

<sup>1</sup> Supplementary Materials: Elucidating the Transcriptome of  
<sup>2</sup> Turkey Hemorrhagic Enteritis Virus

<sup>3</sup>

<sup>4</sup> **Running Title:** Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

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

<sup>6</sup> \*Department of Microbiology and Molecular Biology, Brigham Young University

<sup>7</sup> <sup>1</sup>First-author

<sup>8</sup> <sup>†</sup> Corresponding Author

<sup>9</sup> **Corresponding Author Information**

<sup>10</sup> brian\_poole@byu.edu

<sup>11</sup> Department of Microbiology and Molecular Biology,

<sup>12</sup> 4007 Life Sciences Building (LSB),

<sup>13</sup> Brigham Young University,

<sup>14</sup> Provo, Utah

<sup>15</sup>

<sup>16</sup> **Supplementary Table S1A**

Table 1: Table S1a: Most Transcriptionally Active Regions of THEV at 12h.p.i

Time	Region	Strand	Total Reads	Percentage
12hpi	MLP	+	235	38.8%
12hpi	E2	-	161	26.6%
12hpi	E3	+	104	17.2%
12hpi	E4	-	40	6.6%
12hpi	Unassigned	-,+/-	40	6.6%
12hpi	E1	+	20	3.3%
12hpi	IM	-	5	0.8%

<sup>17</sup> **Supplementary Table S1B**

Table 2: Table S1b: Most Transcriptionally Active Regions of THEV at 24h.p.i

Time	Region	Strand	Total Reads	Percentage
24hpi	MLP	+	52,589	45.7%
24hpi	E3	+	29,209	25.4%
24hpi	E2	-	27,833	24.2%
24hpi	E1	+	2,724	2.4%
24hpi	Unassigned	-,+/-	1,312	1.1%
24hpi	IM	-	744	0.6%

Time	Region	Strand	Total Reads	Percentage
24hpi	E4	-	664	0.6%

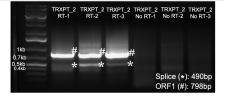
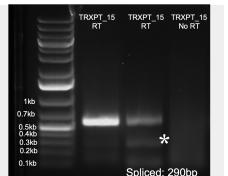
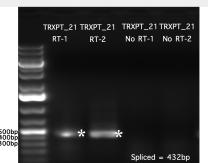
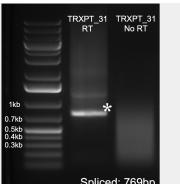
<sup>18</sup> **Supplementary Table S1C**

Table 3: Table S1c: Most Transcriptionally Active Regions of THEV at 72h.p.i

Time	Region	Strand	Total Reads	Percentage
72hpi	MLP	+	1,437,273	67.4%
72hpi	E2	-	304,274	14.3%
72hpi	E3	+	271,392	12.7%
72hpi	E1	+	74,135	3.5%
72hpi	Unassigned	-,+,-	27,680	1.3%
72hpi	IM	-	14,484	0.7%
72hpi	E4	-	3,568	0.2%

