Turkey Hemorrhagic Enteritis Virus transcriptome profiling

Running Title: Novel Insights into Turkey Hemorrhagic Enteritis Virus Transcriptome

Abraham Quaye[1], Bret Pickett, Joel S. Griffitts, Bradford K. Berges, Brian D. Poole[†]

† **Corresponding Author**

Department of Microbiology and Molecular Biology,

Brigham Young University,

A-153ASB Provo, Utah

ABSTRACT

Background: Hemorrhagic enteritis (HE)is a disease affecting 6-12-week-old turkeys characterized by **immunosuppression (IS)**, bloody diarrhea, and up to 60% mortality. This disease is caused by ***Turkey Hemorrhagic Enteritis Virus (THEV)*** of which avirulent strains (THEV-A) that do not cause HE but retain the immunosuppressive ability have been isolated. The THEV-A Virginia Avirulent Strain (VAS) is still used as a live vaccine despite its immunosuppressive properties. ***Our objective is to understand the genetic basis by which VAS brings about IS***. The **transcriptome of THEV** was studied to set the stage for further experimentation with specific viral genes that may mediate IS.

Methods: After infecting a turkey B-cell line (MDTC-RP19) with the VAS vaccine strain, samples in triplicates were collected at 4-, 12-, 24-, and 72-hours post-infection. Total RNA was subsequently extracted, and poly-A-tailed mRNA sequencing done at LC Sciences. The raw sequencing reads were trimmed with the FastQC program, mapped to the THEV genome using Hisat2 and transcripts assembled with StringTie. An in-house script was used to consolidate transcripts from all timepoints, generating the final transcriptome.

Results: A total of ***18.07*** million reads mapped to THEV genome providing good coverage/depth, leaving no regions unmapped. All predicted genes in the genome were represented. In keeping with all adenoviruses, all transcripts were spliced with either with 5'- or 3'-multi exon UTRs hitherto unknown. ThirteenXXXXX novel exons were identified which were validated by PCR and Sanger sequencing. The splicing patterns strongly suggest that there are 3XXXXX main promoters (E1, E3, and major late promoters) driving expression of most of the genes with 2XXXXXX possible minor promoters driving single genes (ORF7 and ORF8).

Conclusions: This RNA-sequencing experiment is the first study of THEV gene expression to date. In keeping with other Adenoviruses, almost all THEV genes are spliced, and several genes are expressed as one transcription unit under a single promoter. This insight into THEV's transcriptome may allow the engineering of the VAS to provide immune protection with less or no associated IS.

INTRODUCTION

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

Virulent strains (THEV-V) and avirulent strains (THEV-A) of THEV are serologically indistinguishable, infecting turkeys, chickens, and pheasants. They infect via the fecal-oral route, and the THEV-V cause different clinical diseases in these birds (2, 11). In turkeys, the THEV-V cause hemorrhagic enteritis (HE), a debilitating acute disease affecting predominantly 6-12-week-old turkeys characterized by immunosuppression (IS), weight loss, intestinal lesions leading to bloody diarrhea, splenomegaly, and up to 60% mortality (11, 12). HE is the most economically significant disease caused by any strain of THEV (11). While the current vaccine strain (a THEV-A isolated from a pheasant, Virginia Avirulent Strain [VAS]) have proven effective at preventing HE in young turkey poults, it still retains the immunosuppressive ability. Thus, vaccinated birds are rendered more susceptible to opportunistic infections and death than unvaccinated cohorts leading to substantial economic losses (11, 13–15). The induced IS also interferes with vaccination schemes for other infections of turkeys (11, 13). To eliminate this immunossupressive side-effect of the vaccine, a thorough understanding of the viral factors (genes) mediating this phenomenon is essential. However, the transcriptome (splicing and gene expression patterns) of THEV has not been characterized, making the investigation of specific viral genes for possible roles in causing IS impractical. To identify such viral factors, a well-characterized transcriptome of THEV is required to set the stage for experimentation with specific

viral genes that may mediate IS.

