# Meta-Analysis of Transcriptomic Datasets for the Investigation of Differential Expression in Hantavirus-Infected Human Tissue

Krapohl J, Pickett BE

## Introduction

*Orthohantavirus*, commonly referred to as hantaviruses, is a viral genus that is found across Asia, Europe, the Americas, and parts of Africa, all within various rodent hosts. Hantavirus is generally well adapted to its rodent host and is unlikely to cause disease. However, when humans encounter infected mice or their feces, virions can potentially cause an infection in a human host. Within humans, infection is far more severe and can cause serious permanent injury or even death. Except for some rare cases of *Andes virus*, spread of hantavirus between human hosts does not occur [1]. However, as cities expand, threat of hantavirus spillover into humans is increasingly becoming a threat. This is especially true in eastern Asia, where thousands of cases of hemorrhagic fever with renal syndrome (HFRS) due to hantavirus infection are reported per year. New world hantaviruses disease, hantavirus cardiopulmonary syndrome (HCPS), is far less common, but has been reported to have up to a 50% fatality rate [2]. Hantavirus infections are characterized by the loss of endothelial barrier integrity leading to severe symptoms and death, as wells as high levels of T-cell proliferation in affected tissues [3]. Interestingly, several experiments have shown that hantavirus infection causes no cytopathic effects *in vitro*. Instead, the highly elevated levels of both T-cells and cytokines in infected tissue indicate that host inflammatory response are responsible for the observed symptoms [4].

While several studies have characterized the proteomics and health outcomes of infected individuals, little has been done to transcriptionally characterize human infection. Further, no studies that examine multiple tissue types in a hantavirus infection have been reported. Evidence of differences in differential gene expression (DE) across tissues would provide evidence that further research is required to understand how DE affects differences in morbidity in human tissues. This study seeks to further examine which host pathways are impacted by infection by combining existing RNA-seq data of hantavirus infection for the analysis of DE *in vivo* and *in vitro*.

## Methods

A search of the Gene Expression Omnibus (GEO) database [5], hosted at the National Center for Biotechnology Information (NCBI), was performed to find RNA-sequencing datasets for hantavirus-infected human tissue. The corresponding sequencing data for six GEO series were retrieved from the Sequence Read Archive (SRA): GSE158712, GSE161354, GSE133319, GSE133634, GSE133751, GSE73410 [6-9]. In total, this represents 73 samples and contains all publicly available datasets. These datasets were generated from cell cultures or patients infected with one of:  Hantaan Virus (HTNV), Dobrava-Belgrade virus (DOBV), Puumala virus (PUUV), Tula Virus (TULV) and Prospect Hill virus (PHV). While analyses were run on each set individually, they were more broadly grouped into in-vitro samples and in-vivo samples. The Automated Reproducible MOdular Workflow for Preprocessing and Differential Analysis of RNA-seq Data (ARMOR) was used to perform quality control, trim reads, map to the reference transcriptome for *Homo sapiens* (Ensembl build GRCh38, release 98), and calculate the log-fold change of differentially expressed genes [10]. DE was defined as significant using a false discovery rate (FDR) adjusted p-value of <0.05. Ensembl identifiers for DE genes identified by ARMOR were converted to their analogous Entrez Gene Identifiers using Bioconductor and Biomart [11-12]. Subsequently, this gene list and corresponding metadata were processed using Signaling Pathway Impact Analysis (SPIA). Using the KEGG, Reactome, Panther, NCI, and BioCarta databases, SPIA identifies perturbed cellular pathways from the DE data [13-16].

