

1      Characterizing the Transcriptome of Turkey Hemorrhagic  
2      Enteritis Virus

3

4      **Running Title:** Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

5      Abraham Quaye<sup>1\*</sup>, Brett Pickett<sup>\*</sup>, Joel S. Griffitts<sup>\*</sup>, Bradford K. Berges<sup>\*</sup>, Brian D. Poole<sup>†\*</sup>

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

7      <sup>1</sup>First-author

8      <sup>†</sup> 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 Hemorrhagic enteritis (HE) is a disease affecting 6-12-week-old turkeys characterized by immunosuppres-  
18 sion (IS) and bloody diarrhea. This disease is caused by *Turkey Hemorrhagic Enteritis Virus* (THEV) of  
19 which avirulent strains (THEV-A) that do not cause HE but retain the immunosuppressive ability have been  
20 isolated. The THEV-A Virginia Avirulent Strain (VAS) is still used as a live vaccine despite its immuno-  
21 suppressive properties. We have performed the first RNA-sequencing experiment characterizing THEV's  
22 transcriptome, yielding the most detailed insight into THEV gene expression, to set the stage for further  
23 experimentation with specific viral genes that may mediate IS. After infecting a turkey B-cell line (MDTC-  
24 RP19) with the VAS vaccine strain, samples in triplicates were collected at 4-, 12-, 24-, and 72-hours  
25 post-infection. Total RNA was subsequently extracted, and poly-A-tailed mRNA sequencing done. After  
26 trimming the raw sequencing reads with the Trim-galore, reads were mapped to the THEV genome using  
27 Hisat2 and transcripts assembled with StringTie. We identified 29 transcripts from our RNA-seq data all  
28 of which consisted of novel exons albeit some exons matched the predicted ORFs. The three predicted  
29 splice junctions were also corroborated in our data. We performed PCR amplification of THEV cDNA,  
30 cloned the PCR products, and Sanger sequencing was used to validate all identified splice junctions. Dur-  
31 ing validation, we identified 5 additional transcripts some of which were further validated by 3'RACE data.  
32 Thus, the transcriptome of THEV consists of 34 unique transcripts with the coding capacity for all predicted  
33 ORFs. However, we found 8 predicted ORFs to be incomplete as either an upstream, in-frame start codon  
34 was identified or additional coding exons were found, making the actual expressed versions of these ORFs  
35 longer. We also identified 7 novel unpredicted ORFs that could be encoded by some transcripts; albeit it  
36 is beyond the scope of this manuscript to investigate whether they are indeed expressed. In keeping with  
37 all Adenoviruses, our data shows that all THEV transcripts are spliced, and organized in transcription units  
38 under the control of their cognate promoter.

39 **INTRODUCTION**

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

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

68 Myriads of studies have elucidated the AdV transcriptome in fine detail (17, 18). However, a large pre-  
69 ponderance of studies focus on MAdVs – specifically human AdVs. Thus, most of the current knowledge

70 regarding AdV gene expression and replication is based on MAdV studies, which is generalized for all other  
71 AdVs (6, 19). MAdV genes are transcribed in a temporal manner; therefore, genes are categorized into five  
72 early transcription units (E1A, E1B, E2, E3, and E4), two intermediate (IM) units (pIX and IVa2), and one  
73 major late unit (MLTU or major late promoter [MLP] region), which generates five families of late mRNAs  
74 (L1-L5) based on the polyadenylation site. An additional gene (UXP or U exon) is located on the reverse  
75 strand. The early genes encode non-structural proteins such as enzymes or host cell modulating proteins,  
76 primarily involved in DNA replication or providing the necessary intracellular niche for optimal replication  
77 while late genes encode structural proteins that act as capsid proteins, promote virion assembly, and direct  
78 genome packaging. The immediate early gene E1A is expressed first, followed by the delayed early  
79 genes, E1B, E2, E3 and E4. Then the intermediate early genes, IVa2 and pIX are expressed followed by  
80 the late genes (6, 17, 18). Noteworthily, the MLP shows basal transcriptional activity during early infection  
81 (before DNA replication), with a comparable efficiency to other early viral promoters, but reaches its max-  
82 imal activity during late infection (after DNA replication). However, during early infection the repertoire of  
83 late transcripts from the MLP is restricted until late infection (6). MAdV makes an extensive use of alterna-  
84 tive RNA splicing to produce a very complex array of mRNAs. All but the pIX mRNA undergo at least one  
85 splicing event. For instance, the MLTU produces over 20 distinct splice variants all of which contain three  
86 non-coding exons at the 5'-end (collectively known as the tripartite leader, TPL) (17, 18). There is also  
87 an alternate 5' three non-coding exons present in varying amounts on a subset of MLTU mRNAs (known  
88 as the x-, y- and z-leaders). Lastly, there is the i-leader exon, which is infrequently included between the  
89 second and third TPL exons, and codes for the i-leader protein (20). Thus, the MLTU produces a complex  
90 repertoire of mRNA with diverse 5' untranslated regions (UTRs) spliced onto different 3' coding exons which  
91 are grouped into five different 3'-end classes (L1-L5) based on polyadenylation site. Each transcription unit  
92 (TU) contains its own promoter driving the expression of all the array of mRNA transcripts produced via  
93 alternative splicing in the unit (6, 17, 18). The promoters are activated at different phases of the infection by  
94 proteins from previously activated TUs. Paradoxically, the early-to-late phase transition during infection re-  
95 quires the L4 genes, 22K and 33K, which should only be available after the transition. However, a promoter  
96 in the L4 region (L4P) that directs the expression of these two proteins independent of the MLP was found,  
97 resolving the paradox (6, 17, 21). During translation of AdV mRNA, recent studies strongly suggest the  
98 potential usage of secondary start codons; adding to what was already a highly complex system for gene  
99 expression (17, 22).

100 High throughput sequencing methods have facilitated the discovery of many novel transcribed regions and  
101 splicing isoforms. It is also a very powerful tool to study alternative splicing under different conditions at an

102 unparalleled depth [(23); (18); Westergren2021]. In this paper, a paired-end deep sequencing experiment  
103 was performed to characterize for the first time the transcriptome of THEV (VAS vaccine strain) during  
104 different phases of the infection, yielding the first THEV splicing map. Our paired-end sequencing allowed  
105 for reading **149** bp long high quality (mean Phred Score of 36) sequences from each end of cDNA fragments,  
106 which were mapped to the genome of THEV.

107 **RESULTS**

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

109 A previous study by Aboeza *et al* showed that almost all THEV transcripts were detectable beginning at  
110 4 hours (24). Therefore, infected MDTC-RP19 cells were harvested at 4-, 12-, 24-, and 72-hours post-  
111 infection(h.p.i) to ensure an amply wide time window to sample all transcripts. Our paired-end RNA se-  
112 quencing (RNA-seq) experiment yielded an average of **107.1** million total reads of **149bp** in length per  
113 time-point, which were simultaneously mapped to both the virus (THEV) and host (*Meleagris gallopavo*)  
114 genomes using the Hisat2 (25) alignment program. A total of **18.1** million reads from all time-points mapped  
115 to the virus genome; this provided good coverage/depth, leaving no regions unmapped. The mapped reads  
116 to the virus genome increased substantially from **432** reads at 4 h.p.i to **16.9** million reads at 72 h.p.i (**Table**  
117 **1, Figure 2a**). From the mapped reads, we identified a total of **2,457** unique THEV splice junctions from all  
118 time-points, with splice junctions from the later time-points being supported by significantly more sequence  
119 reads than earlier time-points. For example all the **13** unique junctions at 4 h.p.i had less than 10 reads  
120 supporting each one, averaging a mere **2.8** reads/junction. Conversely, the **2374** unique junctions at 72 h.p.i  
121 averaged **898.4** reads/junction, some junctions having coverage as high as **322,677** reads. The substantial  
122 increases in splice junction and mapping reads to the THEV genome over time denotes an active infection,  
123 and correlates with our quantitative PCR (qPCR) assay quantifying the total number of viral genome copies  
124 over time (**Figure 2b**).

125 Using StringTie (25), an assembler of RNA-seq alignments into potential transcripts, the mapped reads for  
126 each time point were assembled into transcripts using the genomic location of the predicted THEV ORFs as  
127 a guide. In the consolidated transcriptome, a composite of all unredudant transcripts from all time points,  
128 we counted a total of **30** novel transcripts. Although some exons in some transcripts match the predicted  
129 ORFs exactly, most of our identified exons are longer, spanning multiple predicted ORFs (**Figure 3**).

130 We validated the splice junctions in all transcripts by PCR amplification of viral cDNA, cloning, and Sanger  
131 sequencing (**Supplementary PCR methods**). During validation, we identified 5 additional transcripts some  
132 of which were further validated by 3' Rapid Amplification of cDNA Ends (3'RACE) data. The complete  
133 list of unique splice junctions mapped to THEV's genome has been submitted to the National Center for  
134 Biotechnology Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession  
135 number GSE254416.

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

137 AdV gene expression occurs under exquisite temporal control with each promoter typically producing one or

138 few pre-mRNAs that undergo alternative splicing to yield the manifold repertoire of complex transcripts. To  
139 evaluate the activity of each promoter over time, *StringTie* and *Ballgown* (a program for statistical analysis  
140 of assembled transcriptomes) (26) were used to estimate the normalized expression levels of all transcripts  
141 for each time point in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) units. Very few  
142 unique splice junctions, reads, and transcripts were counted at 4 h.p.i; hence, this time point was excluded  
143 in this analysis.

144 Individually, TRXPT\_21 (DBP) – from the E2 region – was the most significantly expressed at 12 h.p.i.,  
145 comprising about **33.58%** of the total transcripts. Transcripts in the E3 and E4 regions also contributed sig-  
146 nificant proportions, and noticeably, some MLP region transcripts. The later time points were dominated by  
147 the MLP region transcripts – TRXPT\_10 and TRXPT\_14 were the most abundantly expressed at 24 and 72  
148 h.p.i, respectively, as expected (**Figure 4a**). When we performed analysis of the FPKM values of transcripts  
149 per region we found a similar pattern: the E2 region was the most abundantly expressed at 12 h.p.i, after  
150 which the MLP region assumes predominance (**Figure 4b**). Secondly, we estimated relative abundances  
151 of all splice junctions at each time point using the raw reads. For individual junctions, we counted as signif-  
152 icantly expressed only junctions with coverage of at least 1% of the total splice junction reads at the given  
153 time point. At 12 h.p.i, **18** junctions meet the 1% threshold, and were comprised of predominantly early  
154 region (E1, E2, E3, and E4) junctions, albeit the MLTU was the single most preponderant region overall,  
155 constituting **38.8%** of all the junction reads (**Table 2a** and **Supplementary Table 1a**). The top most abun-  
156 dant junctions at 12 h.p.i remained the most significantly expressed at 24 h.p.i also. However, here, the  
157 MLP-derived junctions were unsurprisingly even more preponderant overall, accounting for **45.7%** of all the  
158 junction reads counted (**Table 2b** and **Supplementary Table 1b**). At 72 h.p.i, the trend of increased activity  
159 of the MLP continued as expected; at this time, the MLP region junctions were not only the most abundant  
160 overall – accounting for **67.4%** of all junction reads, – but also contained the most significantly expressed  
161 individual junctions (**Table 2c**, **Supplementary Table 1c** and **Figure 4c**). When we limited this analysis to  
162 only junctions in the final transcriptome, the relative abundances of the junctions for each region over time  
163 was generally similar to the pattern seen with all the junctions included (**Figure 4d**).

