



OPEN

Species-specific transcriptomic changes upon respiratory syncytial virus infection in cotton rats

Britton A. Strickland¹, Seesandra V. Rajagopala², Arash Kamali³, Meghan H. Shilts², Suman B. Pakala², Marina S. Boukhvalova³, Shibu Yooseph⁴, Jorge C. G. Blanco³✉ & Suman R. Das^{1,2}✉

The cotton rat (*Sigmodon*) is the gold standard pre-clinical small animal model for respiratory viral pathogens, especially for respiratory syncytial virus (RSV). However, without a reference genome or a published transcriptome, studies requiring gene expression analysis in cotton rats are severely limited. The aims of this study were to generate a comprehensive transcriptome from multiple tissues of two species of cotton rats that are commonly used as animal models (*Sigmodon fulviventer* and *Sigmodon hispidus*), and to compare and contrast gene expression changes and immune responses to RSV infection between the two species. Transcriptomes were assembled from lung, spleen, kidney, heart, and intestines for each species with a contig N50 > 1600. Annotation of contigs generated nearly 120,000 gene annotations for each species. The transcriptomes of *S. fulviventer* and *S. hispidus* were then used to assess immune response to RSV infection. We identified 238 unique genes that are significantly differentially expressed, including several genes implicated in RSV infection (e.g., *Mx2*, *I27L2*, *LY6E*, *Viperin*, *Keratin 6A*, *ISG15*, *CXCL10*, *CXCL11*, *IRF9*) as well as novel genes that have not previously described in RSV research (*LG3BP*, *SYWC*, *ABEC1*, *IIGP1*, *CREB1*). This study presents two comprehensive transcriptome references as resources for future gene expression analysis studies in the cotton rat model, as well as provides gene sequences for mechanistic characterization of molecular pathways. Overall, our results provide generalizable insights into the effect of host genetics on host-virus interactions, as well as identify new host therapeutic targets for RSV treatment and prevention.

Respiratory syncytial virus (RSV) is the leading cause of lower respiratory tract infection (LRTI) in children below the age of two years, as well as in immunocompromised individuals and the elderly, resulting in 33 million LRTIs, 3.2 million hospital admissions, and nearly 120,000 deaths worldwide each year¹. There is currently no approved RSV vaccine and only one preventative monoclonal antibody (Palivizumab), with use limited to high-risk children due to costs^{2,3}. As 93% of RSV LRTI cases and 99% of RSV mortality occurs in developing countries, the need for an effective vaccine and low-cost preventatives is critical¹. The failure of the formalin-inactivated RSV vaccine in the 1960s, which induced enhanced disease in vaccinees upon encounter with the virus, hampered the development of new RSV vaccines for decades⁴⁻⁶. However, recently there is a renewed effort to develop RSV preventatives, with 14 vaccine candidates and alternative anti-viral strategies against RSV (recombinant antibodies⁷, nanobodies⁸, small molecule inhibitors and analogs⁹) that are at various stages of development¹⁰. This highlights the critical need for an appropriate pre-clinical model for vaccine and drug development against RSV.

The cotton rat (genus *Sigmodon*) is considered the “gold standard” animal model for RSV infection compared to mice and other animals because it is 100-fold more permissive than the majority of laboratory mice to RSV infection and RSV infects both its upper and lower respiratory tracts similar to humans¹¹⁻¹³. Cotton rats have also accurately predicted efficacy of the two FDA-approved RSV therapeutics (RespiGam®, Palivizumab®)¹⁴⁻¹⁷. In addition to RSV, cotton rats have been used to study other human respiratory viruses of significance, i.e., influenza A virus^{18,19}, parainfluenza virus^{20,21}, measles²², human metapneumovirus²³, enterovirus D68²⁴, and human rhinovirus²⁵, due to the broad susceptibility and comparable human disease features²⁶. Unfortunately, studies

¹Department of Pathology Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. ²Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, 1211 21st Avenue South, S2108 Medical Center North, Nashville, TN 37232, USA. ³Sigmovir Biosystems Inc., 9610 Medical Center Drive, Suite 100, Rockville, MD 20850, USA. ⁴Department of Computer Science, Genomics and Bioinformatics Cluster, University of Central Florida, Orlando, FL, USA. ✉email: j.blanco@sigmovir.com; suman.r.das@vumc.org

| Transcriptome assembly | <i>S. fulviventer</i> | <i>S. hispidus</i> |
|---|-----------------------|--------------------|
| Contigs/Isoforms | 620,569 | 592,099 |
| Genes | 587,619 | 559,830 |
| Genes (TPM > 1) | 270,451 | 474,882 |
| Median/Mean Contig Length | 580/1003.01 | 597/1032.27 |
| GCC% | 43.28 | 43.32 |
| Contig N50 | 1604 | 1655 |
| Ex90-N50:#Genes | 2503:108,995 | 2162:240,587 |
| Reads Mapped Back to Assembly | 81.9% | 87.9% |
| BUSCO (vertebrata n = 3354) | 90.2% | 90.4% |
| Annotation | | |
| Identified Coding Sequences | 118,159 | 113,999 |
| Annotated transcripts (<i>BlastX</i> 1e-5) | 118,060 | 117,153 |
| Unique genes (<i>no duplicates</i>) | 18,726 | 18,380 |
| Annotated proteins (<i>BlastP</i> 1e-5) | 31,855 | 30,953 |
| Unique proteins (<i>no duplicates</i>) | 17,154 | 16,887 |
| KEGG Pathways | 90,642 | 89,654 |
| Pfam (protein domains) | 5042 | 6298 |
| TmHMM (transmembrane helices) | 18,846 | 18,027 |
| SignalP (signaling proteins) | 7915 | 7778 |
| EggNOG (orthology relationships) | 2902 | 2205 |

Table 1. De novo assembly and annotations statistics.

comparing transcriptomic changes in cotton rats have been limited due to the lack of publicly available reference genome for any cotton rat species. Previously, we published a lung tissue transcriptome induced by RSV in *S. hispidus*²⁷. However, in that study we had only analyzed expression in one tissue type, and it was limited to only one cotton rat species (*S. hispidus*). As prior studies show, both *S. fulviventer*- and *S. hispidus*-specific differ in disease severity to viral pathogens (i.e., parainfluenza virus²⁸, HIV²⁹) and microbiome community structure³⁰. The main objectives of this study were to develop comprehensive transcriptomes for both species and to compare and contrast gene expression changes upon RSV infection.

To this end, we sequenced total RNA of multiple tissues (lung, spleen, heart, kidney, colon) from healthy animals and generated a comprehensive transcriptome using de novo assembly with functional annotation for two species of cotton rats (*S. fulviventer* and *S. hispidus*). We then infected each species of cotton rat with RSV A/Long strain alongside uninfected controls and used our transcriptome references to determine differentially expressed genes after RSV infection.

Results

RNA extraction and transcriptome sequencing. Total RNA was extracted from sections of spleen, heart, kidney, lung, and colon harvested from 4–6-week-old healthy *S. fulviventer* and *S. hispidus* for comprehensive transcriptome assembly (supplementary table 1). The RNA-seq libraries were sequenced to a depth of 50 million paired-end 2×150 per sample. After sequencing read QC, we pooled data from all healthy tissues (n = 2 per species, lung n = 4 per species) to generate 456 million paired-end reads for *S. fulviventer* (GC content 54.9%) and 465 million paired-end for *S. hispidus* (GC content 53.1%). Transcriptome for each species was generated using de novo assembly, which was then used as a reference database to evaluate RSV-induced transcriptomic changes in lungs of the infected animals compared to uninfected controls.

De novo transcriptome assembly of *S. fulviventer* and *S. hispidus* reads. For each species, sequenced reads from all 5 tissue types were combined. Trinity³¹ was used for de novo assembly of contigs (also referred to as ‘transcripts’) following in silico normalization to achieve an average coverage of 50x. The assembler generated over 1.3 million contigs per species, which were further filtered based on length³² and redundancy^{33,34} (supplementary table 1). The following assembly statistics are shown in Table 1. Each transcript was assigned a unique identifier that designated each ‘gene’ (*S. fulviventer* = 587,619, *S. hispidus* = 559,830) and their alternatively spliced ‘isoforms’ (*S. fulviventer* = 620,569, *S. hispidus* = 592,099). Assembled transcripts from both species had a GC content of 43% and contig N50 > 1600 (which exceeds N50 of other published transcriptome assemblies^{35–37}). The Ex90N50, which is the N50 but limited to the top 90% of total normalized transcripts, was 2503 (*S. fulviventer*, #transcripts = 108,995) and 2162 (*S. hispidus*, #transcripts = 240,587) (Table 1).

