Expression Variations Associated with Influenza Infection Severity in Humans and Mice

Jeffrey Hunter

January 2017

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Thesis of

Jeffrey M. Hunter

"Expression Variations Associated with Influenza Infection Severity in Humans and Mice"

Has been approved

Thesis Advisor – Dr. Shannon McWeeney

Committee Member – Dr. Beth Wilmot

Committee Member – Dr. Melissa Haendel

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Influenza

Influenza is a disease that is estimated by the World Health Organization (WHO) to affect 5-10% of adults and 20-30% of children each year.[1] Of these, there are an estimated 3-5 million severe cases worldwide resulting in between 250-500 thousand deaths a year. These infections are caused by the influenza virus, of which there are three types (A, B, and C). The type A viruses are capable of infecting a wide range of hosts including humans, pigs, horses, and predominantly a wide range of birds[2,3]. Additionally there is a wide range of antigenically distinct subtypes of the A viruses which are thought to arise in part due to the many types of hosts it can present in. These subtypes are categorized by two viral proteins, hemagglutinin (HA) of which there are 15 recorded types and neuraminidase (NA) of which there are 9 types[2,3]. Currently the A(H1N1) and A(H3N2) subtypes are the most common among humans[1]. The B type influenza viruses have historically been responsible for some severe epidemics and continue to circulate among populations annually, the impact of these viruses overall appear to be much lower than type A influenza due to the lower rate at which antigenically different viruses appear in the population[2,3]. This is thought to be potentially influenced by the much more restricted range of hosts as the type B viruses are limited to infecting primarily humans and occasionally seals[3,4]. The type C virus does appear to also be wide spread in human populations, as the majority of adults appear to have antibodies to the virus, and type C influenza is typically known for causing more mild infection symptoms and appears to primarily cause illness in young children[5,6].

1.2 Symptoms

Symptoms of an uncomplicated influenza infection typically include fever, cough, sore throat, muscle pain, and nasal congestion.[1,7] For an influenza infection to be classified as severe,

however, there are four possible criteria. The first criterion is if a patient presents with: clinical or radiological signs of lower respiratory tract disease (e.g. shortness of breath, tachypnea, hypoxia, pneumonia), central nervous system involvement (e.g. encephalopathy, encephalitis), secondary complications (e.g. renal failure, multiple organ failure, septic shock), and other complications (e.g. severe dehydration, rhabdomyoliysis, myocarditis). The second criterion that will earn a severe classification is if the influenza infection exacerbates a preexisting chronic disease in the patient. A patient who has any other condition requiring hospital admission alongside or because of influenza will have their infection classified as severe. Lastly, a patient who initially presents with a mild influenza infection that progresses to meet any of the other criteria is considered to have a severe infection.[7]

As these criteria would indicate, there is a significant amount of variation in the symptoms observed.[8,9] Currently there seem to be a large number of factors contributing to this wide range of symptoms and severity. The strain and quantity of virus a patient is exposed to seems, unsurprisingly, to play a major role in the severity of an influenza infection.[10,11] Environmental factors have also been found to contribute to the occurrences and severity of influenza.[12] Additionally, there are many differences between patients that seem to exhibit a substantial influence on the severity of an infection. Age is a major factor, with the very young and very old showing the highest rates and risk of hospitalization due to influenza.[13,14] Chronic illnesses have also been understandably linked to severe infection rates.[14] Responses to influenza infections have also been found to have a significant genetic component.[15,16]

1.3 Collaborative Cross Mouse Model

The complexity seen underlying influenza infection severity has made it difficult to understand the influence of each individual component. The use of a more controllable model is therefore desirable to allow for identification and quantification of specific influences. Inbred mouse strains have been a frequently used model in biological research, and for influenza specifically, due to both the ease of breeding and the high amount of reproducibility.[11,17,18] Examining these strains has allowed for identification of some genetic factors affecting influenza

infections, and examining effects between strains has reinforced the knowledge of host genetic variations contributing to infection severity.[11] One of the largest disadvantages of the inbred mouse strains as models for human disease is the lack of genetic diversity.[19] Even comparing between strains cannot accurately approximate human outbred genetic diversity. In order to address this issue, the recently established collaborative cross takes 8 parental strains to outbreed in a controlled manner for use in studies.[20] These mice much more accurately reflect observed human population variation, while maintaining much of the reproducibility that was the hallmark of the individual mouse strains.

