J-P. 2012. Hepatitis B virus DNA splicing in Lebanese blood
400 donors and genotype A to E strains: implications for hepatitis B virus DNA quantification
401 and infectivity. J Clin Microbiol 50:3159–3167.
402 24. Chen J, Wu M, Wang F, Zhang W, Wang W, Zhang X, Zhang J, Liu Y, Liu Y, Feng Y, Zheng
403 Y, Hu Y, Yuan Z. 2015. Hepatitis B virus spliced variants are associated with an impaired
404 response to interferon therapy. Sci Rep 5:16459.
405 25. Lam AM, Ren S, Espiritu C, Kelly M, Lau V, Zheng L, Hartman GD, Flores OA, Klumpp
406 K. 2017. Hepatitis B Virus Capsid Assembly Modulators, but Not Nucleoside Analogs,
407 Inhibit the Production of Extracellular Pregenomic RNA and Spliced RNA Variants.
408 Antimicrob Agents Chemother 61.
409 26. Betz-Stablein BD, Töpfer A, Littlejohn M, Yuen L, Colledge D, Sozzi V, Angus P,
410 Thompson A, Revill P, Beerenwinkel N, Warner N, Luciani F. 2016. Single-Molecule
411 Sequencing Reveals Complex Genome Variation of Hepatitis B Virus during 15 Years of
412 Chronic Infection following Liver Transplantation. J Virol 90:7171–7183.
413 27. Hass M, Hannoun C, Kalinina T, Sommer G, Manegold C, Günther S. 2005. Functional
414 analysis of hepatitis B virus reactivating in hepatitis B surface antigen-negative individuals.
415 Hepatology 42:93–103.

28. Candotti D, Lin CK, Belkhiri D, Sakuldamrongpanich T, Biswas S, Lin S, Teo D, Ayob Y, Allain J-P. 2012. Occult hepatitis B infection in blood donors from South East Asia: molecular characterisation and potential mechanisms of occurrence. Gut 61:1744–1753.

29. Soussan P, Tuveri R, Nalpas B, Garreau F, Zavala F, Masson A, Pol S, Brechot C, Kremsdorf D. 2003. The expression of hepatitis B spliced protein (HBSP) encoded by a spliced hepatitis B virus RNA is associated with viral replication and liver fibrosis. J Hepatol 38:343–348.

30. Rosmorduc O, Petit M-A, Pol S, Capel F, Bortolotti F, Berthelot P, Brechot C, Kremsdorf D. 1995. In vivo andin vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology 22:10–19.

31. Suzuki T, Kajino K, Masui N, Saito I, Miyamura T. 1990. Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA. Virology 179:881–885.

32. Soussan P, Pol J, Garreau F, Schneider V, Le Pendeven C, Nalpas B, Lacombe K, Bonnard P, Pol S, Kremsdorf D. 2008. Expression of defective hepatitis B virus particles derived from singly spliced RNA is related to liver disease. J Infect Dis 198:218–225.

33. Duriez M, Mandouri Y, Lekbaby B, Wang H, Schnuriger A, Redelsperger F, Guerrera CI, Lefevre M, Fauveau V, Ahodantin J, Quetier I, Chhuon C, Gourari S, Boissonnas A, Gill U, Kennedy P, Debzi N, Sitterlin D, Maini MK, Kremsdorf D, Soussan P. 2017. Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. J Hepatol 67:687–699.

34. Köck J, Nassal M, Deres K, Blum HE, von Weizsäcker F. 2004. Hepatitis B virus nucleocapsids formed by carboxy-terminally mutated core proteins contain spliced viral

genomes but lack full-size DNA. J Virol 78:13812–13818.

35. Wang Y-L, Liou G-G, Lin C-H, Chen M-L, Kuo T-M, Tsai K-N, Huang C-C, Chen Y-L, Huang L-R, Chou Y-C, Chang C. 2015. The inhibitory effect of the hepatitis B virus singly-spliced RNA-encoded p21.5 protein on HBV nucleocapsid formation. PLoS One 10:e0119625.

36. Chen W-N, Chen J-Y, Lin W-S, Lin J-Y, Lin X. 2010. Hepatitis B doubly spliced protein, generated by a 2.2 kb doubly spliced hepatitis B virus RNA, is a pleiotropic activator protein mediating its effects via activator protein-1- and CCAAT/enhancer-binding protein-binding sites. J Gen Virol 91:2592–2600.