The number of transcripts and their expression levels varied between tissue type, and several transcripts were tissue-specific, with the lung having the greatest number of unique transcripts (*S. fulviventer* = 47,448, *S. hispidus* = 38,270) (Fig. 1A,B). Other transcripts were either found in all 5 tissues (*S. fulviventer* = 216,307, *S. hispidus* = 225,833) or a combination of two or more tissue types (Fig. 1A,B). Many transcripts were differentially expressed between tissues (*S. fulviventer* = 49,215, *S. hispidus* = 55,415) as determined by DESeq2 and visualized

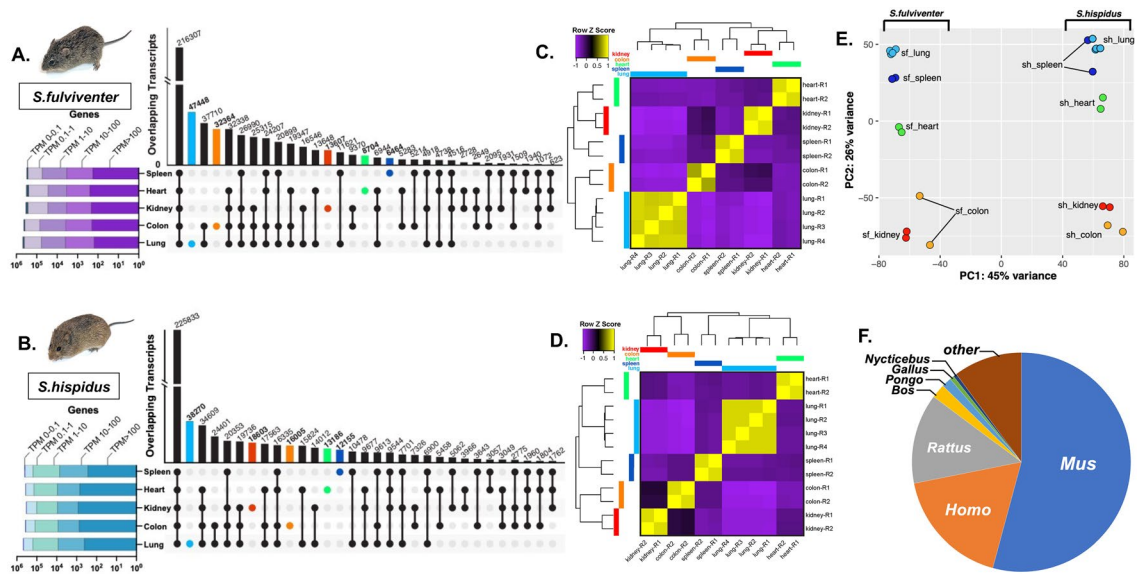


Figure 1. *Sigmoidon* transcriptome assembly using Trinity. **(A,B)** UpSet plot of overlapping and tissue-specific transcripts expressed in the spleen (blue), heart (green), kidney (red), colon (orange), and lung (blue) of both **(A)** *S. fulviventer* ($n=2$) and **(B)** *S. hispidus* ($n=2$). Bars for overlapping sequences are shown in black, while bars for sequences unique to individual organs are shown in corresponding colors. Number of overlapping transcripts are shown above each bar, with individual organ comparison on the y-axis. Number of unique genes per tissue are represented by the horizontal barplot. **(C,D)** Hierarchical clustering and heatmap by Z-score of all differentially expressed transcripts ($Q < 0.001$, $L2fc > 2$, Euclidean distance) within individual tissue samples of both **(C)** *S. fulviventer* and **(D)** *S. hispidus*. **(E)** Principal component analysis (PCA) of expressed transcripts within healthy tissue of both species. **(F)** Taxonomic source of gene annotations determined by NCBI BlastX. Data represents annotations of *S. fulviventer* transcriptome; *S. hispidus* not shown but varies by $\sim 1\%$ in all categories except “other”.

by hierarchical heatmap clustering³⁸ (Fig. 1C,D). The principal component analysis shows significant difference in gene expression patterns between species and different tissue samples, indicated by strong clustering based on species and tissue type (Fig. 1E). Following assembly, transcripts were annotated using protein sequence database (SwissProt) to search for homology against other species. For both species, transcripts shared the highest homology with *Mus musculus* (54%), *Homo sapiens* (17%) and *Rattus rattus* (13%) (Fig. 1F).

To assess the quality of our assembly (Table 1), we mapped the full assembly back to the quality-trimmed reads using Bowtie2³⁹ to find 81.9% and 87.9% of reads were utilized during assembly of the *S. fulviventer* and *S. hispidus* transcriptome respectively, for which $> 70\%$ is indicative of good quality. We also assessed transcriptome completeness by searching for evolutionarily-conserved BUSCO⁴⁰ groups from the vertebrata_odb10 lineage dataset (total $n = 3354$) within our dataset. The *S. fulviventer* assembly had 92.7% complete BUSCOs (Complete:3109 [Complete&Single:785, Complete&Duplicated:2324], Fragmented:154, Missing:91), and *S. hispidus* had 90.2% complete BUSCOs (Complete:3025 [Complete&Single:2599, Complete&Duplicated:426], Fragmented:194, Missing:135). All assembly quality statistics are also shown in Table 1. Final reference transcriptome assemblies for each species are made available as supplementary file 1 (*S. hispidus*) and supplementary file 2 (*S. fulviventer*).

Annotation and functional properties of the *S. fulviventer* and *S. hispidus* multi-tissue transcriptome.

Transcripts were functionally annotated using the Trinotate pipeline (<https://trinotate.github.io/>). Transcriptome references were generated for each species independently. First, cotton rat transcripts were aligned to a non-redundant protein annotation database (UniProt⁴¹)—which contains sequences from 14,132 different taxa (<http://web.expasy.org/docs/relnotes/relnstat.html>)—based on sequence homology via BLASTx⁴² (cutoff e-value $< 1e-05$). We annotated 118,060 transcripts comprising 18,726 unique genes for *S. fulviventer* and 117,153 transcripts comprising of 18,380 unique genes for *S. hispidus* out of the $\sim 600,000$ transcripts assembled for each species (Table 1). Coding regions were then identified based on open reading frames (*S. fulviventer* = 118,159, *S. hispidus* = 113,999). Protein coding transcripts were annotated using BLASTp⁴² (cutoff e-value $< 1e-05$) against the UniProt database⁴¹ to identify $> 30,000$ annotated proteins per species ($\sim 17,000$ proteins after filtering duplicate annotations). Proteins were further annotated based on protein domains (> 5000), transmembrane helices ($> 18,000$), and signaling proteins (> 7700). All annotation statistics can be found in Table 1, and the full annotation report can be viewed in supplementary file 3.

In addition to gene name annotations, BLAST alignment to UniProt also assigned several functional terms to describe genes, including Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways⁴³, Gene Ontology (GO)⁴⁴, and EggNOG orthology⁴⁵. Gene Ontology (GO) terms describe the functional properties of genes and their relationship to one another. One or more GO Terms were assigned to 98.35% *S. fulviventer* annotated transcripts ($n = 116,115$) and 98.45% *S. hispidus* annotated transcripts ($n = 115,332$). GO terms were grouped

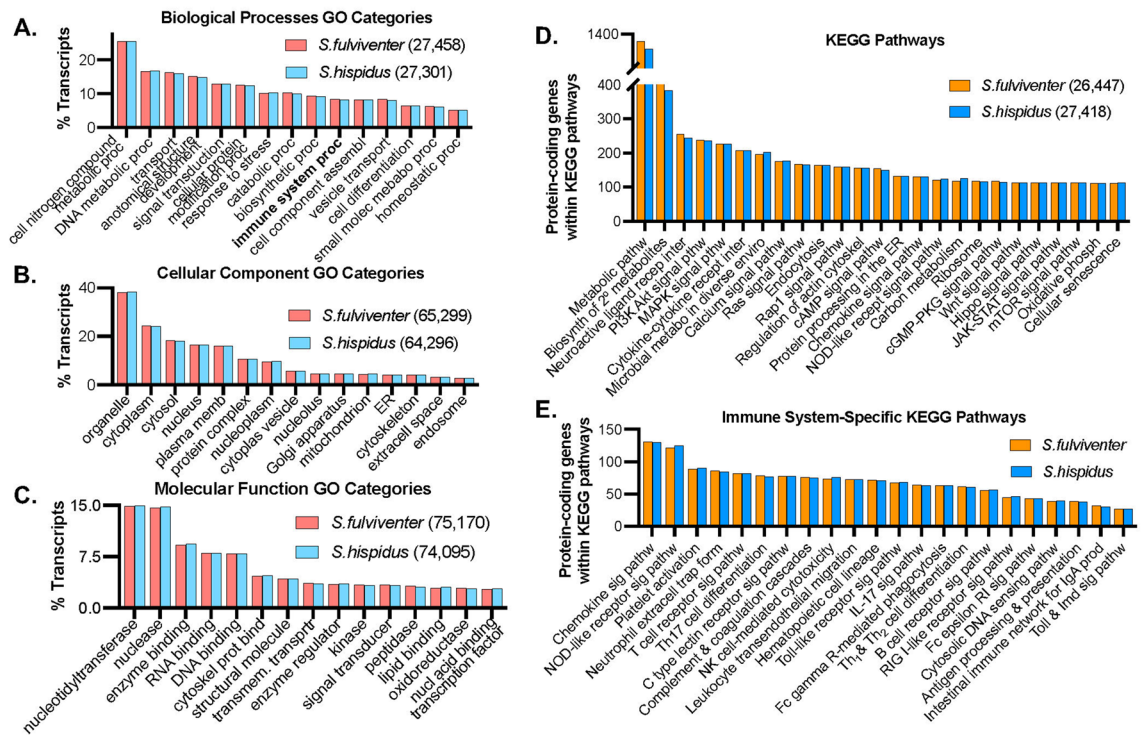


Figure 2. *Sigmodon* transcriptome annotation. (A) Biological processes, (B) cellular components, and (C) molecular function SwissProt Gene Ontology (GO) terms on the x-axis (total GO's within respective categories including overlaps represented in figure legends) with number of GO's (percentage of total ~ 107 k annotated transcripts) on the y-axis; GO's assigned against TrEMBL/SwissProt database using Trinotate v.3.2.2. (D) Top 25 KEGG and (E) immune system-specific pathways on the x-axis with total number of protein-coding genes on the y-axis; pathways assigned using TransDecoder-determined CDS followed by GhostKOALA (<https://www.kegg.jp/ghostkoala/>).

into 3 categories: Biological Processes, Cellular Components, and Molecular Function (Fig. 2). Genes from both *S. fulviventer* and *S. hispidus* had a similar ratio of associated GOs. The top Biological Processes GOs included “cellular nitrogen compound metabolic processes” (~ 25%), “DNA metabolic processes” (~ 17%), “transport” (~ 16%), and “anatomical structure development” (~ 15%) (Fig. 2A). Cellular Component GOs consisted mostly of “organelles” (~ 38%), “cytoplasm” (~ 24%), “cytosol” (~ 18%), and “nucleus” (16%) (Fig. 2B). The most common Molecular Function GOs include “nucleotidyltransferase” (~ 15%), “nuclease” (~ 15%), “enzyme binding” (~ 9%), and “nucleic acid binding” (RNA = 8%, DNA = 8%) (Fig. 2C).