1.4 Ontologies

In order to compare the wealth of information the collaborative cross can provide with other species, it is necessary to have computational tools and methods for accessing and understanding the data obtained. This often involves making use of ontologies, which are structured and controlled vocabularies that describe and annotate knowledge.[21] These ontologies work by defining the vocabulary and structure to be recorded, and by defining the relationships between those objects. These databases are often very specific in scope. Current ontologies tend to be divided by species, and often further by category of information. Thus the Human Phenotype ontology and the Mouse Phenotype ontology are two completely separate entities curated independently. There are also further divisions depending on field, such as the mouse anatomy project and Edinburgh Mouse Anatomy Project which detail mouse anatomical structures in adult and developing mice respectively.[22]

While these ontologies predominantly use their controlled vocabularies to establish terms and relationships that are consistent within a given ontology, there are often much greater differences between these collections. Such differences have drastically limited the ability to integrate data and findings between species, and have compounded the difficulties already arising from different conceptualizations of phenotypes in different species as well as species-specific anatomies.[23] There has been some progress in the last few years on mapping out related terms between species ontologies.[22] The Monarch Initiative, for instance, has been establishing

tools for allowing computation comparisons between species.[24] Such mappings will be necessary to fully integrate the human and mouse influenza infection data

1.5 Research Question and Specific Aims

Using this collaborative cross and available human data, this project sought to answer the following question: what genes and pathways are differentially expressed during mild and severe influenza infections in the collaborative cross mouse model, and do these mirror the differential expression seen in existing human influenza severity data? To guide the examination of this question, three specific aims were used to focus the investigation:

1. Specific Aim 1: Identify public human datasets for contrasting infection responses and calculate gene and pathway expression differences.
2. Specific Aim 2: Identify genetic and gene pathway expression differences related to changes in infection phenotype in the Collaborative Cross
3. Specific Aim 3: Integrate the expression variations to assess similarities and allow for greater context in current and ongoing studies.

2. Materials and Methods

To set up for and analyze the gene expression differences associated with severe influenza infection in mice and humans, a workflow was established that would allow the results for both species to be compared on both a gene and pathway level (Figure 2.1). The human dataset, as the first available, was analyzed first through this process. The mouse expression and plethysmography data was then assessed using the same steps once the data was collected.

