A simple screening approach to prioritize genes for functional analysis identifies a role for IRF7 in the control of RSV disease.

Authors: Jacqueline U. McDonald[1], Myrsini Kaforou[2], Simon Clare[3], Christine Hale[3], Maria Ivanova[1], Derek Huntley[4], Marcus Dorner[5], Victoria J. Wright[2], Michael Levin[2], Federico Martinon-Torres[6], Jethro A. Herberg[2], John S. Tregoning[1,*].

Affiliations:

[1] Mucosal Infection and Immunity Group, Section of Virology, Imperial College London, St Mary's Campus, London, W2 1PG, UK.

[2] Section of Paediatrics, Imperial College London, St Mary's Campus, London, W2 1PG, UK.

[3] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.

[4] Imperial College Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, SW7 2AZ, UK.

[5] Molecular Virology, Section of Virology, Imperial College London, St Mary's Campus, London, W2 1PG, UK.

[6] Department of Paediatrics, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain

* Dr John S. Tregoning, Mucosal Infection and Immunity group, Section of Virology, Imperial College London, St Mary's Campus, London, W2 1PG.

Current address: MI, Janssen, The Netherlands.

Tel +44 20 7594 3176, Fax + 44 20 7594 2699 Email: john.tregoning@imperial.ac.uk

Short title: Simple dataset integration to prioritise gene targets

Summary:

Greater understanding about the functions of host gene products in response to infection is required. Whilst many of these genes enable pathogen clearance, some enhance pathogen growth or contribute to disease symptoms. Many studies have profiled transcriptomic and proteomic responses to infection, generating large datasets, but selecting targets for further study is challenging. Here we propose a novel data-mining approach; combining multiple, heterogeneous datasets to prioritise genes for further study, using respiratory syncytial virus (RSV) infection as a model pathogen with a significant healthcare impact. The assumption was that the more frequently a gene was detected across multiple studies, the more important its role. A literature search was performed to find datasets of genes and proteins that change after RSV infection. The datasets were standardized, collated into a single database, and then panned to determine which genes occurred in multiple datasets, generating a candidate gene list. This candidate gene list was validated using both a clinical cohort and an *in vitro* screen. We identified several genes that were frequently expressed following RSV infection with no assigned function in RSV control, including *IFI27, IFIT3, IFI44L, GBP1, OAS3, IFI44, IRF7* and *ISG15*. Drilling down into the function of these genes, we demonstrate a role in disease for interferon regulatory factor 7 (*IRF7*) which was highly ranked in the list; but not *IRF1*, which was not. Thus we have developed and validated an approach for collating published datasets into a manageable list of candidates, identifying novel targets for future analysis.

Importance:

Making the most of "Big Data" is one of the core challenge of current biology. There is a large array of heterogeneous datasets of host gene responses to infection, but these datasets do not inform us about gene function and require specialized skill sets and training to utilise. Here we describe an approach that combines and simplifies these datasets, distilling this information into a single list of genes commonly upregulated in response to RSV infection, as a model pathogen. Many of the genes in the list have unknown function in RSV disease. We validated the gene list with new clinical, in vitro and in vivo data. This approach allows the rapid selection of genes of interest for further more detailed studies, thus reducing time and costs. Furthermore, the approach is simple to use and widely applicable to a range of diseases.

Tweet: From great complexity comes forth simplicity: combining datasets to find novel gene targets for Respiratory Syncytial Virus (RSV) research.

Introduction

The interpretation of large datasets – Big Data – is one of the challenges of modern biology (1). Several powerful approaches have been developed to derive functional correlates from these large datasets, but they all have some limitations. Analysis of correlated or functionally related groups of genes *en bloc* simplifies analysis (2, 3), however this approach loses gene-level detail, particularly for genes with unknown roles. Systems biology approaches to identify key genes within pathways have been applied to vaccination (4) and infection (5) studies, but there is still a requirement to select individual genes for further analysis. High throughput screens enable the rapid identification of gene functions in an *in vitro* context (6-8), but these screens only investigate the role of genes in the context of individual cells, not in relation to the system as a whole. Recently, programs of work have been developed to systematically target every gene in the mouse genome to define their function (9, 10), nevertheless cost and ethical considerations require a focused selection of the targets of interest. The plethora of data available makes the prioritization of genes for further analysis challenging and often requires specialized skill sets and costly software. We propose a novel and simple approach to integrate published datasets to rapidly identify genes for their function in the control of infection, using freely available software.

We used this novel approach to identify genes involved in the host response to respiratory syncytial virus (RSV) infection as a proof of principle. RSV is a ubiquitous infection in early life and a significant cause of disease (11). Whilst the majority of children are infected with RSV during infancy, only a small proportion (2%) require hospitalization, of whom many have known risk factors including prematurity, congenital heart disease or immunodeficiency. However, the majority of hospitalized children (73-85%) have no known risk factor (12, 13). This phenotypic

variability in host response may reflect the role of host genetic polymorphisms in protection against or potentiation of severe disease (14). Severe RSV disease is associated with perturbation of normal airway function in the lower respiratory tract, but the events leading to the perturbation of airway function after RSV infection are not clear. Indeed the cause of disease may be heterogeneous with viral induced cell death causing disease in some infants and excess local inflammation having a role in others. The broader cellular immune response to RSV has been well dissected, with protective and pathogenic roles assigned to many cell types including macrophages (15, 16), NK (17, 18) and T cells (19, 20), but the molecular immune profile has not been fully explored. A number of studies in the last decade have published host 'omics profiles of RSV disease, identifying signatures of RSV infection, which enable discrimination between RSV and other respiratory viral infections (21, 22). Published associations of genes with the response to RSV are derived from a diverse range of systems, both *in vitro* and *in vivo*, in human and mouse, and based on RNA, DNA and protein data (Table S1). However whilst these studies have identified genes that change in response to infection, they have not defined the functional role of individual genes in RSV disease.

The aim of the current study was to integrate multiple published datasets to prioritise the genes associated with RSV infection and to dissect their function in RSV disease. To achieve this, we employed a combination of *in silico*, *in vitro* and *in vivo* approaches. Our conclusions were consistently supported by a new clinical study. Integrating multiple studies in this fashion increases the confidence in the role of the genes identified. Utilizing this approach we identified *IRF7* as a key gene in the control of RSV. Here, we demonstrate that integration of published 'omic datasets with high throughput studies generates insights into the genetic control of infection.

Results

Meta-analysis of RSV datasets reveals functional pathways in the control of RSV

We developed a novel approach to mine the mass of published gene and protein profile data in order to prioritize genes for functional profiling (Fig. 1). A literature search was performed to identify studies using 'omics tools to analyse the response to RSV infection. Data was collected from multiple published studies of RSV disease, selecting studies that included accessible lists of genes and/ or proteins that were detectable following RSV infection in either human or mouse (Table S1) or identified as significant in genetic studies or genome wide association studies. Due to the heterogeneity of the approaches, when multiple datasets were available in a single study we focused on the 24 hour after infection time points, in primary infection (rather than re-infection) and set an arbitrary cut off at 2-fold increase or decrease in gene expression for transcriptomic studies. Studies were collated in a single database and we then used a custom Perl script (countIDs: https://sourceforge.net/projects/countids/) to parse the file to find genes that were present in multiple studies. Genes were ranked according to their frequency of occurrence and weighted based on the type of study they appeared in: genetic association studies, *ex vivo* human, *in vitro* human or mouse, with more weighting on the human than the murine studies. This subjective weighting score was based on perceived relevance to human infection. Weighting reflected the nature of the input study, not specific data layers, which were treated equally: genes, mRNA and proteins were given equal weight. A candidate list of genes was then generated (Table 1). Using the weighted analysis, the genes that were most commonly reported as being up-regulated were *IFI27, IFIT3, GBP1, IFI44L, OAS3, IFI44* and *CXCL10* (Table 1). No functional role has been previously described for these genes in the control of RSV infection. Fewer genes were down-regulated after RSV infection and they were less uniformly represented

between studies, downregulated genes included *CLC, NDUFS1,* and *PFDN5* (Table 2). For comparison, an unweighted analysis was also performed (Table S2 and S3), these analyses identified similar patterns of genes, with IRF7 relatively higher in the unweighted than in the weighted analyses. Thus we can take published datasets and condense them into a single list, interestingly despite the heterogeneous nature of the input studies, we identified a large overlap in the genes identified.

To visualize the genes and interactions, we plotted the list using the Ingenuity software platform (Fig. 2). Genes with a score of 6 or more were included. The analysis identified a mixture of secreted factors (cytokines and chemokines) and intracellular factors (transcription factors and interferon stimulated genes). Based on our previous experience and published literature, it was of interest to note that extracellular proteins upregulated in response to RSV are often associated with enhanced disease (15, 23-25), whereas intracellular proteins are associated with disease control (26). To simplify the presentation of the interactions, we focused on the interactions of the top 16 upregulated and the top 5 downregulated genes, looking at direct interactions which have been observed experimentally. The main observation from this was that IRF7, which was observed to be upregulated in 9 independent studies (7, 27-32), interacted with several of the other most commonly identified genes. IRF7 was also central when the same data was analysed for canonical pathways: overall, the genes tended to fall into pathways associated with the inflammatory response to viral infection (Table 3).