The majority of KEGG pathways identified belonged to metabolism and signaling, specifically secondary metabolite biosynthesis, ligand/cytokine/metal receptor interactions, and a variety of signaling pathways (Fig. 2D, also see Figure S1A for distribution of “signaling and membrane transport” KEGG pathways). We examined immune-specific KEGG pathways to showcase the large number of relevant genes identified in the de novo transcriptome, including chemokine signaling, innate and adaptive cell receptor signaling, complement cascades, and lymphocyte differentiation (Fig. 2E). KEGG also provides gene mapping against many infection pathways, including infectious agents successfully modeled in cotton rats i.e., influenza¹⁸, HIV-1²⁹, and measles²² (Figure S1B). We mapped gene annotations of both *S. fulviventer* and *S. hispidus* to the host genes associated with influenza infection KEGG pathway, which encompassed 94.3% of the genes required during influenza infection (Fig. 3). All other pathways can be reconstructed and visualized using KEGG's reconstruct tool (<https://www.kegg.jp/kegg/mapper/reconstruct.html>) using data included in the supplementary file 4.

Histopathology and gene expression changes after RSV infection in the lung of *S. fulviventer* and *S. hispidus*.

We intranasally infected *S. fulviventer* and *S. hispidus* with 10^5 PFU of RSV A/Long or PBS (as mock infection control) and harvested lungs after 5 days post infection for analysis. Infection was confirmed by qRT-PCR targeting the RSV G and F mRNA, where significantly higher RNA copy was found in infected animal, whereas uninfected controls did not have amplification of viral RNAs (Fig. 4A). There was no significant difference in viral RNA levels between *S. fulviventer* and *S. hispidus* following 5-day infection ($p = 0.3491$, Tukey's multiple comparisons test). Lung tissue was evaluated for histopathology (Fig. 4B) and blindly scored for 4 inflammatory parameters as described in the methods: peribronchiolitis, perivascularitis, interstitial pneumonia, and alveolitis (Fig. 4C). Each infected animal had a significantly larger cumulative pathology score compared to uninfected controls, indicating successful infection ($Sf p = 0.0020$, $Sh p = < 0.0001$, Tukey's multiple comparisons test) (Fig. 4C). RNA was then sequenced as described, and reads were aligned to the previously annotated reference transcriptome for each species. Following expression normalization, gene expression analysis showed RSV

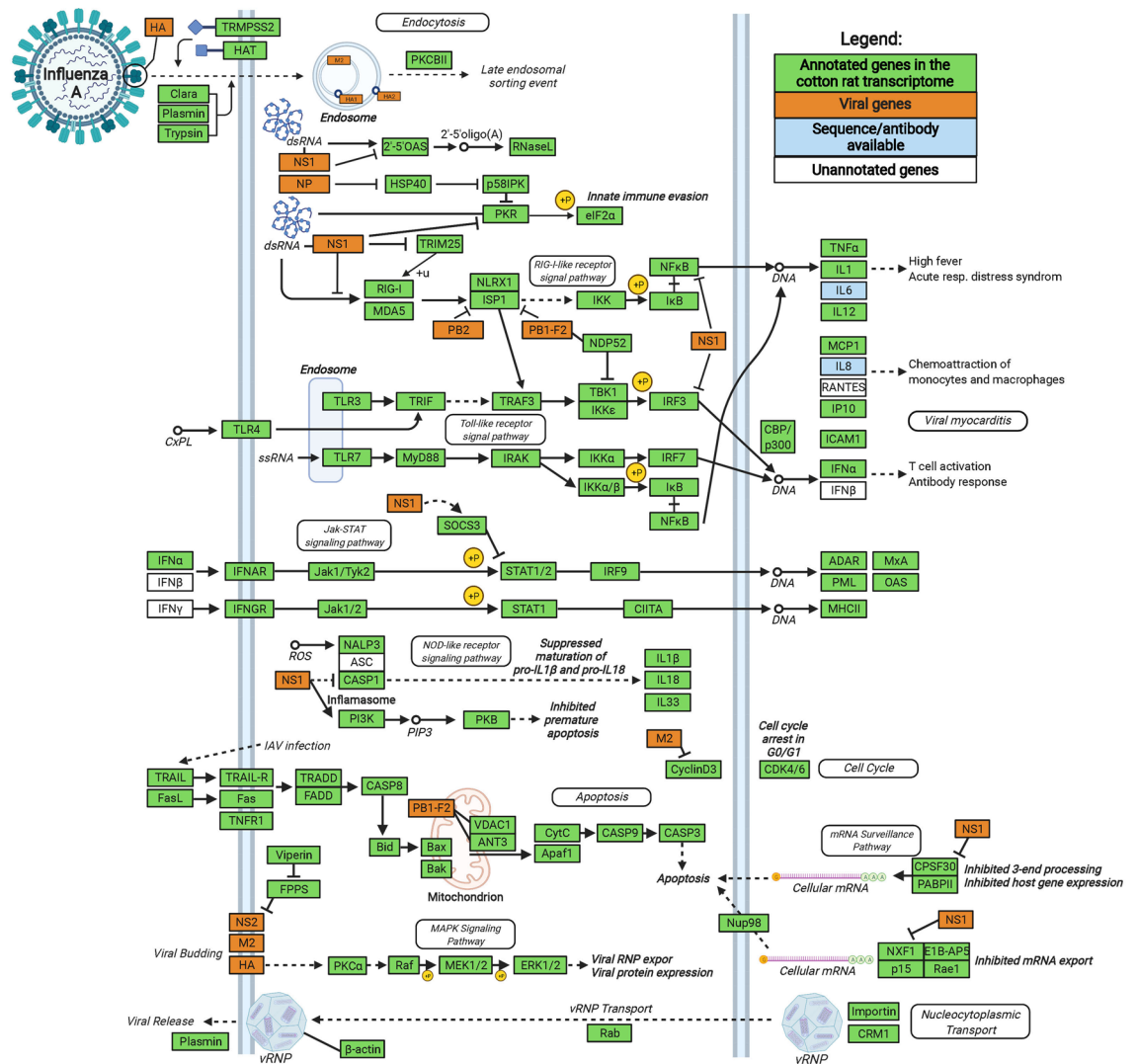


Figure 3. The Influenza Disease Pathway (adapted from Kyoto Encyclopedia of Genes and Genomes [KEGG]). 99 of the 105 essential genes of influenza pathogenesis were successfully assembled and annotated in our de novo transcriptome of *S. fulviventer* and *S. hispidus*. Created with BioRender.

infection significantly altered expression patterns. RSV infection in the lungs of *S. hispidus* resulted in greater alterations in gene expression than that of *S. fulviventer* (Supplemental Figure 2).

Gene differential expression analyses (RSV A/Long vs. mock-infected Controls) identified several genes that were differentially up- and down-regulated in infected lungs, including genes unique and common to each cotton rat species (Fig. 4D). RSV infection in *S. fulviventer* resulted in 325 unique upregulated genes (98 annotated) and 140 unique downregulated genes (20 annotated). RSV infection in *S. hispidus* resulted in 750 upregulated genes (44 annotated) and 245 downregulated genes (8 annotated) (Fig. 4D,E, full data in supplementary file 5). By combining infection data from both species, we found that 276 genes (12 annotated) were upregulated in both species while 9 genes (0 annotated) were downregulated in both species. Only annotated genes for each species were selected for further analysis. Twenty of the top-most differentially expressed annotated genes (selected based on function and Gene Ontology) are listed in Table 2 (*S. fulviventer*) and Table 3 (*S. hispidus*).

Species-specific changes to the RSV-infected lung. In *S. fulviventer*, several genes were downregulated upon infection related to immunoglobulin structure (*HVM44*), transcription factors (*CREB1*, *CRY1*), epithelial structure integrity (*K2C8*), formation and maintenance of tight junctions (*OCLD*), membrane trafficking regulation (*CPNE3*), targeted RNA degradation (*MOV10*), and tissue remodeling and homeostasis (*MFGM*). *GLYC*, which is the RSV G protein, was marked as upregulated in RSV lungs due to the absence of reads in the uninfected group (indicated by DESeq2 basemean = 0). Upregulated host genes were related to the complement system (*CFAB*), immune receptor signaling (*UN93B*), vesicle trafficking (*EXOC5*), interferon-induced antiviral activity (*MX2*, *CXCL10*, *I27L2*, *OAS1A*, *IIGP1*, *ISG15*), prostaglandin metabolism (*PGDH*), and electron transport (*COX1*). These differentially expressed *S. fulviventer* genes following RSV infection are listed in Table 2.