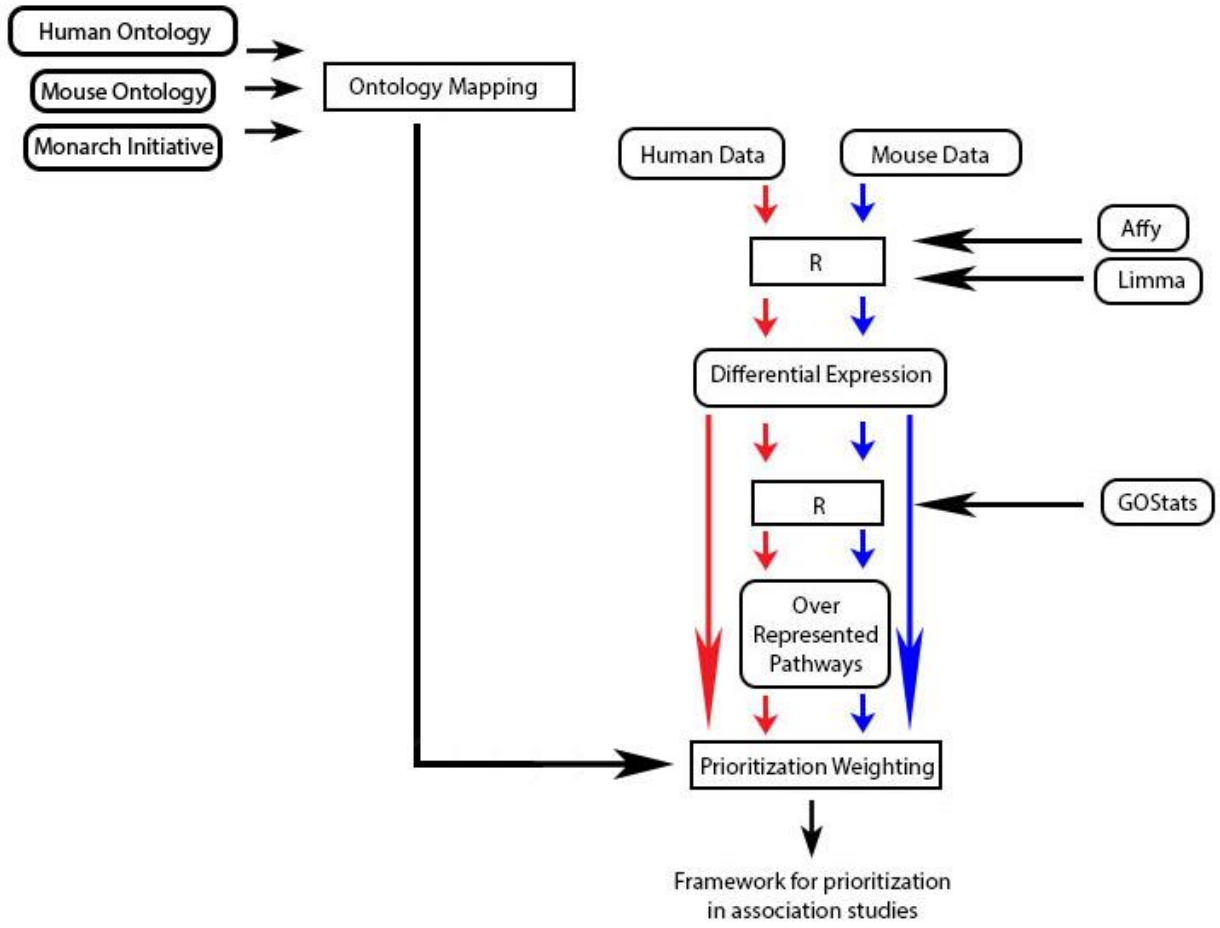

Figure 2.1. Schematic overview of analytical workflow for cross-species comparisons and prioritization

2.1 Human Differential Expression

In order to begin examining genetic differences between hosts experiencing normal verses severe influenza infections, public datasets of humans experiencing differing influenza infection severities were explored and aggregated. The criteria for these datasets were that they contain: influenza infection severity annotation with severity definitions equivalent to the WHO guidelines, raw microarray data available for use and access, and influenza virus data rather than (or in addition to) influenza vaccine data (Table 2.1). The genes identified as significantly differentially expressed in the publications associated with each dataset, once collected and translated to their NCBI Gene IDs, were loaded into R to check for overlap between the genes of each dataset. The Reactome pathways associated with the genes of each dataset were then determined using the reactome.db package to map between the gene IDs and pathway IDs.

2.2 Severe Influenza Infection Phenotyping

To be able to accurately assess the similarities between the results from the human datasets and the mouse models, establishing the criteria to declare an influenza infection as severe was necessary. The main authority on defining a severe infection in humans is the WHO, and the guidelines presented there are very useful and comprehensive. Those guidelines are also framed around clinical presentations, as these are the most commonly available when a human patient is examined (Table 2.2).

Table 2.1. Table of human influenza infection GEO datasets

GEO	PMID	Platform	Samples	Details
GSE61821	25365328	Illumina HumanHT-12 V4.0 expression beadchip	402 (83 mild, 40 moderate, 11 severe, 73 febrile by unknown pathogen)	Examined whole blood
GSE20346	21408152	Illumina HumanHT-12 V3.0 expression beadchip	81 (6 pneumonia patients, 4 severe influenza patients, 18 vaccine patients)	Examined whole blood over 4-5 days in 4 individuals
GSE27131	21781987	Affymetrix Human Gene 1.0 ST Array	21 (7 controls samples and 14 paired samples at day 0 and 6)	Examined peripheral blood at two time points

Table 2.2. Clinical symptoms of normal and severe influenza infection.