164 We also analyzed splice donor and acceptor site nucleotide usage over time to investigate any peculiarities  
165 that THEV may show, generally or over the course of the infection. We found that most splice donor-  
166 acceptor sequences were unsurprisingly the canonical GT-AG nucleotides. However, the splice acceptor-  
167 donor pairing became less specific over time, such that all combinations of nucleotide pairs were eventually  
168 detected (**Figure 5**)

169 **Early Region 1 (E1) transcripts**

170 This region in MAdVs is the first transcribed after successful entry of the viral DNA into the host cell nucleus,  
171 albeit at low levels (18). The host transcription machinery solely mediates the transcription of this region.  
172 After their translation, the E1 proteins in concert with a myriad of host transcription factors activate the other  
173 viral promoters (6).

174 Only two ORFs (ORF1 [sialidase] and Hyd) are predicted in this region; however, we discovered **four** novel  
175 transcripts in this region, which collectively contain **3** unique splice junctions (**Figure 6**). Most of the ORFs  
176 of the novel transcripts are distinct from the predicted ORFs, but they all have the coding potential (CP)  
177 for the predicted Hyd protein as the 3'-most coding sequence (CDS) if secondary start codon usage is  
178 considered as reported for other AdVs (17, 18). The 5'-most CDS of TRXPT\_1 is multi-exonic, encoding  
179 a novel 17.9 kilodalton (kDa), 160 residue [amino acids (aa)] protein (ORF9). From its 5'-most start codon  
180 (SSC), TRXPT\_2 encodes the largest protein in this region – a 64.3 kDa, 580 aa protein (ORF10) with the  
181 same SSC as ORF9 (position 211bp). ORF10 spans almost the entire predicted ORF1 and Hyd, coming  
182 short in two regards: it is spliced from 1655bp to 1964bp (ORF1's C-terminus, including the stop codon), and  
183 it's stop codon (STC; position 2312) is 13 bp short of Hyd's STC. However, it has an SSC 102 bp upstream  
184 and in-frame with ORF1's predicted SSC. Thus, ORF10 shares substantial protein sequence similarity with  
185 ORF1 but not with Hyd, as the SSC of Hyd is not in-frame. Without its splice site removing the ORF1 STC,  
186 TRXPT\_2 would encode a longer variant of ORF1, starting from an upstream SSC. TRXPT\_3 is almost  
187 identical to TRXPT\_1, except for the lack of TRXPT\_1's second exon. Our RNA-seq data shows that all E1  
188 transcripts share the same transcription termination site (TTS; at position 2325bp). However, TRXPT\_3 and  
189 TRXPT\_4 seem to have transcription start sites (TSS) downstream of the TSS of TRXPT\_1 and TRXPT\_2  
190 (E1 TSS; position: 54bp). Given that studies in MAdVs show that E1 mRNAs share not only a common  
191 TTS but also the TSS, and only differ from each other regarding the internal splicing (18), it is likely that  
192 TRXPT\_3 and TRXPT\_4 are incomplete, and their actual TSS just like the TTS are identical for all E1  
193 transcripts. Regardless of the TSS considered for TRXPT\_3, the coding potential (CP) remains unaffected.  
194 Its 5'-most CDS, beginning at 1965bp and sharing the same STC as ORF9, produces a 13.1 kDa, 115  
195 residue protein (ORF4). ORF4 was predicted in an earlier study (27) but was excluded in later studies (1,  
196 12); however, our data suggests it is a bona fide ORF. Unlike TRXPT\_3, the CP of TRXPT\_4 is affected by  
197 the TSS considered; if we consider its unmodified TSS, then its CP is the same as TRXPT\_3 (ORF4 as the  
198 first CDS and Hyd as second CDS using the secondary SSC). However, if we assume that TRXPT\_4 uses  
199 the E1 TSS, then the 5'-most CDS is a distinct, novel, multi-exonic 15.9 kDa, 143 aa protein (ORF11) with  
200 the same SSC as ORF9 and ORF10 but with a unique STC. The splice junctions of all transcripts in this  
201 region (except the junction for TRXPT\_4) were validated by cloning of viral cDNA and Sanger sequencing

202 (Supplementary PCR methods).

203 During the validation of TRXPT\_2, ORF1 was present on the agarose gel (an unspliced band size) and  
204 Sanger sequencing results as a bona fide transcript (**Supplementary PCR methods**). This was corroborated by our 3'RACE experiment, which showed a transcript (TRXPT\_2B) spanning the entire ORF1 and  
205 Hyd ORFs without any splicing, with a poly-A tail immediately after the E1 TTS. The 5'-most CDS of this  
206 transcript (TRXPT\_2B) would encode ORF1. However, TRXPT\_2B has an upstream and in-frame SSC  
207 to the predicted SSC of ORF1, suggesting that the predicted ORF1 CDS is truncated – the actual ORF1  
208 (eORF1) that is expressed shares the same SSC as ORF10, but has a unique STC.

210 **Early Region 2 (E2) and Intermediate Region (IM) transcripts**

211 The E2 TU expressed on the anti-sense strand is subdivided into E2A and E2B and encodes three classical  
212 AdV proteins – pTP and Ad-pol (E2B proteins), and DBP (E2A protein) – essential for genome replication  
213 (17, 18). Unlike MAdV where two promoters (E2-early and E2-late) are known (17), we discovered only a  
214 single TSS (E2 TSS; 18,751bp) from which both E2A and E2B transcription is initiated. However, similar  
215 to MAdVs, E2A and E2B transcripts have distinct TTSs, and the E2B transcripts share the TTS of the IVa2  
216 transcript of the IM region (17, 18) (**Figure 7**).

217 The E2A ORF, DBP is one of three THEV ORFs predicted to be spliced from two exons. The correspond-  
218 ing transcript (TRXPT\_21) found in our data matches this predicted splice junction precisely but with a  
219 non-coding additional exon at the 5'-end (E2-5'UTR) at position 18,684-18,751 bp. Thus, TRXPT\_21 is  
220 a three-exon transcript encoding DBP (380 residues, 43.3 kDa) precisely as predicted. This transcript  
221 (TRXPT\_21) was also corroborated in a 3'RACE experiment. Additionally, from the 3'RACE, a splice vari-  
222 ant of TRXPT\_21 which retains the second intron leading to a 2-exon transcript was found. This transcript  
223 (TRXPT\_21B), albeit longer due to retaining the second intron and possessing a short 3' UTR, encodes a  
224 truncated isoform of DBP (tDBP) because the SSC utilized by TRXPT\_21, is followed shortly by STCs in the  
225 retained intron. The SSC 173 bp downstream of DBP's SSC yields tDBP (a 346 residue, 39.3 kDa product),  
226 which is in-frame of DBP but entirely contained in the second exon. TRXPT\_21 and TRXPT\_21B share a  
227 common TTS but TRXPT\_21B as seen in our 3'-RACE data, extends 39 bp into an adenine-thymine (A-T)  
228 rich sequence before the poly-A tail sequence occur, suggesting this position (16,934bp) as the bona fide  
229 E2A TTS (**Figure 7**).

230 The E2B region transcripts also start with the E2-5'UTR but extend thousands of base pairs downstream to  
231 reach the TTS at 2334bp in the IM region, which is immediately followed by an A-T rich sequence (position  
232 2323-2339bp) where polyadenylation probably occurs. Interestingly, the TTS of the E1 region (position  
233 2,325bp) on the sense strand is also in the immediate vicinity of this A-T rich sequence, which is almost

234 palindromic; hence it likely serves as the polyadenylation signal for both E1 and E2B/IM transcripts. The  
235 E2B transcripts, TRXPT\_6 and TRXPT\_7 are almost identical except for an extra splice junction at the 3'-  
236 end of TRXPT\_6, making TRXPT\_6 a five-exon transcript and TRXPT\_7, four exons (**Figure 7**). TRXPT\_7  
237 has the CP for both classical proteins (pTP and Ad-pol) encoded in this region, of which the pTP ORF is  
238 predicted to be spliced from two exons just like in all other AdVs. The predicted splice junction of pTP  
239 is corroborated by our data; however, the full transcript is markedly longer than the predicted ORF: there  
240 are two novel non-coding 5' exons, the third exon (containing the SSC of pTP) is significantly longer than  
241 predicted, and the last exon containing the bulk of the CDS is more than triple the predicted size of pTP.  
242 The first two exons are 5'-UTRs because the SSC here is immediately followed by STCs; hence, the 5'-  
243 most SSC (position 10,995bp) of the third exon which matches the predicted SSC of pTP is utilized. The  
244 encoded product is identical to the predicted pTP protein (597 residues; 70.5 kDa). If secondary SSC  
245 (secSSC) usage is considered, with SSC at 6768bp and STC at 3430bp, the encoded product is identical  
246 to the predicted Ad-pol (polymerase) ORF (1112 residues; 129.2 kDa). TRXPT\_6 differs from TRXPT\_7 by  
247 containing an extra splice site at 3447-3515bp. However, the CP remains similar to that of TRXPT\_7 except  
248 the Ad-pol encoded from the secSSC is a truncated isoform with a new STC resulting from the splice site.  
249 While both TRXPT\_6 and TRXPT\_7 have the CP for Ad-pol with secSSC usage, in all AdVs studied, the two  
250 proteins (pTP and Ad-pol) are encoded by separate mRNAs with identical first three 5' exons and TTS, but  
251 the splice junction to the terminal exons are different. We checked for a longer splice junction between the  
252 third and fourth (terminal) exons of TRXPT\_7 with our junction validation method (targeted PCR, cloning,  
253 and Sanger sequencing) and discovered a unique splice junction (10,981-7062bp) not found in our RNA-  
254 seq data. If initiated from the E2 TSS and terminated at the E2 TTS, this transcript(TRXPT\_31) would  
255 encode Ad-pol exactly as predicted as its 5'-most CDS (**Figure 7**).  
256 Our RNA-seq data also showed a novel short transcript (TRXPT\_15) entirely nested within the terminal  
257 exon of TRXPT\_7 but with a unique splice site. This transcript is an incomplete construction from the  
258 mapped reads as it contains a truncated CDS. However, we validated this splice junction to be genuine  
259 (**Supplementary PCR methods**).  
260 The IM region is a single-transcript TU, encoding a single classical protein, IVa2. The promoter expressing  
261 this single transcript (TRXPT\_5) is embedded in E2B region and shares a TTS with E2B transcripts (17,  
262 18). TRXPT\_5 is a two-exon transcript spliced exactly as the last splice junction of TRXPT\_6. The first  
263 exon is a UTR, except the last 2 nucleotides, which connect with the first nucleotide of the second exon to  
264 form the 5'-most SSC. This first SSC is 4 codons upstream and in-frame of the predicted IVa2 SSC. Except  
265 for the four extra N-terminus residues, the entire protein sequence is identical to the predicted IVa2.