MAdVs - specifically human adenoviruses - are by far the most extensively studied; hence, most of the current knowledge regarding AdV gene expression and replication is based on MAdV studies, which is used representatively for all other AdVs (6, 16). MAdV genes are transcribed in a temporal manner; therefore, genes are categorized into five early transcription units (E1A, E1B, E2, E3, and E4), two intermediate units (pIX and IVa2), and one major late unit (MLTU), generating five families of late mRNAs (L1-L5). An additional gene (UXP or U exon) is located on the reverse strand. The early genes encode non-structural proteins such as enzymes or host cell modulating proteins, primarily involved in DNA replication or providing the necessary intracellular niche for optimal replication while late genes encode structural proteins. The immediate early gene E1A is expressed first, followed by the the delayed early genes, E1B, E2, E3 and E4. Then the intermediate early genes, IVa2 and pIX are expressed, after which the late genes are expressed (6, 17, 18). MAdV makes an extensive use of alternative RNA splicing to produce a very complex array of mRNAs; all but pIX mRNA undergo at least one splicing event. The MLTU produces over 20 distinct splice variants all of which contain three non-coding exons at the 5'-end (collectively known as the tripartite leader, TPL) (17, 18). There is an alternate 5' three non-coding exons present in varying amounts on a subset of MLTU mRNAs (known as the x-, y- and z-leaders). Lastly, there is the i-leader exon, which is infrequently included between the second and third TPL exons, and codes for the i-leader protein (19). Thus, the MLTU produces a complex repertoire of mRNA with diverse 5'-UTRs, spliced onto different 3' coding exons grouped into five different 3'-end classes (L1-L5). Each transcription unit contains its own promoter that drives the expression of all the array of mRNA transcripts produced via alternative splicing of the genes encoded in the unit(6, 17, 18).

Almost all AdV mRNAs are generated by the excision of one or more introns and most of these introns are located in the 5' or 3' UTRs of pre-mRNA. Thus the viral introns do only in a few cases interrupt the open reading frames (ORFs) (1, 17). The development of high throughput sequencing methods has facilitated the discovery of many novel transcribed regions and splicing isoforms. It is also a very powerful tool to study alternative splicing under different conditions at an unparalleled depth (17, 20). Here, a paired-end deep sequencing experiment was performed to characterize for the first time, the transcriptome of THEV (VAS strain) during different phases of the infection, yielding a complete THEV splicing map. Our paired-end sequencing allowed for reading ***149*** bp long sequences from each end of cDNA fragments. The data generated from paired-end sequencing, utilized in our experiment, should thus be reliable.

RESULTS

Overview of sequencing data and analysis pipeline outputs

A previous study by Zeinab *et al* showed that almost all THEV transcripts were detectable beginning at 4 hours(21); hence, infected MDTC-RP19 cells were collected in triplicates at 4-, 12-, 24-, and 72-hours post-infection(h.p.i) to ensure an amply wide window to sample all transcripts. Our ***149*** bp paired-end RNA sequencing experiment yielded an average of ***106.68*** million total reads per timepoint and the reads mapping to the virus genome increased dramatically from ***6148*** reads at 4 h.p.i to ***16.88*** million reads at 72 h.p.i. Altogether, ***18.07*** million reads from all timepoints mapped to THEV genome providing good coverage/depth, leaving no regions unmapped (**Table 1**).

METHODS

Cell culture and THEV Infection

The Turkey B-cell line (MDTC-RP19, ATCC CRL-8135) was grown as suspension cultures in 1:1 complete Leibovitz's L-15/McCoy's 5A medium with 10% fetal bovine serum (FBS), 20% chicken serum (ChS), 5% tryptose phosphate broth (TPB), and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Streptomycin), at 41°C in a humidified atmosphere with 5% CO_2. When infected, the cells were maintained in 1:1 serum-reduced Leibovitz's L15/McCoy's 5A media (SRLM) with 2.5% FBS, 5% ChS, 1.2% TPB, and 1% antibiotics solution (100 U/mL Penicillin and 100ug/mL Streptomycin). A commercially available HE vaccine was purchased from Hygieia Biological Labs as a source of THEV-A (VAS strain). The stock virus was titrated using an in-house qPCR assay with titer expressed as genome copy number(GCN)/mL. Cells were infected at a multiplicity of infection (MOI) of 100 GCN/cell and samples in triplicates were harvested at 4-, 12-, 24-, and 72-h.p.i for RNA sequencing. A second infection was done but samples in triplicates were harvested at 12-, 24-, 36-, 48-, and 72-h.p.i for PCR validation of novel splice sites.