37. Huang HL, Jeng KS, Hu CP, Tsai CH, Lo SJ, Chang C. 2000. Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. Virology 275:398–410.

38. Tsai K-N, Chong C-L, Chou Y-C, Huang C-C, Wang Y-L, Wang S-W, Chen M-L, Chen C-H, Chang C. 2015. Doubly Spliced RNA of Hepatitis B Virus Suppresses Viral Transcription via TATA-Binding Protein and Induces Stress Granule Assembly. J Virol 89:11406–11419.

39. Pan M-H, Hu H-H, Mason H, Bayliss J, Littlejohn M, Lin Y-L, Su C-Y, Chiang C-T, Chen C-J, Locarnini S, Yang H-I, Revill P. 2018. Hepatitis B splice variants are strongly associated with and are indeed predictive of hepatocellular carcinoma. J Hepatol 68:S474–S475.

40. Bayliss J, Lim L, Thompson AJV, Desmond P, Angus P, Locarnini S, Revill PA. 2013. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. J

460 Hepatol 59:1022–1028.

461 41. Heise T, Sommer G, Reumann K, Meyer I, Will H, Schaal H. 2006. The hepatitis B virus
462 PRE contains a splicing regulatory element. Nucleic Acids Res 34:353–363.

463 42. Chauhan K, Kalam H, Dutt R, Kumar D. 2019. RNA Splicing: A New Paradigm in
464 Host-Pathogen Interactions. J Mol Biol 431:1565–1575.

465 43. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo
466 generator. Genome Res 14:1188–1190.

467 44. Padgett RA. 2012. New connections between splicing and human disease. Trends Genet
468 28:147–154.

469 45. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion
470 for RNA-seq data with DESeq2. Genome Biol 15:550.

471 46. Montalbano R, Honrath B, Wissniowski TT, Elxnat M, Roth S, Ocker M, Quint K, Churin
472 Y, Roederfeld M, Schroeder D, Glebe D, Roeb E, Di Fazio P. 2016. Exogenous hepatitis B
473 virus envelope proteins induce endoplasmic reticulum stress: involvement of cannabinoid
474 axis in liver cancer cells. Oncotarget 7:20312–20323.

475 47. Li X, Pan E, Zhu J, Xu L, Chen X, Li J, Liang L, Hu Y, Xia J, Chen J, Chen W, Hu J, Wang
476 K, Tang N, Huang A. 2018. Cisplatin Enhances Hepatitis B Virus Replication and PGC-1α
477 Expression through Endoplasmic Reticulum Stress. Sci Rep 8:3496.

478 48. Liu X, Green RM. 2019. Endoplasmic reticulum stress and liver diseases. Liver Res
479 3:55–64.

480 49. Sasaki R, Kanda T, Nakamura M, Nakamoto S, Haga Y, Wu S, Shirasawa H, Yokosuka O.
481 2016. Possible Involvement of Hepatitis B Virus Infection of Hepatocytes in the

Attenuation of Apoptosis in Hepatic Stellate Cells. PLoS One 11:e0146314.

50. Wang Q, Lin L, Yoo S, Wang W, Blank S, Fiel MI, Kadri H, Luan W, Warren L, Zhu J, Hiotis SP. 2016. Impact of non-neoplastic vs intratumoural hepatitis B viral DNA and replication on hepatocellular carcinoma recurrence. Br J Cancer 115:841–847.

51. Andrews S. 2016. FastQC: A Quality Control tool for High Throughput Sequence Data.

52. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dökel S, Ralser M, Balzereit D, Lehrach H, Yaspo M-L. 2014. Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics 15:675.

53. Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182.

54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

56. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295.

57. Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. F1000Res 9:304.

58. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118.