Uncomplicated Influenza	**Severe Influenza**
fever	shortness of breath
cough	tachypnea
sore throat	hypoxia
nasal congestion	pneumonia
headache	encephalopathy
muscle pain	encephalitis
malaise	organ failure
Gastrointestinal illness	severe dehydration

2.3 Ontologies

In order to make computational use the WHO clinical presentations, the terms associated with the severe infection symptoms were identified from the Human Phenotype Ontology (Table 2.3). The data used to stratify mouse subjects, however, is overwhelmingly comprised of quantitative measurements that do not always directly relate to the WHO guidelines. To bridge the gap between these, the ontology mappings established by the Monarch Initiative were employed to gather the equivalent terms from the Mammalian Phenotype Ontology (Table 2.3). For each individual phenotype identified and employed, the Monarch Initiative had also collected a list of genes associated with that phenotype for several mammals that were available. The mouse genes associated with the mammalian phenotypes selected were mined from the Monarch Initiative, and used to establish the initial framework for comparing results between species. In the scope of this investigation each ontology term was collected and utilized on its own, rather than including data from their child terms as well. The mouse genes were converted to their human equivalents as annotated by Ensembl through the use of the BioMart package, which allowed the gene identifiers to be translated into their human homologs.[25,26] In instances where no homologs were available, the data was dropped. These now entirely human gene IDs were examined for overlap, then the pathways associated with the human and mammalian ontology datasets were gathered from the reactome.db package to examine for similarities.

Table 2.3. Human and Mammalian Phenotype Ontology terms associated with severe influenza infection.

Human Ontology Term	Human ID	Mammalian ID	Mammalian Ontology term
Tachypnea	HP:0002789	MP_0005426	Tachpnea
Pneumonia	HP:0002090	MP_0001861	Lung inflamation
Encephalopathy	HP:0001298	MP:0013806	Encephalopathy
Encephalitis	HP:0002383	MP_0001847	Brain inflamation
Dehydration	HP:0001944	MP_0001429	Dehydration
myocarditis	HP:0012819	MP:0001856	Myocarditis
Renal insufficiency (kidney failure)	HP:0000083	MP:0003606	Kidney failure
Sepsis	HP:0100806	MP:0005044	Sepsis
Respiratory insufficiency	HP:0002093	MP:0001953	Respiratory failure
Respiratory distress	HP_0002098	MP:0001954	Respiratory distress

2.4 Integration

Once the mouse datasets were available, the phenotypes identified as associated with severe influenza infection were compared to the quantitative measurements gathered from the mice to establish criteria for stratifying the subjects. From the data available from the plethysmography measurements, the frequency of breathing and enhanced pause (Penh) were the most strongly related to these clinical presentations. As the normal breathing rate of mice appears to be between 85-230 breaths per minute, subjects with breath frequencies exceeding this

were considered tachypneic. Additionally, a 1.25-fold change in Penh between infected and controlled mice was used as additional support of the conclusion for severe infection.

To determine the genes differentially expressed in the severe infections, the expression results were examined in R with the use of the packages Affy, Limma, and Oligo. The expression data were processed with RMA normalization and used to create a linear model that was subjected to Empirical Bayes correction to determine the differential expression. Genes were considered significantly differentially expressed if their adjusted p-value was less than or equal to 0.05, while also showing a fold change of at least 1.5. The pathways associated with the differentially expressed genes were then identified in a similar manner to the human genes previously. The mouse genes were first converted to their human homologs in the same manner that the mouse genes obtained from the Mammalian Phenotype Ontology were. The homology data established by Ensembl was applied via the BioMart package to allow for the conversion of the mouse-specific identifiers to their human counterparts.[25,26] If a gene did not have a human homolog, the gene was omitted from the dataset. The homologous genes were then utilized with the same reactome.db package to collect the pathways linked to the mouse differential expression associated with severe influenza infection. The lists of human and mouse results were then compared for overlap on both the gene and pathway level.