Clinical validation of bioinformatic list

Since the list of candidate genes was generated using a literature mining approach, we sought validation using whole blood gene expression data from patients with RSV. We compared the transcriptomic profile of children hospitalized with RSV to age matched healthy controls, using microarrays. The list of genes derived from the literature (Table 1) was compared to the genes that were significantly differentially expressed (SDE) in children with RSV vs. healthy controls and an overlap of 73 out of 130 genes was observed. Of the genes identified as upregulated in the literature derived list, 66 were observed to be significantly upregulated in the clinical study, 45 were not significantly differentially expressed and only 2 were SDE but in the opposite direction (Fig. 3a). The downregulated gene list from literature was smaller but also had a lower proportion of agreement with the clinical study – only 5 of 17 genes were present in both studies (Fig. 3b). These data was then mapped onto the network built using the literature derived list (Fig. 3c), demonstrating that many of the top ranking genes from the literature list, including *IRF7, OAS2, RSAD2, HERC5, ISG15, IFI44, IL1RN, ARG1* and *IFIT3* were in agreement between both methods.

Validation of gene list using in vitro assay

Of note, a number of the genes that we identified as commonly upregulated have no known role in the control of RSV infection. We wished to screen the identified genes for their effect on RSV infection, using a flow cytometry based screen, described by Schoggins et al (7). We screened 39 interferon stimulated genes (ISG) identified from our *in silico* screen and we also included receptors and transcription factors identified as upstream regulators by pathway analysis in previous studies (*IL28RA, IRF1, IRF2, SOCS1, SOCS2, STAT3, TLR3* and *TLR7*). The epithelial cell line, HEp-2, was used as it represents the cell lineages that RSV first encounters during an

infection. HEp-2 cells were transduced with lentiviral vectors expressing each ISG and red fluorescent protein (RFP) in the same vector prior to infection with RSV expressing green fluorescent protein (33). PKR, IFI6 and OASL overexpression reduced GFP expression level (and therefore infection) by more than 75% of the control (Fig. 4). Of the genes identified with the most hits in the *in silico* studies (score >12), reduction in infectivity was as follows: *IFI27* (65.6%+/- 26.6), *IFIT3* (62.5%+/-10.6%), *IFI44L* (67.3%+/-39.7%), *GBP1* (Not in panel), *OAS3* (66.3%+/-36.0%), *IFI44* (74.5%+/-25.5%) and *ISG15* (58.3%+/-11.5%). *IRF7* overexpression led to a 76.7 +/- 11.7% reduction in RSV replication, which demonstrates that in addition to being centrally located in the predicted gene networks from *in silico* analysis, *IRF7* has a role in the control of RSV infection. The *in vitro* data supports a role in viral control for genes identified using our novel screen.

Validation of gene list using in vivo infection model

The overarching aim of the study was to identify new genes of interest for further study: the informatics and in vitro analysis identified IRF7 as being involved in the response to RSV and it has not previously been studied in the context of RSV infection. To validate our screening approach, we compared RSV infection in mice deficient for *IRF7*, a gene identified in our list, with one that is associated with anti-viral responses, but not identified in the list (*IRF1*). *Irf7*$^{-/-}$ and *Irf1*$^{-/-}$ mice were intranasally infected with 5×10^5 PFU of RSV-A (A2 strain) and were monitored daily for weight loss for seven days post-infection. Cohorts of mice were sacrificed on days four and seven post-infection to quantify viral burden and immunological changes over the course of the challenge. Mice were compared to wild type controls on the same background. *Irf7*$^{-/-}$ mice showed significant weight loss on days six and seven post-infection compared to wild

type littermates ($p < 0.01$) (Fig. 5A). There was no difference in weight loss between *Irf1*$^{-/-}$ and wild type controls in (Fig. 5B). RSV viral load was significantly greater in both *Irf7*$^{-/-}$ and *Irf1*$^{-/-}$ mice on day four ($p<0.05$, Fig. 5C, D) but not at day seven post-infection. Cellular infiltrate was quantified over the course of infection, which showed a significant increase in total cells resident in the lungs on day seven post-infection in *Irf7*$^{-/-}$ mice but not *Irf1*$^{-/-}$ ($p < 0.05$, Fig. 5E, F). Flow cytometry revealed an increase in all cellular sub-populations in *Irf7*$^{-/-}$ mice relative to wild type on day seven post-infection. In particular, numbers of total NK cells ($p < 0.05$) were significantly higher in the lungs (Fig. 5G, H), there was no significant difference in the *Irf1*$^{-/-}$ mice. Analysis of inflammatory cytokines present in the lungs revealed differences on day seven post-infection (Fig. 5I, J), with significantly higher levels of IL-1β (lung: $p < 0.05$,) in *Irf1*$^{-/-}$ mice relative to wild type controls.

Discussion

Here we used a novel, integrative approach to identify and characterize genes that are upregulated in response to RSV infection for further analysis. Previous studies have explored genetic signatures to discriminate RSV infection from other viral infections (21, 22). Our approach enabled identification of relevant genes in a hypothesis-free fashion identifying genes with both known and unknown function. We used a novel algorithm that permitted integration of multiple large genetic, gene expression and protein datasets to identify genes consistently upregulated after RSV infection, across multiple model systems. Using this approach we were able to distil down multiple heterogeneous studies into a single list of candidate molecules, generate testable hypotheses and then demonstrate functional relevance. Whilst we focused on

RSV, this approach is broadly applicable to other pathogens for which large sets of gene expression data are available and the data mining program is available as an open source program.

We identified a number of genes with no known function in RSV disease as potential targets for future investigation including *IFI27, IFIT3, IFI44L, GBP1, OAS3, IFI4*4 and *ISG15*. Several of these have reported roles from other infections, but a role has not been reported for RSV infection: IFI27 (also called ISG12) has a pro-inflammatory role by inducing the nuclear export of an anti-inflammatory nuclear receptor NR4A1 (34) and recently has been related to proliferation and cell cycling of human epidermal cells (35), GBP1 is a GTPase with a possible role in actin remodelling (36), IFI44 is anti-proliferative, OAS3 interacts with RNAseL (37) and ISG15 is a ubiquitin like modifier (38), which has been shown to reduce viral growth *in vitro* (39). However some of the most frequently upregulated genes for example IFI44L have no assigned molecular function. We took some of these genes forward into an *in vitro* assay and observed that there was a partial reduction of viral replication for all of the top hits tested (GBP1 was not included as not in the lentiviral panel). We focused on IRF7 for the *in vivo* studies because it gave a strong knockdown *in vitro* and was central to the predicted *in silico* network. IRF7 is an amplificatory molecule responding to pattern recognition receptor detection of viral infection inducing a further cascade of interferons (40) and is identified as the master regulator of type-I interferon-dependent immune responses (41). Previous studies have shown a role for IRF7 in human metapneumovirus (hMPV) (42) and influenza (43) virus infections. A recent study has demonstrated a role for IRF7 in the upregulation of RIG-I in response to RSV infection

237 in vitro (44) but the present study is the first to demonstrate a central role for IRF7 in the control
238 of RSV infection both in vivo and in vitro.
239
240 There are some limitations to this study. First, expression profiling of cells in the peripheral
241 blood has limitations in terms of be representative of responses in the respiratory epithelium,
242 where RSV infects and most of the studies are based on peripheral blood signatures. Due to the
243 heterogeneous nature of the available databases and publications about genes associated with
244 RSV infection, we had to make inclusion decisions which have led to a slight skewing of the
245 gene list. Where multiple gene sets were available we chose to include genes upregulated at 24
246 hours after infection which may have skewed the gene set to the interferon α/β response. Where
247 reported, a cut-off fold change of 2 was used, but this data was not reported in several studies,
248 likewise not all published data had gene lists that could be incorporated into the current study.
249 We chose a system to weight the data for the analysis giving priority to genetic association
250 studies and *in vivo* human data over *in vitro* data and mouse data. This weighting score was a
251 subjective decision based on a perception of the relevance of different data types to human
252 infection. Whilst it over-simplifies the differences both between and within different study types,
253 once the data has been collated, other scoring systems could easily be applied to the same meta-
254 data due to the simplicity of the analysis tools used. It is of note that weighted and unweighted
255 analysis gave similar lists. This demonstrates the power of this tool, because it can be adapted to
256 different questions, integrating heterogeneous datasets, furthermore, the simplicity of the
257 approach means that this can be performed quickly and easily. The *in vitro* screen only contains
258 interferon stimulated genes, which restricted the analysis of genes that were identified but not in
259 this family; for example GBP1. Finally there was only a limited disease phenotype in the control

mice because they are C57BL/6, which are relatively resistant to RSV infection, these mice were used to match the gene knockout animals. One limiting factor for the data mining from our experience is lack of standardization of published data sets: different papers have gene and protein lists in different file formats, with different nomenclatures and many only had lists in tables. Therefore transcription of the data had to be done by hand. A more uniform approach to these datasets would enable more studies to be included in meta-analyses.