## Results

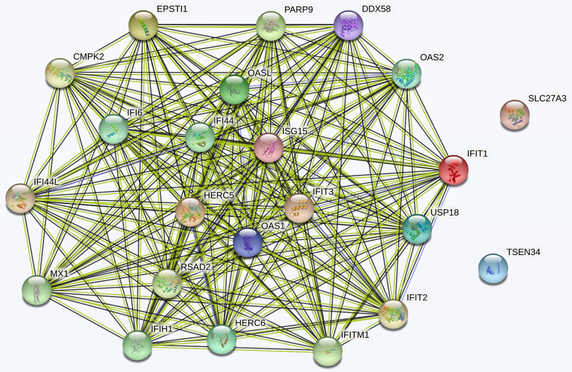


**Fig 1.** Protein-protein interaction network (from the STRING database) of *in vitro* statistically significant DE genes (FDR adjusted p-value <0.05). 25 genes directly interact with each other, while SLC27A3 (Very Long-Chain Acyl-CoA Synthetase) and TSEN34 (tRNA Splicing Endonuclease Subunit 34) remain unconnected [17].

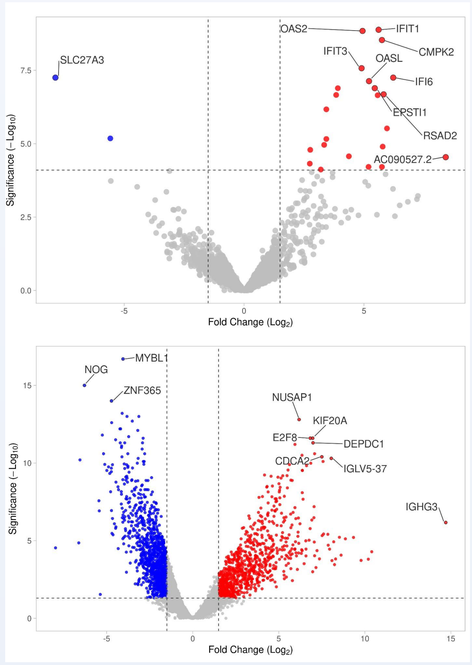


**Fig 2.** Volcano plots of *in vitro* (top) *in vivo* (bottom). Differentially expressed genes (LFC ≥2, FDR adjusted p-value <0.05) are depicted in red (upregulated) and blue (downregulated). In total 26 genes in the *in vitro* dataset and 1134 in the *in vivo* dataset were significant.

**Table 1.** *In vitro* significant cellular pathways identified by SPIA.*

| Name | Status | SourceDB |
|---|---|---|
| Influenza A | Activated | KEGG |
| Herpes simplex infection | Activated | KEGG |
| Measles | Activated | KEGG |
| RIG-I-like receptor signaling pathway | Activated | KEGG |
| Interferon Signaling | Inhibited | Reactome |
| Interferon alpha/beta signaling | Activated | Reactome |
| Cytokine Signaling in Immune system | Inhibited | Reactome |
| Antiviral mechanism by IFN-stimulated genes | Inhibited | Reactome |
| ISG15 antiviral mechanism | Inhibited | Reactome |
| Negative regulators of RIG-I/MDA5 signaling | Activated | Reactome |
| RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways | Inhibited | Reactome |

**Table 2**. *In vivo* significant cellular pathways identified by SPIA.*

| Name | Status | SourceDB |
|---|---|---|
| Cytokine-cytokine receptor interaction | Inhibited | KEGG |
| Systemic lupus erythematosus | Activated | KEGG |
| Cell cycle | Activated | KEGG |
| Carbohydrate digestion and absorption | Activated | KEGG |
| Staphylococcus aureus infection | Activated | KEGG |
| Alcoholism | Activated | KEGG |
| Generic Transcription Pathway | Activated | Reactome |
| Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3 | Activated | Reactome |
| RNA Polymerase I Promoter Opening | Inhibited | Reactome |
| Meiotic recombination | Activated | Reactome |

The top 10 are shown above and 79 pathways were identified in the *in vivo* dataset.

*FDR adjusted p-value <0.05)

## Discussion

As expected, the analysis of the *in vivo* dataset identified substantially more DE genes than what was found in the *in vitro*. Many of the genes identified relate to cellular repair, replication, or other housekeeping functions. In fact, SPIA identified only 4 pathways directly associated with immune response in the *in vivo* dataset that were perturbed, as compared to 11 identified in the *in vitro* dataset.

In the STRING plot, we see that of the 25 genes identified, 23 are highly connected, and are associated with interferon activation. However, SLC27A3 and TSEN34 were found to be significantly dysregulated, but have no known function associated with viral infection. It is possible these genes may have unknown secondary functions [17].

