

Elucidating the Transcriptome of Turkey Hemorrhagic Enteritis Virus

3

Running Title: Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

⁵ Abraham Quaye^{1*}, Bret Pickett^{*}, Joel S. Griffitts^{*}, Bradford K. Berges^{*}, Brian D. Poole^{†*}

⁶*Department of Microbiology and Molecular Biology, Brigham Young University

7 1 First-author

⁸ † Corresponding Author

9 Corresponding Author Information

¹⁰ brian_poole@byu.edu

11 Department of Microbiology and Molecular Biology,

¹² 4007 Life Sciences Building (LSB),

¹³ Brigham Young University,

14 Provo, Utah

15

16 **ABSTRACT**

17 **Background:** Hemorrhagic enteritis (HE) is a disease affecting 6-12-week-old turkeys characterized by *im-*
18 *munosuppression (IS)* and bloody diarrhea. This disease is caused by *Turkey Hemorrhagic Enteritis Virus*
19 (*THEV*) of which avirulent strains (*THEV-A*) that do not cause HE but retain the immunosuppressive ability
20 have been isolated. The *THEV-A* Virginia Avirulent Strain (VAS) is still used as a live vaccine despite its
21 immunosuppressive properties. *Our objective is to understand the genetic basis by which VAS induces*
22 *IS.* The transcriptome of *THEV* was studied to set the stage for further experimentation with specific viral
23 genes that may mediate IS.

24 **Methods:** After infecting a turkey B-cell line (MDTC-RP19) with the VAS vaccine strain, samples in tripli-
25 cates were collected at 4-, 12-, 24-, and 72-hours post-infection. Total RNA was subsequently extracted,
26 and poly-A-tailed mRNA sequencing done. After trimming the raw sequencing reads with the FastQC, reads
27 were mapped to the *THEV* genome using Hisat2 and transcripts assembled with StringTie. An in-house
28 script was used to consolidate transcripts from all time-points, generating the final transcriptome. PCR, gel
29 electrophoresis, and Sanger sequencing were used to validate all identified splice junctions.

30 **Results and Conclusions:** A total of **18.1** million reads mapped to *THEV* genome providing good cover-
31 age/depth, leaving no regions unmapped. All predicted genes in the genome were represented. In keeping
32 with all adenoviruses, all transcripts were spliced with either with 5'- or 3'-multi exon UTRs hitherto un-
33 known. *Thirteen* novel exons were identified which were validated by PCR and Sanger sequencing. The
34 splicing patterns strongly suggest that there are *three* main promoters (E1, E3, and major late promoters)
35 driving expression of most of the genes with *two* possible minor promoters driving single genes (ORF7 and
36 ORF8). This RNA-sequencing experiment is the first study of *THEV* gene expression to date. In keeping
37 with other Adenoviruses, almost all *THEV* genes are spliced, and several genes are expressed as one tran-
38 scription unit under a single promoter. This insight into *THEV*'s transcriptome may allow the engineering of
39 the VAS to provide immune protection with less or no associated IS.

40 **INTRODUCTION**

41 Adenoviruses (AdVs) are non-enveloped icosahedral-shaped DNA viruses, causing infection in virtually all
42 vertebrates. Their double-stranded linear DNA genomes range between 26 and 45kb in size, producing a
43 broad repertoire of transcripts via a highly complex alternative splicing pattern (1, 2). The AdV genome is
44 one of the most optimally economized; both the forward and reverse DNA strands harbor protein-coding
45 genes, making it highly gene-dense. There are 16 genes termed “genus-common” that are homologous in
46 all AdVs; these are thought to be inherited from a common ancestor. All other genes are termed “genus-
47 specific”. “Genus-specific” genes tend to be located at the termini of the genome while “genus-common”
48 genes are usually central (1). This pattern is observed in *Adenoviridae*, *Poxviridae*, and *Herpesviridae* (1,
49 3, 4). The family *Adenoviridae* consists of five genera: *Mastadenovirus* (MAdV), *Aviadenovirus*, *Ataden-
50 ovirus*, *Ictadenovirus*, and *Siadenovirus* (SiAdV) (5, 6). Currently, there are three recognized members
51 of the genus SiAdV: frog adenovirus 1, raptor adenovirus 1, and turkey adenovirus 3 also called turkey
52 hemorrhagic enteritis virus (THEV) (5, 7–10). Members of SiAdV have the smallest genome size (~26 kb)
53 and gene content (~23 genes) of all known AdVs, and many “genus-specific” putative genes of unknown
54 functions have been annotated (see **Figure 1**) (1, 2, 7).

55 Virulent strains (THEV-V) and avirulent strains (THEV-A) of THEV are serologically indistinguishable, infect-
56 ing turkeys, chickens, and pheasants and the THEV-V cause different clinical diseases in these birds (2,
57 11). In turkeys, the THEV-V cause hemorrhagic enteritis (HE), a debilitating acute disease affecting pre-
58 dominantly 6-12-week-old turkeys characterized by immunosuppression (IS), weight loss, intestinal lesions
59 leading to bloody diarrhea, splenomegaly, and up to 80% mortality (11–13). HE is the most economically
60 significant disease caused by any strain of THEV (11). While the current vaccine strain (a THEV-A isolated
61 from a pheasant, Virginia Avirulent Strain [VAS]) have proven effective at preventing HE in young turkey
62 poulets, it still retains the immunosuppressive ability. Thus, vaccinated birds are rendered more susceptible
63 to opportunistic infections and death than unvaccinated cohorts leading to substantial economic losses (11,
64 14–16). The induced IS also interferes with vaccination schemes for other infections of turkeys (11, 14).
65 To eliminate this immunosuppressive side-effect of the vaccine, a thorough investigation of the culprit viral
66 factors (genes) mediating this phenomenon is essential. However, the transcriptome (splicing and gene ex-
67 pression patterns) of THEV has not been characterized, making the investigation of specific viral genes for
68 possible roles in causing IS impractical. A well-characterized transcriptome of THEV is required to enable
69 the next leap forward in THEV research - experimentation with specific viral genes that may mediate IS.

70 Myriads of studies have elucidated the AdV transcriptome in fine detail (17, 18). However, a large pre-

71 ponderance of studies focus on MAdVs - specifically human AdVs - thus, most of the current knowledge
72 regarding AdV gene expression and replication is based on MAdV studies, which is generalized for all other
73 AdVs (6, 19). MAdV genes are transcribed in a temporal manner; therefore, genes are categorized into five
74 early transcription units (E1A, E1B, E2, E3, and E4), two intermediate (IM) units (pIX and IVa2), and one
75 major late unit (MLTU), which generates five families of late mRNAs (L1-L5). An additional gene (UXP or U
76 exon) is located on the reverse strand. The early genes encode non-structural proteins such as enzymes or
77 host cell modulating proteins, primarily involved in DNA replication or providing the necessary intracellular
78 niche for optimal replication while late genes encode structural proteins. The immediate early gene E1A is
79 expressed first, followed by the delayed early genes, E1B, E2, E3 and E4. Then the intermediate early
80 genes, IVa2 and pIX are expressed followed by the late genes (6, 17, 18). MAdV makes an extensive use of
81 alternative RNA splicing to produce a very complex array of mRNAs; all but pIX mRNA undergo at least one
82 splicing event. The MLTU produces over 20 distinct splice variants all of which contain three non-coding
83 exons at the 5'-end (collectively known as the tripartite leader, TPL) (17, 18). There is also an alternate
84 5' three non-coding exons present in varying amounts on a subset of MLTU mRNAs (known as the x-, y-
85 and z-leaders). Lastly, there is the i-leader exon, which is infrequently included between the second and
86 third TPL exons, and codes for the i-leader protein (20). Thus, the MLTU produces a complex repertoire
87 of mRNA with diverse 5'-UTRs spliced onto different 3' coding exons which are grouped into five different
88 3'-end classes (L1-L5). Each transcription unit (TU) contains its own promoter driving the expression of all
89 the array of mRNA transcripts produced via alternative splicing of the genes encoded in the unit(6, 17, 18).
90 Almost all AdV mRNAs are generated by the excision of one or more introns and most of these introns are
91 located in the 5' or 3' UTRs of pre-mRNA. Thus the viral introns scarcely interrupt the open reading frames
92 (ORFs) (1, 18).

93 High throughput sequencing methods have facilitated the discovery of many novel transcribed regions and
94 splicing isoforms. It is also a very powerful tool to study alternative splicing under different conditions at
95 an unparalleled depth (18, 21). In this paper, a paired-end deep sequencing experiment was performed to
96 characterize for the first time, the transcriptome of THEV (VAS vaccine strain) during different phases of the
97 infection, yielding the first THEV splicing map. Our paired-end sequencing allowed for reading **149** bp long
98 high quality (mean Phred Score of 36) sequences from each end of cDNA fragments, which were mapped
99 to the genome of THEV. The generated data from our paired-end sequencing experiment should thus be
100 reliable.