RNA extraction and Sequencing

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

Computational Analysis of RNA Sequencing Data: Mapping and Transcript characterization

Analysis of our sequencing reads were analyzed following a well established general outline described by Pertea *et al* (22). Briefly, sequencing reads were trimmed with the FastQC - version 0.11.9 (23) program to acheive an overall Mean Sequence Quality (Phred Score) of 36. Trimmed reads were mapped to the complete sequence of avirulent turkey hemorrhagic enteritis virus strain Virginia (https://www.ncbi.nlm.nih.

gov/nuccore/AY849321.1/) using `Hisat2 - version 2.2.1` (22) with default settings except increasing the CPU cores to 10 without relying on known splice sites. The generated `BAM` files from each infection timepoint were fed into `StringTie - version 2.2.1` (22) using a `gff3` file from NCBI containing the predicted genes of THEV as a guide. A custom script was used to consolidate all transcripts from all timepoints without redundancy, generating the final transcriptome of THEV.

PCR Validation of Novel Splice Sites and Sanger Sequencing

All splice sites identified in this work are novel except one splice site predicted for pTP, DBP, and 33K each. However, these predictions fell short of the full splicing of these genes and were not experimentally validated. We designed primers that crossed a range of novel exon–exon boundaries for all transcripts in a transcription units with their respective universal primers (supplementary PCR methodsXXXXXXXXX). After first-strand cDNA synthesis with Thermofishers' SuperScript™ III First-Strand Synthesis System (`#18080051`), these primers were used in a targeted PCR experiment, the PCR products were analysed on agarose gels, cloned by traditional restriction enzyme-ligation method and sanger sequenced at `Eton Bioscience, Inc, San Diego, CA` to validate these splice junctions at the sequence level.

3' Rapid Amplification of cDNA Ends (3'RACE)

DISCUSSION/CONCLUSIONS

SCRIPTS AND SUPPLEMENTARY MATERIALS

DATA AVAILABILITY

CODE AVAILABILITY

All the code/scripts written for analysis of the data is available on github (linkXXXXXX)

ACKNOWLEDGMENTS

REFERENCES

1. Davison A, Benko M, Harrach B. 2003. Genetic content and evolution of adenoviruses. The Journal of general virology 84:2895–908.

2. Harrach B. 2008. Adenoviruses: General features, p. 1–9. *In* Mahy, BWJ, Van Regenmortel, MHV (eds.), Encyclopedia of virology (third edition). Book Section. Academic Press, Oxford.

3. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: Toward defining the minimum essential poxvirus genome. Journal of virology 77:7590–7600.

4. McGeoch D, Davison AJ. 1999. Chapter 17 - the molecular evolutionary history of the herpesviruses, p. 441–465. *In* Domingo, E, Webster, R, Holland, J (eds.), Origin and evolution of viruses. Book Section. Academic Press, London.

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 Report of the International Committee on Taxonomy of Viruses 125–141.

6. Guimet D, Hearing P. 2016. 3 - adenovirus replication, p. 59–84. *In* Curiel, DT (ed.), Adenoviral vectors for gene therapy (second edition). Book Section. Academic Press, San Diego.

7. Kovács ER, Benkő M. 2011. Complete sequence of raptor adenovirus 1 confirms the characteristic genome organization of siadenoviruses. Infection, Genetics and Evolution 11:1058–1065.

8. Davison AJ, Wright KM, Harrach B. 2000. DNA sequence of frog adenovirus. J Gen Virol 81:2431–2439.

9. Kovács ER, Jánoska M, Dán Á, Harrach B, Benkő M. 2010. Recognition and partial genome characterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample originating from parus major, a great tit. Journal of Virological Methods 163:262–268.

10. Katoh H, Ohya K, Kubo M, Murata K, Yanai T, Fukushi H. 2009. A novel budgerigar-adenovirus belonging to group II avian adenovirus of siadenovirus. Virus Research 144:294–297.