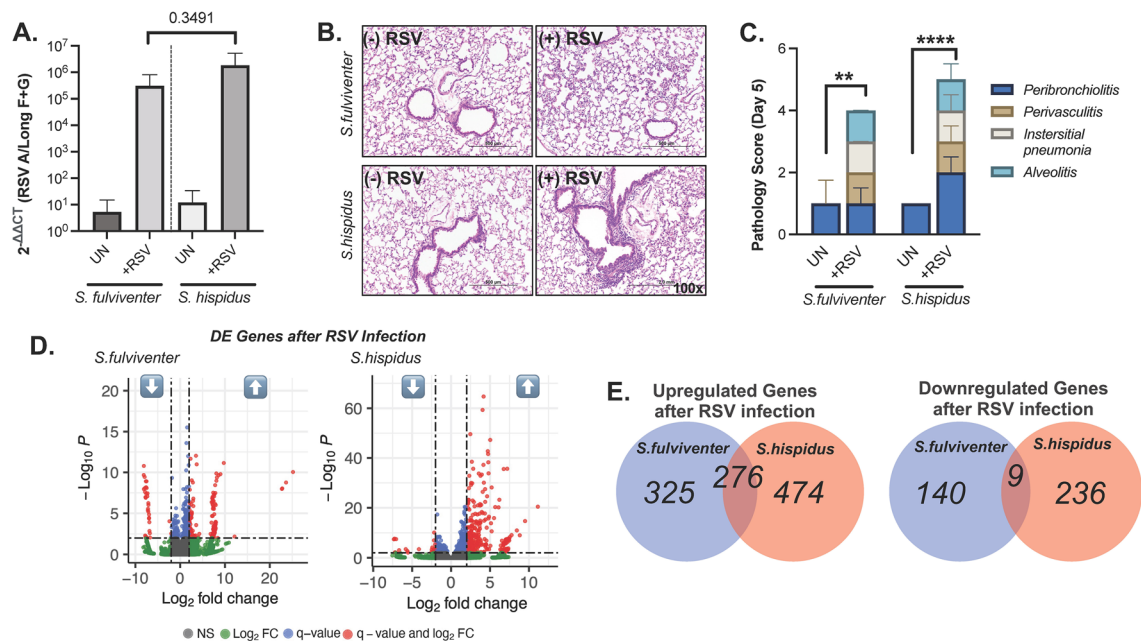


Figure 4. RSV-induced changes to the lung environment. **(A)** Viral titers (RSV G and F proteins) in uninfected and infected lungs determined by qRT-PCR normalized to β -actin; fold change (calculated by $2^{-\Delta\Delta CT}$) on the y-axis. **(B)** Histopathological imaging and **(C)** blindly generated pathology scoring for both healthy ($n = 4$ /species) and RSV-infected ($n = 5$ /species) lung tissue from *S. fulviventer* and *S. hispidus*. **(D,E)** DESeq2 analysis of assembled genes from lung tissues (same tissue imaged from Figures A-B) revealed several genes that are differentially expressed (DE) between healthy and infected tissues ($p < 0.05$, $|2fc| > |1.0|$) of *S. fulviventer* (blue) and *S. hispidus* (red), of which several genes were successfully annotated using Trinotate (as indicated in boxes). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

| Gene | Description | Transcript ID | Log2FC | p value | q value |
|--------|--|----------------------|---------|---------|---------|
| HVM44 | Ig heavy chain V region PJ14 | Sfulv_DN86481_c0_g1 | ↓ -3.51 | 1.7E-04 | 3.4E-02 |
| CREB1 | Cyclic AMP-responsive element-binding protein 1 | Sfulv_DN150233_c0_g1 | ↓ -1.94 | 1.5E-05 | 6.3E-03 |
| CRY2 | Cryptochrome-2 | Sfulv_DN8103_c29_g1 | ↓ -1.69 | 4.8E-08 | 5.8E-05 |
| K2C8 | Keratin, type II cytoskeletal 8 | Sfulv_DN18044_c9_g1 | ↓ -1.54 | 5.4E-07 | 4.7E-04 |
| OCLN | Occludin | Sfulv_DN14349_c3_g1 | ↓ -1.49 | 2.7E-05 | 9.6E-03 |
| CPNE3 | Copine-3 | Sfulv_DN28516_c1_g1 | ↓ -1.27 | 1.6E-04 | 3.4E-02 |
| MOV10 | Putative helicase MOV-10 | Sfulv_DN357483_c0_g1 | ↓ -1.14 | 1.8E-05 | 7.1E-03 |
| MFGM | Lactadherin | Sfulv_DN100738_c3_g1 | ↓ -1.07 | 3.4E-05 | 1.1E-02 |
| ISG15 | Interferon-induced ubiquitin-like protein, 15 kDa | Sfulv_DN2597_c2_g1 | ↑ 1.02 | 1.3E-04 | 2.9E-02 |
| IIGP1 | Interferon-inducible GTPase 1 | Sfulv_DN1249_c0_g1 | ↑ 1.33 | 3.0E-05 | 1.0E-02 |
| OAS1A | 2'-5'-oligoadenylate synthase 1A | Sfulv_DN24798_c0_g1 | ↑ 1.58 | 7.8E-05 | 2.0E-02 |
| I27L2 | Interferon alpha-inducible protein 27-like protein 2 | Sfulv_DN13614_c0_g1 | ↑ 1.96 | 8.0E-13 | 5.2E-09 |
| COX1 | Cytochrome c oxidase assembly protein COX15 homolog | Sfulv_DN104408_c3_g1 | ↑ 2.04 | 4.6E-08 | 5.7E-05 |
| CXCL10 | C-X-C motif chemokine 10 | Sfulv_DN18689_c0_g1 | ↑ 2.60 | 6.1E-09 | 1.0E-05 |
| PGDH | 15-hydroxyprostaglandin dehydrogenase [NAD(+)] | Sfulv_DN56601_c1_g1 | ↑ 2.95 | 5.8E-10 | 1.3E-06 |
| MX2 | Interferon-induced GTP-binding protein Mx2 | Sfulv_DN17914_c1_g1 | ↑ 3.48 | 1.1E-05 | 4.8E-03 |
| EXOC5 | Exocyst complex component 5 | Sfulv_DN7267_c6_g1 | ↑ 4.02 | 4.1E-05 | 1.3E-02 |
| UN93B | Protein unc-93 homolog B1 | Sfulv_DN114523_c0_g1 | ↑ 4.04 | 3.2E-05 | 1.1E-02 |
| CFAB | Complement factor B | Sfulv_DN212381_c0_g2 | ↑ 6.99 | 1.3E-07 | 1.4E-04 |
| GLYC | Major surface glycoprotein G_Orthopneumovirus | Sfulv_DN7526_c1_g1 | ↑ 9.74 | 2.5E-16 | 7.1E-12 |

Table 2. Twenty notable differentially expressed genes in *S. fulviventer* lung following RSV infection. Negative Log2FC = downregulated, positive Log2FC = upregulated. Gene names/descriptions determined by BlastX against UniRef/SwissProt. Transcript IDs correspond to contig names (Sfulv_XXX) in assembly FASTA file. Log2FC = log2FoldChange, q = FalseDiscoveryRate.

| Gene | Description | Transcript ID | Log2FC | p value | q value |
|---------------|---|-----------------------------|---------|---------|---------|
| <i>LMTD1</i> | Lamin tail domain-containing protein 1 | <i>Shisp_DN10656_c2_g1</i> | ↓ -1.04 | 3.7E-04 | 3.5E-02 |
| <i>NLRC5</i> | NLR family CARD domain containing 5 | <i>Shisp_DN1039_c0_g1</i> | ↑ 1.74 | 1.5E-06 | 3.3E-04 |
| <i>GBP4</i> | Guanylate binding protein 4 | <i>Shisp_DN1171_c0_g1</i> | ↑ 1.88 | 2.8E-13 | 2.3E-10 |
| <i>IFIT1</i> | Interferon induced protein with tetratricopeptide repeats 1 | <i>Shisp_DN1078_c15_g1</i> | ↑ 2.01 | 2.6E-06 | 5.3E-04 |
| <i>LY6E</i> | Lymphocyte antigen 6 family member E | <i>Shisp_DN101026_c0_g1</i> | ↑ 2.20 | 2.0E-12 | 1.4E-09 |
| <i>IRF9</i> | Interferon regulatory factor 9 | <i>Shisp_DN12125_c1_g1</i> | ↑ 2.33 | 3.2E-07 | 8.4E-05 |
| <i>C4BPA</i> | C4b-binding protein alpha chain | <i>Shisp_DN107540_c0_g1</i> | ↑ 2.42 | 5.6E-13 | 4.3E-10 |
| <i>CO4A</i> | Collagen alpha-1(IV) chain | <i>Shisp_DN122397_c0_g1</i> | ↑ 2.43 | 3.1E-05 | 4.6E-03 |
| <i>LG3BP</i> | Galectin-3-binding protein | <i>Shisp_DN11955_c0_g1</i> | ↑ 2.45 | 1.1E-12 | 8.1E-10 |
| <i>PSB9</i> | Proteasome subunit beta type-9 | <i>Shisp_DN13011_c0_g1</i> | ↑ 2.52 | 5.2E-10 | 2.3E-07 |
| <i>K2C6A</i> | Keratin, type II cytoskeletal 6A | <i>Shisp_DN12075_c0_g1</i> | ↑ 2.70 | 1.5E-05 | 2.4E-03 |
| <i>HSH2D</i> | Hematopoietic SH2 domain-containing protein | <i>Shisp_DN0_c15_g8</i> | ↑ 2.80 | 3.0E-12 | 2.1E-09 |
| <i>ABEC1</i> | C- > U-editing enzyme APOBEC-1 | <i>Shisp_DN11754_c2_g1</i> | ↑ 3.45 | 3.9E-25 | 1.0E-21 |
| <i>RSAD2</i> | Radical S-adenosyl methionine domain-containing prot. 2 | <i>Shisp_DN112437_c0_g1</i> | ↑ 3.51 | 3.5E-12 | 2.3E-09 |
| <i>CXCL11</i> | C-X-C motif chemokine 11 | <i>Shisp_DN149434_c0_g1</i> | ↑ 3.56 | 8.3E-15 | 8.1E-12 |
| <i>MX2</i> | Interferon-induced GTP-binding protein Mx2 | <i>Shisp_DN140604_c0_g1</i> | ↑ 4.24 | 1.0E-18 | 1.5E-15 |
| <i>CXCL10</i> | C-X-C motif chemokine 10 | <i>Shisp_DN14446_c0_g1</i> | ↑ 4.62 | 6.3E-29 | 2.5E-25 |
| <i>OAS3</i> | 2'-5'-oligoadenylate synthase 3 | <i>Shisp_DN1184_c1_g1</i> | ↑ 5.12 | 1.3E-17 | 1.6E-14 |
| <i>I27L2</i> | Interferon alpha-inducible protein 27-like protein 2 | <i>Shisp_DN12103_c7_g1</i> | ↑ 6.20 | 1.6E-27 | 6.0E-24 |
| <i>SYWC</i> | Tryptophan-tRNA ligase, cytoplasmic | <i>Shisp_DN144721_c0_g1</i> | ↑ 8.44 | 1.1E-14 | 1.1E-11 |

Table 3. Twenty notable differentially expressed genes in *S. hispidus* lung following RSV infection. Negative Log2FC = downregulated, positive Log2FC = upregulated. Gene names/descriptions determined by BlastX against UniRef/SwissProt. Transcript IDs correspond to contig names (*Shisp_XXX*) in assembly FASTA file. Log2FC = log2FoldChange, q = FalseDiscoveryRate.