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

267 The E3 region is wholly contained in the MLTU and encodes proteins involved in modulating and evading  
268 the host immune defenses. In MAdVs, this region contains seven ORFs expressed from several transcripts  
269 which share the same TSS (from the E3 promoter) but have different TTSs (6, 17, 18). However, some  
270 E3 transcripts use the TSS of the MLP. Due to sharing the same TSS, in MAdVs, secSSC usage is heavily  
271 relied on for gene expression in this region except for 12.5K and transcripts using the MLP's TSS, as utilizing  
272 only the first SSC cannot produce all the other transcripts in this TU (17).

273 In THEV, only one ORF (E3) was predicted in this region. However, as the E3 TU is nested in the MLTU,  
274 transcripts from the L4P (100K, 22K, 33K, and pVIII) not only overlap the E3 region transcripts entirely as  
275 seen in our RNA-seq results, but also have their TSS and TTS in practically the same locations (**Figure 8**).  
276 Therefore, we have categorized these two groups together as E3 transcripts.

277 We identified seven novel transcripts here (**TRXPT\_22, TRXPT\_23, TRXPT\_24, TRXPT\_25, TRXPT\_26,**  
278 **TRXPT\_27, TRXPT\_29**) from our RNA-seq data, all originating from two distinct TSSs – one corresponding  
279 to the TSS of the L4P (position 18,230bp) and the other at 18,727bp corresponding the E3 promoter (E3P).  
280 These E3 transcripts collectively have the CP for several predicted THEV ORFs: 100K, 22K, 33K, pVIII, E3,  
281 Fiber (IV), and ORF7 belonging to the MLTU. But some CDSs are nonidentical due to unknown splice sites  
282 or the use of an in-frame upstream SSC. For instance, 33K is one of the few THEV ORFs predicted to be  
283 spliced from two exons; however, we discovered a significantly longer four-exon ORF (e33K) on TRXPT\_24  
284 that contains it almost entirely. The first two exons of e33K were not predicted but the last two match  
285 the predicted exons and the CDS is in-frame, but the first 20bp of the predicted 33K (including the SSC at  
286 20,142bp) is spliced out as part of the second intron of TRXPT\_24. Thus, the bona fide 33K (e33K) is a 19.8  
287 kDa, 171 residue protein spanning four exons instead of the predicted 120 aa protein. TRXPT\_24 also has  
288 the CP for the ORFs, pVIII and a longer variant of E3 (eE3; starting from an in-frame upstream SSC) if we  
289 consider downstream SSC usage. TRXPT\_29 is the shortest transcript in this TU. It is a two-exon transcript,  
290 both exons comprising the CDS. The product of TRXPT\_29 is a novel 73 residue protein (8.3KI) sharing the  
291 SSC of e33K but with a unique STC. TRXPT\_23 being spliced identically as TRXPT\_29 also encodes 8.3KI  
292 from its first SSC. Similarly, TRXPT\_22 also encodes a 73 aa novel protein (8.3KII) from its first SSC that  
293 shares over 80% similarity with 8.3KI, but it differs from 8.3KI at the C-terminus. Considering downstream  
294 SSC usage, both TRXPT\_22 and TRXPT\_23 can encode pVIII and eE3 in that order, but TRXPT\_23 being  
295 longer, has the CP for the Fiber ORF also. As the splice junctions of TRXPT\_22, TRXPT\_23, TRXPT\_24,  
296 and TRXPT\_29 essentially share the same genomic space, their validation was done with a single primer  
297 pair and they were differentiated from each other by cloning and Sanger sequencing (**Supplementary PCR**

298 **methods).**

299 In addition to corroborating the splice junctions for the aforementioned transcripts, the Sanger sequencing  
300 results also showed another splice variant undetected in our RNA-seq transcriptome. This was a three-exon  
301 transcript (TRXPT\_30) with its first and last exons spliced identically as TRXPT\_23, but which also has the  
302 second exon of TRXPT\_24 (**Figure 8**). The first CDS on TRXPT\_30 spans all three exons, producing a  
303 novel 140 residue, 15.7kDa protein (e22K). Interestingly, the last 81 C-terminus residues of e22K are iden-  
304 tical to 22K (89 residues), which is a single-exon ORF predicted to use the same SSC as 33K (20,142bp).  
305 Just as seen for 33K, all the transcripts in this region exclude the first 20bp of 22K (including the SSC) as  
306 part of their introns; therefore, the first 7 residues of 22K are lacking in e22K due to splicing. Hence, we  
307 may consider e22K as a long variant of the predicted 22K ORF. Albeit the TSS and TTS of TRXPT\_30 was  
308 not seen, we presume that they are similar to TRXPT\_23, in which case it would also have the downstream  
309 CP of TRXPT\_23. TRXPT\_25 is the largest transcript in the TU. It also utilizes the L4P TSS but has a  
310 distinct TTS. It is a two-exon transcript, encoding a novel protein (t100K; 543 residues), which is a shorter  
311 isoform of the predicted 100K ORF. Considering secSSC usage on this transcript yields the predicted 22K  
312 ORF precisely. It also has the CP for pVIII and eE3 in that order. Furthermore, during the validation of  
313 TRXPT\_25's splice junction using primers that span its junction (18350-18717bp), we noticed a DNA band  
314 that corresponds to the full unspliced sequence (**Supplementary PCR methods**). As TRXPT\_25 only falls  
315 short of encoding the complete predicted 100K protein due to its splice junction, this band (which we cloned  
316 and validated by Sanger sequencing) suggests that the predicted 100K is indeed expressed. This tran-  
317 script (TRXPT\_25B) although not seen in full, likely shares the same TSS and TTS as TRXPT\_25. Lastly,  
318 TRXPT\_26 and TRXPT\_27 both originate from the E3 TSS but have distinct TTSs. TRXPT\_26 is a three-  
319 exon transcript but the first two are UTRs. It encodes pVIII as the 5'-most ORF and has the CP for eE3 and  
320 Fiber in that order. TRXPT\_27 on the other hand, is only a two-exon transcript but similar to TRXPT\_26,  
321 only the terminal exon contains the CDSs. It encodes Fiber as the 5'-most ORF, and ORF7 downstream  
322 with secSSC usage. TRXPT\_13, which is an L4 transcript that uses the MLP TSS is discussed under the  
323 MLTU transcripts.

324 **Early Region 4 (E4) transcripts**

325 This TU is found at the tail-end (3'-end) of the genome and expressed from the anti-sense strand. Based  
326 on nucleotide position, ORF7 and ORF8 were predicted in this region (1); however, as ORF7 is neither on  
327 the same strand as ORF8 nor transcribed from a promoter in the E4 region, only ORF8 can legitimately  
328 be classified as a transcript in this TU. This is corroborated by our RNA-seq data, as only one transcript  
329 was identified in this region on the anti-sense strand (**Figure 9**). The transcript (TRXPT\_28) spans 25192-

330 26247bp and is spliced at 25701-26055bp, making a two-exon transcript. The second exon fully matches  
331 the predicted ORF8 with 12 extra base pairs at the 3'-end. However, there is a SSC in the first exon at  
332 position 26246bp (192bp upstream of the predicted SSC). The encoded protein from this SSC is in-frame  
333 with the predicted SSC found in the second exon; hence, we consider this protein (eORF8 – 26.4 kDa, 229  
334 aa), a longer isoform of the predicted ORF8, the genuinely expressed ORF with an identical C-terminus to  
335 the predicted ORF8 protein.

336 **Major Late Transcription Unit (MLTU) or MLP Region transcripts**

337 The MLTU transcripts dominate the late phase (i.e, after DNA replication) of the AdV infectious cycle.  
338 The MLP produces all late mRNAs by alternative splicing and alternative polyadenylation of a primary  
339 transcript, grouped into five transcript classes (L1-L5). Most of THEV's coding capacity falls within this  
340 TU. Specifically, about 13 out of the 23 predicted ORFs were assigned to this TU, some of which we have  
341 categorized under the E3 TU instead. Our RNA-seq data revealed 12 transcripts (**TRXPT\_8, TRXPT\_9,**  
342 **TRXPT\_10, TRXPT\_11, TRXPT\_12, TRXPT\_13, TRXPT\_14, TRXPT\_16, TRXPT\_17, TRXPT\_18,**  
343 **TRXPT\_19, TRXPT\_20**) in this TU, the majority of which have the 5' untranslated TPL leader sequence as  
344 seen in all AdVs. For three transcripts (**TRXPT\_16, TRXPT\_17, TRXPT\_18**), a different leader sequence  
345 (sTPL) is used, which differs from the TPL in only one regard: the first TPL exon is substituted for a different  
346 first exon, found between the first and second TPL exons. Also, TRXPT\_20 seems to include only the third  
347 TPL exon (**Figure 10**).