3. Results

3.1 Human Differential Expression

The human influenza datasets were first gathered and analyzed. The genes identified as differentially expressed in these datasets were collected and compared to identify similarities. From these public datasets, 76 genes were identified as significant but there was no overlap between the genes that were significant. The pathways associated with the identified genes were then calculated to allow for greater overlap and applicability to the later steps. There were 151 unique Reactome pathways identified, seven of which were identical. To further help provide insight into differences affecting respiratory distress, two Severe Acute respiratory syndrome (SARS) datasets were compared and assessed for overlap on the gene level. There were 94 genes identified as significant, and similar to the influenza datasets there was no overlap between the genes. The pathways were then calculated as well for use later. 263 unique pathways were found, with 77 of these overlapping. In comparing the 151 influenza pathways and 263 SARS pathways, 96 of these pathways were in common between the two datasets.

3.2 Mammalian Phenotype Ontology

Due to the frequent updates and changes to the Monarch Initiative, the data associated with the Mammalian Phenotype Ontology terms in the Monarch Initiative database were collected on several occasions, with the final being utilized on July 12, 2016. At that time, each of the 10 phenotype terms (Table 2.3) were queried for genes associated with that phenotype. 626 unique genes were found across the datasets linked to the phenotypes, with 256 genes (40.9%) occurring in the gene lists obtained from multiple phenotypes. The pathways associated with the genes in each phenotype's dataset were calculated as well. 2224 pathways were found

across the groups, with 865 pathways being unique. Only 212 (24.5%) of the pathways were exclusively found in the data for a single phenotype term. Due to the high amount of overlap and similarity in these results, the gene lists obtained as associated with each ontology term were combined into a single dataset to be compared with the human and mouse expression data.

3.3 Collaborative Cross Differential Expression

In the collaborative cross mouse data, the influenza infection phenotypes showed by far the strongest differences in breath rate and penh values, lending the analysis to focus on the expression differences between the control and severe influenza infected subjects. The differential expression associated with severe influenza infection was calculated, where 3073 unique genes were found to be differentially expressed with 951 of those genes being repeated between the mouse subjects analyzed. From these the pathways associated with severe influenza infection in infected mice was gathered. These pathways showed 1082 pathways with 771 of these pathways in common. These pathways focused showed, among others, a large number of members that were a part of pathways centered on immune system, cell cycle, signaling, and metabolism.

3.4 Integration

Once the genes and pathways for each dataset were known, they were analyzed for similarities. The genes associated with each pathway were identified and examined to see what overlap existed on both the gene and pathway level. 7495 unique genes were identified with a total of 5373 showing up in multiple places among the human influenza, human SARS, Monarch gene associations, and collaborative cross differential expression results. On the pathway level, 1242 pathways were identified as significantly associated with the respiratory distress in one or more datasets. Of these, only 279 (22.5%) of the pathways were found in only one dataset. Similar to the mouse differential results, the majority of differentially expressed genes were

members of pathways involving the immune system, signaling, metabolism, disease, and cell cycle. Including these different sources of data certainly provided greater coverage in the genes identified within each pathway, with the average number of novel genes found within a pathway across the different datasets being 10.37.

4. Discussion

The ultimate goal of this investigation was to examine the genes and pathways differentially expressed in severe influenza infection in both collaborative cross mice and humans in order to examine their potential similarities. More specifically the desired outcome was to establish a computational framework for comparing between these species that could be used to both place previous work into a broader context incorporating both models and clinical data, as well as guiding the focus of ongoing and future studies into host differences associated with and potentially leading to severe influenza infections. This prioritization framework clearly benefited from the inclusion of the different datasets. Each dataset alone averaged between 1.68 to 7.71 differentially expressed genes in each pathway identified as associated with respiratory distress, while combined the integrated datasets averaged 10.37 genes per pathway with 4.33 genes identified from multiple datasets on average per pathway. In looking at the coverage of the pathways identified, there is some significant variation. On average each pathway identified had 25.4% of their genes show significant differential expression (IQR: 13.6 – 33.3%). There were also 76 pathways that had over half of their member genes identified as significantly associated with severe infections.