In *S. hispidus*, only 1 annotated gene was downregulated: *LMTD1*, which is involved in cell proliferation. Other downregulated genes had annotations indicating inclusion of signaling peptides (via SignalP) and transmembrane regions (via TmHMM) but with unknown function (Supplementary File 5). Several upregulated immune genes in *S. hispidus* were also upregulated in *S. fulviventer* (*I27L2*, *CXCL10*, *MX2*). Most upregulated genes were related to cytokine stimulation of antiviral activity (*OAS3*, *CXCL11*, *IRF9*, *IFIT1*, *NLRC5*), proteasomal degradation (*PSB9*), cytoskeletal reorganization in response to stress (*SYWC*, *K2C6A*), T cell activation (*HSH2D*, *LY6E*), post-transcriptional regulation (*ABEC1*), viral protein degradation (*RSAD2/Viperin*), complement pathway (*CO4A*, *C4BPA*), chemical metabolism (*GBP4*), and cellular defense-related signal transduction (*LG3BP*). These differentially expressed *S. hispidus* genes following RSV infection are presented in Table 3.

Of the differentially expressed genes (DEGs) following RSV infection, 12 DEGs were found to be either up- or down-regulated in both *S. fulviventer* and *S. hispidus*. The chemokine *CXCL10*, *MX2*, and *I27L2* were among the top 10 upregulated genes in both species. Other DEGs within our significance threshold include *GBP2*, *GBP4*, *IRF9*, *NLRC5*, *OAS3*, *PAR14*, *RN213*, *SYWC*, and *IIGP1*. All DEs can be found in Supplementary File 5.

Additionally, differentially expressed Gene Ontology (GO) terms were determined using the GoSeq R package⁴⁶. The top GOs enriched in infected lungs of both *S. fulviventer* (Fig. 5A) and *S. hispidus* (Fig. 5B) were “biological processes” such as response to cytokine and interferon-beta, immune system process, cell surface receptor signaling pathway, and viral process; “molecular functions” such as ion binding, catalytic activity, hydrolase activity, and “cellular components” such as intracellular membrane-bounded organelles and nucleus. Other enriched GOs in infected lungs were related to enhancing the components and function of cells, such as components of the membrane and vesicle, response to stress, regulation of apoptotic processes, regulation of viral genome replication, leukocyte activation, and binding of carbohydrates and other organic compounds. Only *S. hispidus* had enriched GOs in healthy tissues, and these included the cellular component involved in the guanyl-nucleotide exchange factor complex.

Confirmation of differentially expressed genes. To confirm DEGs three de novo assembled genes were selected based on their upregulation on day 5 post infection (*Shisp_DN132151_c0_g1* [*IIGP1*], *Shisp_DN12103_c7_g1* [*I27L2*], *Sfulv_DN158_c1_g1* [*IIGP1*]). qRT-PCR with SYBR green was performed with primers designed to target each gene, normalized to the β -actin housekeeping gene. The three genes included *IIGP1* in both *S. fulviventer* and *S. hispidus* and *I27L2* in *S. hispidus*. qRT-PCR confirmed gene upregulation post RSV infection (i.e., positive fold changes) that correlated with the RNA-Seq data (Supplementary Table 2).

Additionally, to further confirm and validate published differential expression analysis by Rajagopala et al.²⁷, where *S. hispidus* was infected with RSV A/Long in the same facility with the same strain and concentration of virus, we compared with DEGs identified in this study albeit samples were collected in different days. Rajagopala et al.²⁷ study included two time points; Day 4 and 6 post-infection, whereas, in this study we have collected sample only on Day 5. However, the DESeq2 analysis parameters used in both studies are same. Compared to the Rajagopala et al.²⁷, this current study confirmed 4 of the 5 Eukaryotic DEGs (*IFIT1*, *MX2*, *OASL2*, *RSAD2*)

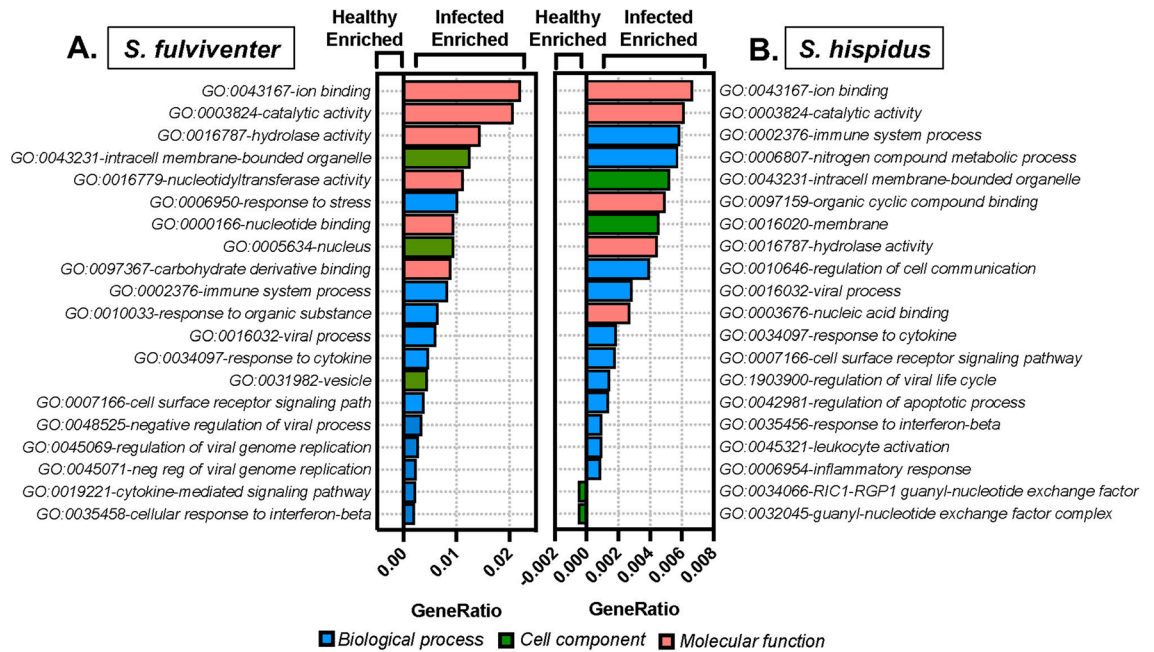


Figure 5. Categorical changes in gene expression after RSV infection in both (A) *S. fulviventer* and (B) *S. hispidus* using Gene Ontology (GO) annotations and GoSeq. SwissProt GO terms on the y-axis (color coordinated based on biological process [blue], cell component [green], or molecular function [red]) with relative expression of GOs on the x-axis (expressed as GeneRatio, which is the proportion of differentially expressed genes in each GO category to the total number of differentially expressed genes in all significant GOs). Positive GeneRatio = enriched in RSV-infected lungs; negative GeneRatio = enriched in healthy lungs (no significant GOs in healthy *S. fulviventer*). All GOs were statistically significant ($p < 0.05$, $q < 0.05$).

that were modified on day 4. Similarly, this study confirmed 29 of the 81 DEGs following infection by RSV (including *CFAB*, *GBP2/4/5/6*, *IFIT1*, *IIGP1*, *K2C6A*, *MX2*, *OASL2*, *RSAD2*; full data in Supplementary File 5) that were modified on day 6.

Discussion

Cotton rats are excellent pre-clinical animal models for the study of infections with RSV and other viruses, but without a published genome for either species, genetic or transcriptomic analysis in cotton rats is limited. In this study, we generated comprehensive transcriptomes from multiple organs of two inbred species of cotton rats: *S. fulviventer* and *S. hispidus*. Comparison of our two transcriptomic references revealed large differences in transcript sequence homology and baseline gene expression levels between each species, although the total number of annotated genes and functional categories (gene ontology, KEGG terms) were similar (Table 1).

Our transcriptome assembly and annotation surpass the quality and completeness standards of previously published transcriptome references in other animals^{35,36,47}. The quality and depth of our transcriptome assembly also improved upon our previous lung transcriptome of *S. hispidus* lung tissues²⁷ due to inclusion of multiple tissues. This resulted in nearly 3 times the number of annotated transcripts (117,153 vs. 38,736) and an additional 6169 unique gene annotations in *S. hispidus*. Furthermore, our application of the transcriptome reference to infer RSV-induced alterations of gene expression exhibits the utility of our reference in the study and treatment of many infections, including RSV, influenza, and other respiratory viruses.

Our analysis confirmed many genes previously implicated in RSV infection and immunity. *RSAD2* (or Viperin) is a significantly upregulated gene upon RSV infection in *S. hispidus* that inhibits RSV filament formation and cell–cell viral transmission without inhibiting viral protein expression⁴⁸. Viperin is also the most upregulated gene in human nasal epithelium following intranasal challenge with Rhinovirus⁴⁹. Another RSV-related gene is *I27L2*, the top upregulated gene in *S. hispidus* and *S. fulviventer* that is also the top differentially expressed gene in preterm infants with severe RSV infection⁵⁰. Other upregulated genes also shown to be involved in RSV infection include chemokines *CXCL10*⁵¹, *CXCL11*⁵², *LY6E*⁵³, *MX2*⁵⁴, *OAS1A*⁵⁵, *ISG15*⁵⁶, *IRP9*⁵⁷, *NLRC5*⁵⁸ and *K2C6A*⁵⁹. We also identified genes involved in the complement system (*CFAB*, *CO4A*, *C4BPA*) and prostaglandin metabolism (*PGDH*), both of which are involved in the host response to RSV^{60,61}. The results from our study supports the importance of these genes in host-mediated protection against RSV and the cotton rat's translational value in viral research.