348 We identified five TTSs (10,549bp, 12,709bp, 16,870bp, 17,891bp, 20,865bp) in this TU, corresponding  
349 to the five late mRNA classes (L1-L5), respectively, as found in all AdVs. L1 mRNAs include TRXPT\_8,  
350 which comprises the TPL (non-coding) and the CDS-containing terminal exon. This transcript encodes  
351 the 52K ORF exactly as predicted with the SSC beginning from the first nucleotide of the terminal exon.  
352 L2 mRNAs include TRXPT\_16, TRXPT\_17, and TRXPT\_18, all of which consist of the sTPL (also non-  
353 coding) followed by their respective terminal exons. TRXPT\_16 encodes pIIla exactly as predicted as the  
354 5'-most ORF, and also has the CP for the ORFs, III and pVII in that order. TRXPT\_17 encodes the ORF, III  
355 (penton), and TRXPT\_18 encodes the ORF pVII exactly as predicted. The L3 mRNAs include TRXPT\_14  
356 and TRXPT\_20, of which TRXPT\_14 utilizes the full TPL whereas TRXPT\_20 uses only the third TPL  
357 exon (TPL3). Both transcripts have the CP for the ORF, hexon (II) but hexon is the only ORF encoded  
358 on TRXPT\_14, whereas the 5'-most ORF on TRXPT\_20 is pX (pre-Mu) followed by pVI and hexon in  
359 that order. L4 mRNAs include TRXPT\_9, TRXPT\_10, TRXPT\_11, and TRXPT\_13 all of which begin with  
360 the TPL followed by three (TRXPT\_9, TRXPT\_10, and TRXPT\_13) or four (TRXPT\_11) coding exons.  
361 These are the largest transcripts found in the transcriptome, each one possessing the CP for several similar

late proteins. Normally, MLTU transcripts encoding particular ORFs splice the TPL onto a splice site just upstream of the ORF to be expressed (17). While this holds true for most MLTU ORFs, several late ORFs (pVI, protease, and ORF7) do not have such close proximity splicing but are contained in larger transcripts such as these L4 mRNAs, strongly suggesting the use of non-standard ribosomal initiation mechanisms such as secSSC utility and ribosome shunting found in other AdVs for their translation (17, 28). TRXPT\_9 and TRXPT\_10 are very similar but not identical. The last exon of TRXPT\_9 seems to be truncated and probably shares the same TTS as the other L4 mRNAs. They are both 6-exon transcripts encoding pVII as the 5'-most ORF (fourth exon) and also have the CP for pX, pVI, hexon, a longer variant of protease (eProt) – uses an upstream in-frame SSC than predicted, and ORF12 (a novel unpredicted 120 aa protein). TRXPT\_10 (and TRXPT\_9 with the L4 TTS) also has the CP for pVIII and eE3. Conversely, TRXPT\_11 is a seven-exon mRNA with hexon as its 5'-most ORF but it also has the CP for eProt, ORF12, e33K, and also pVIII and eE3 in that order. TRXPT\_13 seems to be an E3 ORF utilizing the MLP TSS as it encodes classical L4P genes such as pVIII and eE3 in that order similar to TRXPT\_22 (E3 TU) but lacks TRXPT\_22's novel first ORF (8.3KII).

Lastly, the L5 class includes only TRXPT\_12 which contains the TPL and a coding terminal exon. Its 5'-most ORF is fiber (IV) but it also has the CP for the THEV specific gene, ORF7. TRXPT\_12's CP is identical to TRXPT\_27 of the the E3 TU but they differ in their 5'-UTRs.

379 **DISCUSSION/CONCLUSIONS**

380 While the advent of next-generation sequencing has rendered easier the study of large and complex eu-  
381 karyotic transcriptomes, the study of the smaller and compact viral transcriptomes remains unintuitively  
382 challenging, as several transcripts may have significant overlaps due to genome economization. Char-  
383 acterizing AdV transcriptomes is even more difficult due to the wide array of mRNAs produced via very  
384 complex alternative splicing combined with alternative polyadenylation, all initiated from relatively few pro-  
385 moters. This makes AdV transcriptomes some of the most intricate for a virus. The challenge is further  
386 compounded by the fact that the standard software programs used in the RNA-seq analysis pipelines are  
387 not designed primarily for such compact, gene-dense, and complex transcriptomes as AdVs. Furthermore,  
388 there is no prior transcriptomic studies for THEV. Our approach to properly handle this complex data was  
389 to use standard RNA-seq analysis programs coupled with some custom analysis and validating all splice  
390 junctions with independent methods. Our work provides the first insights into the splicing patterns of THEV,  
391 which is expectedly similar to other MAdVs but with key differences. Our work shows 34 transcripts in  
392 the THEV transcriptome grouped into five TUs, of which the E3 TU shows great complexity of alternative  
393 splicing.

394 An unexpected observation is that the pileup of mapped reads to THEV seems consistently skewed over  
395 similar regions of the genome at all time points. As AdVs gene expression is temporally regulated, we  
396 expected to see unambiguous differences in the pileup of reads over different regions of the genome at  
397 different time points, indicating the different stages of infection. While this could simply mean that the  
398 infection was not well synchronized, we speculate that the temporal gene expression regulation of THEV is  
399 different from MAdVs. This is supported by a previous study stating the same conclusion with its finding that  
400 almost all THEV transcripts were detectable by at 4h.p.i, and by 8h.p.i, mRNA for all predicted ORFs were  
401 present (24). Despite the overall pileup similarity, a close inspection shows that the relative proportions  
402 of reads over some regions show some variation over time. The breakdown of transcripts detected at  
403 different time points in **Figure 3b** seems to support this different temporal regulation of THEV. Specifically,  
404 the MLP of THEV is active significantly earlier in infection – as early as 12h.p.i (**Figure 3b2** and **Table 2a**),  
405 – whereas the late phase shift in MAdVs occurs after 24h.p.i. This also lends credence to our speculation.  
406 However, generally speaking, the overall temporal gene expression regulation known in MAdVs – early  
407 region dominance in earlier time points followed by predominance of the MLTU at later time points – also  
408 holds true for THEV. Further studies would be necessary to establish the precise temporal regulation of  
409 THEV transcription.

410 The use of short read deep sequencing to reconstruct full AdV mRNA structures provides excellent results,  
411 especially for mapping the splice sites. However, due to the substantial overlapping nature of AdV mRNAs  
412 coupled with the fragmentation step in the library preparation protocol, mapping the precise TSS and TTS  
413 of the assembled transcripts is difficult. Also, similar transcripts with substantial overlaps may be assembled  
414 as one longer mRNA, since the short reads alone do not provide enough data for the transcript assembler  
415 (StringTie) to distinguish them. In our results, we see transcripts in the same TU initiated or terminated in  
416 the same approximate area (10-70bp and 1-300bp apart for TSS and TTS, respectively) but not precisely  
417 at the same position. We consider the most upstream TSS or most downstream TTS for the transcripts  
418 involved but we present them unchanged in all the figures shown. Also, by comparison to the more well-  
419 studies MAdV transcriptomes, we think that a few long transcripts in the MLTU (TRXPT\_9, TRXPT\_10, and  
420 TRXPT\_11) are probably a result of fusing some L4P-derived E3 transcripts to the terminal exons of the  
421 bona fide MLTU transcripts by StringTie, making them significantly longer. These mRNAs do not only have  
422 unusually many exons for an AdV, but their last three or four exons are also identical to the L4P-derived  
423 mRNAs. Future studies using long read sequencing technologies are necessary to provide conclusive data  
424 for precisely mapping the TSS and TTS, as well as teasing apart the bona fide structures of the long MLTU  
425 transcripts. Another observation made is that all the TTSs in THEV's transcriptome are in close proximity to  
426 A-T rich sequences which we presume to be polyadenylation signal sequences (PASS). Interestingly, some  
427 of these PASSs are located in the immediate vicinity of two closely-located TTSs expressed on opposite  
428 strands. Namely, the E1 and E2B/IM TTSs have an almost palindromic PASS between them, as do the E4  
429 and TRXPT\_12 and TRXPT\_27.

430 It is well established that adenovirus alternative splicing undergoes a regulated temporal shift in splice site  
431 usage (17, 22, 29).

432 TRXPT\_2 and ORF1 are isoforms.

433 Presumably, if the junction reads were normalized, MLTU would not be predominant at 12hpi. The TTSs  
434 were all in the context of A-T rich sequences; which presumably serve as polyA signals. All splice junctions  
435 were confirmed by cloning and Sanger sequencing of cDNA (**Supplementary PCR methods**). We did not  
436 find the x,y,z or i-leaders for MLP transcripts probably because THEV doesn't use it due to its smaller size  
437 The E3 ORF has an upstream, in-frame SSC.

438 **MATERIALS AND METHODS**

439 **Cell culture and THEV Infection**

440 The Turkey B-cell line (MDTC-RP19, ATCC CRL-8135) was grown as suspension cultures in 1:1 complete  
441 Leibovitz's L-15/McCoy's 5A medium with 10% fetal bovine serum (FBS), 20% chicken serum (ChS), 5%  
442 tryptose phosphate broth (TPB), and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Strepto-  
443 mycin), at 41°C in a humidified atmosphere with 5% CO<sub>2</sub>. Infected cells were maintained in 1:1 serum-  
444 reduced Leibovitz's L15/McCoy's 5A media (SRLM) with 2.5% FBS, 5% ChS, 1.2% TPB, and 1% antibiotics  
445 solution (100 U/mL Penicillin and 100ug/mL Streptomycin). A commercially available HE vaccine was pur-  
446 chased from Hygieia Biological Labs as a source of THEV-A (VAS strain). The stock virus was titrated using  
447 an in-house qPCR assay with titer expressed as genome copy number(GCN)/mL, similar to Mahshoub *et al*  
448 (30) with modifications. Cells were infected in triplicates at a multiplicity of infection (MOI) of 100 GCN/cell,  
449 incubate at 41°C for 1 hour, and washed three times to get rid of free virion particles. Samples in tripli-  
450 cates were harvested at 4-, 12-, 24-, and 72-h.p.i for total RNA extraction. The infection was repeated but  
451 samples in triplicates were harvested at 12-, 24-, 36-, 48-, and 72-h.p.i for PCR validation of novel splice  
452 sites. Still one more independent infection was done at time points ranging from 12 to 168-h.p.i for qPCR  
453 quantification of virus titers.

454 **RNA extraction and Sequencing**

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

463 **Validation of Novel Splice Junctions**

464 All splice junctions identified in this work are novel except one predicted splice site each for pTP, DBP, and  
465 33K, which were corroborated in our work. However, these predicted splice junctions had not been exper-

466 imentally validated hitherto, and we identified additional novel exons, giving the complete picture of these  
467 transcripts. The novel splice junctions discovered in this work using the StringTie transcript assembler were  
468 validated by PCR, cloning, and Sanger Sequencing (**Supplementary PCR methods**). Briefly, we designed  
469 primers that span a range of novel exon-exon boundaries for each specific transcript in a transcription unit  
470 (TU). We designed a universal forward or reverse primers for each respective TU and paired them with  
471 primers binding specific positions in each transcript. Each forward primer contained a KpnI restriction site  
472 and reverse primers, an XbaI site in the primer tails. After first-strand cDNA synthesis of total RNA ex-  
473 tracted from THEV infected MDTC-RP19 cells with SuperScript™ IV First-Strand Synthesis System, these  
474 primers were used in a targeted PCR amplification, the products analyzed with agarose gel electrophoresis  
475 to confirm expected band sizes, cloned by traditional restriction enzyme method, and Sanger sequenced to  
476 validate these splice junctions at the sequence level.