Making use of these data, there are many possible next steps that can be imagined, depending on what a study's goals are. One direction lies with use for prioritization in mouse models. Genes and pathways that were strongly represented in the human datasets would be potentially interesting candidates to focus on. Likewise, a highly represented pathway could be a key component in identifying biomarkers to predict human severe influenza infection. One such pathway would be the Translocation of ZAP-70 to Immunological Synapse (Figure 4.1), part of the immune system pathways. This pathway had 87.5% coverage, with 14 of the 16 genes identified as associated with severe infection. Of these, the CD4 gene was identified in all of the datasets, human influenza, human SARS, mouse, and the genes associated with the Mammalian Phenotype ontologies. This pathway encompasses signaling that is critical in the formation and activation of T-cells in the immune system.[27] The CD4 gene encodes a receptor protein for T-cells that appears to mediate the recognition and binding of antigens.[28] From simply the data collected here this pathway and its member genes would seem like a potentially valuable point of

focus in directing further research. As a key pathway and gene associated with an immune response, the finding is even more compelling as a potential indicator or effector of infection severity. To further approach this, there are many possible avenues. Obtaining both human and mouse data concerning respiratory distress in the absence of an infection would help inform this specific example, as well as the findings as a whole, on whether the results of interest are linked to the infection itself or tied to the respiratory distress used to categorize the severity. Due to the somewhat promiscuous nature of the CD4 gene, which was a member of 22 other pathways as well, it might also be more advantageous or informative to focus on some of the members that had less representation across the pathways. Ensuring platforms that will measure all the genes of interest from this pathway could help either reinforce other members of this pathway as potentially key, or suggest that some may only be species-specific and of less importance to future studies.

While the integration of these data into a prioritization framework was successful and allows for greater insight into the similarities and potential avenues for research going forward, there were certain limitations to this process as well. The most overarching of these would be the restriction to the WHO severe influenza infection criteria which broadly establishes respiratory distress as the main phenotype for classifying a severe infection. The translation of these phenotypes and their corresponding ontologies to the quantitative mouse measurements certainly affected how the mouse subjects were stratified and studied. The WHO criteria also requires reliance on multiple clinicians to make the same diagnoses both within and across the different studies, and can broadly encompass a variety of symptoms. These differences in phenotypes presented by the patients, in addition to potential differences in both the influenza infection as well as other possible exposures was unfortunately unavoidable. Lastly, there were temporal differences between the datasets employed. The human datasets were collected at a single time point, almost inevitably at varying times post infection. The mouse data, however, was a time series that was much more rigorously controlled.

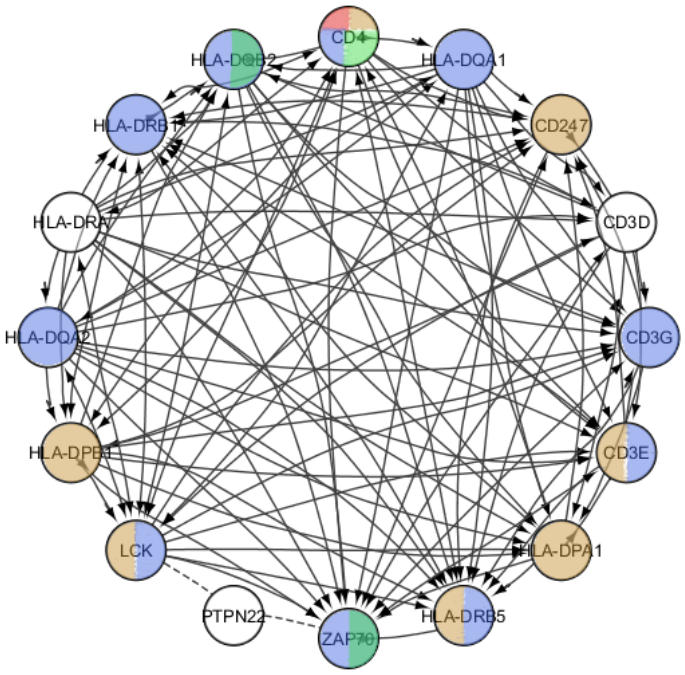

Figure 4.1 Network Diagram of Genes from the Translocation of ZAP-70 to Immunological Synapse Pathway. Genes colored with red were found in the human influenza datasets, those colored with orange were identified in the human SARS datasets, blue indicated genes found in the mouse expression datasets, and green indicates those found from the genes associated with the Mammalian Phenotype ontologies.