101 **RESULTS**

102 **Overview of sequencing data and analysis pipeline outputs**

103 A previous study by Zeinab *et al* showed that almost all THEV transcripts were detectable beginning at
104 4 hours (22). Therefore, infected MDTC-RP19 cells were harvested at 4-, 12-, 24-, and 72-hours post-
105 infection(h.p.i) to ensure an amply wide time window to sample all transcripts. Our paired-end RNA se-
106 quencing (RNA-seq) experiment yielded an average of **107.1** million total reads of **149bp** in length per
107 time-point, which were simultaneously mapped to both the virus (THEV) and host (*M.gallipavo*) genomes
108 using the Hisat2 (23) alignment program. A total of **18.1** million reads from all time-points mapped to the
109 virus genome; this provided good coverage/depth, leaving no regions unmapped. The mapped reads to
110 the virus genome increased substantially from **432** reads at 4 h.p.i to **16.9** million reads at 72 h.p.i (**Table**
111 **1, Figure 2a**). From the mapped reads, we identified an overall total of **2,457** unique THEV splice junctions
112 from all time-points, with splice junctions from the later time-points being supported by significantly more
113 sequence reads than earlier time-points. For example all the **13** unique junctions at 4 h.p.i had less than
114 10 reads supporting each one, averaging a mere **2.8** reads/junction. Conversely, the **2374** unique junc-
115 tions at 72 h.p.i averaged **898.4** reads/junction, some junctions having coverage as high as **322,677** reads.
116 The substantial increases in splice junctions and mapping reads to the THEV genome over time denotes
117 an active infection, and correlates with our quantitative PCR (qPCR) assay quantifying the total number of
118 viral genome copies over time (**Figure 2b**). Using StringTie (23), an assembler of RNA-seq alignments
119 into potential transcripts, the mapped reads for each time-point were assembled into transcripts using the
120 genomic location of the predicted THEV ORFs as a guide. In the consolidated transcriptome, a composite
121 of all unredudant transcripts from all time points, we counted a total of **28** transcripts all of which are novel,
122 and using ~~3'-Rapid Amplification of cDNA Ends (3'RACE)~~ and other methods, we further identified #####
123 unique splice variants. Although some exons in some transcripts match the predicted ORFs exactly, most
124 of our identified exons are longer, spanning multiple predicted ORFs (**Figure 3**). The complete list of unique
125 splice junctions mapped to THEV's genome has been submitted to the National Center for Biotechnology
126 Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under **accession no. XXXXXX**.

127 **Changes in THEV splicing profile over time**

128 AdV gene expression occurs under exquisite temporal control, supervised by designated promoters for
129 each transcription unit (TU) or region. Each promoter typically produce one or few pre-mRNAs that undergo
130 alternative splicing to yield the manifold repertoire of complex transcripts characteristic of AdVs (17, 18).
131 To evaluate the activity of each promoter over time, Firstly, Ballgown, a program for statistical analysis of
132 assembled transcriptomes (24) was used to estimate and normalize expression levels of all transcripts for

133 each time point in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) units. Very few
134 unique splice junctions, reads, and transcripts were counted at 4 h.p.i; hence, this time point was excluded
135 in this analysis.

136 Individually, TRXPT_21 (DBP) - from the E2 region - was the most significantly expressed at 12 h.p.i,
137 comprising about **33.58%** of the total transcripts. Transcripts in the E3 and E4 regions also contributed
138 significant proportions, and noticeably, some MLP region transcripts. The later time points were dominated
139 by the MLP region; transcripts TRXPT_10 and TRXPT_14 were most abundantly expressed at 24 and 72
140 h.p.i, respectively, as expected (**Figure 4a**). When we performed analysis of the FPKM values of transcripts
141 per region we found a similar pattern: the E2 region was the most abundantly expressed at 12 h.p.i, after
142 which the MLP region assumes predominance (**Figure 4b**). Secondly, we estimated relative abundances
143 of all splice junctions for each time point using the raw reads. We counted as significantly expressed only
144 junctions with coverage of at least 1% of the total splice junction reads counted at the given time point.
145 At 12 h.p.i, **18** junctions meet the 1% threshold, and were comprised of predominantly early region (E1,
146 E2, E3, and E4) junctions, albeit the MLP was the single most preponderant region overall, constituting
147 **38.8%** of all the junctions reads. (**Table 2a**). The levels of the top most abundant junctions at 12 h.p.i were
148 maintained also 24 h.p.i as the most significantly expressed. However, here, the MLP-derived junctions
149 were unsurprisingly even more preponderant overall, accounting for **45.7%** of all the junction reads counted
150 (**Table 2b**). At 72 h.p.i, the trend of increased activity of the MLP continued as expected; at this time, the
151 MLP-derived junctions were not only the most abundant overall - accounting for **67.4%** of all junctions reads -
152 but also contained the most significantly expressed individual junctions (**Table 2c**. Also see **Supplementary**
153 **Tables 1a-c; Figure 4c**). When we limited this analysis to only junctions in the final transcriptome, the
154 relative abundances of the junctions for each region over time was generally similar to the pattern seen with
155 all the junctions included (**Figure 4d**).
156 Furthermore, we analyzed splice donor and acceptor site nucleotide usage over time to investigate any
157 peculiarities that THEV may show, generally or over the course of the infection. We found that most splice
158 donor-acceptor sequences were unsurprisingly the canonical GU-AG nucleotides.

159 **Early Region 1 (E1) transcripts.** This region in MAdVs is the first transcribed after successful entry of
160 the viral DNA into the host cell nucleus, albeit at low levels (18). The host transcription machinery solely
161 mediates the transcription of this region. After their translation, the E1 proteins in concert with a myriad of
162 host transcription factors activate the other viral promoters (6). Only two ORFs (ORF1 [sialidase] and Hyd)
163 are predicted in this region; however, we discovered **four** novel transcripts in this region, which collectively
164 contain **3** unique splice junctions (**Figure 5**). Most of the encoded proteins of the novel transcripts are

165 distinct from the predicted ORFs, although they all have the potential to encode the Hyd protein as the
166 3'-most coding sequence (CDS) if the first start codon (SSC) is skipped. The 5'-most CDS of TRXPT_1
167 is multi-exonic, producing a 17.9 kilodalton (kDa) protein of 160 residues [amino acids (aa)]. The CDS
168 begins in the first exon, starting at position 211, spans the second exon, and terminates in the third exon
169 at position 2312. From the 5'-most SSC, TRXPT_2 encodes the largest protein in this region — a 64.3
170 kDa, 580 aa protein with the same SSC as TRXPT_1 (position 211). This CDS spans almost the entire
171 predicted ORF1 and Hyd, coming short in two regards: it is spliced from 1655 to 1964 (ORF1's C-terminus,
172 including the stop codon), and its stop codon (STC; position 2312) is 13 bp short of the Hyd STC. However,
173 it has an SSC 102 bp upstream and in-frame with ORF1's predicted SSC. The N-terminus of TRXPT_2
174 CDS therefore, shares substantial protein sequence similarity with ORF1 but Hyd and TRXPT_2 are not
175 in-frame; hence no protein sequence similarity. TRXPT_3 is almost identical to TRXPT_1, except for the
176 lack of TRXPT_1's second exon. From our analysis, TRXPT_3 and TRXPT_4 seem to have transcription
177 start sites (TSS) downstream of the TSS of TRXPT_1 and TRXPT_2; however, given that studies in MAdVs
178 show that E1 mRNAs share a common 5' and 3' positions, only differing from each other regarding the
179 introns (18), it is likely that TRXPT_3 and TRXPT_4 are incomplete, and the TSS just like the transcription
180 termination site (TTS) are identical for all E1 transcripts. Regardless of the TSS considered for TRXPT_3,
181 the coding potential remains unaffected. Its 5'-most CDS, beginning at 1965 and sharing the same STC as
182 TRXPT_1 and TRXPT_2, produces a 13.1 kDa, 115 residue protein. This CDS (ORF4) was predicted in
183 an earlier study (25) but was excluded in later studies (1, 12); however, our data suggests it is a bona fide
184 ORF. The coding potential of TRXPT_4 is affected by the TSS considered; if we consider its unmodified
185 TSS, then its coding potential is the same as TRXPT_3 (ORF4 as the first CDS and Hyd as second CDS if
186 the first SSC is skipped). However, if we assume that TRXPT_4 shares the same TSS as TRXPT_1, then
187 the 5'-most CDS is a distinct multi-exonic 15.9 kDa, 143 aa protein with the same SSC as TRXPT_1 and
188 TRXPT_2 but with a unique STC. All splice junctions of the transcripts in this region (except the junction for
189 TRXPT_4) have been validated by cloning and Sanger sequencing of cDNA (**Figure 5b; supplementary**
190 **PCR methods**). Finally, during our validation of TRXPT_2, ORF1 was present on the agarose gel and
191 Sanger sequencing results as a bona fide transcript (**supplementary PCR methods**). This is corroborated
192 by 3'-RACE experiment, which shows transcripts spanning the entire ORF1 and Hyd ORFs without any
193 splicing with a poly-A tail at the TTS of transcripts in this region. The 5'-most CDS of this transcript would
194 encode ORF1. Given that the SSC of the predicted ORF1 is in-frame but downstream of TRXPT_2 SSC,
195 it suggests that the predicted ORF1 CDS is incomplete; it shares the same SSC and TSS, and TTS as
196 TRXPT_2, but has a unique STC.

