Supplementary Materials: Elucidating the Transcriptome of Turkey Hemorrhagic Enteritis Virus

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

Abraham Quaye1\*, Bret Pickett\*, Joel S. Griffitts\*, Bradford K. Berges\*, Brian D. Poole\*

\*Department of Microbiology and Molecular Biology, Brigham Young University  
1First-author  
 Corresponding Author

**Corresponding Author Information**  
[brian\_poole@byu.edu](mailto:brian_poole@byu.edu)  
Department of Microbiology and Molecular Biology,  
4007 Life Sciences Building (LSB),  
Brigham Young University,  
Provo, Utah

## Supplementary Table S1A

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% |

## Supplementary Table S1B

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 | + | 29,208 | 25.4% |
| 24hpi | E2 | - | 27,833 | 24.2% |
| 24hpi | E1 | + | 2,724 | 2.4% |
| 24hpi | Unassigned | -,+ | 1,313 | 1.1% |
| 24hpi | IM | - | 744 | 0.6% |
| 24hpi | E4 | - | 664 | 0.6% |

## Supplementary Table S1C

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

| **Time** | **Region** | **Strand** | **Total Reads** | **Percentage** |
| --- | --- | --- | --- | --- |
| 72hpi | MLP | + | 1,436,199 | 67.3% |
| 72hpi | E2 | - | 304,191 | 14.3% |
| 72hpi | E3 | + | 271,310 | 12.7% |
| 72hpi | E1 | + | 74,135 | 3.5% |
| 72hpi | Unassigned | -,+ | 28,921 | 1.4% |
| 72hpi | IM | - | 14,482 | 0.7% |
| 72hpi | E4 | - | 3,568 | 0.2% |