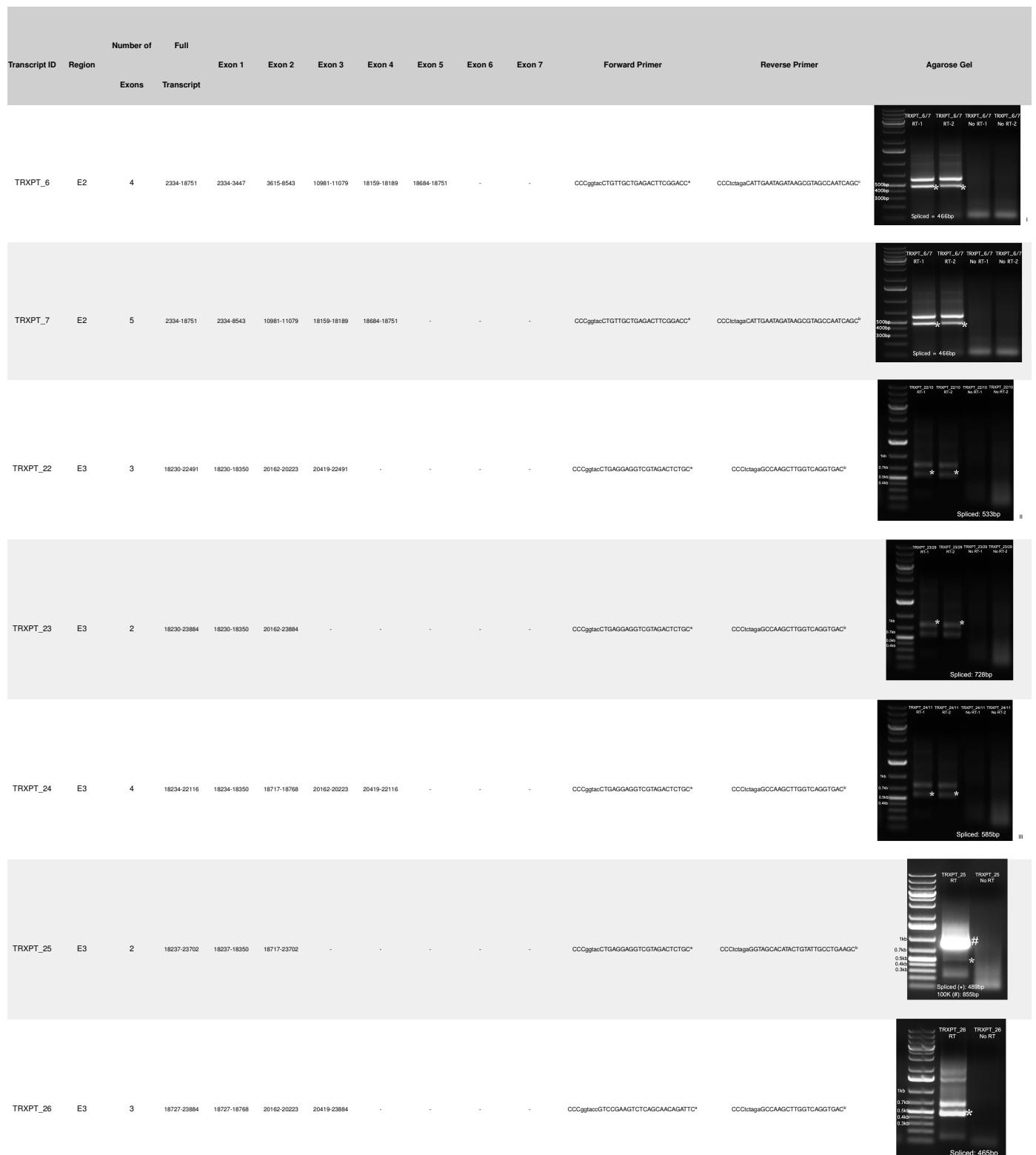
19 **Supplementary PCR Methods**

**Table 4: Agarose Gels Showing PCR Amplification of THEV cDNA With Gene-Specific Primers**

Transcript ID	Region	Number of Exons	Full Transcript	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7	Forward Primer	Reverse Primer	Agarose Gel
TRXPT_1	E1	3	54-2325 54-304 1616-1655 1964-2325	-	-	-	-	-	-	-	CCGgtaccTTCTGTTGAATTGGGGCGG <sup>a</sup>	CCGctttagCGTCCAGTAGTCAGGAATTCTAGTG <sup>b</sup>	
TRXPT_2	E1	2	54-2325 54-1655 1964-2325	-	-	-	-	-	-	-	CCGgtaccGAGGCCCTGTGGATTGTGC <sup>c</sup>	CCGctttagCGTCCAGTAGTCAGGAATTCTAGTG <sup>d</sup>	
TRXPT_3	E1	2	225-2325 225-304 1964-2325	-	-	-	-	-	-	-	CCGgtaccCATTCCCCTACACGGTGTG <sup>e</sup>	CCGctttagCGTCCAGTAGTCAGGAATTCTAGTG <sup>f</sup>	
TRXPT_4	E1	2	271-2303 271-304 1616-2303	-	-	-	-	-	-	-	CCGgtaccGTCATCACAACTGACCTTGTGTC <sup>g</sup>	CCGctttagCGTCCAGTAGTCAGGAATTCTAGTG <sup>h</sup>	Not Validated
TRXPT_15	E2	2	6206-6878 6206-6551 6843-6878	-	-	-	-	-	-	-	CCGgtaccCTTTAAAATCAAGCCTATTGGCTTGTAA <sup>i</sup>	CCGctttagGTGTCATTGTCAGCTGTGTTGAGTAG <sup>j</sup>	
TRXPT_21	E2	3	16973-18751 16973-18087 18159-18189 18684-18751	-	-	-	-	-	-	-	CCGgtaccCTGTTGCTGAGACTTCGGACC <sup>k</sup>	CCGctttagGAACCCAGATAATTGGCTCCAAGG <sup>l</sup>	