Understanding more about the functions of the genes that are most commonly upregulated following RSV infection may give us insight into pathways to control disease after infection. A number of papers have proposed genetic signatures of RSV infection; here we show that the genes that are commonly upregulated as a result of RSV infection are characteristic of inflammation and viral control. There is an ongoing debate as to whether inflammation or viral mediated pathology is the primary cause of disease after RSV infection. In the context of this, one of the striking features of the pathway analysis was a divergence in effect on disease outcome of the gene products localized to the extracellular space and those found in the nucleus and cytoplasm. Animal studies suggest that the extracellular proteins enhance disease following RSV infection (45) by increasing inflammation, whilst the intracellular proteins reduce disease by decreasing viral replication. In support of this we have recently shown that the anti-viral gene, IFITM3, is important in the prevention of RSV infection (26, 46). In the *in vitro* studies over-expression of IFITM3 led to a 68 +/- 20.6 % reduction in RSV replication. Furthermore, overexpression of the chemokines CCL4, CCL5 and CCL8 had little effect on viral replication. This would suggest that boosting the anti-viral response without increasing inflammation would be a good strategy to control RSV disease. One potential target to achieve this would be the type

I IFN response, which if boosted may increase the transcription of anti-viral genes. However, recent studies have shown for RSV (47) and influenza (48) type I IFN contribute to inflammation and disease after viral infection, suggesting a sweet spot of IFN production where either too little or too much can both lead to a disease state. Targeting the host response may be particularly beneficial as it is not necessarily specific to the pathogen and is less likely to induce anti-microbial resistance.

There several are other approaches for integrating heterogeneous data for immunology research (49). Tool selection to use depends on the desired outcome, ranging from analysis of TCR and antibody repertoire analysis to network analysis and visualization. Some tools enable the integration of gene expression data with DNA variation (eQTL) (50) or against epigenetic status of the same gene (51), both of which enable greater understanding of the processes underpinning gene regulation and expression in the immune response. For higher level analysis, data can be merged using network analysis tools to see novel interactions between genes (52). These are sophisticated models underpinned by statistical techniques, requiring specialized skill sets and data analysis software to perform rigorously. Our approach is not. It requires no specialist informatics skills or software (a how to guide is in Fig S1): it uses excel, but would work using any spreadsheet program, and a free Perl script. Additionally, it performs a different function prioritizing genes for further study: this tool generates a hypothesis free, candidate list which can be investigated further. That the genes were biologically plausible and affected RSV infection in vitro and in vivo validates it as a down-selection tool. In conclusion here we describe a novel literature data mining approach for candidate gene prioritization. It has the benefit of simplicity and is broadly applicable to a range of infectious diseases.

Materials and Methods

Meta-analysis of papers and in silico analysis

We selected papers published prior to July 2015, using the search terms "RSV AND Microarray OR transcriptome OR genetic or proteome" on PubMed; other studies were also included from previous literature searches (Table S1). Papers were not included if the data set described in the study, with fold change and time point data was not easily accessible. Individual genes were included for analysis if they were reported to have a greater than 2 fold change than the reference group in the specific study. Gene lists were harvested from the published literature and collated in a single excel file. For analysis, genes were given a subjective weighting score based on the type of study they were collated from (in brackets): human genetic studies (4), human *in vivo* transcriptomic studies (3), human *in vitro* studies (2), murine studies (1). A custom Perl script: countIDs (available at: https://sourceforge.net/projects/countids/) was written to parse the gene list file generated, to assign a weighted score to each gene, generating an output gene list. Fig. 1 has a flowchart of the process and a step by step guide to the approach is in Fig S1. For visualization genes with an assigned score of 6 or more were included for Ingenuity pathway analysis (IPA – Qiagen).

Validation in Clinical Cohort

We established a case-control group comprising 27 RSV patients and 80 healthy controls. Whole blood (2.5ml) was collected into PAXgene blood RNA tubes (PreAnalytiX, Germany), incubated for 2 hours, frozen at -20°C within 6 hours of collection, before storage at -80°C. RNA was extracted using PAXgene blood RNA kits (PreAnalytiX, Germany) according to the manufacturer's instructions. The integrity and yield of the total RNA was assessed using an

Agilent 2100 Bioanalyser and a NanoDrop 1000 spectrophotometer. After quantification and quality control, biotin-labelled cRNA was prepared using Illumina TotalPrep RNA Amplification kits (Applied Biosystems) from 500ng RNA. Labelled cRNA was hybridized overnight to Human HT-12 V4 Expression BeadChip arrays (Illumina). After washing, blocking and staining, the arrays were scanned using an Illumina BeadArray Reader according to the manufacturer's instructions. Using Genome Studio software the microarray images were inspected for artefacts and QC parameters were assessed. Data was analysed using 'R' Language and Environment for Statistical Computing (R) 3.1.2 (53). Mean raw intensity values for each transcript were transformed to a logarithmic scale (base 2), corrected for local background intensities and normalized using robust spline normalization. We identified the transcripts that were significantly differentially expressed between the RSV infected children and the healthy control group with an adjusted *P*-value < 0.05, using a linear model for transcript expression. The functions lmFit and eBayes in the R package limma were used to calculate statistics.

Validation In vitro

The ISG library and screen were performed using a modified version of the assay described previously (7). $2x10^5$ HEp-2 cells were seeded into 96 well plates overnight prior to transfection with 10^5 of the individual Red-fluorescent protein Interferon Stimulated Gene (RFP-ISG) lentiviruses in DMEM supplemented with 20 mM HEPES and 4 mg/ml polybrene by spinoculation at 1,000 rpm for 1 hour. Twenty four hours later the cells were infected with green fluorescent protein expressing RSV (RSV-GFP) (33) and 24 hours after that the cells were harvested for analysis by flow cytometry. Live/dead discrimination was performed by the addition of live/dead fixable Aqua dead cell stain kit (Molecular probes) prior to acquisition of

data on an LSRFortessa (BD). Viral infection was determined based on the percentage of GFP: RFP double positive cells, relative infectivity in each well was normalized to the samples transfected with control lentivirus only. Data was analysed using CyAn ADP Summit 4.3.

Validation in vivo using gene knockout mice

Sex matched 8-10 week old wild type C57bl6N, *Irf1*$^{tm1a(EUCOMM)Wtsi}$ and *Irf7*$^{tm1(KOMP)Wtsi}$ mice (54) (Wellcome Trust Sanger Institute), were maintained in accordance with UK Home Office regulations, UK Animals Scientific Procedures Act 1986 under the project license PPL 80/2596. Animals were supplied with food and water *ad libitum* and were monitored daily for signs of illness. Founder mice were phenotyped through pipelines at the Wellcome Trust Sanger Institute as described previously (9, 10).

RSV in vivo studies

RSV strain A2 (from Prof P. Openshaw, Imperial College London) was grown in HEp-2 cells and viral titre determined by plaque assay. Mice were infected intranasally (i.n.) with 5 x 10^5 PFU in a volume of 100µl under isoflurane anaesthesia. Weight was measured daily to monitor disease severity. Lungs were removed, the smaller lobe was snap frozen in liquid nitrogen for RNA extraction and the remainder was homogenized by passage through 100-µm cell strainers (Falcon). Red blood cells in the lung sample were lysed in ammonium chloride buffer, and the remaining cells resuspended in RPMI medium with 10% fetal calf serum. Viable cell numbers were determined by trypan blue exclusion and lung cell types were differentiated by flow cytometry as described previously (55). In brief, cells were suspended in Fc block (Anti-CD16/32, BD) in PBS-1% BSA and stained with surface antibodies CD3-FITC (BD, Oxford

UK), CD4-APC (BD), CD8-APC Alexa750 (Invitrogen, Paisley, UK) NK1.1-PerCP-Cy5.5 (BD) and CD19-eflour450 (eBioscience, Hatfield, UK). Cells were run on a BD FACS Aria II. Singlet, lymphocyte cells were defined based on their size, side scatter and doublet discrimination, and then analysed for immune phenotype based the cell surface markers. RSV viral load was measured by quantitative RT-PCR for the RSV L gene using primers and probes previously described (19), with L gene copy number determined using a RSV L gene standard and presented relative to μg lung RNA. The cytokines IL-1β and IFNγ in lung were quantified using duosets from R and D systems.

Statistics

Analysis was performed by weighted Student's t-tests using GraphPad Prism 6.0.

Data availability

Data from the clinical study has been uploaded to http://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE80179.