## Supplementary PCR Methods

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 | - | - | - | - | CCCggtaccTTCTGT TTGAATTGTGGGCGGa | CCCtctagaCGTCCA GTAGTCAGGAATTCTAGTGb |  |
| TRXPT\_2 | E1 | 2 | 54-2325 | 54-1655 | 1964-2325 | - | - | - | - | - | CCCggtaccGAGGCCT GTTGGAATTGTTGCa | CCCtctagaCGTCCA GTAGTCAGGAATTCTAGTGb |  |
| TRXPT\_3 | E1 | 2 | 225-2325 | 225-304 | 1964-2325 | - | - | - | - | - | CCCggtacCATTTCCC GTACACGGTGTTGa | CCCtctagaCGTCCA GTAGTCAGGAATTCTAGTGb |  |
| TRXPT\_4 | E1 | 2 | 271-2303 | 271-304 | 1616-2303 | - | - | - | - | - | CCCggtaccGTCATCA CAACTGACCTTGTCGTCa | CCCtctagaCGTCCA GTAGTCAGGAATTCTAGTGb | Not Validated |
| TRXPT\_15 | E2 | 2 | 6206-6878 | 6206-6551 | 6843-6878 | - | - | - | - | - | CCCggtacCCTTTAAA ATCAAGCCTATTGGTCTTGTAACa | CCCtctagaGTGTCATT GTCTACGCTGTTGTAGTAGb |  |
| TRXPT\_21 | E2 | 3 | 16973-18751 | 16973-18087 | 18159-18189 | 18684-18751 | - | - | - | - | CCCggtacCTGT TGCTGAGACTTCGGACCa | CCCtctagaGAACCC AGATATTGGCTCCAAGGb |  |
| TRXPT\_31 | E2 | 4 | 2334-18751 | 2334-7062 | 10981-11079 | 18159-18189 | 18684-18751 | - | - | - | CCCggtacCTAGTGGC AGTGTTCGAAGATTCCa | CCCtctagaCATTGCAGG TATGAATTGCGGAGTAGb |  |
| TRXPT\_6 | E2 | 4 | 2334-18751 | 2334-3447 | 3615-8543 | 10981-11079 | 18159-18189 | 18684-18751 | - | - | CCCggtacCTGT TGCTGAGACTTCGGACCa | CCCtctagaCATTGAATA GATAAGCGTAGCCAATCAGCc | I |
| TRXPT\_7 | E2 | 5 | 2334-18751 | 2334-8543 | 10981-11079 | 18159-18189 | 18684-18751 | - | - | - | CCCggtacCTGT TGCTGAGACTTCGGACCa | CCCtctagaCATTGAATA GATAAGCGTAGCCAATCAGCb |  |
| TRXPT\_22 | E3 | 3 | 18230-22491 | 18230-18350 | 20162-20223 | 20419-22491 | - | - | - | - | CCCggtacCTGA GGAGGTCGTAGACTCTGCa | CCCtctagaGCCA AGCTTGGTCAGGTGACb | II |
| TRXPT\_23 | E3 | 2 | 18230-23884 | 18230-18350 | 20162-23884 | - | - | - | - | - | CCCggtacCTGA GGAGGTCGTAGACTCTGCa | CCCtctagaGCCA AGCTTGGTCAGGTGACb |  |
| TRXPT\_24 | E3 | 4 | 18234-22116 | 18234-18350 | 18717-18768 | 20162-20223 | 20419-22116 | - | - | - | CCCggtacCTGA GGAGGTCGTAGACTCTGCa | CCCtctagaGCCA AGCTTGGTCAGGTGACb | III |
| TRXPT\_25 | E3 | 2 | 18237-23702 | 18237-18350 | 18717-23702 | - | - | - | - | - | CCCggtacCTGA GGAGGTCGTAGACTCTGCa | CCCtctagaGGTAGCACA TACTGTATTGCCTGAAGCb |  |
| TRXPT\_26 | E3 | 3 | 18727-23884 | 18727-18768 | 20162-20223 | 20419-23884 | - | - | - | - | CCCggtaccGTC CGAAGTCTCAGCAACAGATTCa | CCCtctagaGCCAAG CTTGGTCAGGTGACb |  |
| TRXPT\_27 | E3 | 2 | 18727-25168 | 18727-18768 | 22492-25168 | - | - | - | - | - | CCCggtaccGTC CGAAGTCTCAGCAACAGATTCa | CCCtctagaTGCAAT GCTAATCCTCCTGCTGb |  |
| TRXPT\_29 | E3 | 2 | 18230-20732 | 18230-18350 | 20162-20732 | - | - | - | - | - | CCCggtacCTGAGGAG GTCGTAGACTCTGCa | CCCtctagaGCCAAG CTTGGTCAGGTGACb | IV |
| TRXPT\_28 | E4 | 2 | 25192-26247 | 25192-25701 | 26055-26247 | - | - | - | - | - | CCCggtaccGGACAC GTGTTCGTTAGAGAACCa | CCCtctagaCAGTG CAATCCGACGCTCTGb |  |
| TRXPT\_5 | IM | 2 | 2334-3678 | 2334-3447 | 3615-3678 | - | - | - | - | - | CCCggtaccTCTGGTGAGA TCTTCCAAACAGAAAGa | CCCtctagaCGCAA CCTGTAGGTCCGATTACb |  |
| TRXPT\_10 | MLP | 7 | 4226-22116 | 4226-4360 | 7454-7531 | 7754-7807 | 12238-18350 | 20162-20223 | 20419-22116 | - | CCCggtaccGCTCATCATC CAGTTCTAAATTTCTCTCTGCa | CCCtctagaCCTACTC TACGTCTCTTAGCAGCc | V |
| TRXPT\_11 | MLP | 6 | 4226-22116 | 4226-4360 | 7454-7531 | 7754-7807 | 13610-18350 | 18717-18768 | 20162-20223 | 20419-22116 | CCCggtaccGCTCATCATC CAGTTCTAAATTTCTCTCTGCa | CCCtctagaGCTTCAG TATTAGCAGCTGCACAACCc | VI |