TRXPT_31	E2	4	2334-18751 2334-7062 10981-11079 18159-18189 18684-18751	-	-	-	-	-	-	-	CCGgtaccCTAGTGGCACTGTTGAAAGATTC <sup>m</sup>	CCGctttagCATTGCAGGTATGAATTGGGGAGTAG <sup>n</sup>	

<sup>a</sup>Primer binds inside first exon; <sup>b</sup>Primer binds inside terminal exon; <sup>c</sup>Primer binds inside fourth exon; <sup>d</sup>Agarose gel identical to TRXPT\_7 due to identical splicing; <sup>e</sup>Agarose gel identical to last 3 exons of TRXPT\_10 due to identical splicing; <sup>f</sup>Agarose gel identical to last 4 exons of TRXPT\_11 due to identical splicing;

<sup>g</sup>Identical splicing; <sup>h</sup>Agarose gel identical to TRXPT\_23 due to identical splicing; <sup>i</sup>Agarose gel identical to TRXPT\_9 due to identical splicing; <sup>j</sup>Agarose gel identical to TRXPT\_14 due to identical splicing;



<sup>a</sup>Primer binds inside first exon; <sup>b</sup>Primer binds inside terminal exon; <sup>c</sup>Primer binds inside fourth exon; <sup>i</sup>Agarose gel identical to TRXPT\_7 due to identical splicing; <sup>ii</sup>Agarose gel identical to last 3 exons of TRXPT\_10 due to identical splicing; <sup>iii</sup>Agarose gel identical to last 4 exons of TRXPT\_11 due to identical splicing;