CRY2 (downregulated in *S. fulviventer*) has also been indirectly attributed to RSV disease severity. *CRY2* is a transcription factor that modulates circadian rhythms through suppression of *BMAL1*, a putative regulator of cellular factors essential for viral replication⁶². Analysis of *BMAL1*-/- primary cells revealed low-*BMAL1* expression enhances susceptibility to influenza⁶³, parainfluenza virus 3, and RSV⁶⁴. *BMAL1* also shows seasonal variation in humans with lowest levels in winter months during peak respiratory virus season⁶². While this gene

will need further investigation in its role in RSV pathogenesis, this is the first association of modulated expression during RSV infection.

The cotton rat mimics the respiratory infections in humans due to the presence of human homologous genes that are absent in other laboratory rodents (e.g. *Mus musculus*), i.e., interferon-stimulated *MX* genes⁶⁵. Our analysis is consistent with the upregulation of *MX2* in both *S. fulviventer* and *S. hispidus* (as well as *MX1*, *MX1B*, and *MX3* in *S. fulviventer*) upon RSV infection as previously described⁶⁶, highlighting the importance of this model in capturing interferon-induced immune response to RSV and other viruses. Additionally, our analysis captures *IIGP1* upregulation in *S. fulviventer* and *S. hispidus*. *IIGP1* is another human interferon-stimulated gene with similar importance and function of *Mx* genes, but it is only present in mice as a less-effective paralog *IGPT*⁶⁷. A study on *IGPT*-deficient mice found no increased susceptibility or interferon-induced cytokine production towards intracellular pathogens such as *Listeria* and cytomegalovirus, which suggests a different upstream mechanism in mice⁶⁸. We, for the first time, successfully annotated *IIGP1* in cotton rats and show its modulation during RSV infection. While functional validation studies are needed, results from our study further argue the translational utility of the cotton rat as a model for RSV and other respiratory viruses compared to mice, who are deficient in *IIGP1*.

CREB1 (cAMP-responsive element-binding protein 1) is a widely studied transcriptional factor that modulates many immune genes and was downregulated in *S. fulviventer* upon RSV infection. *CREB1* promotes anti-inflammatory responses such as inhibition of NF- κ B activity, the induction of IL-10, and the generation of Tregs⁶⁹. *CREB1* activation was observed to be a driver of vaccine efficacy in HIV-1 vaccines in nonhuman primates by recruiting CD4+ T cells and B cells to the site of antigen presentation⁷⁰. Future functional studies are needed to further understand whether there is an association between RSV and *CREB1*.

One more interesting upregulated gene in both *S. fulviventer* and *S. hispidus*, *SYWC* is an interferon-induced activator of ERK, Akt, and eNOS pathways and cytoskeletal reorganization in response to stress and is a serum marker for pulmonary tuberculosis⁷¹. While this study is the first association of *SYWC* with RSV infection, further study may uncover it as a marker for severe RSV infection in tissue and serum.

While we have presented a comprehensive overview of the cotton rat transcriptome, we recognize some limitations in our study. Previously published data from Pletneva et al. has shown that different RSV strains and isolates have differential induction of interferon activated genes in cotton rats⁶⁶. While the study by Pletneva et al. only includes *S. hispidus*, we expect that this is also true for *S. fulviventer* due to similar susceptibility to RSV A/Long. To this end, we recognize that our conclusions are limited to RSV A/Long, and transcriptomic comparison of RSV variants should be considered for future studies. Further, we identified multiple genes that have not been previously implicated in RSV infection, such as (*LG3BP*, *SYWC*, *ABEC1*, *IIGP1*, *CREB1*). While our analysis suggests potential unknown mechanisms of pathogenesis, we recognize that association of these genes with RSV infection is severely limited in the absence of experimental data. Such efforts are beyond the scope of this paper on the reference cotton rat transcriptome but will be considered for future studies.

Conclusion

This report highlights the utility of cotton rat transcriptome to understand infection-induced transcriptomic changes in the absence of an available genome references. Both *S. hispidus* and *S. fulviventer* transcriptome references have been made publicly available at BioProject PRJNA816878 for additional research. Additionally, identification of gene sequences allows for better understanding of cotton rat genetics and generation of molecular tools, such as qRT-PCR primers and probes targeting genes of interest, recombinant protein, antibodies, and other assays. Differential gene expression analysis revealed host species specific differences upon RSV infection. As the development of RSV therapeutics calls for a well-developed, robust pre-clinical models for RSV, we hope our transcriptome references and gene associations with RSV can accelerate biomedical interventions against this pathogen of significant public health importance.

Methods

Animals, infection, and tissue collection. Four- to six-week-old *S. hispidus* (n=8 male, 1 female) and *S. fulviventer* (n=8 male, 1 female) cotton rats (~100 g) were obtained from the inbred colony maintained at Sigmovir Biosystems, Inc. Cotton rats in the colony were seronegative by ELISA to adventitious respiratory viruses (i.e., Pneumonia Virus of Mice, Rat parvovirus, Rat coronavirus, Sendai virus). Each species was randomly split into 2 groups: RSV-infected (n=5) and Uninfected controls (n=4). Transcriptome assembly and annotation was performed on all tissues (whole lung, large intestine, heart, spleen, and kidney) from 2 males within each uninfected group. Female animals were only included in RSV infected groups. To avoid fighting, all animals were individually housed in large polycarbonate cages and fed a diet of standard rodent chow and water ad libitum. All animal procedures followed NIH and USDA guidelines approved by the Sigmovir Biosystems, Inc. IACUC. Study design, analysis, and reporting of methods and results are presented in accordance with the ARRIVE guidelines.

The Long strain of RSV A/Long (ATCC Cat. # VR-26) was propagated in HEp-2 cells, and viral titer was determined using plaque assay. Cotton rats were intranasally inoculated under isoflurane anesthesia with 10⁵ plaque forming units in 100 μ L of either RSV suspension or PBS vehicle. Animals were sacrificed by carbon dioxide inhalation at day 5 of infection. Whole lung, large intestine, heart, spleen, and kidney were dissected from all animals and frozen in RNA-later (Invitrogen AM7021) for processing and sequencing at VUMC.

Lung histopathology and viral titers. Lungs (right lobe) were dissected and inflated with 10% neutral buffered formalin and immersed in formalin for fixation. Lungs were embedded in paraffin blocks, sectioned, and stained with hematoxylin and eosin (H&E). Pathologists were blinded to the study group, and slides were

examined for 4 parameters of pulmonary immune cell infiltration and inflammation as previously described⁷²: peribronchiolitis (bronchioles), perivasculitis (small blood vessels), interstitial pneumonia (alveolar walls), and alveolitis (alveolar space) (listed from least to greatest severity). Each condition was assigned a score from 0 to 4, where 0 is no pathology and 4 is max pathology. Scores correspond to the percentage of pathology present in the field of view pictured in Fig. 4A (0 = 0%, 1 = 5%, 2 = 25%, 3 = 75%, and 4 = 100%). Cumulative pathology scores were calculated by summing of the median score for each condition.

Viral RNA copy numbers were determined using Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) from RNA extracted from lung (lingular lobe) homogenates (see next section *RNA extraction and cDNA library preparation* for extraction methods). Following extraction, cDNA was generated using 1 µg of total RNA and the SuperScript™ III Reverse Transcriptase (Invitrogen™) kit according to the manufacturer's instructions. cDNA was diluted 1:5, and 3 µL was added to qPCR reactions using iQ™ SYBR® Green Supermix (10 µL total). Reactions for each target were performed in duplicate for each sample using primers (IDT, 250 nM final concentration) targeting the G and F mRNA of RSV and β-actin, which have been previously published and validated for use in both species of cotton rats⁷³. No-template-controls and a positive control (RNA extracted from RSV A/Long viral stock used for infection) were run on each plate. CT values for both G and F were averaged and normalized to β-actin. Results were calculated using the $2^{-\Delta\Delta CT}$ method⁷⁴. Figures and statistical analysis were performed using ANOVA/Tukey's multiple comparisons test in Prism 8.

RNA extraction and cDNA library preparation. RNA was prepared as previously described²⁷. In summary, small sections of cotton rat lung (lingular lobe), large intestines (~20 mm colon, flushed with sterile PBS), spleen, kidney, and heart were homogenized using a NextAdvance Bullet Blender® with RED RINO™ tubes containing 3.2 mm stainless steel beads (SSB32) and 2 × volume of QIAzol® Lysis Regent (Qiagen, cat. no. 79306) at max speed for 3 min. RNA was extracted from homogenates using RNeasy® Plus Universal Mini kit (Qiagen, cat. No. 73404) via additional QIAzol® (total volume 900 µL), gDNA eliminator, and chloroform with column DNase digestion as recommend by the manufacturer's protocol. RNA quality was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). Host rRNA was removed using the NEBNext rRNA Depletion Kit v2 (Cat no: NEB E7400X). The cDNA libraries were prepared using 1 µg total RNA from each sample using the NEBNext Ultra II RNA Library Prep Kit for Illumina (Cat no: NEB E7805L) following the manufacturers protocol for intact RNA (RIN > 7), AMPure XP Beads for cleanup steps (Beckman, cat. No. A63881), and NEBNext Multiplex Oligos for Illumina (Set 1, cat no: E7600S) for pooling samples. Sequencing was performed using Illumina NovaSeq6000 2 × 150 bp sequencing at the Vanderbilt Technologies for Applied Genomics (VANTAGE) core facility.

Transcriptome assembly. Data was then parsed into individual samples by barcode. Adaptor sequences, low quality (minimum Phred 33), and short (< 75 bp) reads were removed using Trimmomatic (version 0.36, "ILLUMINACLIP: NEB_multiplexoligos.fa:2:30:10 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75")⁷⁵. Only paired-end reads which passed the quality threshold as described in⁷⁶ were retained.