Acknowledgments: We thank Alex Bukreyev UTMB, and Peter Collins NIH for the RSV-GFP, John Schoggins and Charles Rice for the ISG library and Gordon Dougan for advice and discussion. The data for the clinical cohort was collected by the IRIS consortium: **Imperial College London (UK)** Stuart Gormley, Jethro A. Herberg, Myrsini Kaforou, Michael Levin; Hannah Shailes, Victoria J. Wright, **Micropathology Ltd (UK)** Colin Fink, Ed Sumner; **Hospital Clínico Universitario de Santiago (Spain)** Miriam Cebey Lopez, Antonio Salas, Antonio Justicia Grande, Irene Rivero, Alberto Gómez Carballa, Jacobo Pardo Seco, José María Martinón Sánchez, Lorenzo Redondo Collazo, Carmen Rodríguez-Tenreiro, Lucia Vilanova Trillo, Federico Martinón Torres, **Southampton (UK)** Saul Faust, Jenni McCorkill, Sanjay Patel. **Funding**: The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° [115308] Biovacsafe,

400 resources of which are composed of financial contribution from the European Union's Seventh Framework
401 Programme (FP7/2007-2013) and EFPIA members' in kind contribution. Animal resources are funded from the
402 Wellcome Trust via grant number 098051. **Author contributions**: JT, SC designed the research; JM, CH, JT, SC
403 performed the *in vivo* studies; JT, SC, JH wrote the paper, MI, MD performed the *in vitro* studies, MK, JH, VW,
404 FMT and ML performed and analysed the clinical studies, DH, JM, MK performed the *in silico* studies.
405

References

1. **Marx V.** 2013. Biology: The big challenges of big data. Nature **498:**255-260.
2. **Xiong Q, Ancona N, Hauser ER, Mukherjee S, Furey TS.** 2012. Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. Genome Res **22:**386-397.
3. **Chaussabel D, Baldwin N.** 2014. Democratizing systems immunology with modular transcriptional repertoire analyses. Nat Rev Immunol **14:**271-280.
4. **Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, Pirani A, Gernert K, Deng J, Marzolf B, Kennedy K, Wu H, Bennouna S, Oluoch H, Miller J, Vencio RZ, Mulligan M, Aderem A, Ahmed R, Pulendran B.** 2009. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol **10:**116-125.
5. **Litvak V, Ratushny AV, Lampano AE, Schmitz F, Huang AC, Raman A, Rust AG, Bergthaler A, Aitchison JD, Aderem A.** 2012. A FOXO3-IRF7 gene regulatory circuit limits inflammatory sequelae of antiviral responses. Nature **490:**421-425.
6. **Pichlmair A, Kandasamy K, Alvisi G, Mulhern O, Sacco R, Habjan M, Binder M, Stefanovic A, Eberle CA, Goncalves A, Burckstummer T, Muller AC, Fauster A, Holze C, Lindsten K, Goodbourn S, Kochs G, Weber F, Bartenschlager R, Bowie AG, Bennett KL, Colinge J, Superti-Furga G.** 2012. Viral immune modulators perturb the human molecular network by common and unique strategies. Nature **487:**486-490.
7. **Schoggins JW, MacDuff DA, Imanaka N, Gainey MD, Shrestha B, Eitson JL, Mar KB, Richardson RB, Ratushny AV, Litvak V, Dabelic R, Manicassamy B, Aitchison JD, Aderem A, Elliott RM, Garcia-Sastre A, Racaniello V, Snijder EJ, Yokoyama WM, Diamond MS, Virgin HW, Rice CM.** 2014. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. Nature **505:**691-695.
8. **Karlas A, Machuy N, Shin Y, Pleissner KP, Artarini A, Heuer D, Becker D, Khalil H, Ogilvie LA, Hess S, Maurer AP, Muller E, Wolff T, Rudel T, Meyer TF.** 2010. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature **463:**818-822.
9. **Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, Gailus-Durner V, Greenaway S, Houghton R, Karp N, Leblanc S, Lengger C, Maier H, Mallon AM, Marschall S, Melvin D, Morgan H, Pavlovic G, Ryder E, Skarnes WC, Selloum M, Ramirez-Solis R, Sorg T, Teboul L, Vasseur L, Walling A, Weaver T, Wells S, White JK, Bradley A, Adams DJ, Steel KP, Hrabe de Angelis M, Brown SD, Herault Y.** 2012. Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mamm Genome **23:**600-610.
10. **White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, Bussell JN, Salisbury J, Clare S, Ingham NJ, Podrini C, Houghton R, Estabel J, Bottomley JR, Melvin DG, Sunter D, Adams NC, Sanger Institute Mouse Genetics P, Tannahill D, Logan DW, Macarthur DG, Flint J, Mahajan VB, Tsang SH, Smyth I, Watt FM, Skarnes WC, Dougan G, Adams DJ, Ramirez-Solis R, Bradley A, Steel KP.** 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell **154:**452-464.

11. **Tregoning JS, Schwarze J.** 2010. Respiratory viral infections in infants: causes, clinical symptoms, virology, and immunology. Clin Microbiol Rev **23:**74-98.
12. **Murray J, Bottle A, Sharland M, Modi N, Aylin P, Majeed A, Saxena S, Medicines for Neonates Investigator G.** 2014. Risk factors for hospital admission with RSV bronchiolitis in England: a population-based birth cohort study. PLoS One **9:**e89186.
13. **Garcia CG, Bhore R, Soriano-Fallas A, Trost M, Chason R, Ramilo O, Mejias A.** 2010. Risk factors in children hospitalized with RSV bronchiolitis versus non-RSV bronchiolitis. Pediatrics **126:**e1453-1460.
14. **Drysdale SB, Milner AD, Greenough A.** 2012. Respiratory syncytial virus infection and chronic respiratory morbidity - is there a functional or genetic predisposition? Acta Paediatr **101:**1114-1120.
15. **Pribul PK, Harker J, Wang B, Wang H, Tregoning JS, Schwarze J, Openshaw PJ.** 2008. Alveolar macrophages are a major determinant of early responses to viral lung infection but do not influence subsequent disease development. J Virol **82:**4441-4448.
16. **Kolli D, Gupta MR, Sbrana E, Velayutham TS, Hong C, Casola A, Garofalo RP.** 2014. Alveolar Macrophages Contribute to the Pathogenesis of hMPV Infection While Protecting Against RSV Infection. Am J Respir Cell Mol Biol doi:10.1165/rcmb.2013-0414OC.
17. **Harker JA, Yamaguchi Y, Culley FJ, Tregoning JS, Openshaw PJ.** 2014. Delayed sequelae of neonatal respiratory syncytial virus infection are dependent on cells of the innate immune system. J Virol **88:**604-611.
18. **Li F, Zhu H, Sun R, Wei H, Tian Z.** 2012. Natural killer cells are involved in acute lung immune injury caused by respiratory syncytial virus infection. J Virol **86:**2251-2258.
19. **Tregoning JS, Yamaguchi Y, Harker J, Wang B, Openshaw PJ.** 2008. The role of T cells in the enhancement of respiratory syncytial virus infection severity during adult reinfection of neonatally sensitized mice. J Virol **82:**4115-4124.
20. **Welliver TP, Garofalo RP, Hosakote Y, Hintz KH, Avendano L, Sanchez K, Velozo L, Jafri H, Chavez-Bueno S, Ogra PL, McKinney L, Reed JL, Welliver RC, Sr.** 2007. Severe human lower respiratory tract illness caused by respiratory syncytial virus and influenza virus is characterized by the absence of pulmonary cytotoxic lymphocyte responses. The journal of infectious diseases **195:**1126-1136.
21. **Herberg JA, Kaforou M, Gormley S, Sumner ER, Patel S, Jones KD, Paulus S, Fink C, Martinon-Torres F, Montana G, Wright VJ, Levin M.** 2013. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. J Infect Dis **208:**1664-1668.
22. **Mejias A, Dimo B, Suarez NM, Garcia C, Suarez-Arrabal MC, Jartti T, Blankenship D, Jordan-Villegas A, Ardura MI, Xu Z, Banchereau J, Chaussabel D, Ramilo O.** 2013. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. PLoS Med **10:**e1001549.
23. **Harker JA, Godlee A, Wahlsten JL, Lee DC, Thorne LG, Sawant D, Tregoning JS, Caspi RR, Bukreyev A, Collins PL, Openshaw PJ.** 2010. Interleukin 18 coexpression during respiratory syncytial virus infection results in enhanced disease mediated by natural killer cells. J Virol **84:**4073-4082.