| TRXPT\_12 | MLP | 4 | 4226-25168 | 4226-4360 | 7454-7531 | 7754-7807 | 22492-25168 | - | - | - | CCCggtaccGCTCATCATC CAGTTCTAAATTTCTCTCTGCa | CCCtctagaTTTCC AGCTGAAGCCTGGAGb |  |
| TRXPT\_13 | MLP | 6 | 4279-22116 | 4279-4360 | 7454-7531 | 7754-7807 | 18717-18768 | 20162-20223 | 20419-22116 | - | CCCggtaccGCTCATCATC CAGTTCTAAATTTCTCTCTGCa | CCCtctagaGCCAAG CTTGGTCAGGTGACb |  |
| TRXPT\_14 | MLP | 4 | 4304-16870 | 4304-4360 | 7454-7531 | 7754-7807 | 13610-16870 | - | - | - | CCCggtaccGCTCATCATC CAGTTCTAAATTTCTCTCTGCa | CCCtctagaGCTTCAGT ATTAGCAGCTGCACAACCb |  |
| TRXPT\_16 | MLP | 4 | 6934-12709 | 6934-6969 | 7454-7531 | 7754-7807 | 9430-12709 | - | - | - | CCCggtaccGGATCTC CAGATTCTGGTCTGTGa | CCCtctagaGCCT GTCCAACAACCTGCb |  |
| TRXPT\_17 | MLP | 4 | 6934-12709 | 6934-6969 | 7454-7531 | 7754-7807 | 11001-12709 | - | - | - | CCCggtaccGGATCTC CAGATTCTGGTCTGTGa | CTCCCCATCTAGAC CTTTCATCTAACTGb |  |
| TRXPT\_18 | MLP | 4 | 6934-12709 | 6934-6969 | 7454-7531 | 7754-7807 | 12238-12709 | - | - | - | CCCggtaccGGATCTC CAGATTCTGGTCTGTGa | CCCtctagaGTTCTC CGTCTTCTACGTCGTGb |  |
| TRXPT\_19 | MLP | 2 | 7401-7836 | 7401-7531 | 7754-7836 | - | - | - | - | - |  |  | N/A |
| TRXPT\_20 | MLP | 2 | 7765-16856 | 7765-7807 | 12466-16856 | - | - | - | - | - | CCCggtaccGAGGATTTGA AGCCAATTATCCTTCAACGa | CCCtctagaCTGCA GGCACAACAGGTGb |  |
| TRXPT\_8 | MLP | 4 | 4226-10549 | 4226-4360 | 7454-7531 | 7754-7807 | 8570-10549 | - | - | - | CCCggtaccGCTCATCAT CCAGTTCTAAATTTCTCTCTGCa | CCCtctagaCCTATC ATCTGGCAATTCCGGTATGb |  |
| TRXPT\_9 | MLP | 6 | 4226-20484 | 4226-4360 | 7454-7531 | 7754-7807 | 12238-18768 | 20162-20223 | 20419-20484 | - | CCCggtaccGCTCATCAT CCAGTTCTAAATTTCTCTCTGCa | CCCtctagaCCTACT CTACGTCTCTTAGCAGCc |  |
| aPrimer binds inside first exon; bPrimer binds inside terminal exon; cPrimer binds inside fourth exon; IAgarose gel identical to TRXPT\_7 due to identical splicing; IIAgarose gel identical to last 3 exons of TRXPT\_10 due to identical splicing; IIIAgarose gel identical to last 4 exons of TRXPT\_11 due to identical splicing; IVAgarose gel dentical to TRXPT\_23 due to identical splicing; VAgarose gel identical to TRXPT\_9 due to identical splicing; VIAgarose gel identical to TRXPT\_14 due to identical splicing; | | | | | | | | | | | | | |

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

## Supplementary Computational Analysis

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

StringTie was also used to estimate the normalized expression levels (FPKM) of all the transcripts and Ballgown in R was used to perform statistical analysis and comparisons of the transcript expression levels, which instructive in understanding the temporal regulation THEV gene expression.

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

We used RegTools to extract and analyze the splice junctions in the BAM files. The command regtools junctions extract provides a wealth of information about all the splice sites in the BAM file provided such as: the start and end positions, the strand, and number of reads supporting the splice junctions. The command regtools junctions annotate gives even more information such as: the splice site donor-acceptor sequences and transcripts/genes that overlap the junction. These information was the basis for estimating and comparing the splicing activity of different regions (TUs) of THEV over time. Also, Samtools was also used to count the total sequencing reads for all replicates at each time point. 