About 921 million paired-end reads from healthy tissues were utilized for de novo transcriptome assembly using the Trinity v2.13.1 software package³¹, with default parameters of 50 × coverage. An additional 314.5 million paired-end reads from viral-infected lungs were used for experimental analysis. Contigs were assembled for *S. fulviventer* and *S. hispidus* by first clustering individual contigs/transcripts into 'genes' followed by construction of de Bruijn graphs to report full length, alternatively spliced isoforms. All transcripts were then filtered by transcript length cutoff of < 200 bp using SeqKit³², contaminant annotation (Blast annotation of "virus", "bacteria", or "fungi"), sequence similarity of 95% or greater using CD-HIT³³, and removal of redundant mRNAs picked by the EvidentialGene tr2aacds pipeline³⁴. Assembly statistics of final transcripts, such as mean number of transcripts, transcript length, mean transcript length, N50, are listed in Table 1. Raw reads and assembly data will be deposited in SRA under BioProject PRJNA816878 upon acceptance of the manuscript for publication.

The RSEM package was used for quantifying and normalizing gene and isoform abundances from paired-end RNA-Seq data⁷⁷. RSEM enables accurate transcript quantification per sample and sample type without a reference genome. TPM (transcript per million mapped reads) was calculated using the RSEM package with Bowtie2 aligner, and a cutoff of TPM > 1 was used to filter the low-quality assembled transcripts to use for differential expression analysis. The Ex90N50 was calculated using the Trinity *contig_ExN50_statistic.pl* script.

Transcriptome characterization. The Trinotate v3.2.2 annotation pipeline was used to annotate transcripts from both *S. hispidus* and *S. fulviventer*. BlastX and BlastP (e-value cutoff of 1e-05)⁴² were used to find sequence homology of individual contigs and protein coding regions (determined by Transdecoder, <https://github.com/TransDecoder/TransDecoder>) against the UniProt/SwissProt database⁴¹ (data found in supplementary file 6). KEGG terms⁴³, Gene Ontology terms⁴⁴, and EggNOG terms⁴⁵ from the Swissprot⁴¹ database alignment with BLAST⁴². Other tools within the pipeline annotated transcripts based on protein domains via Pfam⁷⁸, transmembrane helices via TmHMM⁷⁹, and signaling proteins⁸⁰.

Differential gene expression and GO terms. Lungs isolated from RSV-infected and Uninfected groups were processed and sequenced as described above. The Trinotate pipeline was used to annotate all reads from experimental groups against our reference. Normalized expression levels of transcripts across tissues were determined using Salmon⁸¹ and the TPM (transcripts per million) metric; only transcripts with a TPM > 1 were used for downstream differential expression analysis (*S. fulviventer* = 270,451, *S. hispidus* = 474,882). Raw count matrices at the gene level for each species can be found in supplementary file 7 DESeq2 package³⁸ was used within the pipeline to compare experimental group (*RSV-infected lungs*) containing biological replicates with the

corresponding control group (*uninfected lungs*). Genes with a $p < 0.05$, adjusted $p < 0.05$ (“q”/false discovery rate/Benjamini-Hochberg), and a \log_2 fold change $> |1|$ were treated as differentially expressed. We used the goseq package in Bioconductor to detect differentially abundant GO terms⁴⁴. GOs with a $p < 0.05$ and adjusted $p < 0.05$ (“q”/false discovery rate/Benjamini-Hochberg) were treated as differentially expressed.

Validation of differentially expressed genes. RNA extracted from the lung tissue was used for qRT-PCR assays. qRT-PCR was performed in duplicate using the SuperScript™ III Reverse Transcriptase (Invitrogen™) kit and iQ™ SYBR® Green Supermix as described above. qRT-PCR primers were designed for 3 randomly selected up-regulated genes based on de novo assembly and differential expression analysis (*Shisp_DN132151_c0_g1*, *Shisp_DN12103_c7_g1*, *Sfulv_DN158_c1_g1*). Primers were designed using Primer3Plus⁸² and are reported in Supplementary Table 2. No-template-controls were run on each plate. Technical replicate CT values were averaged for each gene and normalized to β -actin housekeeping gene. Results were calculated as fold change induction over uninfected lungs using the $2^{-\Delta\Delta CT}$ method⁷⁴. Additionally, to further confirm differential expression of genes, we compared the results of our current study to our previous annotation of the *S. hispidus* lung transcriptome upon RSV infection²⁷ using the same DESeq2 significance parameters ($p < 0.05$, $q < 0.05$, $l2fc > |1.0|$).

Ethics approval and consent to participate. No human subjects participated in this study. All animal procedures followed NIH and USDA guidelines and were approved by the Sigmovir Biosystems, Inc. IACUC.

Data availability

The sequencing data will be deposited at the NCBI Short Read Archive (SRA) upon manuscript acceptance under BioProject PRJNA816878. For other details, please contact corresponding authors for specific data requests.

Received: 27 May 2022; Accepted: 5 September 2022

Published online: 04 October 2022

References

- Shi, T. *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* **390**, 946–958 (2017).
- The I-RVSG. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics* **102**, 531–537 (1998).
- Xing, Y. & Proesmans, M. New therapies for acute RSV infections: where are we? *Eur J Pediatr* **178**, 131–138 (2019).
- Fulginiti, V. A. *et al.* Respiratory virus immunization. I. A field trial of two inactivated respiratory virus vaccines; an aqueous trivalent parainfluenza virus vaccine and an alum-precipitated respiratory syncytial virus vaccine. *Am. J. Epidemiol.* **89**, 435–448 (1969).
- Chin, J., Magoffin, R. L., Shearer, L. A., Schieble, J. H. & Lennette, E. H. Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population. *Am. J. Epidemiol.* **89**, 449–463 (1969).
- Kapikian, A. Z., Mitchell, R. H., Chanock, R. M., Shvedoff, R. A. & Stewart, C. E. An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine. *Am. J. Epidemiol.* **89**, 405–421 (1969).
- Endt, K. *et al.* A recombinant MVA-based RSV vaccine induces T-cell and antibody responses that cooperate in the protection against RSV infection. *Front Immunol.* **13**, 841471 (2022).
- Detalle, L. *et al.* Generation and characterization of ALX-0171, a potent novel therapeutic nanobody for the treatment of respiratory syncytial virus infection. *Antimicrob Agents Chemother.* **60**, 6–13 (2016).
- Tang, W. *et al.* Small molecule inhibits respiratory syncytial virus entry and infection by blocking the interaction of the viral fusion protein with the cell membrane. *FASEB J.* **33**, 4287–4299 (2019).
- Shan, J., Britton, P. N., King, C. L. & Booy, R. The immunogenicity and safety of respiratory syncytial virus vaccines in development: A systematic review. *Influenza Other Respir Viruses* **15**, 539–551 (2021).
- Prince, G. A., Jenson, A. B., Horswood, R. L., Camargo, E. & Chanock, R. M. The pathogenesis of respiratory syncytial virus infection in cotton rats. *Am. J. Pathol.* **93**, 771–791 (1978).
- Graham, B. S., Perkins, M. D., Wright, P. F. & Karzon, D. T. Primary respiratory syncytial virus infection in mice. *J. Med. Virol.* **26**, 153–162 (1988).
- Grievies, J. L., Yin, Z., Durbin, R. K. & Durbin, J. E. Acute and chronic airway disease after human respiratory syncytial virus infection in cotton rats (*Sigmodon hispidus*). *Comp. Med.* **65**, 315–326 (2015).
- Rodriguez, W. J. *et al.* Respiratory syncytial virus (RSV) immune globulin intravenous therapy for RSV lower respiratory tract infection in infants and young children at high risk for severe RSV infections: Respiratory Syncytial Virus Immune Globulin Study Group. *Pediatrics* **99**, 454–461 (1997).
- Prince, G. A., Curtis, S. J., Yim, K. C. & Porter, D. D. Vaccine-enhanced respiratory syncytial virus disease in cotton rats following immunization with Lot 100 or a newly prepared reference vaccine. *J. Gen. Virol.* **82**, 2881–2888 (2001).
- Prince, G. A. *et al.* Enhancement of respiratory syncytial virus pulmonary pathology in cotton rats by prior intramuscular inoculation of formalin-inactivated virus. *J. Virol.* **57**, 721–728 (1986).
- Ottolini, M. G., Curtis, S. R., Mathews, A., Ottolini, S. R. & Prince, G. A. Palivizumab is highly effective in suppressing respiratory syncytial virus in an immunosuppressed animal model. *Bone Marrow Transplant* **29**, 117–120 (2002).
- Blanco, J. C. *et al.* Receptor characterization and susceptibility of cotton rats to avian and 2009 pandemic influenza virus strains. *J. Virol.* **87**, 2036–2045 (2013).
- Ottolini, M. G. *et al.* The cotton rat provides a useful small-animal model for the study of influenza virus pathogenesis. *J. Gen. Virol.* **86**, 2823–2830 (2005).
- Ottolini, M. G., Porter, D. D., Hemming, V. G. & Prince, G. A. Enhanced pulmonary pathology in cotton rats upon challenge after immunization with inactivated parainfluenza virus 3 vaccines. *Viral Immunol.* **13**, 231–236 (2000).
- Ottolini, M. G., Porter, D. D., Blanco, J. C. & Prince, G. A. A cotton rat model of human parainfluenza 3 laryngotracheitis: Virus growth, pathology, and therapy. *J. Infect. Dis.* **186**, 1713–1717 (2002).
- Pfeuffer, J., Puschel, K., Meulen, V., Schneider-Schaulies, J. & Niewiesk, S. Extent of measles virus spread and immune suppression differentiates between wild-type and vaccine strains in the cotton rat model (*Sigmodon hispidus*). *J. Virol.* **77**, 150–158 (2003).
- Hamelin, M. E. *et al.* Pathogenesis of human metapneumovirus lung infection in BALB/c mice and cotton rats. *J. Virol.* **79**, 8894–8903 (2005).