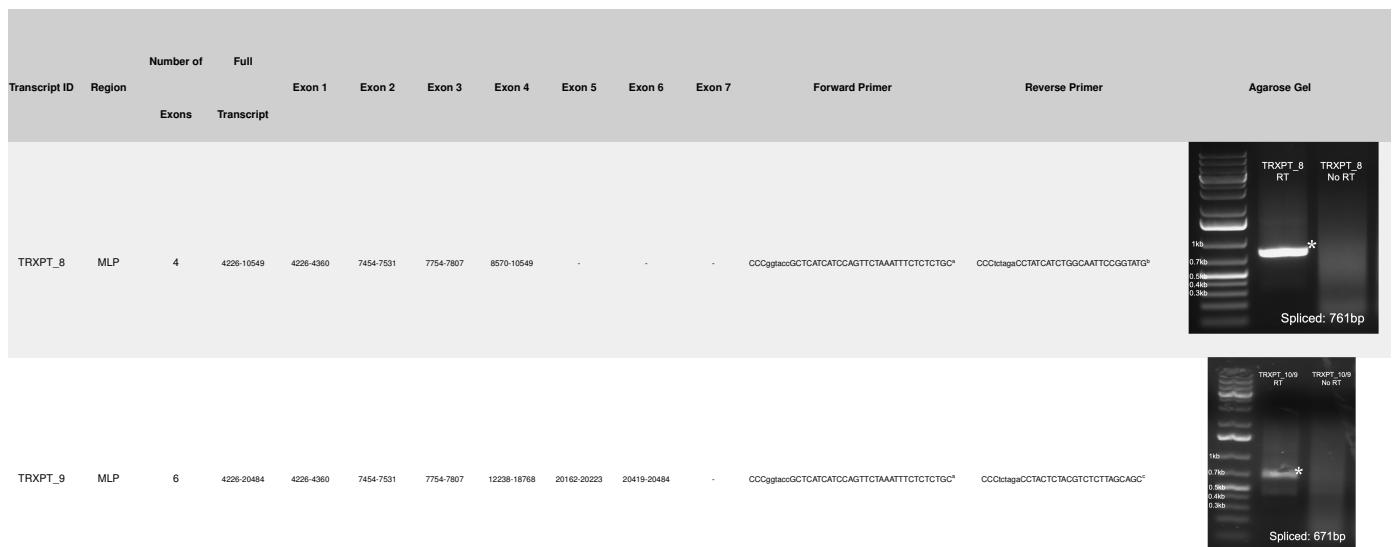
identical splicing; <sup>iv</sup>Agarose gel identical to TRXPT\_23 due to identical splicing; <sup>v</sup>Agarose gel identical to TRXPT\_9 due to identical splicing; <sup>vi</sup>Agarose gel identical to TRXPT\_14 due to identical splicing;



Transcript ID	Region	Number of Exons		Full Transcript							Forward Primer	Reverse Primer	Agarose Gel	
		Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7						
TRXPT_13	MLP	6	4279-22116	4279-4360	7454-7531	7754-7807	18717-18768	20162-20223	20419-22116	-	CCGgtaccGGCTCATCATCAGTTCTAAATTCTCTCTGC*	CCCTctagaGCCAAGCTTGGTCAGGTGAC*		Spliced: 660bp
TRXPT_14	MLP	4	4304-16870	4304-4360	7454-7531	7754-7807	13610-16870	-	-	-	CCGgtaccGGCTCATCCAGTTCTAAATTCTCTCTGC*	CCCTctagaGCTTCAGTATTAGCAGCTGCACAACC*		Spliced: 697bp
TRXPT_16	MLP	4	6934-12709	6934-6969	7454-7531	7754-7807	9430-12709	-	-	-	CCGgtaccGGATCTCCAGATTCTGGTCTGTG*	CCCTctagaGCCTGTCACAACTGCG*		Spliced: 447bp
TRXPT_17	MLP	4	6934-12709	6934-6969	7454-7531	7754-7807	11001-12709	-	-	-	CCGgtaccGGATCTCCAGATTCTGGTCTGTG*	CTCCCCATCTAGACCTTTCTACTG*		Spliced: 430bp
TRXPT_18	MLP	4	6934-12709	6934-6969	7454-7531	7754-7807	12238-12709	-	-	-	CCGgtaccGGATCTCCAGATTCTGGTCTGTG*	CCCTctagaGTTCTCGTCTTACGTGCG*		Spliced: 498bp
TRXPT_19	MLP	2	7401-7836	7401-7531	7754-7836	-	-	-	-	-	-	-		N/A
TRXPT_20	MLP	2	7765-16856	7765-7807	12466-16856	-	-	-	-	-	CCGgtaccGAGGATTGAAGCCAATTACCTCAACG*	CCCTctagaCTGCAGGACAAACAGGTG*		Spliced: 205bp