While the integration of these data into a prioritization framework was successful and allows for greater insight into the similarities and potential avenues for research going forward, there were certain limitations to this process as well. The most overarching of these would be the restriction to the WHO severe influenza infection criteria which broadly establishes respiratory distress as the main phenotype for classifying a severe infection. The translation of these phenotypes and their corresponding ontologies to the quantitative mouse measurements certainly affected how the mouse subjects were stratified and studied. The WHO criteria also requires reliance on multiple clinicians to make the same diagnoses both within and across the different studies, and can broadly encompass a variety of symptoms. These differences in phenotypes presented by the patients, in addition to potential differences in both the influenza infection as well as other possible exposures was unfortunately unavoidable. Lastly, there were

temporal differences between the datasets employed. The human datasets were collected at a single time point, almost inevitably at varying times post infection. The mouse data, however, was a time series that was much more rigorously controlled.

Ultimately, this project has provided a robust framework for comparing expression differences associated with severe influenza infection in both mice and humans. It's also emphasized the need for improvements in relating measurements and methods between clinical presentations and model organism measurements. Still, the combined and integrated results should allow for greater focus in both clinical and model research into drivers and predictors of severe infection in the ongoing effort to reduce the burden of influenza.

5. References

1 Organization, W. H. *WHO | Influenza (Seasonal)*, <http://www.who.int/mediacentre/factsheets/fs211/en/> (2014).

2 Suarez, D. L. & Schultz-Cherry, S. Immunology of avian influenza virus: a review. *Developmental & Comparative Immunology* **24**, 269-283, doi:http://dx.doi.org/10.1016/S0145-305X(99)00078-6 (2000).

3 Hay, A. J., Gregory, V., Douglas, A. R. & Lin, Y. P. The evolution of human influenza viruses. *Philosophical Transactions of the Royal Society of London. Series B* **356**, 1861-1870, doi:10.1098/rstb.2001.0999 (2001).

4 Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Martina, B. E. E., Bestebroer, T. M. & Fouchier, R. A. M. Influenza B Virus in Seals. *Science* **288**, 1051-1053, doi:10.1126/science.288.5468.1051 (2000).

5 Matsuzaki, Y. *et al.* Antigenic and Genetic Characterization of Influenza C Viruses Which Caused Two Outbreaks in Yamagata City, Japan, in 1996 and 1998. *Journal of Clinical Microbiology* **40**, 422-429, doi:10.1128/jcm.40.2.422-429.2002 (2002).

6 JENNINGS, R. RESPIRATORY VIRUSES IN JAMAICA: A VIROLOGIC AND SEROLOGIC STUDY: 3. HEMAGGLUTINATION-INHIBITING ANTIBODIES TO TYPE B AND C INFLUENZA VIRUSES IN THE SERA OF JAMAICANS. *American Journal of Epidemiology* **87**, 440-446 (1968).

7 Organization, W. H. WHO guidelines for pharmacological management of pandemic influenza A (H1N1) 2009 and other influenza viruses. *Part II: review of evidence. Revised February*, 21-22 (2010).

8 Tsalik, E. L. *et al.* Clinical presentation and response to treatment of novel influenza A H1N1 in a university-based summer camp population. *Journal of Clinical Virology* **47**, 286-288, doi:10.1016/j.jcv.2009.12.012.

9 Pabst, D., Kuehn, J., Schuler-Luettmann, S., Wiebe, K. & Lebiedz, P. Acute Respiratory Distress Syndrome as a presenting manifestation in young patients infected with H1N1 influenza virus. *European Journal of Internal Medicine* **22**, e119-e124, doi:10.1016/j.ejim.2011.08.014.