#### 477 **3' Rapid Amplification of cDNA Ends (3'-RACE)**

478 We performed a rapid amplification of sequences from the 3' ends of mRNAs (3'-RACE) experiment us-  
479 ing a portion of the extracted total RNA of infected MDTC-RP19 cells used for the RNA-seq experiment  
480 as explained above. We followed the protocol described by Green *et al* (31) with modifications. Briefly,  
481 1ug of total RNA was reverse transcribed to cDNA using SuperScript™ IV First-Strand Synthesis System  
482 following the manufacturing instructions using an adapter-primer with a 3'-end poly(T) and a 5'-end BamHI  
483 restriction site. A gene-specific sense primer with a 5'-end KpnI restriction site paired with an anti-sense  
484 adapter-primer with a 5'-end BamHI site were used to amplify target sections of the cDNA using Invitrogen's  
485 Platinum™ Taq DNA polymerase High Fidelity, following manufacturer's instructions. The PCR amplicons  
486 were restriction digested, cloned, and Sanger sequenced.

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

488 Our sequence reads were analyzed following a well established protocol described by Pertea *et*  
489 *al* (25), using Snakemake - version 7.24.0 (32), a popular workflow management system to  
490 drive the pipeline. Briefly, sequencing reads were trimmed with the Trim-galore - version  
491 0.6.6 (33) program to achieve an overall Mean Sequence Quality (Phred Score) of 36. Trimmed  
492 reads were mapped simultaneously to the complete genomic sequence of avirulent turkey hemor-  
493 rhagic enteritis virus (<https://www.ncbi.nlm.nih.gov/nuccore/AY849321.1/>) and *Meleagris gallopavo*  
494 (<https://www.ncbi.nlm.nih.gov/genome/?term=Meleagris+gallopavo>) using Hisat2 - version 2.2.1 (25)

495 with default settings. The generated alignment (BAM) files from each infection time point were filtered  
496 for reads mapping to the THEV genome using Samtools – version 1.16.1 and fed into StringTie –  
497 version 2.2.1 (25) to assemble the transcripts, using a GTF annotation file derived from a GFF3 annotation  
498 file obtained from NCBI, which contains the predicted ORFs of THEV as a guide. GFFCOMPARE – version  
499 0.12.6 was used to merge all transcripts from all time points without redundancy and using a custom R  
500 script, adenovirus transcripts units (regions) were assigned to each transcript, generating the transcriptome  
501 of THEV. StringTie set to expression estimation mode was used to calculate FPKM scores for all  
502 transcripts after which Ballgown – version 2.33.0 in R was used to perform the statistical analysis on the  
503 transcript expression levels. Samtools was also used to count the total sequencing reads for all replicates  
504 at each time point and Regtools – version 1.0.0 was used to count all junctions, the reads supporting  
505 them, and extract all other information related to the junction. See **Supplementary Computational**  
506 **Analysis** for the details of transcript expression level estimations and splice junction read counts.

507 **SUPPLEMENTARY MATERIALS**

508 **DATA AVAILABILITY**

509 The raw sequence data (FastQ), transcript expression counts, and total unique junctions have been de-  
510 posited at the National Center for Biotechnology Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE254416.

512 Data is available on request by contacting the designated corresponding author

513 **CODE AVAILABILITY**

514 All the code/scripts in the entire analysis pipeline are available on github ([https://github.com/Abraham-Quaye/thev\\_transcriptome](https://github.com/Abraham-Quaye/thev_transcriptome))

516 **ACKNOWLEDGMENTS**

517 LC Sciences - RNA sequencing was done here

518 Eton Bioscience, Inc, San Diego, CA - All Sanger sequencing validations was done here BYU high

519 performance computing systems - Memory intensive analysis were run here.

520 REFERENCES

- 521 1. Davison A, Benko M, Harrach B. 2003. Genetic content and evolution of adenoviruses. *The Journal*  
522 of general virology
- 523 2. Harrach B. 2008. Adenoviruses: General features, p. 1–9. *In* Mahy, BWJ, Van Regenmortel, MHV  
524 (eds.), *Encyclopedia of virology* (third edition). Book Section. Academic Press, Oxford.
- 525 3. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: Toward  
526 defining the minimum essential poxvirus genome. *Journal of virology* 77:7590–7600.
- 527 4. McGeoch D, Davison AJ. 1999. Chapter 17 - the molecular evolutionary history of the herpesviruses,  
528 p. 441–465. *In* Domingo, E, Webster, R, Holland, J (eds.), *Origin and evolution of viruses*. Book  
Section. Academic Press, London.
- 529 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*  
530 *Report of the International Committee on Taxonomy of Viruses* 125–141.
- 531 6. Guimet D, Hearing P. 2016. 3 - adenovirus replication, p. 59–84. *In* Curiel, DT (ed.), *Adenoviral*  
532 *vectors for gene therapy* (second edition). Book Section. Academic Press, San Diego.
- 533 7. Kovács ER, Benkő M. 2011. Complete sequence of raptor adenovirus 1 confirms the characteristic  
534 genome organization of siadenoviruses. *Infection, Genetics and Evolution* 11:1058–1065.
- 535 8. Davison AJ, Wright KM, Harrach B. 2000. DNA sequence of frog adenovirus. *J Gen Virol* 81:2431–  
536 2439.
- 537 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  
538 originating from parus major, a great tit. *Journal of Virological Methods* 163:262–268.
- 539 10. Katoh H, Ohya K, Kubo M, Murata K, Yanai T, Fukushi H. 2009. A novel budgerigar-adenovirus  
540 belonging to group II avian adenovirus of siadenovirus. *Virus Research* 144:294–297.
- 541 11. Beach NM. 2006. Characterization of avirulent turkey hemorrhagic enteritis virus: A study of the  
542 molecular basis for variation in virulence and the occurrence of persistent infection. Thesis.

- 543 12. Beach NM, Duncan RB, Larsen CT, Meng XJ, Sriranganathan N, Pierson FW. 2009. Comparison of  
544 12 turkey hemorrhagic enteritis virus isolates allows prediction of genetic factors affecting virulence.  
545 J Gen Virol 90:1978–85.
- 546
- 547 13. Gross WB, Moore WE. 1967. Hemorrhagic enteritis of turkeys. Avian Dis 11:296–307.
- 548
- 549 14. Rautenschlein S, Sharma JM. 2000. Immunopathogenesis of haemorrhagic enteritis virus (HEV) in  
550 turkeys. Dev Comp Immunol 24:237–46.
- 551 15. Larsen CT, Domermuth CH, Sponenberg DP, Gross WB. 1985. Colibacillosis of turkeys exacerbated  
552 by hemorrhagic enteritis virus. Laboratory studies. Avian Dis 29:729–32.
- 553 16. Dhama K, Gowthaman V, Karthik K, Tiwari R, Sachan S, Kumar MA, Palanivelu M, Malik YS, Singh  
554 RK, Munir M. 2017. Haemorrhagic enteritis of turkeys – current knowledge. Veterinary Quarterly  
555 37:31–42.
- 556
- 557 17. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. 2020. Deep splicing plasticity  
558 of the human adenovirus type 5 transcriptome drives virus evolution. Communications Biology 3:124.
- 559 18. Zhao H, Chen M, Pettersson U. 2014. A new look at adenovirus splicing. Virology 456-457:329–341.
- 560
- 561 19. Wolfrum N, Greber UF. 2013. Adenovirus signalling in entry. Cell Microbiol 15:53–62.
- 562
- 563 20. Falvey E, Ziff E. 1983. Sequence arrangement and protein coding capacity of the adenovirus type 2  
564 "i" leader. Journal of Virology 45:185–191.
- 565 21. Morris SJ, Scott GE, Leppard KN. 2010. Adenovirus late-phase infection is controlled by a novel L4  
566 promoter. Journal of Virology 84:7096–7104.
- 567
- 568 22. Westergren Jakobsson A, Segerman B, Wallerman O, Bergström Lind S, Zhao H, Rubin C-J, Pet-  
569 tersson U, Akusjärvi G. 2021. The human adenovirus 2 transcriptome: An amazing complexity of  
570 alternatively spliced mRNAs. Journal of Virology 95.

- 565 23. 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  
566 cells. *Nature* 489:101–108.
- 567 24. Aboeza Z, Mabsoub H, El-Bagoury G, Pierson F. 2019. In vitro growth kinetics and gene expression  
568 analysis of the turkey adenovirus 3, a siadenovirus. *Virus Research* 263:47–54.
- 569 25. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of  
570 RNA-seq experiments with HISAT, StringTie and ballgown. *Nature Protocols* 11:1650–1667.
- 571 26. Jack Fu [Aut], Alyssa C. Frazee [Aut, Cre], LeonardoCollado-Torres [Aut], Andrew E. Jaffe [Aut],  
572 Jeffrey T. Leek[Aut, Ths]. 2017. Ballgown. Bioconductor.
- 573 27. 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,  
574 hemorrhagic enteritis virus. *Virology* 249:307–315.
- 575 28. Yueh A, Schneider RJ. 1996. Selective translation initiation by ribosome jumping in adenovirus-  
576 infected and heat-shocked cells. *Genes & Development* 10:1557–1567.
- 577 29. Akusjarvi G. 2008. Temporal regulation of adenovirus major late alternative RNA splicing. *Frontiers  
578 in Bioscience Volume:5006*.
- 579 30. Mabsoub HM, Evans NP, Beach NM, Yuan L, Zimmerman K, Pierson FW. 2017. Real-time PCR-  
based infectivity assay for the titration of turkey hemorrhagic enteritis virus, an adenovirus, in live  
580 vaccines. *Journal of Virological Methods* 239:42–49.
- 581 31. Green MR, Sambrook J. 2019. Rapid amplification of sequences from the 3' ends of mRNAs: 3'-  
582 RACE. *Cold Spring Harbor Protocols* 2019:pdb.prot095216.