\*Primer binds inside first exon; ^Primer binds inside terminal exon; ^Primer binds inside fourth exon; ^Agarose gel identical to TRXPT\_7 due to identical splicing; ^Agarose gel identical to last 3 exons of TRXPT\_10 due to identical splicing; ^Agarose gel identical to last 4 exons of TRXPT\_11 due to identical splicing; ^Agarose gel identical to TRXPT\_23 due to identical splicing; ^Agarose gel identical to TRXPT\_9 due to identical splicing; ^Agarose gel identical to TRXPT\_14 due to identical splicing;

identical splicing; ^Agarose gel identical to TRXPT\_20 due to identical splicing; ^Agarose gel identical to TRXPT\_20 due to identical splicing;



\*Primer binds inside first exon; ^Primer binds inside terminal exon; ^Primer binds inside fourth exon; ^Agarose gel identical to TRXPT\_7 due to identical splicing; ^Agarose gel identical to last 3 exons of TRXPT\_10 due to identical splicing; ^Agarose gel identical to last 4 exons of TRXPT\_11 due to identical splicing;

identical splicing; ^Agarose gel identical to TRXPT\_23 due to identical splicing; ^Agarose gel identical to TRXPT\_9 due to identical splicing; ^Agarose gel identical to TRXPT\_14 due to identical splicing;

20 In the table above, the restriction sites in the primer tails are shown in lowercase letters. All the primer  
 21 melting temperatures (TMs) are 58-60°C using a hot start Taq DNA polymerase. The PCR reaction mix was  
 22 done per manufacturer's instructions. The PCR cycling conditions were as follows: Initial denaturation – 95°C  
 23 for 1 minute; cyclical denaturation – 95°C for 30 seconds, annealing – variable temperature (53°C-56°C) for  
 24 30 seconds, primer extension – 68°C for variable time, and final elongation – 68°C for 5 minutes. We used  
 25 35 cycles of amplification.

## 26 Supplementary Computational Analysis

27 Snakemake v7.24.0 was used to manage our entire workflow. A graph of the main steps in our pipeline  
 28 generated with Snakemake is shown below. Our trimmed RNA-seq reads were mapped to the genome of  
 29 *M. gallopano* (with THEV's genome as one of its chromosomes) using Hisat2, to generate the alignment  
 30 (BAM) files and StringTie used to assemble the transcriptome with a GTF annotation file containing the  
 31 predicted THEV ORFs as a guide. The GTF annotation file was derived from a GFF3 annotation file obtained  
 32 from NCBI using Agat – version 1.0.0, a program for converting between many different file formats used  
 33 in bioinformatics. However, the NCBI GFF3 annotation file itself was first modified to remove all unimportant  
 34 features, leaving only the ORFs.

35 StringTie was also used to estimate the normalized expression levels (FPKM) of all the transcripts and

<sup>36</sup> Ballgown in R was used to perform statistical analysis and comparisons of the transcript expression levels,  
<sup>37</sup> which instructive in understanding the temporal regulation THEV gene expression.

<sup>38</sup> In these steps above, each sample (replicate of each time point) was processed independently and merged  
<sup>39</sup> only in the final transcriptome assembly or during analysis with Ballgown. In the subsequent steps described  
<sup>40</sup> below, all samples for each time point were processed together.

<sup>41</sup> We used RegTools to extract and analyze the splice junctions in the BAM files. The command regtools  
<sup>42</sup> junctions extract provides a wealth of information about all the splice sites in the BAM file provided  
<sup>43</sup> such as: the start and end positions, the strand, and number of reads supporting the splice junctions.

<sup>44</sup> The command regtools junctions annotate gives even more information such as: the splice site  
<sup>45</sup> donor-acceptor sequences and transcripts/genes that overlap the junction. These information was the  
<sup>46</sup> basis for estimating and comparing the splicing activity of different regions (TUs) of THEV over time.

<sup>47</sup> Also, Samtools was also used to count the total sequencing reads for all replicates at each time point.