197 **Early Region 2 (E2) and Intermediate Region (IM) transcripts.** The E2 TU expressed on the reverse
198 strand, is subdivided into E2A and E2B and encodes three classical AdV proteins: pTP and Ad-pol (E2B
199 proteins), and DBP (E2A protein) (17, 18). Unlike MAdV where two promoters (E2-early and E2-late) were
200 discovered (17), we discovered only a single promoter from which both E2A and E2B transcription is ini-
201 tiated. However, similar to MAdVs, E2A and E2B transcripts have distinct TTSs, and the E2B transcripts
202 share the TTS of the IVa2 transcript of the IM region (17, 18) (**Figure 6**). The E2A ORF, DBP is one of
203 three ORFs predicted to be spliced from two exons, with the CDS spanning both exons. The correspond-
204 ing transcript (TRXPT_21) found in our data matches this predicted splicing pattern exactly but with an
205 additional non-coding exon at the 5'-end (E2-5'UTR); hence, a three-exon transcript. The encoded protein
206 (DBP; 380 residues, 43.3 kDa) is identical. This transcript was also corroborated in a 3'-RACE experiment.
207 Additionally, from the 3'-RACE, a splice variant of TRXPT_21 which retains the second intron leading to a
208 2-exon transcript was found. This transcript (TRXPT_21B), albeit longer due to retaining the second intron
209 and possessing a short 3' UTR, encodes a truncated isoform of DBP because the first SSC utilized by
210 TRXPT_21, is followed shortly by STCs in the retained intron, and does not yield any viable product. The
211 SSC 173 bp downstream of TRXPT_21's SSC yields a 346 residue, 39.3 kDa product, which is in-frame
212 of DBP but wholly contained in the second exon starting at 18,013 bp and terminating just like DBP at
213 16,973 bp. TRXPT_21 and TRXPT_21B share a common TSS but TRXPT_21B as seen in our 3'-RACE
214 data, extends 39 bp into an adenine-thymine-rich sequence where the polyadenylation occurs. Hence, we
215 suggest that the two transcripts also share the same TTS at 16,934 bp (**Figure 6**).

216 The IM region is a single-transcript TU, encoding a single classical protein, IVa2. The promoter expressing
217 this single transcript (TRXPT_5) is embedded in E2B region and shares a TTS with E2B transcripts (17,
218 18). TRXPT_5 is a two-exon transcript located on the reverse strand spliced at 3447-3615. The first exon
219 is an untranslated region (UTR), except the last 2 nucleotides, which connect with the first nucleotide of
220 second exon to form the 5'-most SSC. This first SSC is 4 codons upstream and in-frame of the predicted
221 IVa2 SSC. Regardless of the SSC considered, the encoded protein (IVa2) is largely unaffected. Except
222 for the four extra residues at the N-terminus (considering the 5'-most SSC), the entire protein sequence is
223 identical.

224 The splice junction of TRXPT_5 add_trxpts from E2 were confirmed by cloning and Sanger sequencing of
225 cDNA (**supplementary PCR methods**).

226 **Early Region 3 (E3) transcripts.**

227 **Early Region 4 (E4) transcripts.** This transcription unit (TU) is the found at the tail-end (3'-end) of the
228 genome, on the reverse strand. Based on nucleotide position, ORF7 and ORF8 were predicted in this

229 region (1); however, as ORF7 is neither on the same strand as ORF8 nor transcribed from a promoter in
230 the E4 region, only ORF8 can legitimately be classified as a transcript in this TU. This is corroborated by
231 our RNA-seq data, as only one transcript was identified in this region on the reverse strand (**Figure 8**). The
232 transcript (TRXPT_28) spans 25192-26247 and is spliced at 25701-26055; hence, a two-exon transcript.
233 The second exon fully matches the predicted ORF8 with 12 extra base pairs at the 3'-end; however, the
234 encoded protein is an exact match. There is a SSC in the first exon at position 26246 (second nucleotide
235 of the transcript). The encoded protein from this SSC is in-frame with the SSC of ORF8 in the second
236 exon; hence, the C-terminus of this longer protein (26.4 kDa, 229 aa) would be identical to the predicted
237 ORF8 protein. ~~The splice junction of TRXPT_28 was validated by cloning and Sanger sequencing of cDNA~~
238 (**supplementary PCR methods**).

239 **Major Late Promoter Region (MLP) transcripts.**

240 **DISCUSSION/CONCLUSIONS**

241 In the original study where the ORFs of THEV were predicted, ORF4 was predicted in the E1 region span-
242 ning the Hyd gene. However, later studies predicted and preferred Hyd instead of ORF4; hence, the current
243 prediction map. However, this study shows that while both Hyd and ORF4 may be both expressed, ORF4 is
244 most likely the bona fide gene. For fig2a: There is a dramatic increase of mean coverage/depth from **2.42**
245 at 4 h.p.i to **95,042** at 72 h.p.i, strongly demonstrating an active infection. Unexpectedly, the pileup of reads
246 seems consistently skewed over similar regions of the genome. We could speculate that the temporal gene
247 expression regulation of THEV is different from MAdVs or this could simply mean that the infection was not
248 well synchronized. However, the relative proportions over these similar regions shows some variation over
249 time. For fig2b: titer reaching a plateau at 120 h.p.i, probably due to high cell death TRXPT_2 and ORF1
250 are isoforms Presumeably, if the junction reads were normalized, MLP would not be predominant at 12hpi.

251 **MATERIALS AND METHODS**

252 **Cell culture and THEV Infection**

253 The Turkey B-cell line (MDTC-RP19, ATCC CRL-8135) was grown as suspension cultures in 1:1 complete
254 Leibovitz's L-15/McCoy's 5A medium with 10% fetal bovine serum (FBS), 20% chicken serum (ChS), 5%
255 tryptose phosphate broth (TPB), and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Strepto-
256 mycin), at 41°C in a humidified atmosphere with 5% CO₂. Infected cells were maintained in 1:1 serum-
257 reduced Leibovitz's L15/McCoy's 5A media (SRLM) with 2.5% FBS, 5% ChS, 1.2% TPB, and 1% antibiotics
258 solution (100 U/mL Penicillin and 100ug/mL Streptomycin). A commercially available HE vaccine was pur-
259 chased from Hygieia Biological Labs as a source of THEV-A (VAS strain). The stock virus was titrated using
260 an in-house qPCR assay with titer expressed as genome copy number(GCN)/mL, similar to Mahshoub *et*
261 *al* (26) with modifications. Cells were infected at a multiplicity of infection (MOI) of 100 GCN/cell and sam-
262 ples in triplicates were harvested at 4-, 12-, 24-, and 72-h.p.i for RNA-seq. The infection was repeated but
263 samples in triplicates were harvested at 12-, 24-, 36-, 48-, and 72-h.p.i for PCR validation of novel splice
264 sites. Still one more independent infection was done at time points ranging from 12 to 168-h.p.i for qPCR
265 quantification of virus titers.

266 **RNA extraction and Sequencing**

267 Total RNA was extracted from infected cells using Thermofishers' RNAqueous™-4PCR Total RNA Isolation
268 Kit (#AM1914) per manufacturer's instructions. An agarose gel electrophoresis was performed to check
269 RNA integrity. The RNA quantity and purity was initially assessed using nanodrop, and RNA was used only
270 if the A260/A280 ratio was 2.0 ± 0.05 and the A260/A230 ratio was >2 and <2.2. Extracted total RNA sam-
271 ples were sent to LC Sciences, Houston TX for poly-A-tailed mRNA sequencing where RNA integrity was
272 checked with Agilent Technologies 2100 Bioanalyzer High Sensitivity DNA Chip and poly(A) RNA-
273 seq library was prepared following Illumina's TruSeq-stranded-mRNA sample preparation protocol.
274 Paired-end sequencing was performed on Illumina's NovaSeq 6000 sequencing system.

275 **Validation of Novel Splice Junctions**

276 All splice junctions identified in this work are novel except one predicted splice site each for pTP and DBP,
277 which were corroborated in our work. However, these predicted splice junctions had not been experimen-
278 tally validated hitherto, and we identified additional novel exons, giving the complete picture of these tran-

279 scripts. The novel splice junctions in this work discovered in the assembled transcripts using the StringTie
280 transcript assembler which we validated by PCR and Sanger Sequencing are shown in supplementary PCR
281 methods Table 1. Briefly, we designed primers that crossed a range of novel exon-exon boundaries for each
282 specific transcript in a transcription unit (TU) paired with their respective universal primers for the TU (~sup-
283 plementary PCR methods). Each forward primer contained a KpnI restriction site and reverse primers, an
284 XbaI site. After first-strand cDNA synthesis with SuperScript™ III First-Strand Synthesis System, these
285 primers were used in a targeted PCR amplification, the products analyzed with agarose gel electrophoresis
286 to confirm expected band sizes, cloned by traditional restriction enzyme method, and Sanger sequenced to
287 validate these splice junctions at the sequence level.

288 **3' Rapid Amplification of cDNA Ends (3'-RACE)**

289 We performed a rapid amplification of sequences from the 3' ends of mRNAs (3'-RACE) experiment using
290 a portion of the extracted total RNA of infected MDTC-RP19 cells used for the RNA-seq experiment as
291 explained above. We followed the protocol described by Green *et al* (27) with modifications. Briefly, 1ug
292 of total RNA was reverse transcribed to cDNA using SuperScript™ IV First-Strand Synthesis System fol-
293 lowing the manufacturing instructions using an adapter-primer with a 3'-end poly(T) and a 5'-end BamHI
294 restriction site. A gene-specific sense primer with a 5'-end KpnI restriction site paired with an anti-sense
295 adapter-primer with a 5'-end BamHI site were used to amplify target sections of the cDNA using Invitrogen's
296 Platinum™ Taq DNA polymerase High Fidelity, following manufacturer's instructions. The PCR amplicons
297 were restriction digested, cloned, and Sanger sequenced.