504 59. Hahne F, Ivanek R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor.
505 Methods Mol Biol 1418:335–351.
506 60. Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with
507 applications to RNA splicing signals. J Comput Biol 11:377–394.
508 61. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F.
509 2016. The Ensembl Variant Effect Predictor. Genome Biol 17:122.
510 62. Jian X, Boerwinkle E, Liu X. 2014. In silico prediction of splice-altering single nucleotide
511 variants in the human genome. Nucleic Acids Res 42:13534–13544.
512 63. Pervouchine DD, Knowles DG, Guigó R. 2013. Intron-centric estimation of alternative
513 splicing from RNA-seq data. Bioinformatics 29:273–274.
514 64. Zytnicki M. 2017. mmquant: how to count multi-mapping reads? BMC Bioinformatics
515 18:411.
516 65. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers
517 differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids
518 Res 43:e47.
519 66. Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large
520 gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.
521 67. Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths
522 toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res
523 37:1–13.
524 68. Torsten Hothorn and Kurt Hornik. 2019. exactRankTests: Exact Distributions for Rank and
525 Permutation Tests. Comprehensive R Archive Network (CRAN).

69. R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

70. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

71. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 41:D991–5.

## FIGURE LEGENDS

**Fig 1. HBV genotypes expressed a wide variety of spliced transcript isoforms. (A)** Proportions of the HBV spliced transcripts. Only the spliced transcripts present in both biological replicates are shown. **(B)** Relative abundance of the HBV splice variants in genotype A to D. See also Supplementary Table S1 and S2. TPM, Transcripts Per kilobase Million mapped reads.

**Fig 2. Distinct splicing profiles were observed across the HBV genotypes.** Lollipop plot indicates the positions of splice sites relative to the *EcoR*I site of genotype C2. Blue and red colors indicate 5′ and 3′ splice sites, respectively). SP and pSP denote the known and putative spliced pgRNA transcripts, respectively (splice variants' panel). Black dotted lines denote the positions of initiation codons of C, P, preS1, and X reading frames (ORFs' panel). Heatmap shows the abundance of HBV spliced transcripts. Read coverage is shown in gray (coverage panel, y-axis). Arcs represent RNA-seq reads mapped across the splice junctions (supporting read counts in red color). Only the splice junctions supported by at least 100 reads are shown. Blue and red vertical lines indicate the MaxEntScan scores of the 5′ and 3′ splice sites, respectively (coverage panel). A positive MaxEntScan score predicts a good splice site sequence context whereas a negative score predicts a poor splice site sequence context. Three main scenarios were observed. ① The presence and absence of spliced reads at position 2087 were predicted by MaxEntScan scores, in which reads were found to map across the 5′ splice sites with strong positive scores (B2 and C2) but not those with strong negative scores (A2 and D3). ② Varying spliced read counts could not be explained by similar scores. ③ Most spliced reads

were mapped across a weak splice donor site. See also Supplementary Table S3. ORFs, open reading frames.

**Fig 3. HBV 5′ splice sites are more likely to be spliced than 3′ splice sites. (A)** More spliced reads were mapped across the 5′ splice sites of HBV than 3′ splice sites. Similar results were obtained from Welch two-sample t-test (one-sided) and permutation test (e.g. *p*-values of 0.04 and 0.06 were obtained for genotype B2, respectively). Solid black lines indicate median values. **(B)** Completeness of splicing at the 5′ and 3′ splice sites. Only the splice sites supported by at least 10 reads were included for comparison. coSI, completed Splicing Index.

**Fig 4. Most frequently used splice sites differed between HBV and the host.** The nucleotide frequencies surrounding the splice sites are represented by the spliced reads. Exon boundaries are shaded in gray. Only the splice sites supported by at least 10 reads were included.

**Fig 5. MA plots show differential gene expression between the HBV treated and control samples.** Normalized counts indicate the counts divided by the normalization factors (as computed using the DESeq2 default function). Red points denote the FDR-adjusted *p*-value of <0.05. Unfilled triangles denote the genes that have undergone twofold changes in expression. MA plot, log ratio versus mean expression (log scale). See also the related Supplementary Table S6.

**Fig 6. Significantly dysregulated genes in the HBV treated cells.** A total of 12 genes were differentially expressed in the B2 treated sample. Circled numbers denote the ranking based on FDR-adjusted *p*-values. See also Supplementary Table S6 and S7.

