Gene Expression Meta-Analysis Identifies Molecular Changes Associated with SARS-CoV Infection in Lungs

Amber Park[1] and Laura K. Harris[1*]

[1]Davenport University, Harris Interdisciplinary Research, Grand Rapids, MI, USA

*** Correspondence:**
Laura K. Harris
laura.harris@davenport.edu

Keywords: coronavirus, gene expression, meta-analysis, gene set enrichment analysis, GSEA, SARS, SARS-CoV.

Abstract

Background: Severe Acute Respiratory Syndrome (SARS) corona virus (SARS-CoV) infections are a serious public health threat because of their pandemic-causing potential. This work uses mRNA expression data to predict genes associated with SARS-CoV infection through an innovative meta-analysis examining gene signatures (*i.e.*, gene lists ranked by differential gene expression between SARS and mock infection).

Methods: This work defines 29 gene signatures representing SARS infection across seven strains with established mutations that vary virulence (infectious clone SARS (icSARS), Urbani, MA15, ΔORF6, BAT-SRBD, ΔNSP16, and ExoNI) and host (human lung cultures and/or mouse lung samples) and examines them through Gene Set Enrichment Analysis (GSEA). To do this, first positive and negative icSARS gene panels were defined from GSEA-identified leading-edge genes between 500 genes from positive or negative tails of the GSE47960-derived icSARSvsmock signature and the GSE47961-derived icSARSvsmock signature, both from human cultures. GSEA then was used to assess enrichment and identify leading-edge icSARS panel genes in the other 27 signatures. Genes associated with SARS-CoV infection are predicted by examining membership in GSEA-identified leading-edges across signatures.

Results: Significant enrichment (GSEA $p<0.001$) was observed between GSE47960-derived and GSE47961-derived signatures, and those leading-edges defined the positive (233 genes) and negative (114 genes) icSARS panels. Non-random (null distribution $p<0.001$) significant enrichment ($p<0.001$) was observed between icSARS panels and all verification icSARSvsmock signatures derived from human cultures, from which 51 over- and 22 under-expressed genes were shared across leading-edges with 10 over-expressed genes already being associated with icSARS infection. For the icSARSvsmock mouse signature, significant, non-random enrichment (both $p<0.001$) held for only the positive icSARS panel, from which nine genes were shared with icSARS infection in human cultures. Considering other SARS strains, significant ($p<0.01$), non-random ($p<0.002$) enrichment was observed across signatures derived from other SARS strains for the positive icSARS panel. Five positive icSARS panel genes, CXCL10, OAS3, OASL, IFIT3, and XAF1, were found in mice and human signatures.

Conclusion: The GSEA-based meta-analysis approach used here identified genes with and without reported associations with SARS-CoV infections, highlighting this approach's predictability and usefulness in identifying genes that have potential as therapeutic targets to preclude or overcome SARS infections.

1 Background

Human β-coronaviruses (CoV) are enveloped, positive-sense RNA viruses that infect humans and a variety of animal species (Li et al., 2019). Human CoV infections typically cause mild upper respiratory distress, referred to as the common cold, and are generally non-lethal (Sims et al., 2013; Li et al., 2019). However, in 2002 a novel CoV was found to cause the potentially life-threatening disease, severe acute respiratory syndrome (SARS) with the initial outbreak of SARS-CoV infecting an estimated 8,400 people with over a 9% mortality rate (Boulos, 2004; Lai, 2005; Yount et al., 2005; Sims et al., 2013). Limited therapeutic options to treat SARS infections were available at the time. Ribavirin and corticosteroids were the cornerstones of treatment during the SARS outbreak despite Ribavirin showing a lack of efficacy in several studies and corticosteroid efficacy requiring further investigations in controlled trials (Lai, 2005). In 2012, another highly lethal CoV causing Middle East respiratory syndrome (MERS) emerged with an over 30% mortality rate (Bahadur et al., 2020; Petrosillo et al., 2020). MERS treatment options were limited also, with most therapies being existing drugs for other disease indications shown to target MERS-CoV replication *in vitro* (Modjarrad, 2016). Fortunately, both the SARS-CoV and MERS-CoV outbreaks were quickly contained through aggressive infection control measures, though efforts are still ongoing to control the pandemic of SARS-CoV2, a closely related coronavirus strain with almost 80% genome similarity to SARS-CoV. SARS-CoV2 is the causative agent of coronavirus disease 2019 (COVID-19), which is currently ravishing the globe and already responsible for hundreds of thousands of deaths worldwide (Yi et al., 2020). Therefore, an improved understanding of the molecular changes driving SARS infections can assist in the development of new therapies to fight COVID-19 and other future β-coronavirus outbreaks.