298 **Computational Analysis of RNA Sequencing Data: Mapping and Transcript characterization**

299 Analysis of our sequence reads were analyzed following a well established protocol described by Pertea
300 *et al* (23), using SNAKE MAKE 7.24.0 to drive the pipeline. Briefly, sequencing reads were trimmed with the
301 FastQC - version 0.11.9 (28) program to achieve an overall Mean Sequence Quality (Phred Score)
302 of 36. Trimmed reads were mapped to the complete sequence of avirulent turkey hemorrhagic enteritis
303 virus strain Virginia (<https://www.ncbi.nlm.nih.gov/nuccore/AY849321.1/>) and *Meleagris gallopavo* (<https://www.ncbi.nlm.nih.gov/genome/?term=Meleagris+gallopavo>) using Hisat2 - version 2.2.1 (23) with de-
304 fault settings without relying on known splice sites. The generated BAM files from each infection time-point
305 were filtered for reads mapping to the THEV genome and fed into StringTie - version 2.2.1 (23) using
306 a gff3 file from NCBI cont m. aining the predicted ORFs of THEV as a guide. A custom script was used

³⁰⁸ to consolidate all transcripts from all time-points without redundancy, generating the final transcriptome of
³⁰⁹ THEV.

³¹⁰ **SCRIPTS AND SUPPLEMENTARY MATERIALS**

³¹¹ **DATA AVAILABILITY**

³¹² **CODE AVAILABILITY**

- ³¹³ All the code/scripts written for analysis of the data are available on github (https://github.com/Abraham-Quaye/thev_transcriptome)
- ³¹⁴

315 **ACKNOWLEDGMENTS**

316 LC Sciences - RNA sequencing was done here Eton Bioscience, Inc, San Diego, CA - All Sanger se-
317 quencing validations was done here

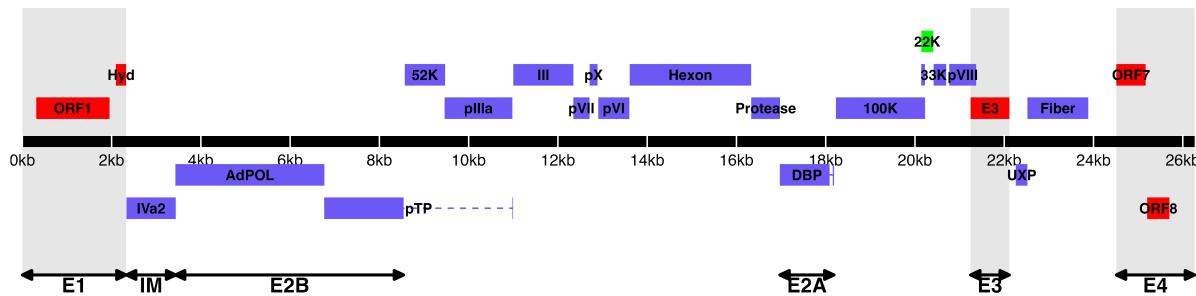
318 REFERENCES

- 319 1. Davison A, Benko M, Harrach B. 2003. Genetic content and evolution of adenoviruses. *The Journal*
320 of general virology
- 321 2. Harrach B. 2008. Adenoviruses: General features, p. 1–9. In Mahy, BWJ, Van Regenmortel, MHV
322 (eds.), *Encyclopedia of virology* (third edition). Book Section. Academic Press, Oxford.
- 323 3. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: Toward
324 defining the minimum essential poxvirus genome. *Journal of virology* 77:7590–7600.
- 325 4. McGeoch D, Davison AJ. 1999. Chapter 17 - the molecular evolutionary history of the herpesviruses,
326 p. 441–465. In Domingo, E, Webster, R, Holland, J (eds.), *Origin and evolution of viruses*. Book
Section. Academic Press, London.
- 327 5. Harrach B, Benko M, Both GW, Brown M, Davison AJ, Echavarría M, Hess M, Jones M, Kajon A,
Lehmkuhl HD, Mautner V, Mittal S, Wadell G. 2011. Family adenoviridae. *Virus Taxonomy: 9th*
328 Report of the International Committee on Taxonomy of Viruses 125–141.
- 329 6. Guimet D, Hearing P. 2016. 3 - adenovirus replication, p. 59–84. In Curiel, DT (ed.), *Adenoviral*
330 vectors for gene therapy (second edition). Book Section. Academic Press, San Diego.
- 331 7. Kovács ER, Benkő M. 2011. Complete sequence of raptor adenovirus 1 confirms the characteristic
332 genome organization of siadenoviruses. *Infection, Genetics and Evolution* 11:1058–1065.
- 333 8. Davison AJ, Wright KM, Harrach B. 2000. DNA sequence of frog adenovirus. *J Gen Virol* 81:2431–
334 2439.
- 335 9. Kovács ER, Jánoska M, Dán Á, Harrach B, Benkő M. 2010. Recognition and partial genome char-
acterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample
336 originating from parus major, a great tit. *Journal of Virological Methods* 163:262–268.
- 337 10. Katoh H, Ohya K, Kubo M, Murata K, Yanai T, Fukushi H. 2009. A novel budgerigar-adenovirus
338 belonging to group II avian adenovirus of siadenovirus. *Virus Research* 144:294–297.
- 339 11. Beach NM. 2006. Characterization of avirulent turkey hemorrhagic enteritis virus: A study of the
340 molecular basis for variation in virulence and the occurrence of persistent infection. Thesis.

- 341 12. Beach NM, Duncan RB, Larsen CT, Meng XJ, Sriranganathan N, Pierson FW. 2009. Comparison of
12 turkey hemorrhagic enteritis virus isolates allows prediction of genetic factors affecting virulence.
342 J Gen Virol 90:1978–85.
- 343 13. Gross WB, Moore WE. 1967. Hemorrhagic enteritis of turkeys. Avian Dis 11:296–307.
- 344
- 345 14. Rautenschlein S, Sharma JM. 2000. Immunopathogenesis of haemorrhagic enteritis virus (HEV) in
346 turkeys. Dev Comp Immunol 24:237–46.
- 347 15. Larsen CT, Domermuth CH, Sponenberg DP, Gross WB. 1985. Colibacillosis of turkeys exacerbated
348 by hemorrhagic enteritis virus. Laboratory studies. Avian Dis 29:729–32.
- 349 16. Dhama K, Gowthaman V, Karthik K, Tiwari R, Sachan S, Kumar MA, Palanivelu M, Malik YS, Singh
350 RK, Munir M. 2017. Haemorrhagic enteritis of turkeys – current knowledge. Veterinary Quarterly
37:31–42.
- 351 17. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. 2020. Deep splicing plasticity
352 of the human adenovirus type 5 transcriptome drives virus evolution. Communications Biology 3:124.
- 353 18. Zhao H, Chen M, Pettersson U. 2014. A new look at adenovirus splicing. Virology 456-457:329–341.
- 354
- 355 19. Wolfrum N, Greber UF. 2013. Adenovirus signalling in entry. Cell Microbiol 15:53–62.
- 356
- 357 20. Falvey E, Ziff E. 1983. Sequence arrangement and protein coding capacity of the adenovirus type 2
358 "i" leader. Journal of Virology 45:185–191.

- 359 21. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W,
Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. 2012. Landscape of transcription in human
360 cells. *Nature* 489:101–108.
- 361 22. Aboeza Z, Mabsoub H, El-Bagoury G, Pierson F. 2019. In vitro growth kinetics and gene expression
362 analysis of the turkey adenovirus 3, a siadenovirus. *Virus Research* 263:47–54.
- 363 23. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of
364 RNA-seq experiments with HISAT, StringTie and ballgown. *Nature Protocols* 11:1650–1667.
- 365 24. Jack Fu [Aut], Alyssa C. Frazee [Aut, Cre], LeonardoCollado-Torres [Aut], Andrew E. Jaffe [Aut],
366 Jeffrey T. Leek[Aut, Ths]. 2017. Ballgown. Bioconductor.
- 367 25. Pitcovski J, Mualem M, Rei-Koren Z, Krispel S, Shmueli E, Peretz Y, Gutter B, Gallili GE, Michael A,
Goldberg D. 1998. The complete DNA sequence and genome organization of the avian adenovirus,
368 hemorrhagic enteritis virus. *Virology* 249:307–315.
- 369 26. Mabsoub HM, Evans NP, Beach NM, Yuan L, Zimmerman K, Pierson FW. 2017. Real-time PCR-
370 based infectivity assay for the titration of turkey hemorrhagic enteritis virus, an adenovirus, in live
vaccines. *Journal of Virological Methods* 239:42–49.
- 371 27. Green MR, Sambrook J. 2019. Rapid amplification of sequences from the 3' ends of mRNAs: 3'-
372 RACE. *Cold Spring Harbor Protocols* 2019:pdb.prot095216.
- 373 28. 2015. FastQC.
374