Interestingly, SLC27A3 is also one of 3 genes downregulated in the *in vitro* dataset, along with TRGV3 (T cell receptor gamma variable 3) and IGHG2 (immunoglobulin heavy constant gamma 2). In contrast, IGHG2 and IGHG3 are highly upregulated *in vivo*. The presence of IGHG2 in endothelial tissue in our data could be an artifact of our data processing, however recent studies have shown the IgG can be produced in some endothelial tissues and cancer lines [18].

SPIA results showed that some interferon pathways are inhibited while others are activated. This may indicate an incomplete inhibition of the interferon response, as hantavirus is adapted to a murine host, not a human one. Further investigation of genes in this pathway may further elucidate hantavirus pathogenesis.

## Future Directions

As several genes of interest have been identified, it would be worthwhile to look at differential expression of these genes *in vitro* using multiple cell lines and viral species. This may increase our understanding of the differences in pathogenesis between HCPS and HFRS causing strains of hantavirus. In addition, this would illuminate how hantavirus has varying success in masking infection between tissue types.

## Acknowledgements

We would like to thank the Office of Research Computing at Brigham Young University for their support on this project.

## References

1. Martinez, V.P., et al., *Person-to-person transmission of Andes virus.* Emerg Infect Dis, 2005. **11**(12): p. 1848-53.
2. Zhang, R., et al., *The changing epidemiology of hemorrhagic fever with renal syndrome in Southeastern China during 1963-2020: A retrospective analysis of surveillance data.* PLoS Negl Trop Dis, 2021. **15**(8): p. e0009673.
3. Hepojoki, J., A. Vaheri, and T. Strandin, *The fundamental role of endothelial cells in hantavirus pathogenesis.* Front Microbiol, 2014. **5**: p. 727.
4. Vaheri, A., et al., *Uncovering the mysteries of hantavirus infections.* Nat Rev Microbiol, 2013. **11**(8): p. 539-50.
5. Orjuela, S., et al., *ARMOR: An Automated Reproducible MOdular Workflow for Preprocessing and Differential Analysis of RNA-seq Data.* G3 (Bethesda), 2019. **9**(7): p. 2089-2096.
6. Li, Y., et al., *Rapid humoral immune responses are required for recovery from haemorrhagic fever with renal syndrome patients.* Emerg Microbes Infect, 2020. **9**(1): p. 2303-2314.
7. Zhang, J., et al., *The Presence of Circulating Nucleated Red Blood Cells Is Associated With Disease Severity in Patients of Hemorrhagic Fever With Renal Syndrome.* Front Med (Lausanne), 2021. **8**: p. 665410.
8. Lu, S., et al., *RNA-Seq Revealed a Circular RNA-microRNA-mRNA Regulatory Network in Hantaan Virus Infection.* Front Cell Infect Microbiol, 2020. **10**: p. 97.
9. Witkowski, P.T., et al., *Infection of human airway epithelial cells by different subtypes of Dobrava-Belgrade virus reveals gene expression patterns corresponding to their virulence potential.* Virology, 2016. **493**: p. 189-201.
10. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.
11. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nat Methods, 2015. **12**(2): p. 115-21.
12. Durinck, S., et al., *BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis.* Bioinformatics, 2005. **21**(16): p. 3439-40.
13. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
14. Jassal, B., et al., *The reactome pathway knowledgebase.* Nucleic Acids Res, 2020. **48**(D1): p. D498-d503.
15. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Res, 2017. **45**(D1): p. D183-d189.
16. Schaefer, C.F., et al., *PID: the Pathway Interaction Database.* Nucleic Acids Res, 2009. **37**(Database issue): p. D674-9.
17. Szklarczyk, D., et al., *STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Research, 2018. **47**(D1): p. D607-D613.
18. Zhao, Y., et al., *Immunoglobulin g (IgG) expression in human umbilical cord endothelial cells.* J Histochem Cytochem, 2011. **59**(5): p. 474-88.