Many studies have been done to elucidate molecular changes associated with a SARS infection with a few studies specifically examined gene expression changes. These studies use differential expression analysis of individual genes using statistical methods such as fold change and/or T-test p-value to identify genes of interest (Yoshikawa et al., 2010; Gralinski et al., 2013; Mitchell et al., 2013; Sims et al., 2013; Totura et al., 2015; Menachery et al., 2018). This approach generates an exhaustive list of several thousand genes that meet pre-established statistical criteria (*e.g.*, fold change <2 and/or T-test p-value<0.05). Functional enrichment was performed to interpret generated gene lists using a variety of bioinformatic approaches including network analysis (Gralinski et al., 2013; Mitchell et al., 2013; Totura et al., 2015; Gralinski et al., 2018) and/or hypergeometric test-based approaches (*e.g.*, Fisher's Exact Test) that utilize established gene sets from public knowledgebases like Gene Ontology (Yoshikawa et al., 2010; Gralinski et al., 2013; Sims et al., 2013; Menachery et al., 2018). These studies identified several molecular changes associated with SARS infection including increased immune and inflammatory response as seen by increased expression of inflammatory mediators IL-6, IL-8, CXCL10, and IFN-λ in SARS-infected human lung cultures and mouse lung samples (Yoshikawa et al., 2010; Totura et al., 2015; Gralinski et al., 2018). Hypergeometric test-based approaches are known to be limited because only genes that meet an established cut-off (*e.g.*, T-test p-value<0.05) are considered (Subramanian et al., 2005). To overcome this limitation by considering all genes in an expression dataset, other studies utilized Gene Set Enrichment Analysis (GSEA) to calculate the enrichment of an established gene set from a public knowledgebase (*e.g.*, MSigDB, Blood Transcriptional Modules, and/or Kyoto Encyclopedia of Genes

This is a provisional file, not the final typeset article

and Genomes (KEGG) pathways) in a gene signature, which is defined as a gene list ranked by differential expression between mock- and SARS-infected samples by an appropriate statistical method (Mitchell et al., 2013; Gardinassi et al., 2020). Findings from these studies included positively enriched modules associated with antiviral IFN, cell cycle and proliferation, and monocytes and dendritic cells in peripheral blood mononuclear cells from SARS patients (Gardinassi et al., 2020). Further, by combining GSEA with a network analysis approach, regulatory genes associated with the pathogenicity of SARS have been identified (Mitchell et al., 2013), demonstrating that a GSEA-based approach is also good for gene candidate identification. While reported studies have identified molecular changes associated with SARS infection, the lack of available treatment options suggests that further examinations are needed.

In a prior study to identify novel genes associated with macrolide resistance in *Streptococcus pneumoniae*, we demonstrated a GSEA-approach for gene identification that compared differential gene expression between mRNA expression datasets (Goad and Harris, 2018). This study successfully identified known and novel genes though it was limited due to incomplete genome annotation, a common problem for many microbial genomes (Kolker et al., 2004; Sivashankari and Shanmughavel, 2006; Mohan and Venugopal, 2012; Bharat Siva Varma et al., 2015; Ijaq et al., 2015; Islam et al., 2015; School et al., 2016; Omeershffudin and Kumar, 2019). Applying our GSEA-approach to mRNA expression data derived from a species with a more completely annotated genome, such as human or mouse, overall results would improve. Therefore, in this paper we use the GSEA-approach previously used on *Streptococcus pneumoniae* to identify gene expression changes associated with SARS-CoV infection in human lung cell cultures and mouse lung samples. Using GSEA in this way generates a list of gene candidates associated with SARS-CoV infection which is similar in nature to gene candidate lists generated previously by using single-gene analysis. Rather than using a network analysis approach as done previously to refine our gene candidate list and allow for prioritization of gene candidates for future experimental examination, we performed a GSEA-based meta-analysis that incorporates over 20 different gene signatures. We hypothesize that therapeutically targeting gene candidates identified from our nested GSEA meta-analysis has the potential to improve treatment options for SARS-CoV and other β-coronavirus infections.