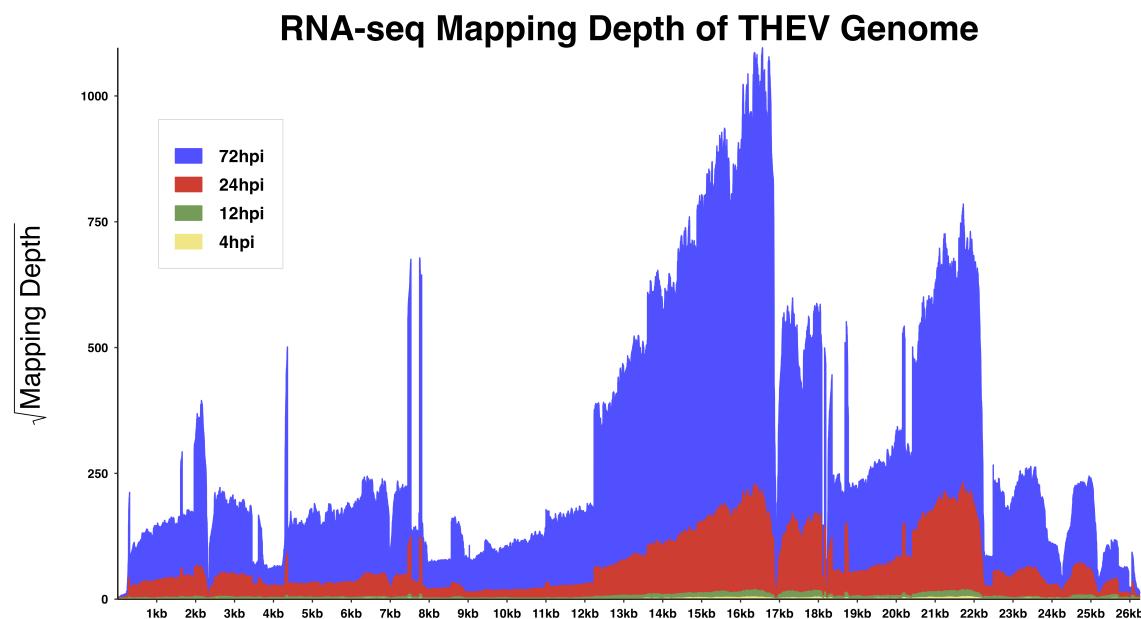
375 **TABLES AND FIGURES**



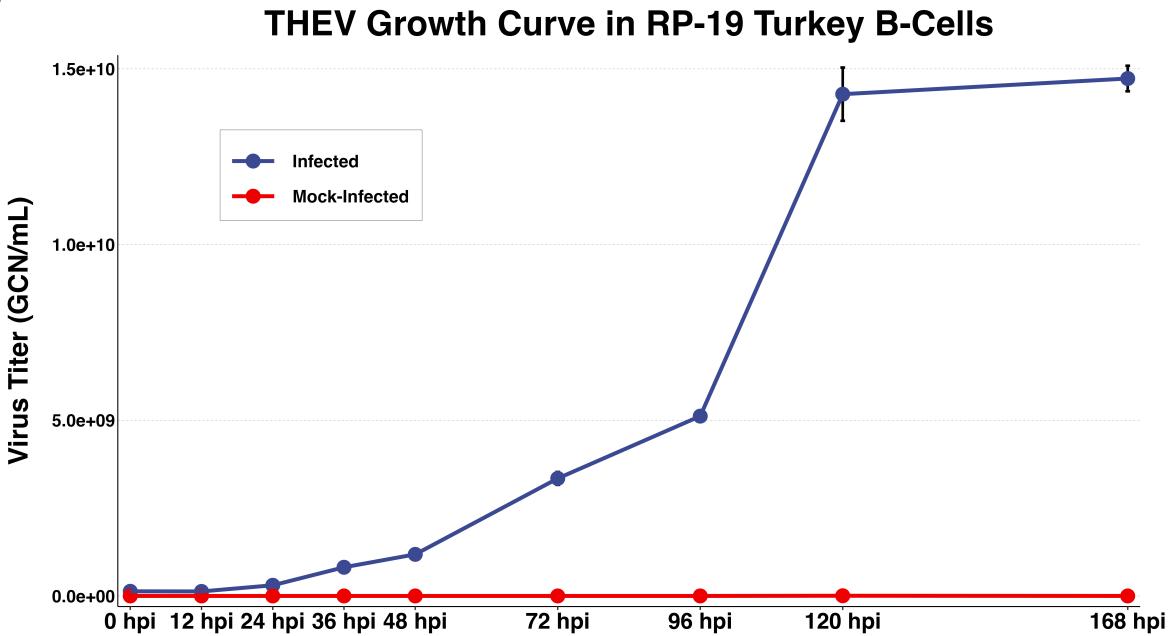
376

377 **Figure 1. Genomic map of THEV avirulent strain annotated ORFs.** The central horizontal line repre-
 378 presents the double-stranded DNA marked at 5kb intervals as white line breaks. Blocks represent viral genes.
 379 Blocks above the DNA line are transcribed rightward, those below are transcribed leftward. pTP, DBP
 380 and 33K predicted to be spliced are shown as having tails. Shaded regions indicate regions containing
 381 "genus-specific" genes (colored red). Genes colored in blue are "genus-common". Gene colored in light
 382 green is conserved in all but Atadenoviruses. The UXP (light blue) is an incomplete gene present in almost
 383 all AdVs. Regions comprising the different transcription units are labelled at the bottom (E1, E2A, E2B, E3,
 384 E4, and IM); the unlabeled regions comprise the MLTU.