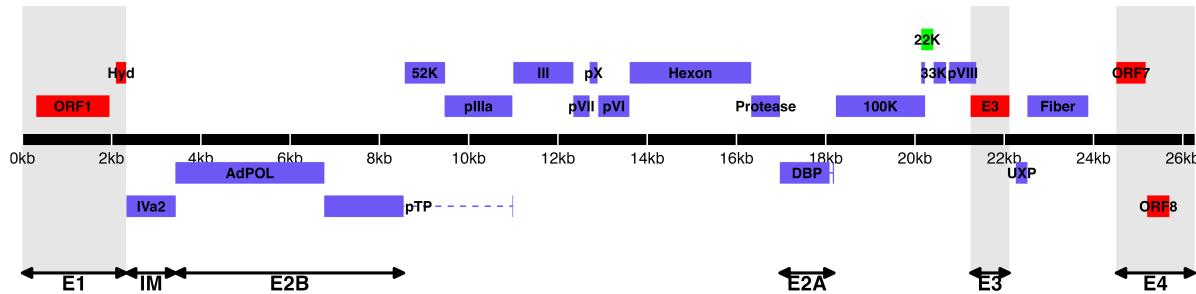
- 583 32. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok  
SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. 2021. Sustainable data  
analysis with snakemake. *F1000Research* 10:33.

584

585 33. Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B, Hulselmans G, Sclal-  
mons. 2023. *FelixKrueger/TrimGalore*: v0.6.10 - add default decompression path. Zenodo.

586

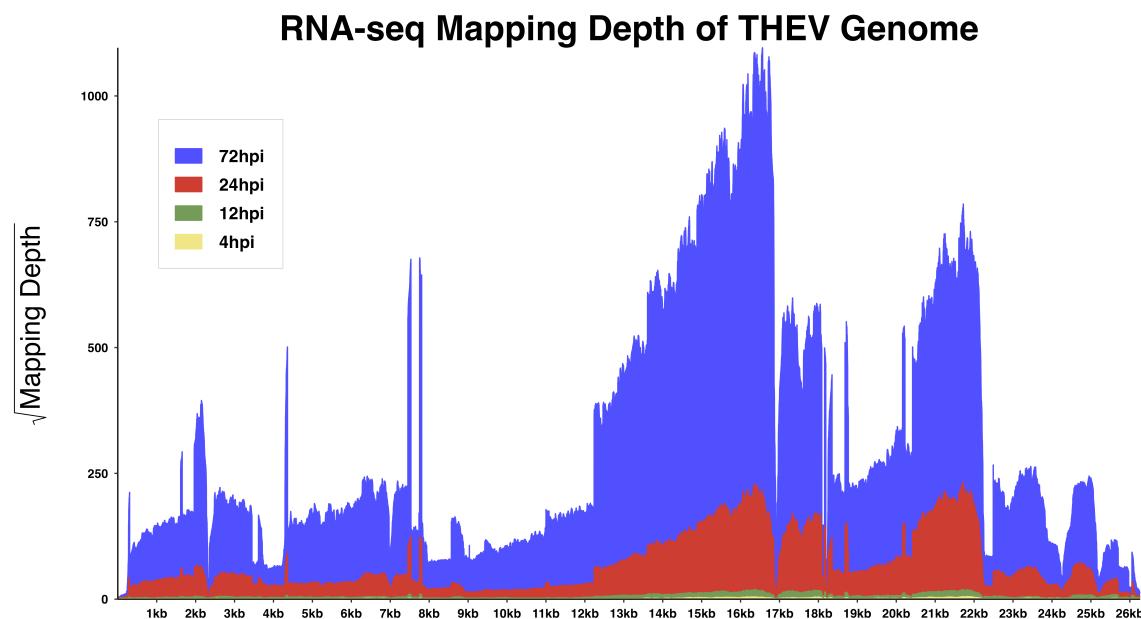
587 **TABLES AND FIGURES**



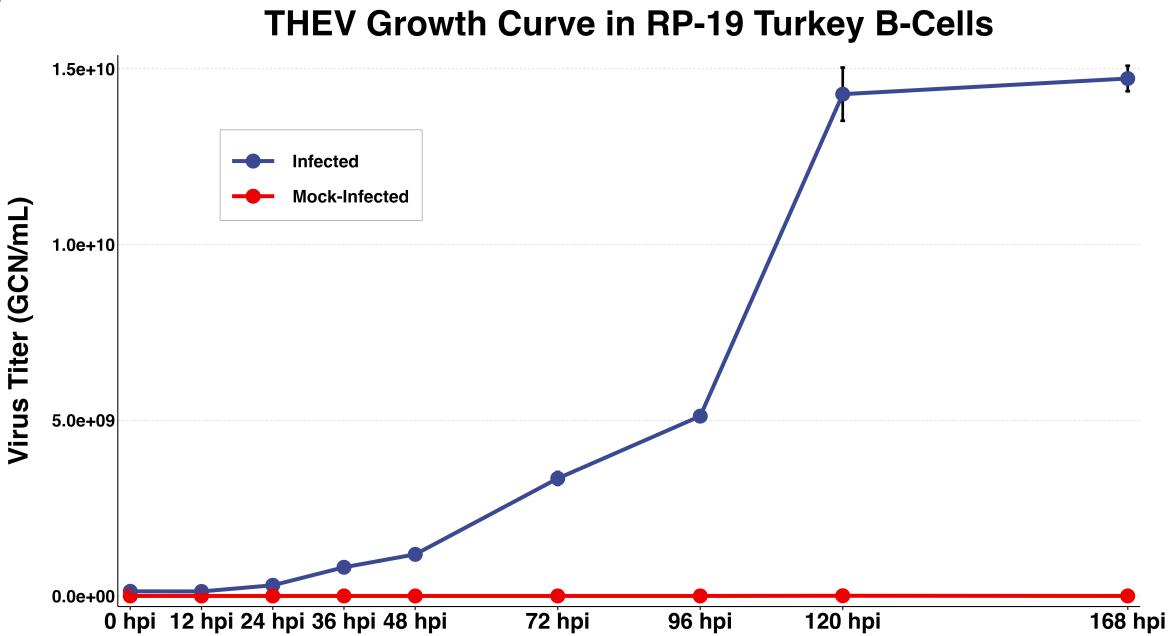
588

589 **Figure 1. Predicted ORF map of THEV virulent strain.** The central horizontal line represents the  
 590 double-stranded DNA marked at 5kb intervals as white line breaks. Blocks represent viral genes. Blocks  
 591 above the DNA line are transcribed rightward, those below are transcribed leftward. pTP, DBP and  
 592 33K predicted to be spliced are shown as having tails. Shaded regions indicate regions containing  
 593 "genus-specific" genes (colored red). Genes colored in blue are "genus-common". Gene colored in light  
 594 green is conserved in all but Atadenoviruses. The UXP (light blue) is an incomplete gene present in almost  
 595 all AdVs. Regions comprising the different transcription units are labelled at the bottom (E1, E2A, E2B,  
 596 E3, and E4); the unlabeled regions comprise the MLTU.

A



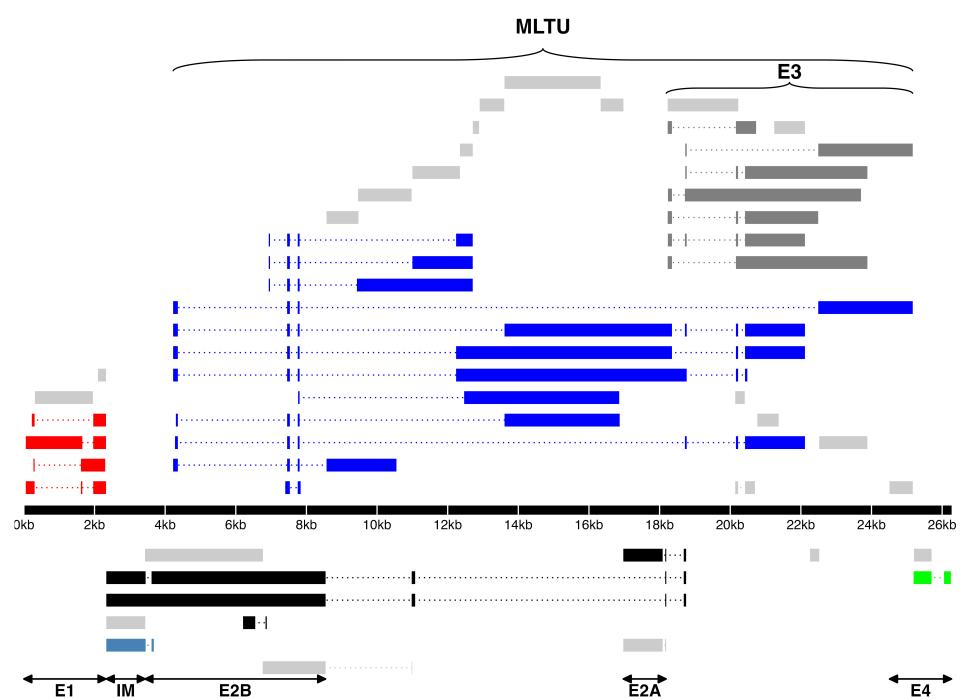
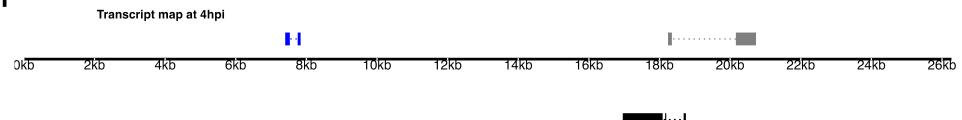
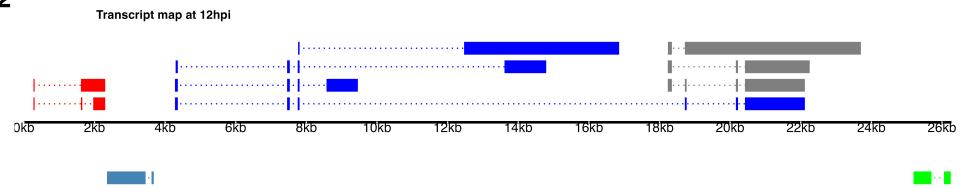
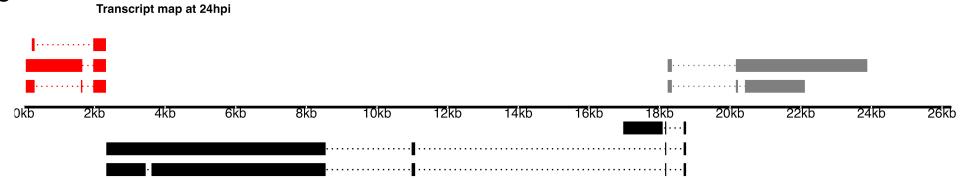
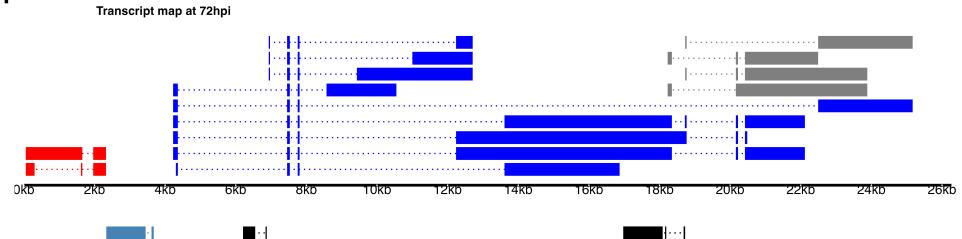
B