2 Methods

2.1 mRNA Expression Resources

To identify molecular changes associated with SARS infection in human lung cell cultures, we searched the Gene Expression Omnibus (GEO) repository (Edgar et al., 2002; Barrett et al., 2011; Barrett et al., 2013) to find seven datasets for use in our study (Table 1). Since most datasets contained samples infected with an infectious clone of SARS-CoV Urbani (icSARS), we selected the published SuperSeries GSE47963, which contained three independent datasets (GSE47960, GSE47961, and GSE47962) that examined gene expression in icSARS- and mock-infected human primary tracheobronchial epithelium cells (Mitchell et al., 2013), to identify and verify gene expression changes associated with icSARS infection mechanisms. These three datasets also contained gene expression data for cells infected with icSARS-dORF6, the icSARS strain with a genetic deletion causing a lack of expression of ORF6 which encodes an interferon antagonist protein that increases virulence by blocking human nuclear translocation (Yount et al., 2005; Mitchell et al., 2013; Sims et al., 2013), or SARS-BatSRBD, a SARS-CoV like virus isolated from bats that was synthetically modified to contain the spike receptor binding domain (SRBD) from the wild type Urbani strain to allow for infection of human and non-human primate cells (Becker et al., 2008; Mitchell et al., 2013). To extend our verification of identified gene expression changes associated

with icSARS infection into another lung cell type, we selected unpublished datasets GSE37827 and GSE48142 that examined gene expression in icSARS- and mock-infected human lung adenocarcinoma 2B-4 cells. GSE37827 also contained SARS-BatSRBD infected samples, and GSE48142 had samples infected with mutant strains containing either a genetic modification of NSP16 (deltaNSP16), which encodes a non-structural 2'O methyltransferase whose disruption increases sensitivity to type 1 and 3 interferon responses (Menachery et al., 2018), or ExoNI, which has no formal description in public knowledgebases . All gene expression data for these five datasets were profiled on the commercial probe name version of the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F platform (GEO GPL6480). Further, we selected GSE33267 because it had mock-, icSARS-, and dORF6-infected samples of Calu-3 cells, from which 2B-4 is a clonal derivative. GSE33267 was profiled on the feature number version of the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (GEO GPL4133). To compare identified changes associated with icSARS infection to changes associated with the icSARS's parent Urbani SARS strain, we selected GSE17400 since it had samples of 2B-4 with mock or Urbani infection (Yoshikawa et al., 2010). GSE17400 was profiled on the Affymetrix Human Genome U133 Plus 2.0 Array (GEO GPL570).

To determine if the molecular changes associated with SARS infection observed in human lung cell cultures are reproducible in an *in vivo* mouse model, we selected four datasets that examined lung samples from mock- or SARS-infected 20-week-old C57BL6 mice. GSE50000 contained samples from mice infected with icSARS, SARS-BatSRBD, or MA15, a mouse adapted SARS-CoV strain, at two different inoculation doses (10^4 and 10^5). GSE33266 had samples over an inoculation dose range (10^2, 10^3, 10^4, and 10^5). GSE49262 had samples from mice infected with MA15 (10^5 inoculation dose) or dORF6 mutant strains, and GSE49263 had samples infected with MA15 (10^5 inoculation dose) or dNSP16 mutant strains. All gene expression data for these datasets were profiled on the Probe Name version of the commercial Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F platform (GPL7202), except for GSE33266 which was profiled on the Feature Number version of the same platform (GPL4134).