10 de Jong, M. D. *et al.* Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat Med* **12**, 1203-1207, doi:http://www.nature.com/nm/journal/v12/n10/suppinfo/nm1477_S1.html (2006).

11 Boon, A. C. M. *et al.* Host Genetic Variation Affects Resistance to Infection with a Highly Pathogenic H5N1 Influenza A Virus in Mice. *Journal of Virology* **83**, 10417-10426, doi:10.1128/jvi.00514-09 (2009).

12 Fang, L.-Q. *et al.* Environmental Factors Contributing to the Spread of H5N1 Avian Influenza in Mainland China. *PLoS ONE* **3**, e2268, doi:10.1371/journal.pone.0002268 (2008).

13 Rothberg, M. B., Haessler, S. D. & Brown, R. B. Complications of Viral Influenza. *The American journal of medicine* **121**, 258-264, doi:http://dx.doi.org/10.1016/j.amjmed.2007.10.040 (2008).

14 Van Kerkhove, M. D. *et al.* Risk Factors for Severe Outcomes following 2009 Influenza A (H1N1) Infection: A Global Pooled Analysis. *PLoS Med* **8**, e1001053, doi:10.1371/journal.pmed.1001053 (2011).

15 Zaas, A. K. *et al.* Gene Expression Signatures Diagnose Influenza and Other Symptomatic Respiratory Viral Infection in Humans. *Cell host & microbe* **6**, 207-217, doi:10.1016/j.chom.2009.07.006 (2009).

16 Albright, F. S., Orlando, P., Pavia, A. T., Jackson, G. G. & Albright, L. A. C. Evidence for a Heritable Predisposition to Death Due to Influenza. *Journal of Infectious Diseases* **197**, 18-24, doi:10.1086/524064 (2008).

17 Staeheli, P., Grob, R., Meier, E., Sutcliffe, J. G. & Haller, O. Influenza virus-susceptible mice carry Mx genes with a large deletion or a nonsense mutation. *Molecular and Cellular Biology* **8**, 4518-4523, doi:10.1128/mcb.8.10.4518 (1988).

18 Karupiah, G., Chen, J.-H., Mahalingam, S., Nathan, C. F. & MacMicking, J. D. Rapid Interferon γ–dependent Clearance of Influenza A Virus and Protection from Consolidating Pneumonitis in Nitric Oxide Synthase 2–deficient Mice. *The Journal of Experimental Medicine* **188**, 1541-1546, doi:10.1084/jem.188.8.1541 (1998).

19 Aylor, D. L. *et al.* Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Research* **21**, 1213-1222, doi:10.1101/gr.111310.110 (2011).

20 CTC (2004), C., G. A., et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* **36**, 1133-1137 (2004).

21 Beißbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464-1465, doi:10.1093/bioinformatics/bth088 (2004).

22 Mungall, C. *et al.* Integrating phenotype ontologies across multiple species. *Genome Biology* **11**, R2 (2010).

23 Köhler, S. *et al.* Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. *F1000Research* **2**, 30, doi:10.12688/f1000research.2-30.v2 (2013).

24 Washington, N. L. *et al.* Linking Human Diseases to Animal Models Using Ontology-Based Phenotype Annotation. *PLoS Biol* **7**, e1000247, doi:10.1371/journal.pbio.1000247 (2009).

25 Herrero, J. *et al.* Ensembl comparative genomics resources. *Database* **2016**, doi:10.1093/database/bav096 (2016).

26 Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research* **43**, W589-W598, doi:10.1093/nar/gkv350 (2015).

27 van Leeuwen, J. E. M. & Samelson, L. E. T cell antigen-receptor signal transduction. *Current Opinion in Immunology* **11**, 242-248, doi:http://dx.doi.org/10.1016/S0952-7915(99)80040-5 (1999).

28 Doyle, C. & Strominger, J. L. Interaction between CD4 and class II MHC molecules mediates cell adhesion. *Nature* **330**, 256-259 (1987).