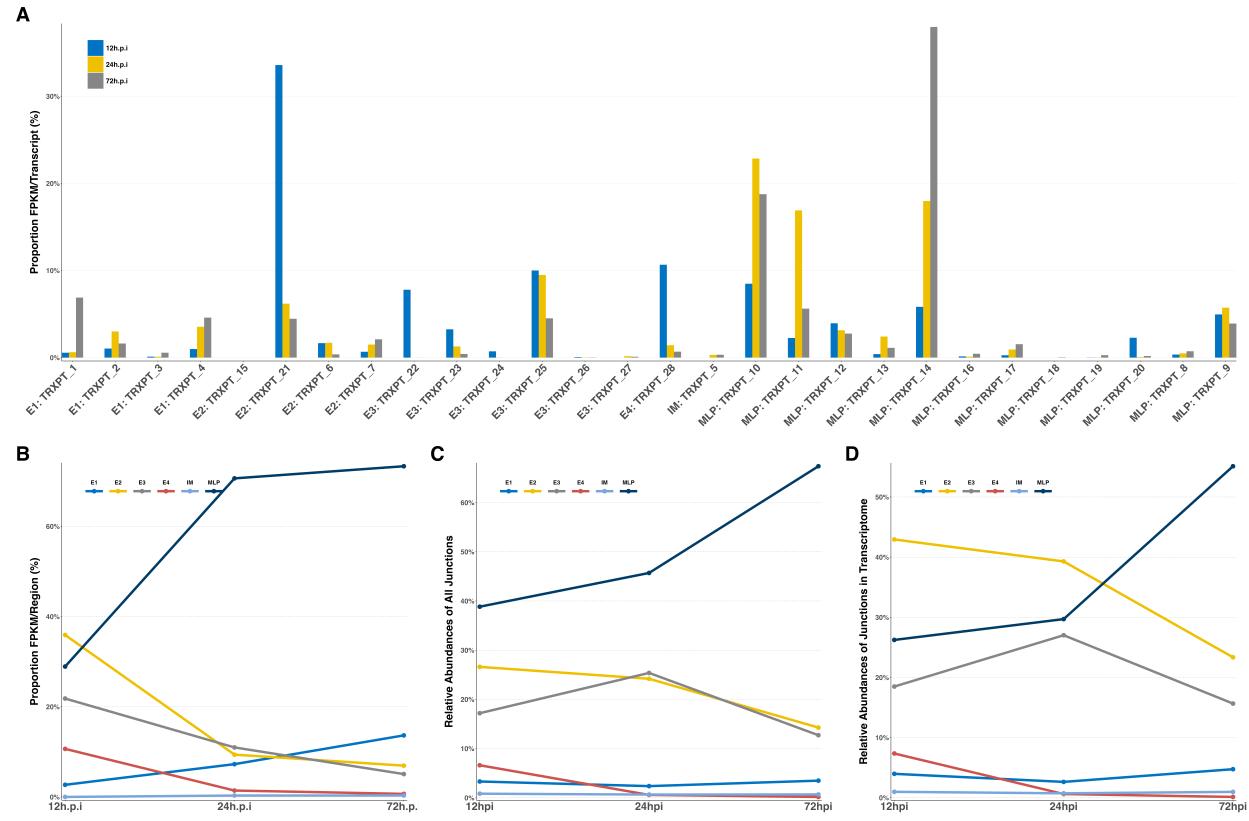
597

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

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

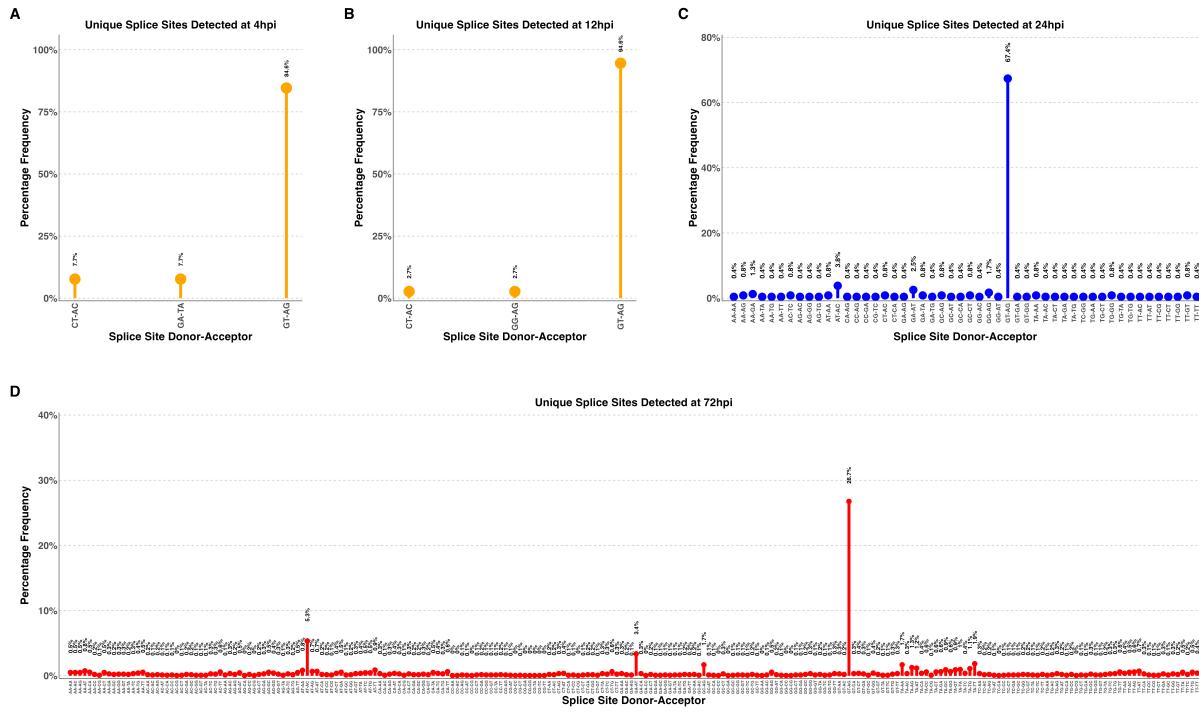
**A****B1****B2****B3****B4**

606 **Figure 3. a) Transcriptome of THEV from RNA-seq.** THEV transcripts assembled from all time points  
 607 by StringTie are unified forming this final transcriptome (splicing map). Transcripts belonging to the same  
 608 transcription unit (TU) are located in close proximity on the genome and are color coded and labeled in this  
 609 figure as such. The organization of TUs in the THEV genome is unsurprisingly similar to MAdVs; however,  
 610 the MAdV genome shows significantly more transcripts. The TUs are color coded: E1 transcripts - red, E2  
 611 - black, E3 - dark grey, E4 - green, MLTU - blue. Predicted ORFs are also indicated here, colored light grey.  
 612 **b) THEV transcripts identified at given time points.** Transcripts are color coded as explained in (a).

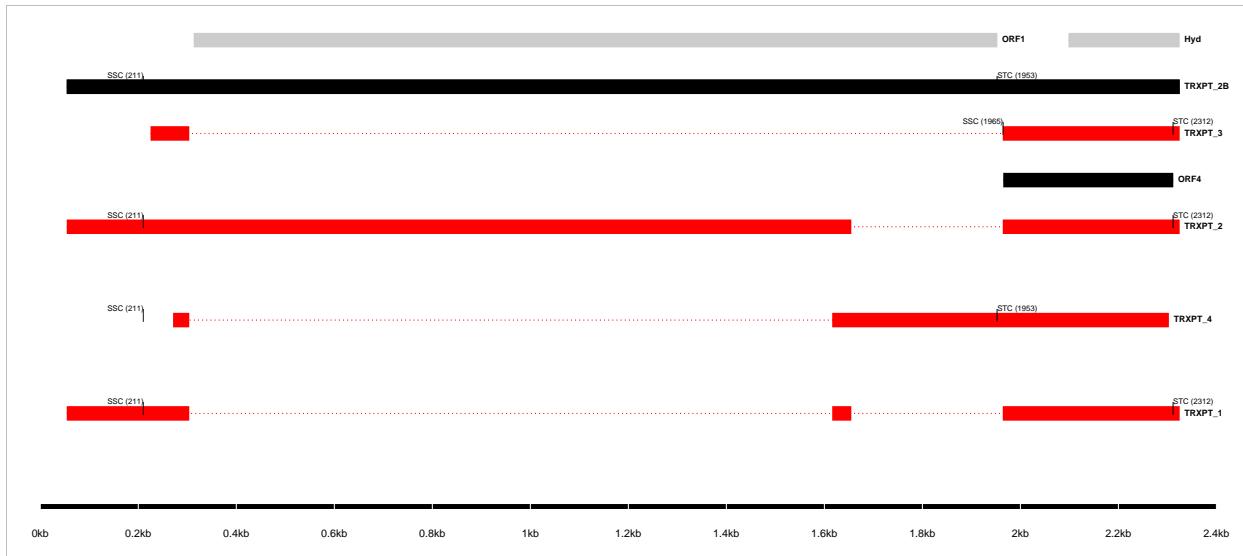


613 **Figure 4: Changes in splicing and expression profile of THEV over time. a) Normalized (FPKM)**  
 614 *expression levels of transcripts over time.* The expression levels (FPKM) of individual transcripts as a  
 615 percentage of the total expression of all transcripts at each time point are indicated. Only transcripts from  
 616 our RNA-seq data are included here. **b) Normalized (FPKM) expression levels of transcripts by region over**  
 617 *time.* The expression levels of each region/TU as a percentage of the total expression of all transcripts at  
 618 each time point are indicated. Region expression levels were calculated by summing up the FPKMs of all  
 619 transcripts categorized in that region. **c) Relative abundances of all splice junctions grouped by region/TU**  
 620 *over time.* After assigning all **2,457** unique junctions to a TU and the total junction reads counted at each  
 621 time point for each region, the total junction reads for each TU plotted as percentage of all junction reads at  
 622 each time point.