All datasets measured gene expression over a time course, so we used time-matched samples collected at 48hrs post-infection unless otherwise stated. Expression data provided by GEO for all datasets except GSE49262 and GSE49263 were unlogged intensities, so we z-scored across all samples within the dataset for normalization prior to use when appropriate. We converted probes to Entrez ID for each gene using the GEO provided platform data table when necessary. Probes with no provided Entrez ID were removed from analysis. If multiple probes with the same Entrez ID existed, the probe with the highest coefficient of variance across duplicate probes was selected.

2.2 Defining Gene Signatures

From these datasets, we measured differential gene expression by Welch's two-sample T-test score of z-score or log2 normalized values. We then defined 29 gene signatures (*i.e.*, gene lists ranked by differential gene expression between SARS- and mock-infected samples), which included 17 gene signatures from human lung cultures and 12 signatures from mouse lung samples (Table 2). Specifically, we defined seven icSARSvsmock signatures (six in human cultures, one in mouse samples), one SARSvsmock signature with the Urbani strain in human cultures, eight MA15vsmock signatures doses in mouse samples representing varying inoculation, five ORF6vsmock signatures (four in human cultures, one in mouse samples), five BATSRBDvsmock signatures (four in human cultures, one in mouse samples), two NSP16vsmock signatures (one in human cultures, one in mouse samples), and one ExoNIvsmock signature from human cultures. Genes that are over-expressed in

This is a provisional file, not the final typeset article

SARS- compared to mock-infected samples (*e.g.*, positive T-score) fall within the positive tail of the gene signature while under-expressed genes (*e.g.*, negative T-score) fall in the negative tail (Figure 1A). Genes with no substantial change in expression between SARS- and mock-infected samples (*e.g.*, T-score around 0) are located toward the middle of the gene signature. Therefore, genes that fall within the tails of a gene signature likely change in response to a specific SARS infection. We note that three signatures were substantially skewed (*i.e.*, rank of genes where T-score crosses from positive to negative is in top or bottom quartile of signature). For substantially skewed signatures, we adjust all T-scores in the signature by the T-score of the gene at mid-point so that genes in the signature are balanced between positive and negative T-scores.

2.3 Overview of Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) is a statistical method that estimates enrichment between a query gene set (*i.e.*, unranked list of genes) and a reference gene signature (Subramanian et al., 2005). GSEA uses the statistical metric used to rank genes in the reference signature (*e.g.*, T-score) to calculate a running summation enrichment score where hits (*i.e.*, matches between query set and reference signature) increase the enrichment score proportional to the ranking statistical metric (*e.g.*, T-score) and a miss (*i.e.*, non-matches between query set and reference signature) decreases the enrichment score. From this, GSEA determines a maximum enrichment score for the specific query set and reference signature. Leading-edge genes contribute to reaching the maximum enrichment score, indicating leading-edge genes are associated with cellular response to a specific β-coronavirus infection. Further, GSEA calculates a normalized enrichment score (NES) from 1000 permutations of the reference signature to estimate the significance of enrichment between a specific query set and reference signature. This work used the javaGSEA Desktop Application release 3.0 version of GSEA available from Broad Institute.

2.4 Identification of icSARS Associated Genes

To identify gene expression changes associated with icSARS infection, we generated two icSARS gene panels (Figure 1B). To do this, we selected 500 genes from the positive and negative tails from the GSE47960-derived icSARSvsmock gene signature and used them to form two individual query gene sets. GSEA compared each query gene set to the GSE47961-derived icSARSvsmock gene signature (reference). Leading-edge genes from each analysis were used to define the two icSARS gene panels, one panel per tail. Pathway enrichment analysis was performed on both icSARS gene panels using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8.

2.5 Verification of icSARS Gene Panels

To verify the icSARS gene panels, we performed GSEA between icSARS gene panels and GSE47962-derived, GSE37827-derived, GSE48142-derived, and GSE33267-derived icSARSvsmock signatures (Figure 2A). To assess if results generated from GSEA could be achieved randomly, we randomly selected 1000 gene panels consisting of either 233- or 114-genes, to match the number of genes in the positive and negative icSARS panels, respectively, selected from the GPL6480 platform used to define gene signatures (queries) for GSEA against GSE47962-derived, GSE37827-derived, GSE48142-derived and GSE33267-derived icSARSvsmock signatures (references). These analyses generated a null distribution of NES to which we compared the NES achieved by icSARS gene panels for each reference gene signature and count the number of equal or better NES to estimate significance (*i.e.*, distribution p-value). Histogram data and associated graphs (*e.g.*, distribution curves and box and whiskers plot) were calculated using XLStat version 2020.3 (XLSTAT, 2013; Addinsoft, 2019).