**Fig 7. RNA-seq analysis of the HBV and host transcriptomes.** Quality control (QC) check on the paired-end RNA-seq libraries was carried out using FastQC. Adapter sequences were trimmed from the RNA-seq reads using Skewer. Trimmed reads were aligned to the human and HBV genomes using STAR in 2-pass mode. Duplicated and multi-mapped reads were discarded from the Binary Alignment Map (BAM) files using SAMtools. Post-alignment QC check was performed using Picard tools. Transcriptome assembly and quantification were carried out using StringTie, with a post-processing step focusing on the HBV spliced transcript isoforms. Splice site sequence contexts were scored using MaxEntScan. Completeness of RNA splicing was evaluated using IPSA. Reads mapped to human genes were quantified using mmquant, followed by differential gene expression analysis using DESeq2. A list of differentially expressed genes was submitted to the DAVID webserver for functional annotation analysis. coSI, completed Splicing Index.

**Fig 1**

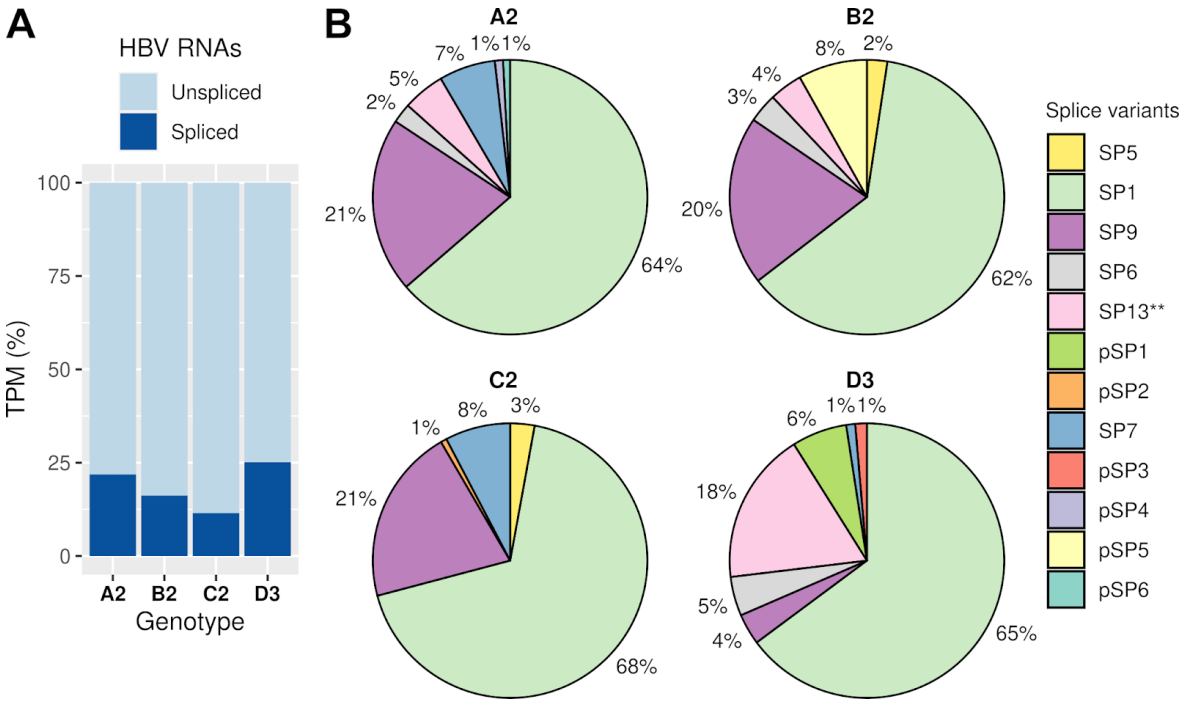


**Fig 2**

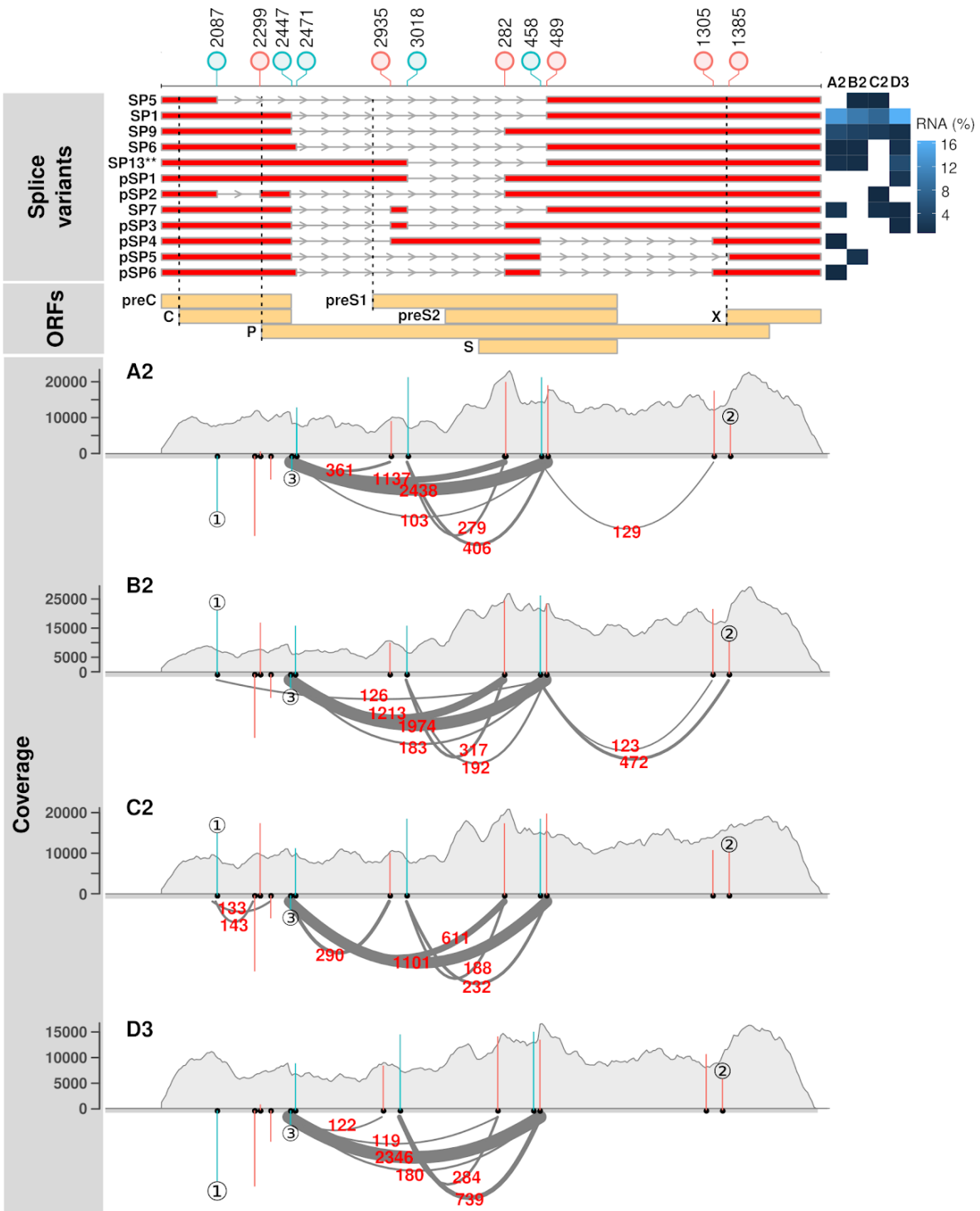


**Fig 3**

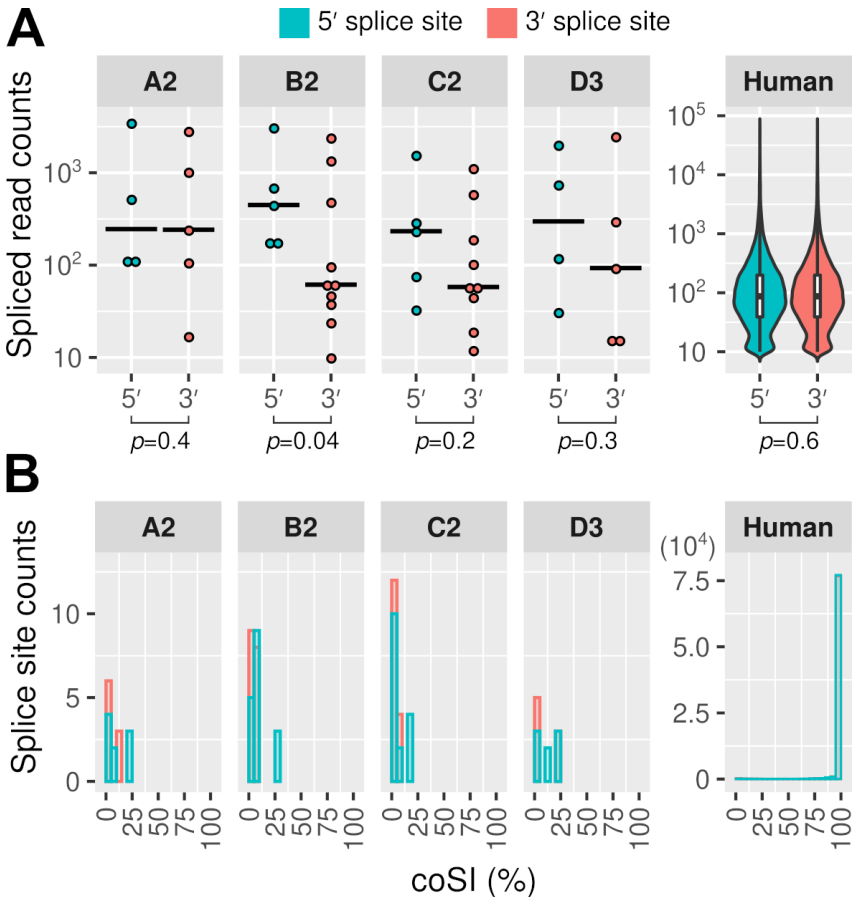


**Fig 4**

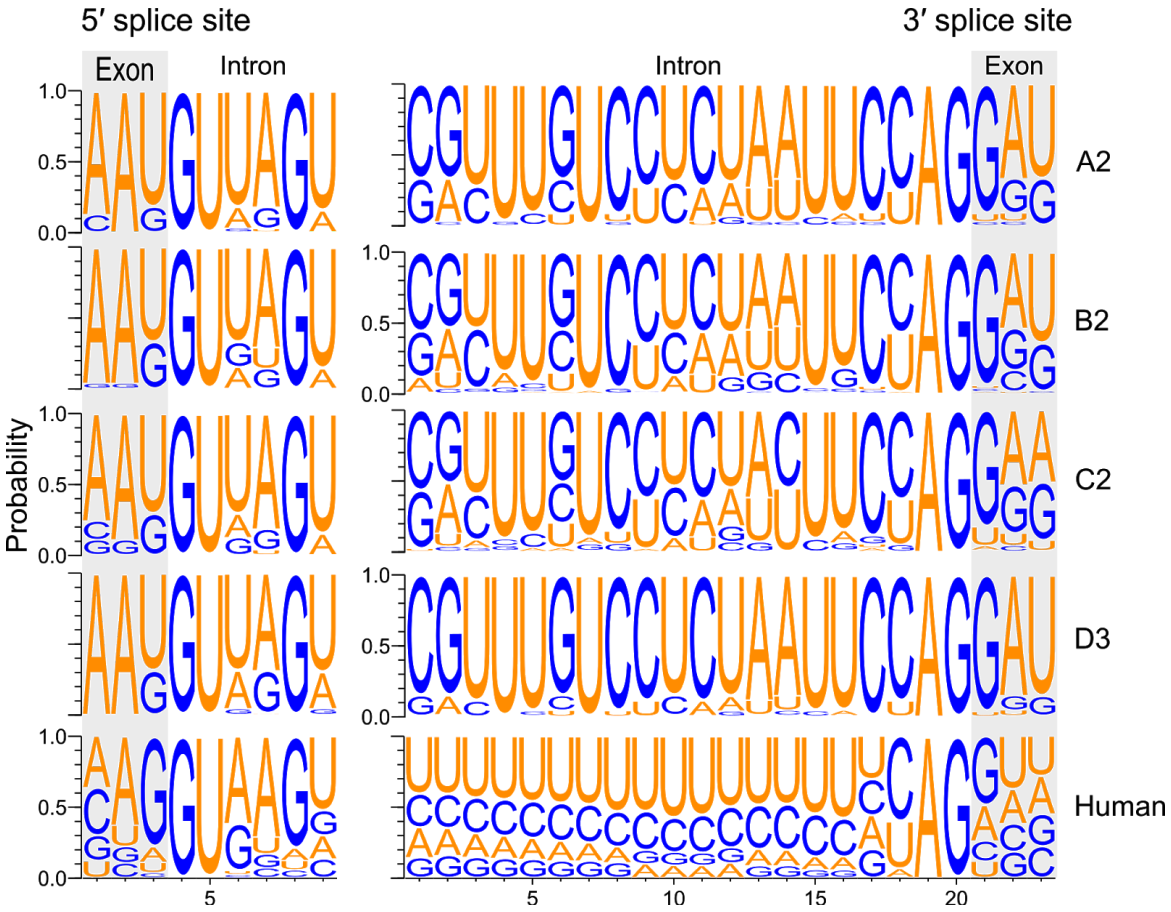
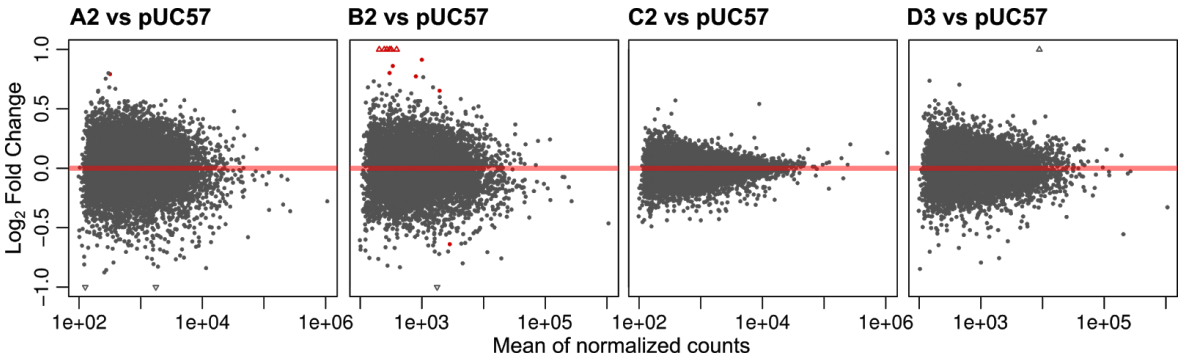
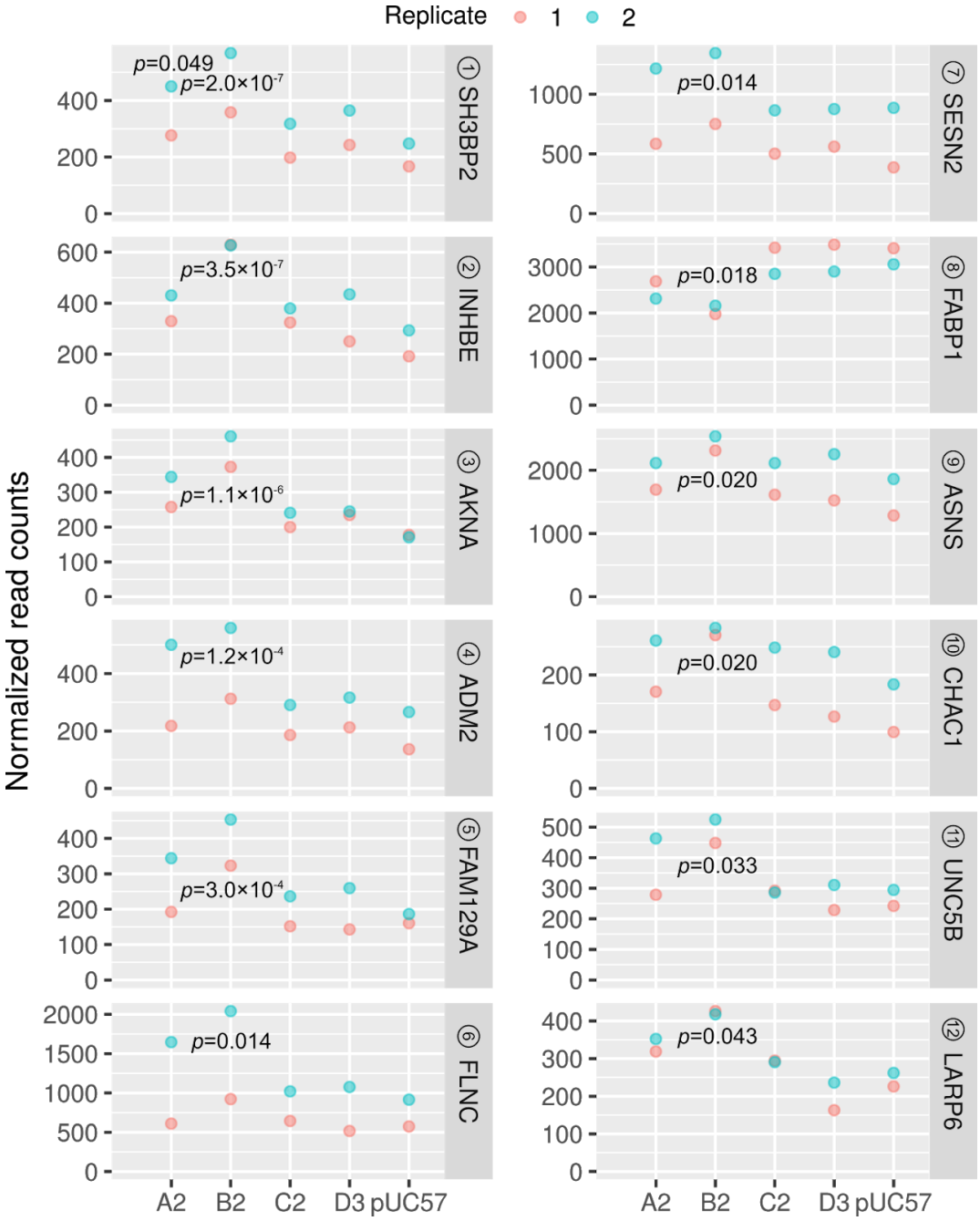