A



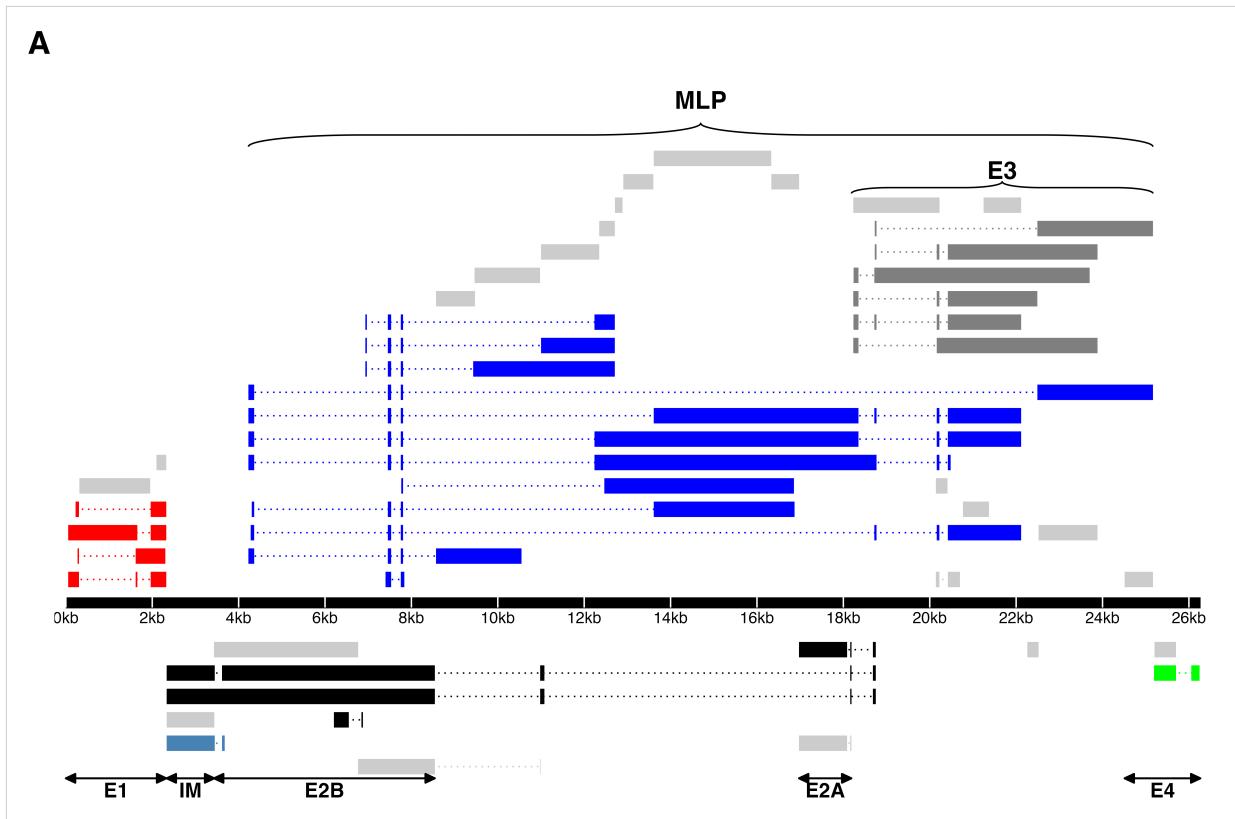
B

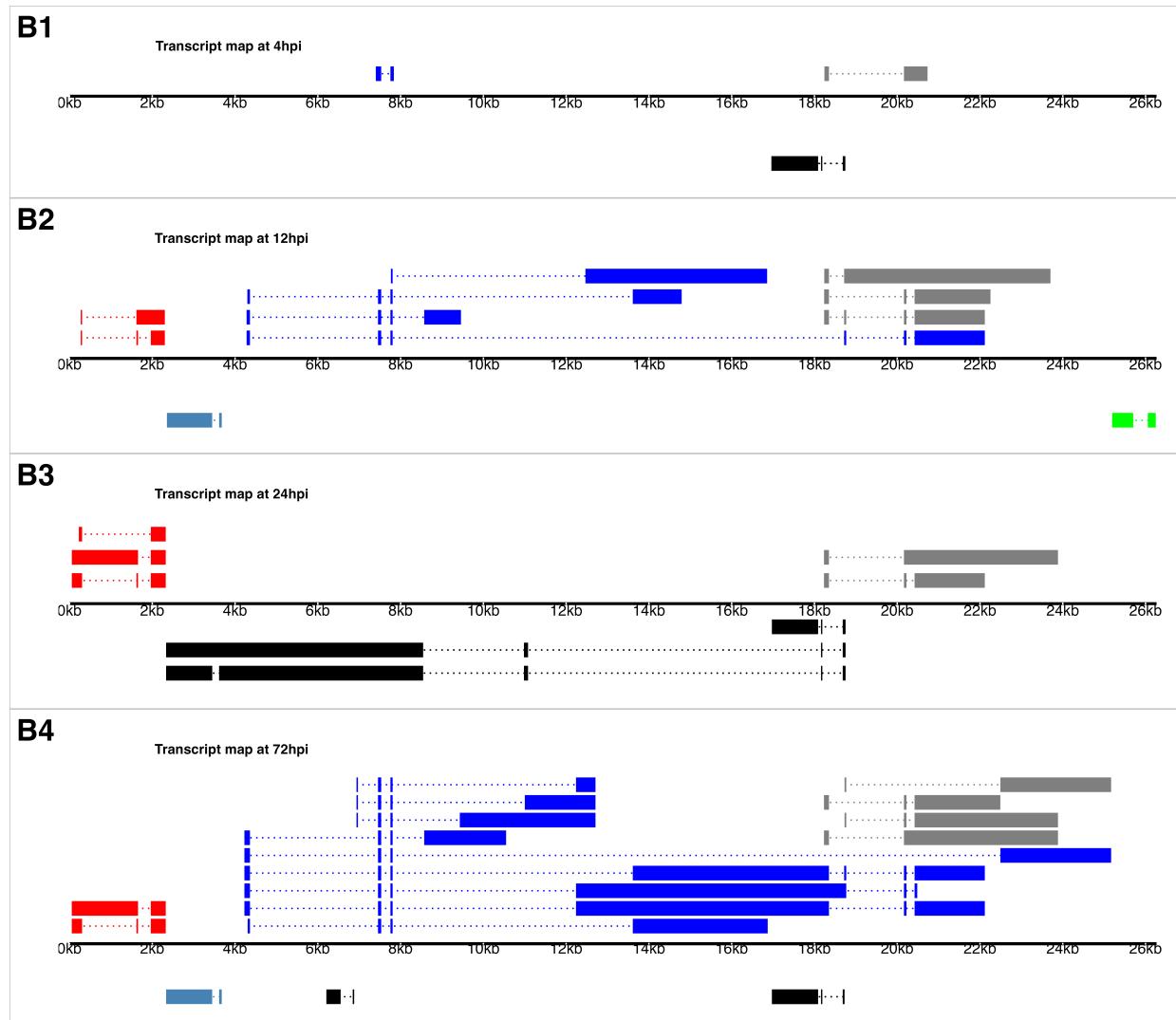


385

386 **Figure 2: Increasing levels of THEV over time. a) Per base coverage of sequence reads mapping**
 387 **to THEV genome by time point.** The pileup of mRNA reads mapping to THEV genome at the base-pair
 388 level for each indicated time point. b) **Growth curve of THEV (VAS vaccine strain) in MDTC-RP19 cell**
 389 **line.** Virus titers were quantified with a qPCR assay. There is no discernible increase in virus titer up 12
 390 h.p.i, after which there is a steady increase in virus titer is measured. The virus titer expands exponentially

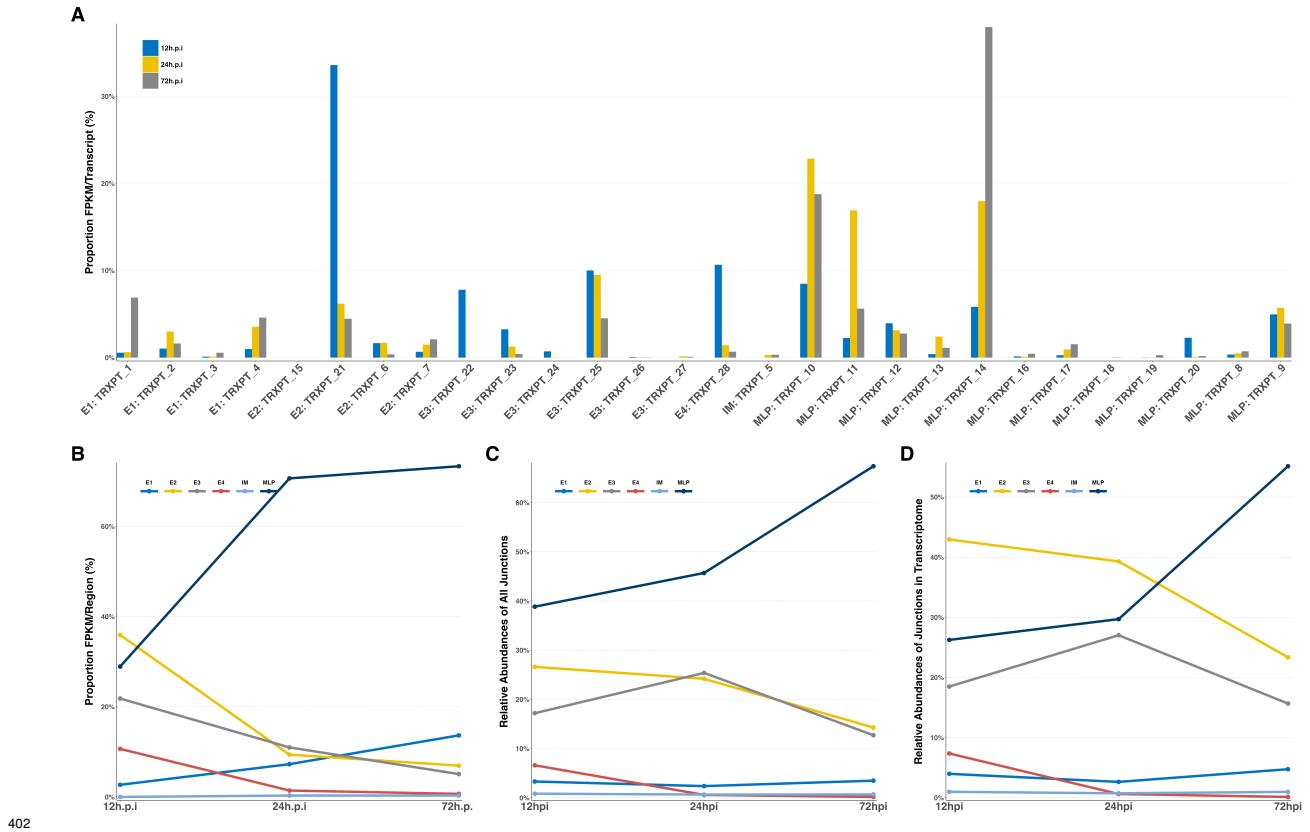
³⁹¹ beginning from 48 h.p.i, increasing by orders of magnitude before reaching a plateau at 120 h.p.i. GCN:
³⁹² genome copy number.



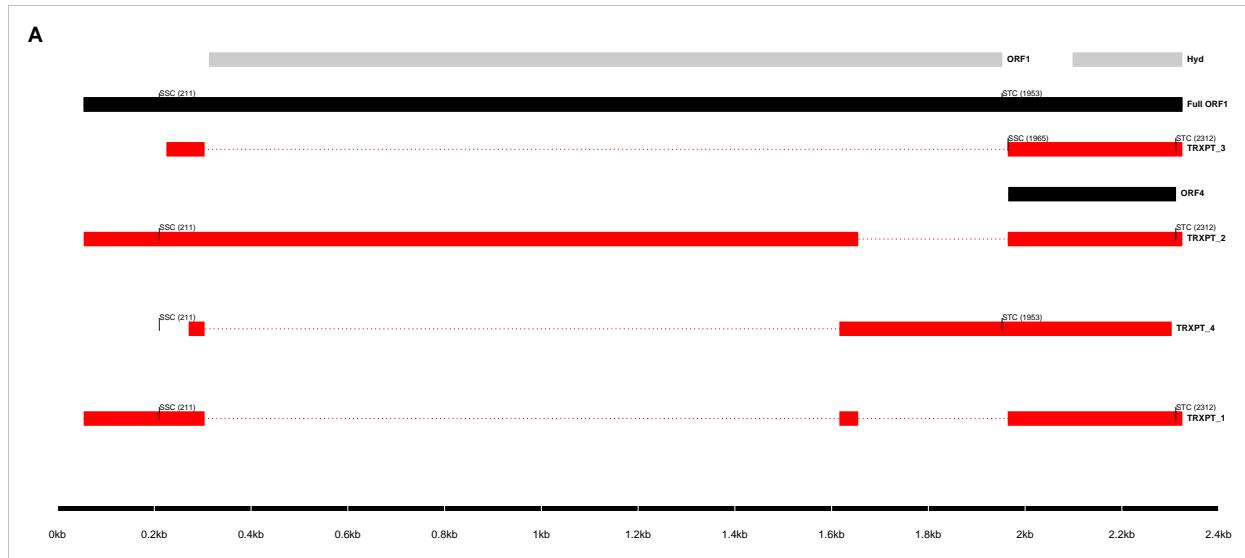


394

395 **a) Full transcriptome of THEV.** THEV transcripts assembled from all time points by StringTie
 396 are unified forming this final transcriptome (splicing map). Transcripts belonging to the same transcription
 397 unit (TU) are located in close proximity on the genome and are color coded and labeled in this figure as
 398 such. The organization of TUs in the THEV genome is unsurprisingly similar to MAdVs; however, the MAdV
 399 genome shows significantly more transcripts. The TUs are color coded: E1 transcripts - red, E2 - black, E3
 400 - dark grey, E4 - green, MLP - blue. Predicted ORFs are also indicated here, colored light grey. **b) THEV**
 401 **transcripts identified at given time points.** Transcripts are color coded as explained in **a**.