24. **Matthews SP, Tregoning JS, Coyle AJ, Hussell T, Openshaw PJ.** 2005. Role of CCL11 in eosinophilic lung disease during respiratory syncytial virus infection. J Virol **79:**2050-2057.
25. **Tregoning JS, Pribul PK, Pennycook AM, Hussell T, Wang B, Lukacs N, Schwarze J, Culley FJ, Openshaw PJ.** 2010. The chemokine MIP1alpha/CCL3 determines pathology in primary RSV infection by regulating the balance of T cell populations in the murine lung. PLoS One **5:**e9381.
26. **Everitt AR, Clare S, McDonald JU, Kane L, Harcourt K, Ahras M, Lall A, Hale C, Rodgers A, Young DB, Haque A, Billker O, Tregoning JS, Dougan G, Kellam P.** 2013. Defining the range of pathogens susceptible to Ifitm3 restriction using a knockout mouse model. PLoS One **8:**e80723.
27. **Wong TM, Boyapalle S, Sampayo V, Nguyen HD, Bedi R, Kamath SG, Moore ML, Mohapatra S, Mohapatra SS.** 2014. Respiratory syncytial virus (RSV) infection in elderly mice results in altered antiviral gene expression and enhanced pathology. PLoS One **9:**e88764.
28. **Ravi LI, Li L, Sutejo R, Chen H, Wong PS, Tan BH, Sugrue RJ.** 2013. A systems-based approach to analyse the host response in murine lung macrophages challenged with respiratory syncytial virus. BMC Genomics **14:**190.
29. **Hennus MP, Janssen R, Pennings JL, Hodemaekers HM, Kruijsen D, Jansen NJ, Meyaard L, van Vught AJ, Bont LJ.** 2012. Host response to mechanical ventilation for viral respiratory tract infection. Eur Respir J **40:**1508-1515.
30. **Ioannidis I, McNally B, Willette M, Peeples ME, Chaussabel D, Durbin JE, Ramilo O, Mejias A, Flano E.** 2012. Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. J Virol **86:**5422-5436.
31. **Pennings JL, Schuurhof A, Hodemaekers HM, Buisman A, de Rond LC, Widjojoatmodjo MN, Luytjes W, Kimpen JL, Bont L, Janssen R.** 2011. Systemic signature of the lung response to respiratory syncytial virus infection. PLoS ONE **6:**e21461.
32. **Janssen R, Pennings J, Hodemaekers H, Buisman A, van Oosten M, de Rond L, Ozturk K, Dormans J, Kimman T, Hoebee B.** 2007. Host transcription profiles upon primary respiratory syncytial virus infection. J Virol **81:**5958-5967.
33. **Zhang L, Peeples ME, Boucher RC, Collins PL, Pickles RJ.** 2002. Respiratory syncytial virus infection of human airway epithelial cells is polarized, specific to ciliated cells, and without obvious cytopathology. J Virol **76:**5654-5666.
34. **Papac-Milicevic N, Breuss JM, Zaujec J, Ryban L, Plyushch T, Wagner GA, Fenzl S, Dremsek P, Cabaravdic M, Steiner M, Glass CK, Binder CJ, Uhrin P, Binder BR.** 2012. The interferon stimulated gene 12 inactivates vasculoprotective functions of NR4A nuclear receptors. Circ Res **110:**e50-63.
35. **Hsieh WL, Huang YH, Wang TM, Ming YC, Tsai CN, Pang JH.** 2015. IFI27, a novel epidermal growth factor-stabilized protein, is functionally involved in proliferation and cell cycling of human epidermal keratinocytes. Cell Prolif **48:**187-197.
36. **Ostler N, Britzen-Laurent N, Liebl A, Naschberger E, Lochnit G, Ostler M, Forster F, Kunzelmann P, Ince S, Supper V, Praefcke GJ, Schubert DW, Stockinger H, Herrmann C, Sturzl M.** 2014. Gamma interferon-induced guanylate binding protein 1 is a novel actin cytoskeleton remodeling factor. Mol Cell Biol **34:**196-209.

37. **Ibsen MS, Gad HH, Thavachelvam K, Boesen T, Despres P, Hartmann R.** 2014. The 2'-5'-Oligoadenylate Synthetase 3 Enzyme Potently Synthesizes the 2'-5'-Oligoadenylates Required for RNase L Activation. J Virol **88:**14222-14231.
38. **Zhao C, Denison C, Huibregtse JM, Gygi S, Krug RM.** 2005. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. Proceedings of the National Academy of Sciences of the United States of America **102:**10200-10205.
39. **Gonzalez-Sanz R, Mata M, Bermejo-Martin J, Alvarez A, Cortijo J, Melero JA, Martinez I.** 2016. ISG15 is upregulated in respiratory syncytial virus infection and reduces virus growth through protein ISGylation. J Virol doi:10.1128/jvi.02695-15.
40. **Ning S, Pagano JS, Barber GN.** 2011. IRF7: activation, regulation, modification and function. Genes Immun **12:**399-414.
41. **Honda K, Yanai H, Negishi H, Asagiri M, Sato M, Mizutani T, Shimada N, Ohba Y, Takaoka A, Yoshida N, Taniguchi T.** 2005. IRF-7 is the master regulator of type-I interferon-dependent immune responses. Nature **434:**772-777.
42. **Spann KM, Loh Z, Lynch JP, Ullah A, Zhang V, Baturcam E, Werder RB, Khajornjiraphan N, Rudd P, Loo YM, Suhrbier A, Gale M, Jr., Upham JW, Phipps S.** 2014. IRF-3, IRF-7, and IPS-1 promote host defense against acute human metapneumovirus infection in neonatal mice. Am J Pathol **184:**1795-1806.
43. **Crotta S, Davidson S, Mahlakoiv T, Desmet CJ, Buckwalter MR, Albert ML, Staeheli P, Wack A.** 2013. Type I and type III interferons drive redundant amplification loops to induce a transcriptional signature in influenza-infected airway epithelia. PLoS Pathog **9:**e1003773.
44. **Fang L, Choudhary S, Tian B, Boldogh I, Yang C, Ivanciuc T, Ma Y, Garofalo RP, Brasier AR.** 2015. Ataxia telangiectasia mutated kinase mediates NF-kappaB serine 276 phosphorylation and interferon expression via the IRF7-RIG-I amplification loop in paramyxovirus infection. J Virol **89:**2628-2642.
45. **Openshaw PJ, Tregoning JS.** 2005. Immune responses and disease enhancement during respiratory syncytial virus infection. Clin Microbiol Rev **18:**541-555.
46. **Zhang W, Zhang L, Zan Y, Du N, Yang Y, Tien P.** 2015. Human respiratory syncytial virus infection is inhibited by IFN-induced transmembrane proteins. J Gen Virol **96:**170-182.
47. **Goritzka M, Durant LR, Pereira C, Salek-Ardakani S, Openshaw PJ, Johansson C.** 2014. Alpha/beta interferon receptor signaling amplifies early proinflammatory cytokine production in the lung during respiratory syncytial virus infection. J Virol **88:**6128-6136.
48. **Davidson S, Crotta S, McCabe TM, Wack A.** 2014. Pathogenic potential of interferon alphabeta in acute influenza infection. Nat Commun **5:**3864.
49. **Kidd BA, Peters LA, Schadt EE, Dudley JT.** 2014. Unifying immunology with informatics and multiscale biology. Nat Immunol **15:**118-127.
50. **Shabalin AA.** 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics **28:**1353-1358.
51. **Pepke S, Wold B, Mortazavi A.** 2009. Computation for ChIP-seq and RNA-seq studies. Nat Methods **6:**S22-32.
52. **Arazi A, Pendergraft WF, 3rd, Ribeiro RM, Perelson AS, Hacohen N.** 2013. Human systems immunology: hypothesis-based modeling and unbiased data-driven approaches. Semin Immunol **25:**193-200.

53. **team Rcd.** 2011. R: a language and environment for statistical computing.
54. **Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A.** 2011. A conditional knockout resource for the genome-wide study of mouse gene function. Nature **474:**337-342.
55. **Russell RF, McDonald JU, Ivanova M, Zhong Z, Bukreyev A, Tregoning JS.** 2015. Partial Attenuation of Respiratory Syncytial Virus with a Deletion of a Small Hydrophobic Gene Is Associated with Elevated Interleukin-1beta Responses. J Virol **89:**8974-8981.

Fig. Legends

Fig. 1. Flowchart of gene selection method.

Fig. 2. Pathway analysis of genes upregulated after RSV infection. The top 130 genes identified by literature search shown by predicted function and sub cellular location. Top candidates indicated by bold outline, upregulated genes in red, downregulated in green. Interactions between the top 16 upregulated and top 5 downregulated genes based on known interactions in the Ingenuity Pathway Analysis (IPA) knowledgebase.

Fig. 3. Validation of bioinformatic screen in patient cohort. Significantly differentially expressed (SDE) genes from the clinical cohort were compared with the literature derived gene list. Overlaps between the gene lists expressed as pie charts, with directionality of agreement indicated for genes upregulated (A) or downregulated (B) in the literature derived list. Relative expression data from RSV infected patients overlaid on gene network derived from literature list (C), upregulated genes in red, downregulated in green, shading represents differential expression, bold outline from literature gene list.

Fig. 4. Flow cytometry confirmation of inhibitory function of genes identified *in silico*. HEp-2 cells were transduced with lentiviral vectors expressing genes of interest from the *in silico* screen. 24 hours later the cells were infected with RSV expressing GFP. Cells were harvested at 48 hours post infection and expression relative to control lentiviral transfected wells was assessed. Bars represent mean of n=3 +/- SEM. Red bars represent the top upregulated genes from the literature gene list.

Fig. 5. IRF7 but not IRF1 is important in the control of RSV infection. *IRF7*$^{-/-}$ (A, C, E, G, I) or *IRF1*$^{-/-}$ (B, D, F, H, I) were infected with $5x10^5$ PFU RSV A2 and compared to wild type

controls on the same background. Mice were weighed daily and weight changes recorded as a percentage of original weight (**A, F**). Lungs were excised and viral load calculated by qPCR on days 4 and 7 post-infection (**B, G**). Total cell counts from lung (**C, H**) were calculated, along with totals of CD3, CD4 and CD8 (T cells), CD19 (B cells) and $DX5^+$ (NK cells) (**D, I**) measured in lung by flow cytometry on day 7 post infection. Levels of the inflammatory cytokines IL-1β and IFNγ in lung (**E, J**) were measured by ELISA on day 7 post infection. Results show means ± S.E.M. (n > 5). Statistical significance was assessed by Student's *t*-test (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

Supplemental Table Captions

Table S1. Datasets mined for study. Papers published prior to July 2015 were selected, using the search terms "RSV AND Microarray OR transcriptome OR genetic or proteome" on PubMed. Table incorporates subjective weighting score based on the type of study they were collated from (in brackets): human genetic studies (4), human in vivo transcriptomic studies (3), human in vitro studies (2), murine studies (1). Sample type, time point and analysis method also included.