24. Patel, M. C. *et al.* Enterovirus D-68 infection, prophylaxis, and vaccination in a novel permissive animal model, the cotton rat (*Sigmodon hispidus*). *PLoS ONE* **11**, e0166336 (2016).
25. Blanco, J. C. *et al.* Prophylactic antibody treatment and intramuscular immunization reduce infectious human rhinovirus 16 load in the lower respiratory tract of challenged cotton rats. *Trials Vaccinol* **3**, 52–60 (2014).
26. Boukhvalova, M. S., Prince, G. A. & Blanco, J. C. The cotton rat model of respiratory viral infections. *Biologicals* **37**, 152–159 (2009).
27. Rajagopala, S. V. *et al.* Cotton rat lung transcriptome reveals host immune response to Respiratory Syncytial Virus infection. *Sci. Rep.* **8**, 11318–11318 (2018).
28. Porter, D. D., Prince, G. A., Hemming, V. G. & Porter, H. G. Pathogenesis of human parainfluenza virus 3 infection in two species of cotton rats: *Sigmodon hispidus* develops bronchiolitis, while *Sigmodon fulviventer* develops interstitial pneumonia. *J. Virol.* **65**, 103–111 (1991).
29. Langley, R. J., Prince, G. A. & Ginsberg, H. S. HIV type-1 infection of the cotton rat (*Sigmodon fulviventer* and *S. hispidus*). *Proc. Natl. Acad. Sci. USA* **95**, 14355–14360 (1998).
30. Strickland, B. A. *et al.* Microbial community structure and composition is associated with host species and sex in *Sigmodon* cotton rats. *Anim. Microbiome* **3**, 29 (2021).
31. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
32. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
33. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
34. Gilbert D. Gene-omes built from mRNA seq not genome DNA. *7th annual arthropod genomics symposium, Notre Dame* 2013.
35. Moreno-Santillan, D. D., Machain-Williams, C., Hernandez-Montes, G. & Ortega, J. De Novo transcriptome assembly and functional annotation in five species of bats. *Sci. Rep.* **9**, 6222 (2019).
36. Thunders, M., Cavanagh, J. & Li, Y. De novo transcriptome assembly, functional annotation and differential gene expression analysis of juvenile and adult *E. fetida*, a model oligochaete used in ecotoxicological studies. *Biol. Res.* **50**, 7 (2017).
37. Rajagopala, S. V. *et al.* Cotton rat lung transcriptome reveals host immune response to Respiratory syncytial virus infection. *Sci. Rep.* **8**, 1–12 (2018).
38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–550 (2014).
39. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
40. Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
41. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).
42. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
43. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
44. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
45. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309–D314 (2019).
46. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14 (2010).
47. Cardoso-Silva, C. B. *et al.* De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS ONE* **9**, e88462 (2014).
48. Jumat, M. R. *et al.* Viperin protein expression inhibits the late stage of respiratory syncytial virus morphogenesis. *Antiviral Res.* **114**, 11–20 (2015).
49. Zaas, A. K. *et al.* Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* **6**, 207–217 (2009).
50. Gao, J. *et al.* IFI27 may predict and evaluate the severity of respiratory syncytial virus infection in preterm infants. *Hereditas* **158**, 3 (2021).
51. Lindell, D. M., Lane, T. E. & Lukacs, N. W. CXCL10/CXCR3-mediated responses promote immunity to respiratory syncytial virus infection by augmenting dendritic cell and CD8(+) T cell efficacy. *Eur. J. Immunol.* **38**, 2168–2179 (2008).
52. Vissers, M. *et al.* Antibodies enhance CXCL10 production during RSV infection of infant and adult immune cells. *Cytokine* **76**, 458–464 (2015).
53. Li, L., Ni, Y. A., Song, Z., Yi, Z. & Wang, F. Identification of pathogenic genes and transcription factors in respiratory syncytial virus. *BMC Pediatr* **21**, 27 (2021).
54. Pletneva, L. M., Haller, O., Porter, D. D., Prince, G. A. & Blanco, J. C. Interferon-inducible Mx gene expression in cotton rats: cloning, characterization, and expression during influenza viral infection. *J Interferon Cytokine Res* **26**, 914–921 (2006).
55. Behera, A. K., Kumar, M., Lockey, R. F. & Mohapatra, S. S. 2'-5' Oligoadenylate synthetase plays a critical role in interferon-gamma inhibition of respiratory syncytial virus infection of human epithelial cells. *J. Biol. Chem.* **277**, 25601–25608 (2002).
56. Gonzalez-Sanz, R. *et al.* ISG15 is upregulated in respiratory syncytial virus infection and reduces virus growth through protein ISGylation. *J. Virol.* **90**, 3428–3438 (2016).
57. Fink, K. *et al.* IFNbeta/TNFalpha synergism induces a non-canonical STAT2/IRF9-dependent pathway triggering a novel DUOX2 NADPH oxidase-mediated airway antiviral response. *Cell Res.* **23**, 673–690 (2013).
58. Guo, X. *et al.* Respiratory syncytial virus infection upregulates NLRC5 and major histocompatibility complex class I expression through RIG-I induction in airway epithelial cells. *J. Virol.* **89**, 7636–7645 (2015).
59. Shirato, K., Ujiike, M., Kawase, M. & Matsuyama, S. Increased replication of respiratory syncytial virus in the presence of cytokeratin 8 and 18. *J. Med. Virol.* **84**, 365–370 (2012).
60. Edwards, K. M., Snyder, P. N. & Wright, P. F. Complement activation by respiratory syncytial virus-infected cells. *Arch. Virol.* **88**, 49–56 (1986).
61. Hashimoto, K. *et al.* Signaling through the prostaglandin I2 receptor IP protects against respiratory syncytial virus-induced illness. *J. Virol.* **78**, 10303–10309 (2004).
62. Zhuang, X., Rambhatla, S. B., Lai, A. G. & McKeating, J. A. Interplay between circadian clock and viral infection. *J. Mol. Med. (Berl)* **95**, 1283–1289 (2017).
63. Edgar, R. S. *et al.* Cell autonomous regulation of herpes and influenza virus infection by the circadian clock. *Proc Natl Acad Sci USA* **113**, 10085–10090 (2016).
64. Majumdar, T., Dhar, J., Patel, S., Kondratov, R. & Barik, S. Circadian transcription factor BMAL1 regulates innate immunity against select RNA viruses. *Innate Immun.* **23**, 147–154 (2017).
65. Boukhvalova, M. S., Prince, G. A. & Blanco, J. C. G. The cotton rat model of respiratory viral infections. *Biolog. J. Int. Assoc. Biol. Standard.* **37**, 152–159 (2009).

66. Pletneva, L. M., Haller, O., Porter, D. D., Prince, G. A. & Blanco, J. C. G. Induction of type I interferons and interferon-inducible Mx genes during respiratory syncytial virus infection and reinfection in cotton rats. *J. Gen. Virol.* **89**, 261–270 (2008).
67. Uthaiah, R. C., Praefcke, G. J., Howard, J. C. & Herrmann, C. IIGP1, an interferon-gamma-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization. *J. Biol. Chem.* **278**, 29336–29343 (2003).
68. Taylor, G. A. *et al.* Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IIGP. *Proc. Natl. Acad. Sci. USA* **97**, 751–755 (2000).
69. Wen, A. Y., Sakamoto, K. M. & Miller, L. S. The role of the transcription factor CREB in immune function. *J. Immunol.* **185**, 6413–6419 (2010).
70. Tomalka, J. A. *et al.* The transcription factor CREB1 is a mechanistic driver of immunogenicity and reduced HIV-1 acquisition following ALVAC vaccination. *Nat. Immunol.* **22**, 1294–1305 (2021).
71. De Groote, M. A. *et al.* Discovery and validation of a six-marker serum protein signature for the diagnosis of active pulmonary tuberculosis. *J. Clin. Microbiol.* **55**, 3057–3071 (2017).
72. Blanco, J. C., Boukhvalova, M. S., Pletneva, L. M., Shirey, K. A. & Vogel, S. N. A recombinant anchorless respiratory syncytial virus (RSV) fusion (F) protein/monophosphoryl lipid A (MPL) vaccine protects against RSV-induced replication and lung pathology. *Vaccine* **32**, 1495–1500 (2014).
73. Boukhvalova, M. S., Prince, G. A. & Blanco, J. C. Respiratory syncytial virus infects and abortively replicates in the lungs in spite of preexisting immunity. *J. Virol.* **81**, 9443–9450 (2007).
74. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
77. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323–323 (2011).
78. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021).
79. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
80. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
81. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
82. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

Acknowledgements

We thank Mr. Charles Smith and Ms. Martha Malache, ABSL2 technicians at Sigmovir Biosystems Inc., for breeding, handling, and care of cotton rats. We thank Angela Jones and others at VANTAGE for QC and sequencing our samples.

Author contributions

B.A.S., S.V.R., S.Y., J.C.G.B., and S.R.D. contributed to the study design. A.K. and M.S.B. contributed to the animal experiments and sample collection. B.A.S. contributed to sample processing and sequencing. B.A.S. contributed to bioinformatic and statistical analyses with consultation from S.V.R., M.H.S., S.B.P., and S.Y. J.C.G.B. and S.R.D. obtained the research funding supporting this study. B.A.S. and S.R.D. wrote the initial version of the manuscript, and all authors reviewed and approved the final version. All authors have reviewed the manuscript and consent for publication.

Funding

This work was supported by funds from National Institute of Allergy and Infectious Diseases (under award numbers R21AI142321-02S1, R21AI142321, R21AI154016, and R21AI149262); This work was supported by grants AI163543, AI109926, AI125215 to JCGB; the Centers For Disease Control and Prevention (CDC) 75D3012110094; the National Heart, Lung, and Blood Institute (under award numbers K23HL148638 and R01HL146401) and the Vanderbilt Technologies for Advanced Genomics Core (grant support from the National Institutes of Health under award numbers UL1RR024975, P30CA68485, P30EY08126, and G20RR030956). The contents are solely the responsibility of the authors and do not necessarily represent official views of the funding agencies.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19810-4>.

Correspondence and requests for materials should be addressed to J.C.G.B. or S.R.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022