403 **Figure 4: Changes in splicing and expression profile of THEV over time.** **a)** Expression levels of
 404 transcripts over time. **b)** Expression levels of transcripts by region over time. **c)** Relative abundances of all
 405 splice junctions over time. **d)** Relative abundances of junctions in transcriptome.

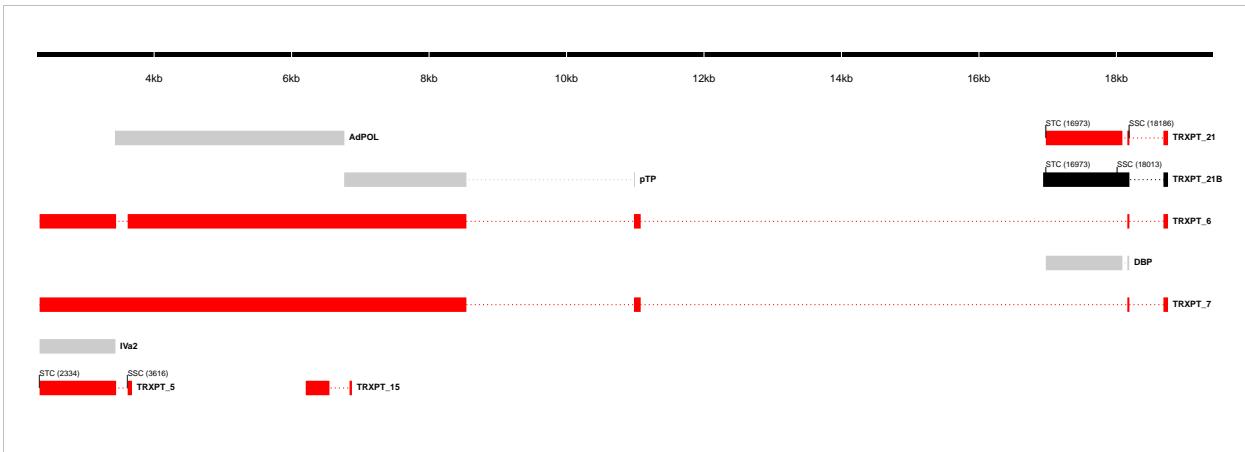


B

Transcript ID	Splice Junction					Strand	Junction Reads				Coding Potential	
	Start	End	Intron Length	Splice Donor-Acceptor			4h.p.i	12h.p.i	24h.p.i	72h.p.i		
TRXPT_1 TRXPT_4	304	1616	1313bp	GT-AG		+	0	9	1019	25041	Validated*	Hyd_iso_1, ORF4_novel
TRXPT_3	304	1964	1661bp	GT-AG		+	0	2	168	1588	Validated	Hyd_iso_2
TRXPT_2 TRXPT_1	1655	1964	310bp	GT-AG		+	0	9	1395	38491	Validated	ORF1_novel_iso, Hyd_iso_1

406 * Not validated for TRXPT_4

407 **Figure 5: E1 region transcripts. a)** The splice map of the E1 transcription unit. Exons are depicted as
 408 boxes connected by introns (dotted lines). Transcripts from RNA-seq data are colored red, predicted ORFs
 409 are colored grey, and the full ORF1 transcript and previously annotated ORF4 are colored black. Each
 410 transcript or ORF is labelled with its name to the right. The start codon (SSC) and stop codon (STC) of
 411 the 5'-most CDS of each transcript is indicated with the nucleotide position in brackets. The region of the
 412 virus is depicted at the bottom as a black line with labels of the nucleotide positions for reference **b)** The
 413 sequence reads covering the splice junctions are indicated with information about their validation status
 414 using cloning and Sanger sequencing.



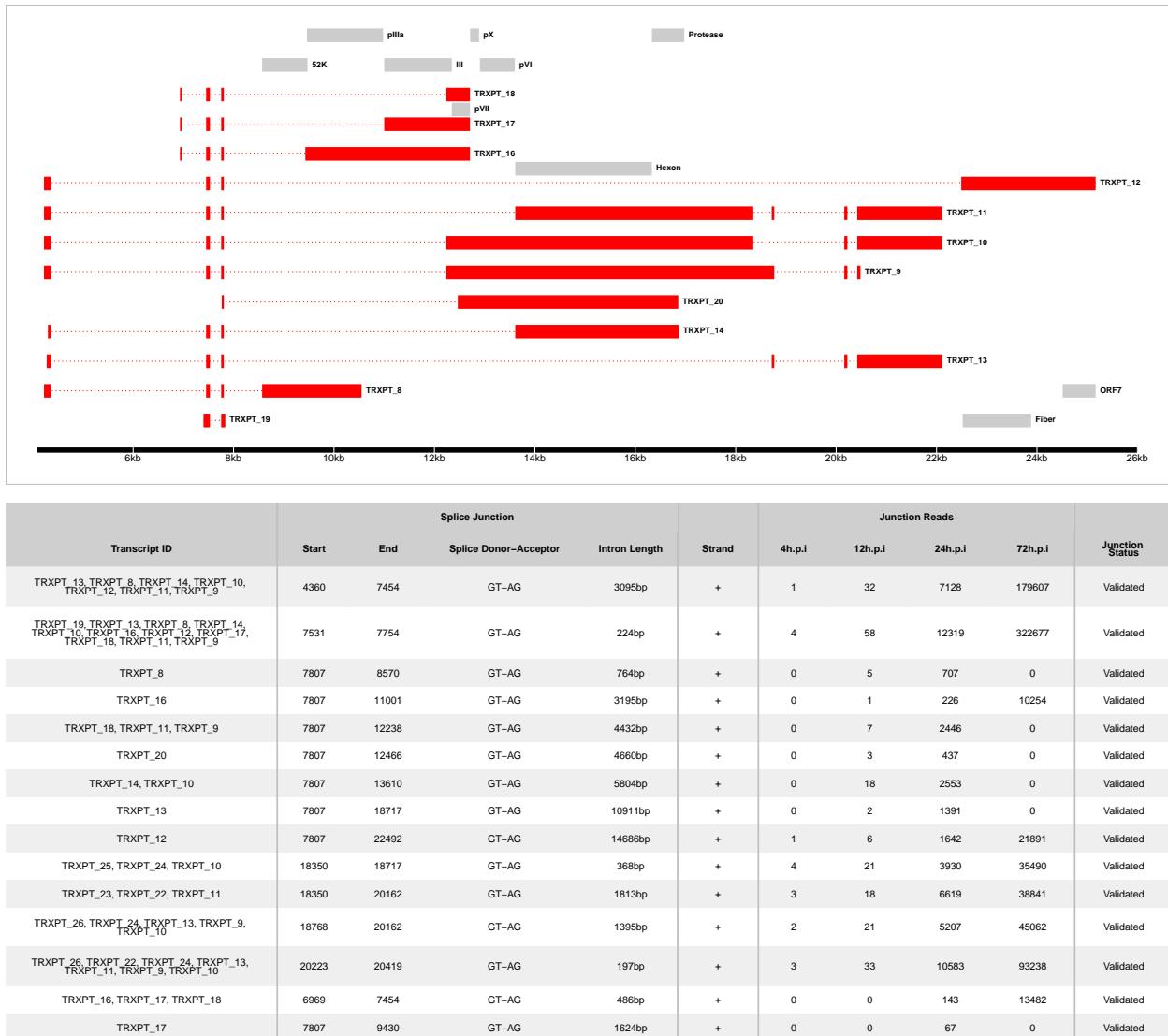
Transcript ID	Splice Junction					Junction Reads					Junction Status
	Start	End	Splice Donor-Acceptor	Intron Length	Strand	region	4h.p.i	12h.p.i	24h.p.i	72h.p.i	
TRXPT_5, TRXPT_7	3447	3615	GT-AG	169bp	-	IM, E2	1	5	720	13422	Validated
TRXPT_6, TRXPT_7	11079	18159	GT-AG	7081bp	-	E2	0	2	0	0	Validated
TRXPT_21	18087	18159	GT-AG	73bp	-	E2	9	103	0	0	Validated
TRXPT_21, TRXPT_6, TRXPT_7	18189	18684	CT-AC, GT-AG	496bp	-	E2	0	111	18794	156037	Validated
TRXPT_6, TRXPT_7	8543	10981	GT-AG	2439bp	-	E2	0	0	298	850	Validated
TRXPT_15	6551	6843	GT-GC	293bp	-	E2	0	0	0	6	Unvalidated*

415 *Incomplete transcript

416 **Figure 6: E2 and IM region transcripts.** **a)** The splice map of the E1 and IM transcription units. Exons
 417 are depicted as boxes connected by introns (dotted lines). Red transcripts are generated from RNA-seq
 418 data and predicted ORFs are colored grey. Each transcript or ORF is labelled with its name to the right.
 419 The start codon (SSC) and stop codon (STC) of the 5'-most CDS of each transcript is indicated with the
 420 nucleotide position in brackets. The region of the virus is depicted at the bottom as a black line with labels
 421 of the nucleotide positions for reference **b)** The sequence reads covering the splice junctions are indicated
 422 with information about their validation status using cloning and Sanger sequencing.