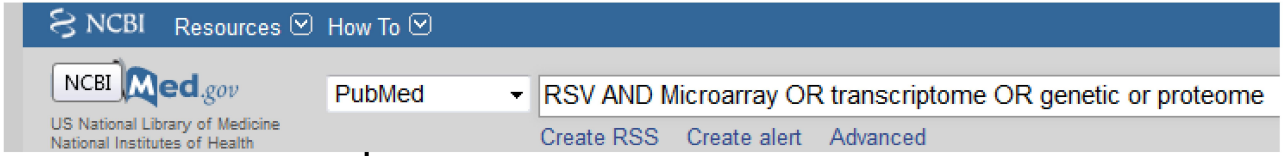
Table S2 Unweighted analysis – upregulated genes. Genes were collated from multiple studies of RSV; a two fold increase in expression compared to reference group in the study data was collated from, was used as a cut off, where available. Genes were analysed for multiple hits by a custom PERL script.

Table S3 Unweighted analysis – downregulated genes. Genes were collated from multiple studies of RSV; a two fold decrease in expression compared to reference group in the study data

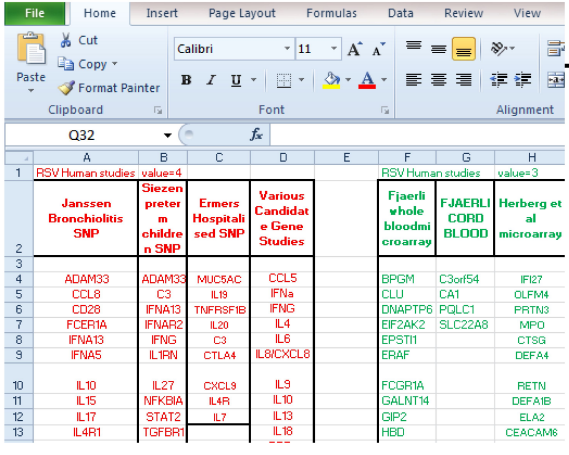
was collated from, was used as a cut off, where available. Genes were analysed for multiple hits by a custom PERL script.

Fig. S1. Try this at home. Step by step guide to using our analysis approach.

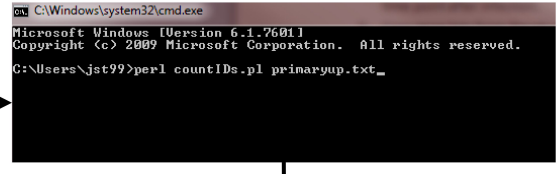
Step 1. Literature search



Step 2. Curate published gene lists into single excel database



Step 3. Run countIDs in Perl



Step 4. Generate candidate list

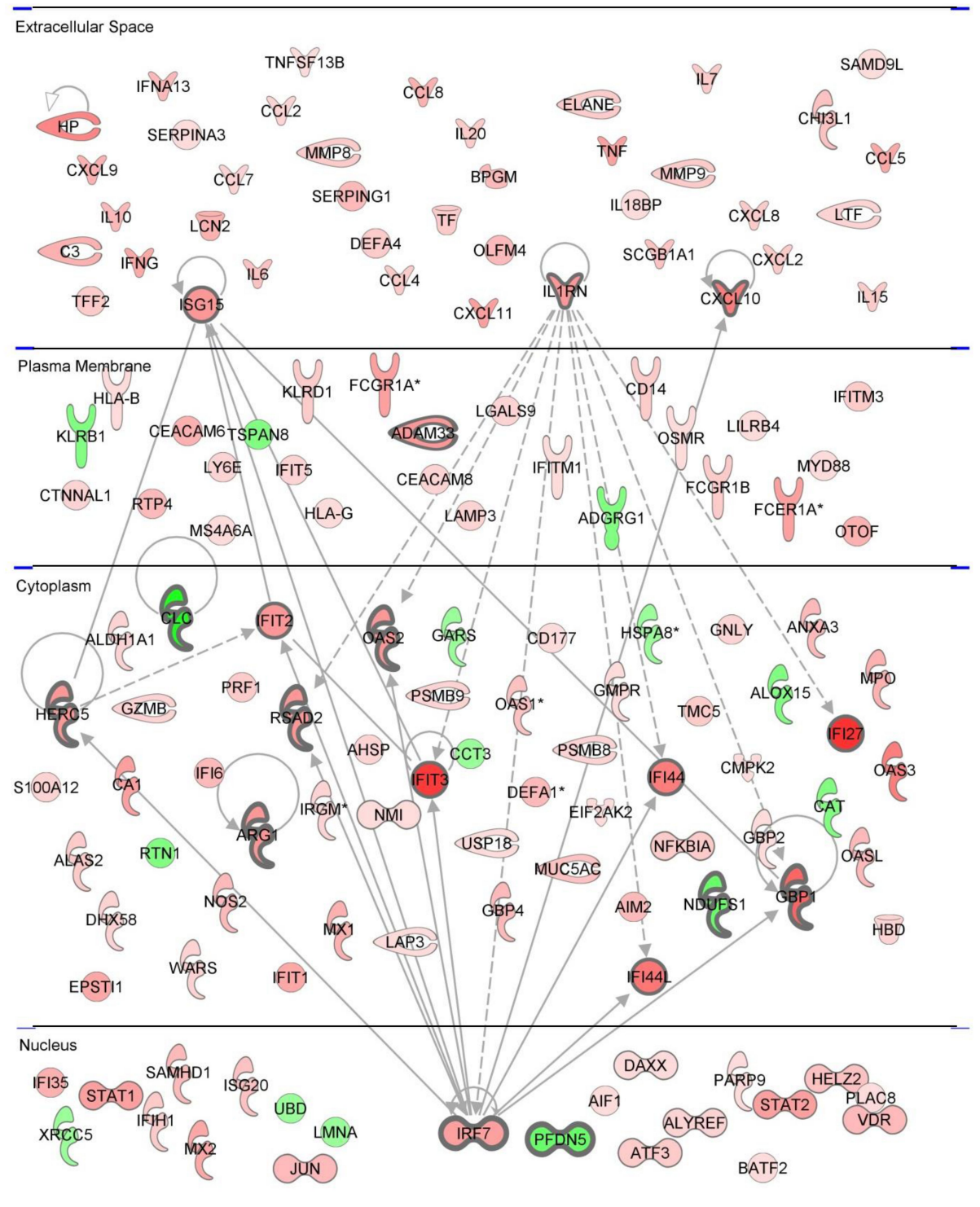
Weighted Score	Gene
24	IFI27
23	IFIT3
18	GBP1
16	IFI44L
15	OAS3, IFI44, CXCL10
14	HP
12	ADAM33, OAS2, IL1RN, IFIT2, HERC5 ARG1, ISG15