11. Beach NM. 2006. Characterization of avirulent turkey hemorrhagic enteritis virus: A study of the molecular basis for variation in virulence and the occurrence of persistent infection. Thesis.

12. Gross WB, Moore WE. 1967. Hemorrhagic enteritis of turkeys. Avian Dis 11:296–307.

13. Rautenschlein S, Sharma JM. 2000. Immunopathogenesis of haemorrhagic enteritis virus (HEV) in turkeys. Dev Comp Immunol 24:237–46.

14. Larsen CT, Domermuth CH, Sponenberg DP, Gross WB. 1985. Colibacillosis of turkeys exacerbated by hemorrhagic enteritis virus. Laboratory studies. Avian Dis 29:729–32.

15. Dhama K, Gowthaman V, Karthik K, Tiwari R, Sachan S, Kumar MA, Palanivelu M, Malik YS, Singh RK, Munir M. 2017. Haemorrhagic enteritis of turkeys – current knowledge. Veterinary Quarterly 37:31–42.

16. Wolfrum N, Greber UF. 2013. Adenovirus signalling in entry. Cell Microbiol 15:53–62.

17. Zhao H, Chen M, Pettersson U. 2014. A new look at adenovirus splicing. Virology 456-457:329–341.

18. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. 2020. Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. Communications Biology 3:124.

19. Falvey E, Ziff E. 1983. Sequence arrangement and protein coding capacity of the adenovirus type 2 "i" leader. Journal of Virology 45:185–191.

20. 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 cells. Nature 489:101–108.

21. Aboezz Z, Mahsoub H, El-Bagoury G, Pierson F. 2019. In vitro growth kinetics and gene expression analysis of the turkey adenovirus 3, a siadenovirus. Virus Research 263:47–54.

22. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. Nature Protocols 11:1650–1667.

23. 2015. FastQC.

TABLES AND FIGURES

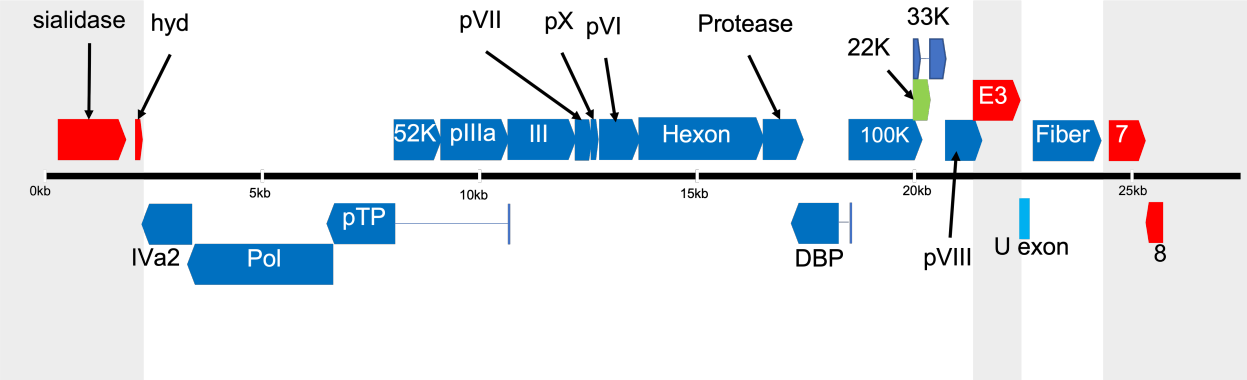

Figure 1. ***Genomic map of THEV avirulent strain***. The central horizontal line represents the double-stranded DNA marked at 5kb intervals as white line breaks. Blocks represent viral genes. Blocks above the DNA line are transcribed rightward, those below are transcribed leftward. pTP, DBP and 33K predicted to be spliced are shown as having tails. Shaded regions indicates regions containing genes (colored red) that are not conserved across adenovirus genera (i.e. “genus-specific”). Genes colored in blue are “genus-common”. Gene colored in light green is conserved in all but Atadenoviruses. The U exon (light blue) is an incomplete gene present in almost all AdVs.