423 **Figure 8: E4 region transcripts. a)** The splice map of the E4 transcription unit. Exons are depicted
 424 as boxes connected by introns (dotted lines). The transcript from RNA-seq data is colored red and the
 425 predicted ORF, grey. The transcript and ORF are labelled with their names to the right. The start codon
 426 (SSC) and stop codon (STC) of the 5'-most CDS is indicated with the nucleotide position in brackets. The
 427 region of the virus is depicted at the bottom as a black line with labels of the nucleotide positions for
 428 reference **b)** The sequence reads covering the splice junction are indicated.
 429



430

431 **Figure 9: MLP region transcripts. a)**

Table 1: Table 1: Overview of sequencing results

Metric	4h.p.i	12h.p.i	24h.p.i	72h.p.i	Total
Total reads	1.17e+08	7.63e+07	1.20e+08	1.15e+08	4.28e+08
Mapped (Host)	1.04e+08	6.79e+07	1.06e+08	8.38e+07	3.62e+08
Mapped (THEV)	4.32e+02	6.70e+03	1.18e+06	1.69e+07	1.81e+07
Mean Per Base Coverage/Depth	2.42	37.71	6,666.96	95,041.7	101,749
Total unique splice junctions	13	37	236	2374	2,457
Junction coverage Total (at least 1 read)	37	605	115075	2132806	2.25e+06
Junction coverage Mean reads	2.8	16.4	487.6	898.4	351.3
Junction coverage (at least 10 reads)	0	13	132	1791	1,936
Junction coverage (at least 100 reads)	0	1	53	805	859
Junction coverage (at least 1000 reads)	0	0	18	168	186

Table 2: Table 2a: Most abundant splice junctions at 12h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice		Region	Reads	Intron Length	Reads_Percentage
					Acceptor-	Donor				
12hpi	-	18,087	18,159	GT-AG	T-A		E2	103	72 bp	103 (17%)
12hpi	+	18,189	18,684	CT-AC	T-A		MLP	97	495 bp	97 (16%)
12hpi	+	7,531	7,754	GT-AG	T-A		MLP	58	223 bp	58 (9.6%)
12hpi	-	25,701	26,055	GT-AG	T-A		E4	37	354 bp	37 (6.1%)
12hpi	+	20,223	20,419	GT-AG	T-A		E3	33	196 bp	33 (5.5%)
12hpi	+	4,360	7,454	GT-AG	T-A		MLP	32	3,094 bp	32 (5.3%)
12hpi	-	18,751	20,668	GT-AG	T-A		E2	22	1,917 bp	22 (3.6%)
12hpi	+	18,350	18,717	GT-AG	T-A		E3	21	367 bp	21 (3.5%)
12hpi	+	18,768	20,162	GT-AG	T-A		E3	21	1,394 bp	21 (3.5%)
12hpi	+	7,807	13,610	GT-AG	T-A		MLP	18	5,803 bp	18 (3%)
12hpi	+	18,350	20,162	GT-AG	T-A		E3	18	1,812 bp	18 (3%)
12hpi	-	18,189	18,684	GT-AG	T-A		E2	14	495 bp	14 (2.3%)
12hpi	-	18,751	21,682	GT-AG	T-A		E2	10	2,931 bp	10 (1.7%)
12hpi	+	304	1,616	GT-AG	T-A		E1	9	1,312 bp	9 (1.5%)
12hpi	+	1,655	1,964	GT-AG	T-A		E1	9	309 bp	9 (1.5%)
12hpi	-	18,087	18,163	GT-AG	T-A		E2	8	76 bp	8 (1.3%)
12hpi	+	7,807	12,238	GT-AG	T-A		MLP	7	4,431 bp	7 (1.2%)
12hpi	+	7,807	22,492	GT-AG	T-A		MLP	6	14,685 bp	6 (1%)

Table 3: Table 2b: Most abundant splice junctions at 24h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice		Region	Reads	Intron Length	Reads_Percentage
					Acceptor-	Donor				
24hpi	-	18,087	18,159	GT-AG	T-A		E2	18,825	72 bp	18,825 (16.4%)
24hpi	+	18,189	18,684	CT-AC	T-A		MLP	17,670	495 bp	17,670 (15.4%)
24hpi	+	7,531	7,754	GT-AG	T-A		MLP	12,319	223 bp	12,319 (10.7%)
24hpi	+	20,223	20,419	GT-AG	T-A		E3	10,583	196 bp	10,583 (9.2%)
24hpi	+	4,360	7,454	GT-AG	T-A		MLP	7,128	3,094 bp	7,128 (6.2%)
24hpi	+	18,350	20,162	GT-AG	T-A		E3	6,619	1,812 bp	6,619 (5.8%)
24hpi	+	18,768	20,162	GT-AG	T-A		E3	5,207	1,394 bp	5,207 (4.5%)
24hpi	+	18,350	18,717	GT-AG	T-A		E3	3,930	367 bp	3,930 (3.4%)
24hpi	-	18,751	20,668	GT-AG	T-A		E2	3,870	1,917 bp	3,870 (3.4%)
24hpi	+	7,807	13,610	GT-AG	T-A		MLP	2,553	5,803 bp	2,553 (2.2%)
24hpi	+	7,807	12,238	GT-AG	T-A		MLP	2,446	4,431 bp	2,446 (2.1%)
24hpi	+	7,807	22,492	GT-AG	T-A		MLP	1,642	14,685 bp	1,642 (1.4%)
24hpi	+	1,655	1,964	GT-AG	T-A		E1	1,395	309 bp	1,395 (1.2%)
24hpi	+	7,807	18,717	GT-AG	T-A		MLP	1,391	10,910 bp	1,391 (1.2%)
24hpi	-	18,189	18,684	GT-AG	T-A		E2	1,124	495 bp	1,124 (1%)
24hpi	-	18,751	21,128	GT-AG	T-A		E2	1,124	2,377 bp	1,124 (1%)
24hpi	+	20,223	20,894	GT-AG	T-A		E3	1,208	671 bp	1,208 (1%)

Table 4: Table 2c: Most abundant splice junctions at 72h.p.i

Timepoint	Strand	Start	End	Splice_Site	Splice		Region	Reads	Intron Length	Reads_Percentage
					Acceptor-	Donor				
72hpi	+	7,531	7,754	GT-AG	T-A		MLP	322,677	223 bp	322,677 (15.1%)
72hpi	+	4,360	7,454	GT-AG	T-A		MLP	179,607	3,094 bp	179,607 (8.4%)
72hpi	-	18,087	18,159	GT-AG	T-A	E2		161,336	72 bp	161,336 (7.6%)
72hpi	+	18,189	18,684	CT-AC	T-A		MLP	146,425	495 bp	146,425 (6.9%)
72hpi	+	20,223	20,419	GT-AG	T-A	E3		93,238	196 bp	93,238 (4.4%)
72hpi	+	7,807	13,610	GT-AG	T-A		MLP	81,420	5,803 bp	81,420 (3.8%)
72hpi	+	7,807	12,238	GT-AG	T-A		MLP	77,616	4,431 bp	77,616 (3.6%)
72hpi	+	18,768	20,162	GT-AG	T-A	E3		45,062	1,394 bp	45,062 (2.1%)
72hpi	+	1,655	1,964	GT-AG	T-A	E1		38,491	309 bp	38,491 (1.8%)
72hpi	+	18,350	20,162	GT-AG	T-A	E3		38,841	1,812 bp	38,841 (1.8%)
72hpi	+	18,350	18,717	GT-AG	T-A	E3		35,490	367 bp	35,490 (1.7%)
72hpi	+	304	1,616	GT-AG	T-A	E1		25,041	1,312 bp	25,041 (1.2%)
72hpi	-	18,751	20,668	GT-AG	T-A	E2		26,338	1,917 bp	26,338 (1.2%)
72hpi	+	7,807	12,904	GT-AG	T-A		MLP	21,946	5,097 bp	21,946 (1%)
72hpi	+	7,807	22,492	GT-AG	T-A		MLP	21,891	14,685 bp	21,891 (1%)