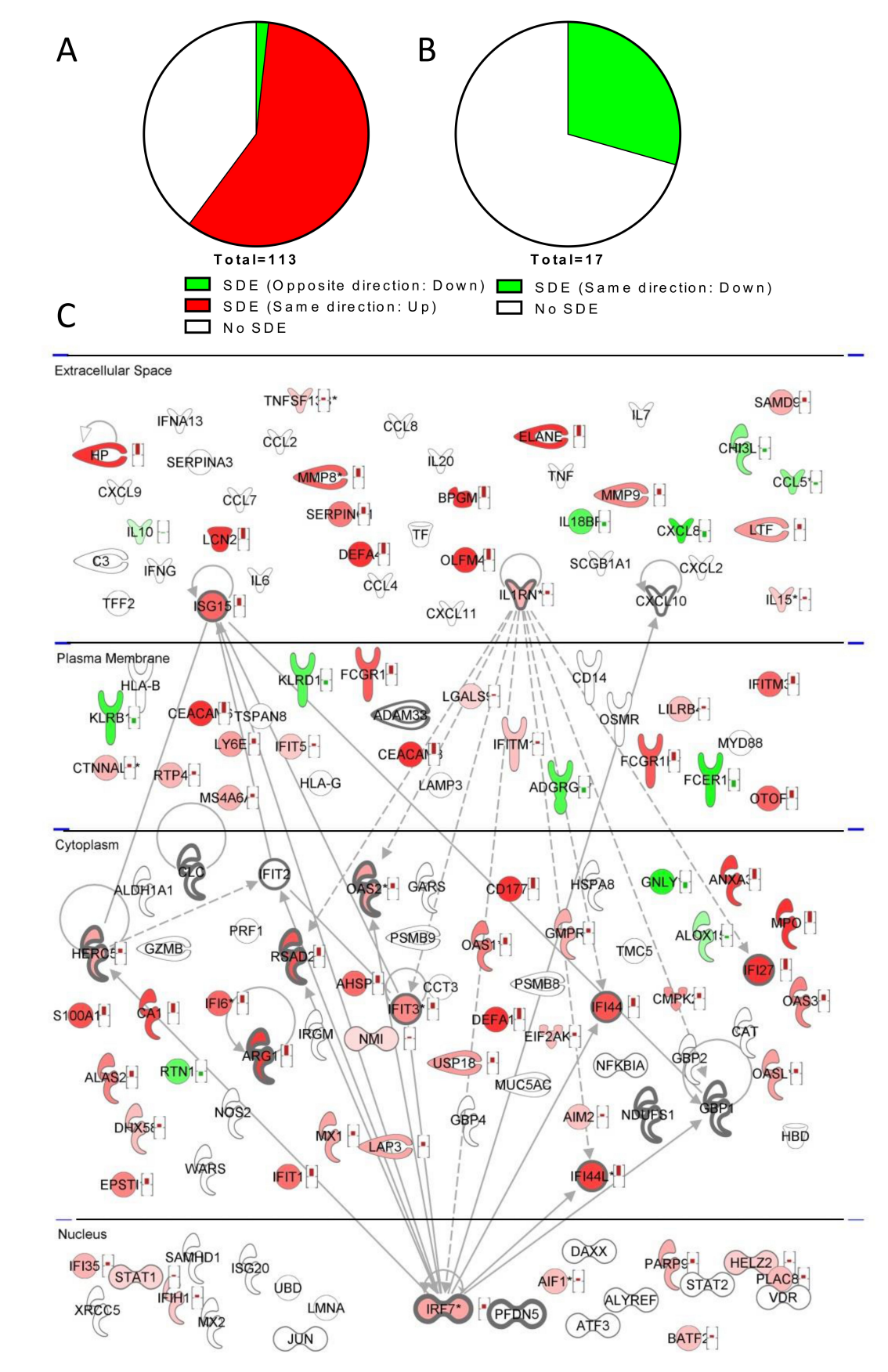
Step 5. Validate:

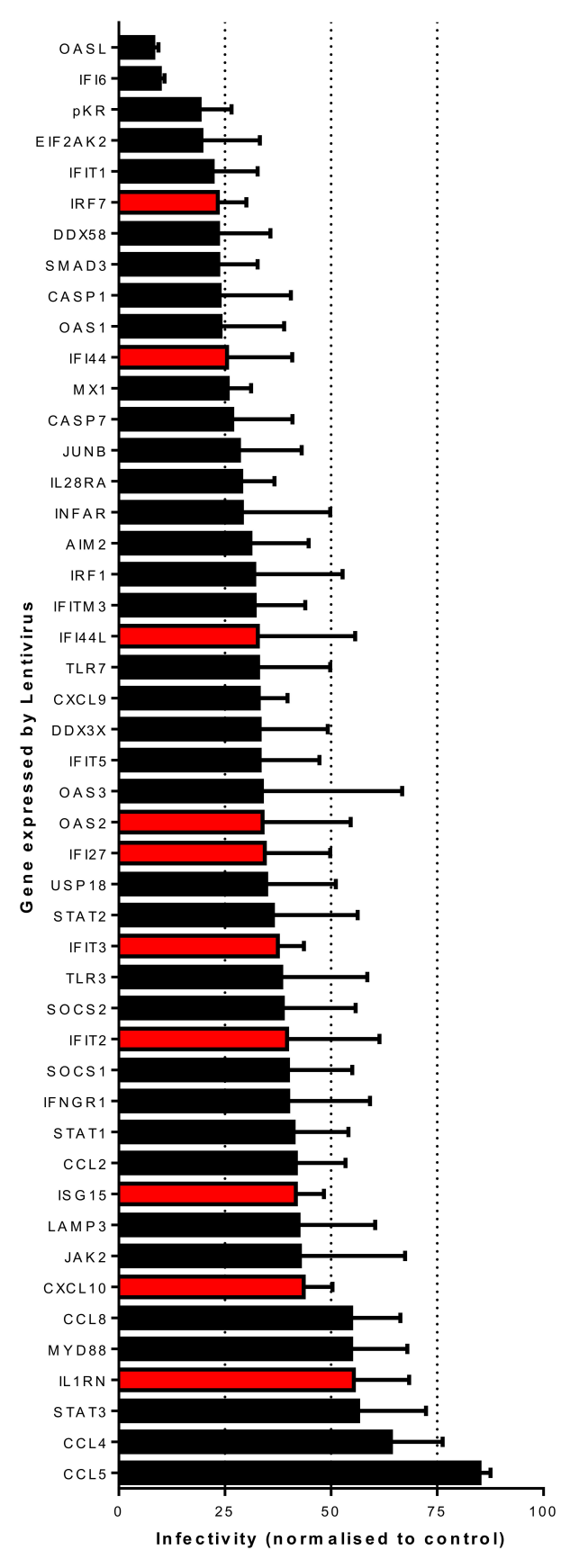
In vitro

In vivo

Clinical


Extracellular Space
TNFSF13B
IFNA13
CCL8
IL7
SAMD9L
HP
CCL2
ELANE
CHI3L1
SERPINA3
IL20
TNF
CCL5
CXCL9
MMP8
MMP9
CCL7
BPGM
SERPING1
IL18BP
IL10
LCN2
TF
CXCL8
LTF
C3
DEFA4
OLFM4
SCGB1A1
IFNG
IL6
CXCL2
CCL4
IL1RN
TFF2
ISG15
CXCL11
CXCL10
IL15
Plasma Membrane
KLRD1
FCGR1A*
CD14
IFITM3
HLA-B
LGALS9
KLRB1
CEACAM6
TSPAN8
ADAM33
OSMR
LILRB4
LY6E
IFIT5
IFITM1
MYD88
CEACAM8
FCGR1B
CTNNAL1
RTP4
HLA-G
LAMP3
ADGRG1
FCER1A*
MS4A6A
OTOF
Cytoplasm
CLC
IFIT2
ALDH1A1
OAS2
GARS
CD177
HSPA8*
GNLY
ANXA3
PRF1
PSMB9
OAS1*
GMPR
ALOX15
MPO
HERC5
GZMB
RSAD2
TMC5
IFI27
AHSP
CCT3
PSMB8
CMPK2
OAS3
S100A12
CA1
IFI6
IFIT3
IFI44
DEFA1*
CAT
IRGM*
NMI
EIF2AK2
ARG1
USP18
NFKBIA
GBP2
OASL
ALAS2
RTN1
MUC5AC
NOS2
AIM2
NDUFS1
GBP1
DHX58
GBP4
HBD
MX1
LAP3
WARS
IFIT1
IFI44L
EPSTI1
Nucleus
IFI35
SAMHD1
DAXX
PARP9
HELZ2
STAT1
ISG20
AIF1
STAT2
PLAC8
UBD
IFIH1
ALYREF
VDR
XRCC5
LMNA
IRF7
PFDN5
ATF3
MX2
BATF2
JUN


A
Total=113
B
Total=17
SDE (Opposite direction: Down)
SDE (Same direction: Up)
No SDE
SDE (Same direction: Down)
No SDE
C
Extracellular Space
Plasma Membrane
Cytoplasm
Nucleus
IFNA13
TNFSF1
CCL2
CCL8
IL7
SAMD9
HP
SERPINA3
ELANE
CHI3L1
CXCL9
MMP8
IL20
TNF
CCL5
CCL7
BPGM
MMP9
SERPIN
IL18BP
IL10
LCN2
TF
CXCL8
LTF
C3
IFNG
DEFA4
OLFM4
SCGB1A1
CXCL2
IL6
CCL4
TFF2
ISG15
IL1RN
CXCL10
IL15
CXCL11
HLA-B
KLRD1
FCGR1
CD14
IFITM3
KLRB1
CEACAM
TSPAN8
ADAM33
LGALS9
OSMR
LILRB
LY6E
IFIT5
CEACAM
IFITM1
MYD88
CTNNAL1
RTP4
HLA-G
LAMP3
ADGRG
FCGR1
FCER1
MS4A6A
QTOF
ALDH1A1
CLC
FIT2
OAS2
GARS
CD177
HSPA8
GNLY
ANXA
PRF1
PSMB9
OAS1
GMPR
MPO
HERC5
GZMB
RSAD2
TMC5
ALOX15
AHSP
CCT3
PSMB8
IFI27
S100A1
CA1
IFI6
IFIT3
DEFA1
IFI44
CMPK1
OAS3
IRGM
NMI
EIF2AK
CAT
ARG1
USP18
NFKBIA
GBP2
ALAS2
RTN1
MUC5AC
OASL
NOS2
AIM2
NDUFS1
GBP1
DHX58
MX1
LAP3
GBP4
HBD
WARS
IFIT1
EPSTI1
IFI44L
SAMHD1
DAXX
IFI35
STAT1
ISG20
AIF1
PARP9
HELZ2
STAT2
PLAC8
UBD
IFIH1
ALYREF
VDR
XRCC5
LMNA
IRF7
PFDN5
ATF3
MX2
JUN
BATF2