623 each time point is indicated. Note that the junction read counts are not normalized. **d) Relative abundances**  
 624 *of junctions in transcriptome grouped by region/TU over time*. This is identical to **(c)**, except that only the  
 625 junctions found in the full transcriptome obtained from the RNA-seq data were included.



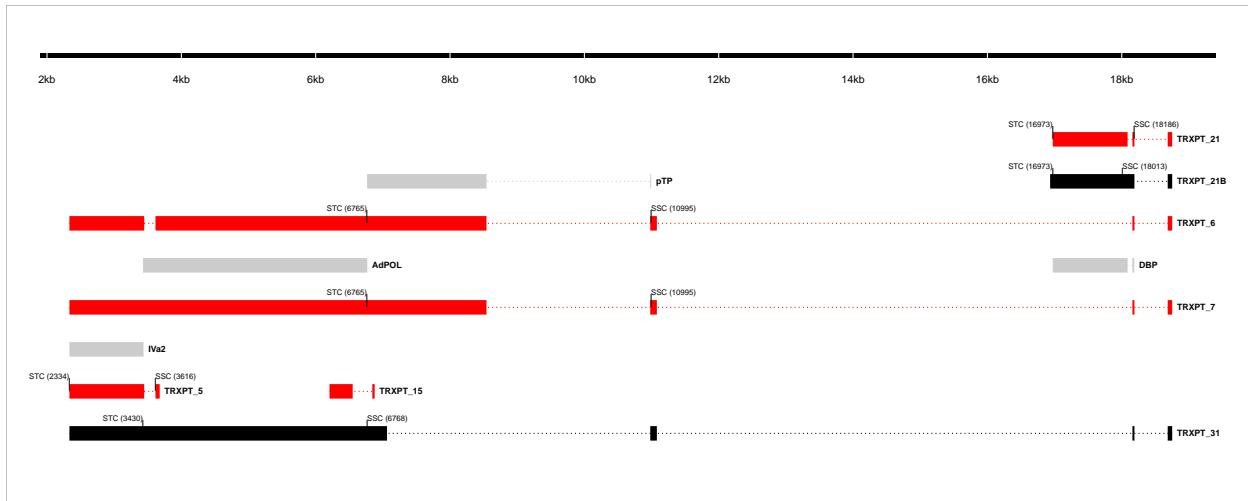
626  
 627 **Figure 5: Changes in splice donor-acceptor nucleotides over time.** The splice donor-acceptor  
 628 nucleotides of THEV just like other AdVs is mostly the canonical GT-AG. At early time points (4h.p.i and  
 629 12h.p.i [(a) and (b)]) the junction nucleotides used appear to be well scrutinized or restricted, utilizing  
 630 mostly the canonical splice nucleotides. However, as the infection progresses to the late stages (24h.p.i  
 631 and 72h.p.i [(c) and (d)]), the selectivity of specific splice acceptor-donor pairs seems to degenerate  
 632 significantly, such that all combinations of nucleotides are utilized.



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

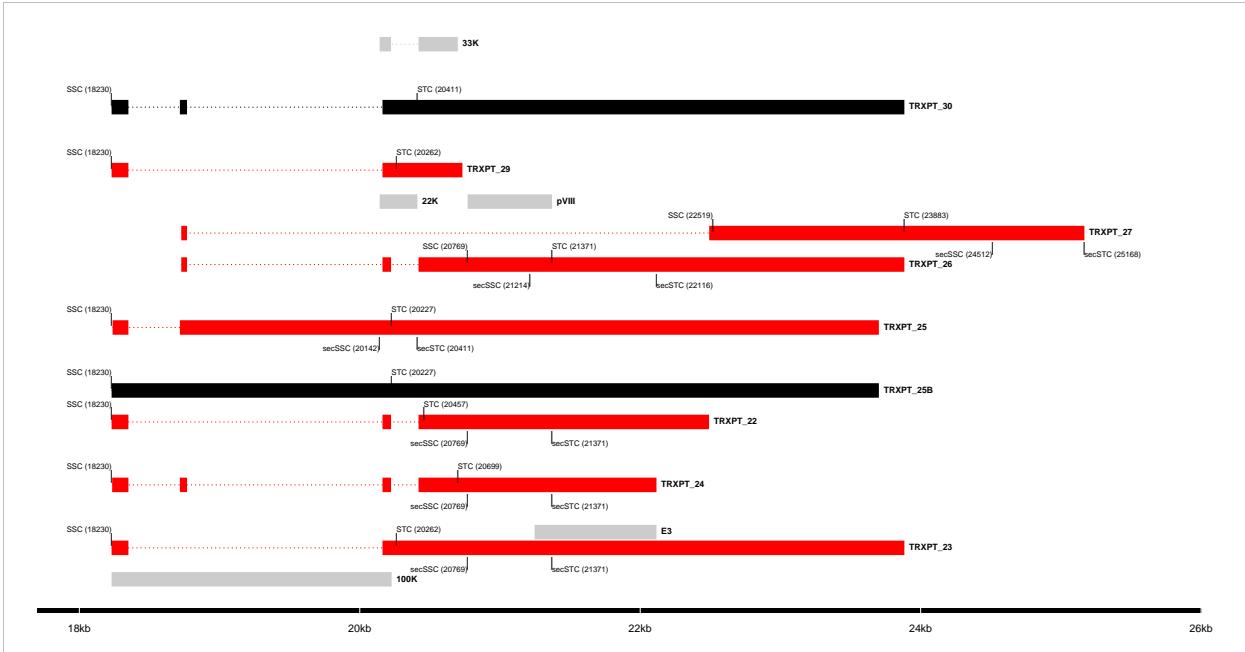
633 \*Not validated for TRXPT\_4

634 **Figure 6: The splice map of the E1 transcription unit (TU).** Exons are depicted as boxes connected by  
 635 introns (dotted lines). Transcripts from RNA-seq data are colored red, predicted ORFs are colored grey, and  
 636 transcripts or ORFs discovered by other means are colored black. Each transcript or ORF is labelled with  
 637 its name to the right. The start codon (SSC) and stop codon (STC) of the 5'-most CDS of each transcript  
 638 is indicated with the nucleotide position in brackets. The region of the virus is depicted at the bottom as a  
 639 black line with labels of the nucleotide positions for reference. The table shows sequence reads covering  
 640 the splice junctions with information about their validation status using cloning and Sanger sequencing.



Transcript ID	Splice Junction				Strand	region	Junction Reads				Junction Status
	Start	End	Splice Donor-Acceptor	Intron Length			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	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	Validated

641 **Figure 7: The splice map of the E2 and IM TUs.** Exons are depicted as boxes connected by introns  
 642 (dotted lines). Red transcripts are generated from RNA-seq data and predicted ORFs are colored grey.  
 643 TRXPT\_21B discovered by 3'RACE is colored black. Each transcript or ORF is labelled with its name to  
 644 the right. The SSC and STC of the 5'-most CDS of each transcript is indicated with the nucleotide position  
 645 in brackets. The region of the virus is depicted at the bottom as a black line with labels of the nucleotide  
 646 positions for reference. The table shows sequence reads covering the splice junctions with information  
 647 about their validation status 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_25, TRXPT_24, TRXPT_10	18350	18717	GT-AG	368bp	+	E3, MLP	4	21	3930	35490	Validated
TRXPT_23, TRXPT_22, TRXPT_11	18350	20162	GT-AG	1813bp	+	E3, MLP	3	18	6619	38841	Validated
TRXPT_26, TRXPT_24, TRXPT_13, TRXPT_11, TRXPT_9, TRXPT_10	18768	20162	GT-AG	1395bp	+	E3, MLP	2	21	5207	45062	Validated
TRXPT_26, TRXPT_22, TRXPT_24, TRXPT_13, TRXPT_11, TRXPT_9, TRXPT_10	20223	20419	GT-AG	197bp	+	E3, MLP	3	33	10583	93238	Validated
649 TRXPT_27	18768	22492	GT-AG	3725bp	+	E3	0	0	101	1950	Validated

650 **Figure 8: The splice map of the E3 TU.** Exons are depicted as boxes connected by introns (dotted  
 651 lines). Red transcripts are generated from RNA-seq data and predicted ORFs are colored grey. Transcripts  
 652 discovered by other means are colored black. Each transcript or ORF is labelled with its name to the right.  
 653 The start codon (SSC) and stop codon (STC) of the 5'-most CDS of each transcript is indicated with the  
 654 nucleotide position in brackets. Similarly, the secondary SSC (secSSC) and secondary STC (secSTC)  
 655 are shown. The region of the virus is depicted at the bottom as a black line with labels of the nucleotide  
 656 positions for reference. The table shows sequence reads covering the splice junctions with information  
 657 about their validation status using cloning and Sanger sequencing.



659 **Figure 9: The splice map of the E4 TU.** Exons are depicted as boxes connected by introns (dotted lines).  
 660 The transcript from RNA-seq data is colored red and the predicted ORF, grey. The transcript and ORF are  
 661 labelled with their names to the right. The start codon (SSC) and stop codon (STC) of the 5'-most CDS  
 662 is indicated with the nucleotide position in brackets. The region of the virus is depicted at the bottom as a  
 663 black line with labels of the nucleotide positions for reference. The table shows sequence reads covering  
 664 the splice junction with its validation status using cloning and Sanger sequencing.



665 **Figure 10: The splice map of the MLTU.** Exons are depicted as boxes connected by introns (dotted lines).  
 666 The transcripts from our RNA-seq data are colored red and the predicted ORFs, grey. The transcripts and  
 667 ORFs are labelled with their names to the right. The start codon (SSC) and stop codon (STC) of the 5'-most  
 668 CDS of each transcript is indicated with the nucleotide position in brackets. Similarly, the secondary SSC  
 669 (secSSC) and secondary STC (secSTC) are shown. The region of the virus is depicted at the bottom as a  
 670 black line with labels of the nucleotide positions for reference. The table shows sequence reads covering  
 671 the splice junctions with information about their validation status using cloning and Sanger sequencing.  
 672

Table 1: Table 1: Overview of sequencing results

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

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

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

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

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

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

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