**Fig 5**



667 **Fig 6**




668

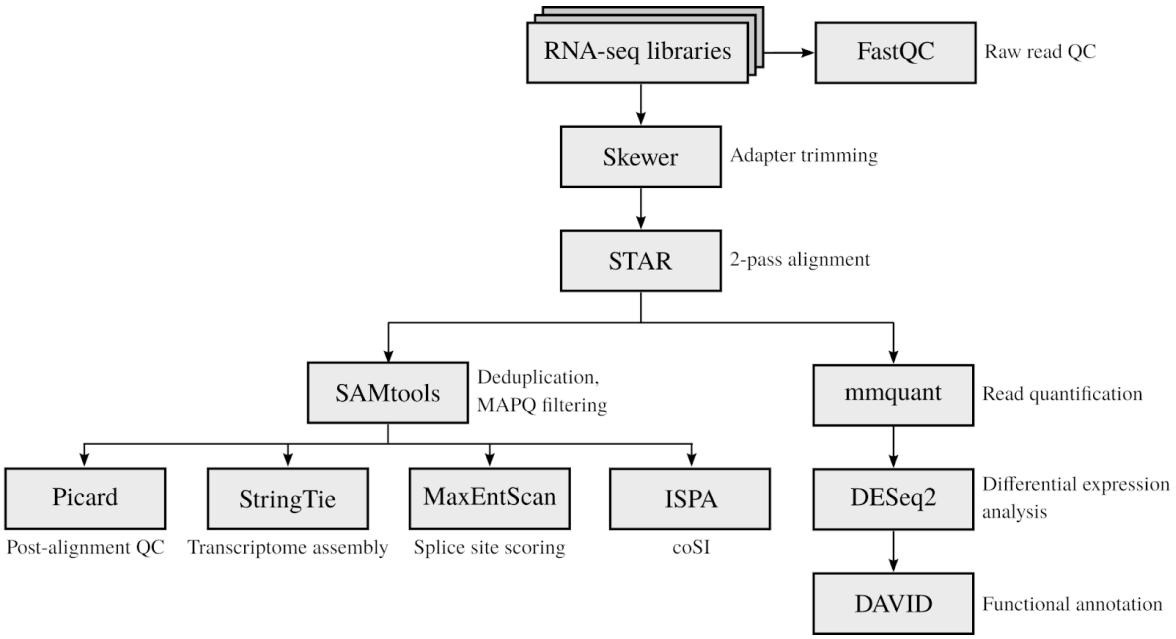
669

670

671

672 **Fig 7**

673




674

675