OASL
IFI6
pKR
EIF2AK2
IFIT1
IRF7
DDX58
SMAD3
CASP1
OAS1
IFI44
MX1
CASP7
JUNB
IL28RA
INFAR
AIM2
IRF1
IFITM3
IFI44L
TLR7
CXCL9
DDX3X
IFIT5
OAS3
OAS2
IFI27
USP18
STAT2
IFIT3
TLR3
SOCS2
IFIT2
SOCS1
IFNGR1
STAT1
CCL2
ISG15
LAMP3
JAK2
CXCL10
CCL8
MYD88
IL1RN
STAT3
CCL4
CCL5
Gene expressed by Lentivirus
0
25
50
75
100
Infectivity (normalised to control)

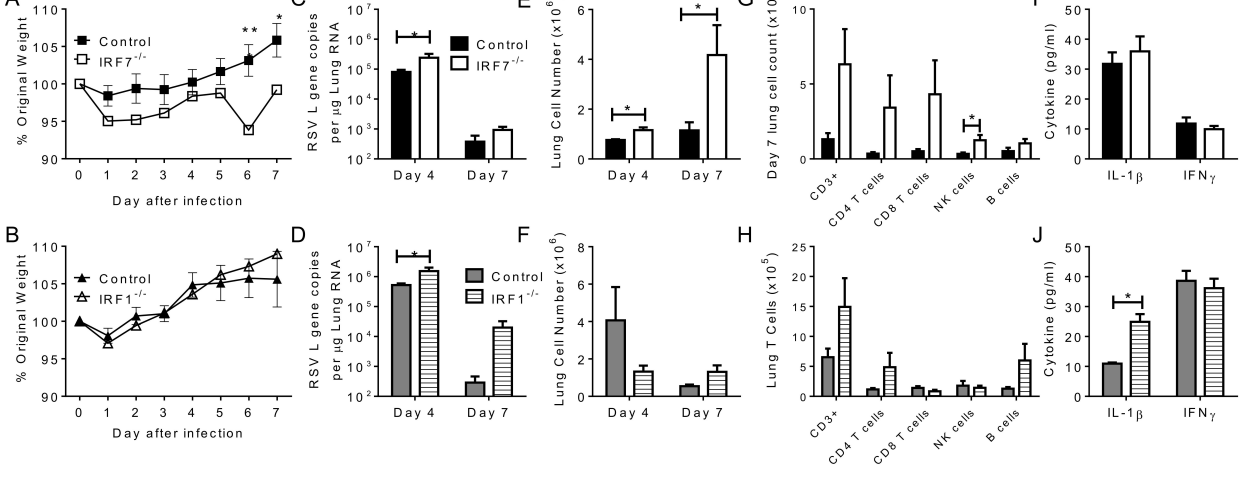

A
% Original Weight
Control
IRF7-/-
Day after infection
B
% Original Weight
Control
IRF1-/-
Day after infection
C
RSV L gene copies per µg Lung RNA
Control
IRF7-/-
Day 4
Day 7
D
RSV L gene copies per µg Lung RNA
Control
IRF1-/-
Day 4
Day 7
E
Lung Cell Number (x10^6)
Day 4
Day 7
F
Lung Cell Number (x10^6)
Day 4
Day 7
G
Day 7 lung cell count (x10^5)
CD3+
CD4 T cells
CD8 T cells
NK cells
B cells
H
Lung T Cells (x10^5)
CD3+
CD4 T cells
CD8 T cells
NK cells
B cells
I
Cytokine (pg/ml)
IL-1β
IFNγ
J
Cytokine (pg/ml)
IL-1β
IFNγ

Table 1. Most frequently upregulated genes following RSV infection. Genes were collated from multiple studies of RSV; a cut-off of a two fold increase in expression, compared to reference group in the study from which the data was collated, was used where available. Genes were weighted on the basis of the study they were collated from - Human Genetic Studies (4), Human *in vivo* Microarray studies (3), Human *in vitro* microarray (2), Mouse studies (1). After weighting, genes were analyzed for multiple hits by a custom PERL script.

Weighted Score	**Gene**
24	IFI27
23	IFIT3
18	GBP1
16	IFI44L
15	OAS3, IFI44, CXCL10
14	HP
12	ADAM33, OAS2, IL1RN, IFIT2, HERC5, RSAD2/Viperin, ARG1, ISG15
11	CXCL11, STAT2, TNF, FCER1A, IRF7, FCGR1A, CA1, STAT1
10	IFNG, IFIT1, CCL5, EPSTI1, MX2
9	CEACAM6, C3, CCL8, CXCL9, TRAC, IFI35, MX1, MPO, LCN2, OTOF
8	VDR, BPGM, IFI6, IL10, ANXA3, OLFM4, SAMHD1, SERPING1, DEFA1, IFNA13, RTP4, NOS2A, AIM2, JUN, OASL, GBP4
7	SCGB1A1, ISG20, CHI3L1, OAS1A, MUC5AC, PRIC285, IL6
6	IFITM3, IL20, MMP8, DEFA4, OASL2, ATF3, CXCL2, TMC5, TF, HBBP1, FCGR1B, IFIH1, CXCL8, CCL4, IL15, PRF1, ALAS2, NFKBIA, MSP, ELA2, KLRD1, IL7, MMP9, CD14
5	DEFA3, THOC4, CEACAM8, IFIT5, LAMP3, ERAF, IFNA, ALDH1A1, LGALS9, GZMB, LTF, CCL7, HBM, OAS1, CTNNAL1, WARS, LY6E, HBD, IGTP, S100A12, PSMB8, DHX58, IFI1, IFI47, CCL2, PSMB9, GNLY
4	EIF2AK2, MYD88, BATF2, CMPK2, GMPR, LILRB4, OSMR, IIGP1 TNFSF13B, DAXX, HLA-G, HSPA8, IL18BP, NMI, HLA-B, SAMD9L, CD177, IIGP2, LAP3, GBP2, USP18, PLAC8, MS4A6D, FCGR1, IFITM1, PARP9, AIF1, SERPINA3G, IFI202B

Table 2 **Most frequently downregulated genes following RSV infection.** Genes were collated from multiple studies of RSV; a cut-off of a two fold decrease in expression was used where available. Genes were weighted on the basis of the study they were collated from - Human Genetic Studies (4), Human *in vivo* Microarray studies (3), Human *in vitro* microarray (2), Mouse studies (1). After weighting, genes were analyzed for multiple hits by a custom PERL script.

Weighted Score	**Gene**
9	CLC
6	NDUFS1, PFDN5
5	RTN1, CAT, FCER1A, TSPAN8, ALOX15, GPR56, KLRB1
4	XRCC5, LMNA, UBD, CCT3, HSPA8, GARS

Table 3. **Canonical Pathways**. Ingenuity pathway analysis was applied to the top scoring genes

	Ingenuity Canonical Pathways	**Molecules**
1	Interferon Signaling	OAS1,IFIT1,IFNG,IFITM1,STAT1,IFNA1/IFNA13,IFIT3,STAT2,MX1,IFI35,IFITM3,PSMB8
2	Activation of IRF by Cytosolic Pattern Recognition Receptors	JUN,DHX58,STAT2,IFIT2,IL6,NFKBIA,IRF7,STAT1,TNF,IFNA1/IFNA13,ISG15,IFIH1,IL10
3	Communication between Innate and Adaptive Immune Cells	IL15,TNFSF13B,IFNG,CCL5,CXCL10,HLA-G,IL6,CXCL8,IL1RN,TNF,IFNA1/IFNA13,IL10,HLA-B,CCL4
4	Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza	IL15,IL1RN,IFNG,CCL5,CXCL10,TNF,IFNA1/IFNA13,CCL2,CCL4,IL6,CXCL8
5	Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	EIF2AK2,OAS1,IFNG,C3,MYD88,CCL5,OAS2,IL6,CXCL8,IRF7,TNF,IFNA1/IFNA13,OAS3,IFIH1,IL10
6	Granulocyte Adhesion and Diapedesis	CXCL9,CCL8,CCL7,CCL5,MMP9,CXCL10,CXCL2,CXCL8,CXCL11,IL1RN,TNF,CCL2,MMP8,CCL4
7	Agranulocyte Adhesion and Diapedesis	CXCL9,CCL8,CCL7,CCL5,MMP9,CXCL10,CXCL2,CXCL8,CXCL11,IL1RN,TNF,CCL2,MMP8,CCL4
8	Role of Cytokines in Mediating Communication between Immune Cells	IL15,IL1RN,IFNG,TNF,IFNA1/IFNA13,IL20,IL10,IL6,CXCL8
9	Differential Regulation of Cytokine Production in Intestinal Epithelial Cells by IL-17A and IL-17F	IFNG,CCL5,TNF,CCL2,IL10,CCL4,LCN2
10	Dendritic Cell Maturation	FCGR1B,IL15,MYD88,FCGR1A,STAT2,IL6,NFKBIA,IL1RN,STAT1,TNF,IFNA1/IFNA13,IL10,HLA-B