2.6 Comparison to icSARS-induced Gene Expression Changes in Mice and Across Other SARS Strains

To examine differential gene expression of genes from the icSARS gene panels in mice, we performed GSEA between icSARS gene panels (queries) and the GSE17400-derived icSARSvsmock signature from mock- or icSARS-infected mouse lung samples (reference, Figure 2B). To compare gene signatures across infections caused by other SARS strains, we performed GSEA between icSARS gene panels (queries) and 22 signatures from samples from cells infected with strains other than icSARS, such as Urbani, MA15, SARS-BATSRBD, and mutants in ORF6, ExoNI, or NSP16 (references, Figure 2B). Eleven of these gene signatures came from mouse samples and 11 signatures came from human lung cultures with all SARS strains except ExoNI having strain-matched samples (*i.e.*, wild-type and mutant) in both mouse samples and human lung cultures. Random modelling for all comparisons were done as previously described. Any icSARS panel genes that were not included the dataset's platform still received consideration when examining across identified leading-edge genes. Heat maps were generated by Morpheus, https://software.broadinstitute.org/morpheus.

3 Results

3.1 Gene Signature Approach Identified Gene Expression Changes Associated with icSARS Infection for Human Lung Epithelium Cells *in vitro*

To identify genes associated with response to an icSARS infection, we began by defining the GSE47960-derived icSARSvsmock and GSE47961-derived icSARSvsmock gene signatures (Table 2) We used GSE47960-derived icSARSvsmock to generate two gene sets containing the 500 most differentially expressed genes from the positive and negative tails of GSE47960-derived icSARSvsmock (T-score >2.9 and <-3.2 for positive and negative tails, respectively) , capturing maximum coverage of the signature that was allowable by GSEA (Subramanian et al., 2005). To assess similarity between these two gene signatures, we calculated enrichment using GSEA between either GSE47960-derived icSARSvsmock positive or negative tail gene sets (individual queries) and the GSE47961-derived icSARSvsmock (reference), and achieved NES=3.28 and NES=-1.87 for positive and negative tail query gene sets, respectively, both with a GSEA p-value<0.001. We defined separate positive and negative icSARS gene panels from the 233 and 114 leading-edge genes identified (Supplemental Material STable 1), representing over- and under-expressed genes associated with icSARS infection.

Among genes in positive icSARS panel, we found 13 genes (5.5% of the panel) had previously reported associations with icSARS infections via single-gene analysis with minimum fold change of 2.0 and maximum false discovery rate-corrected p-value of 0.05 between mock and icSARS time-matched samples (Mitchell et al., 2013), demonstrating the predictive ability of our gene signature approach. There were no genes in the negative icSARS gene panel with established associations with icSARS infection. We also identified genes with no prior association with icSARS infection. These genes included 12 genes had a zinc finger in addition to the two that were already reported (5.2% of the panel), six genes encoding interferon induced proteins (2.6%), five genes from the solute carrier family (2.1%), and six genes encoding for uncharacterized proteins (2.6%) in the positive icSARS gene panel. In the negative icSARS gene panel, we found three zinc finger protein genes (2.6%), three genes from the solute carrier family (2.6%), and five genes encoding for uncharacterized proteins (4.4%). We speculate that these identified genes without previously reported associations with icSARS infections in human lung epithelium cultures are also associated with an icSARS infection.

This is a provisional file, not the final typeset article

3.2 Enrichment of icSARS Gene Panels and Specific icSARS Panel Genes Verified in Independent Datasets

To verify our icSARS gene panels are associated with response to an icSARS infection, we used GSEA to calculate enrichment between our icSARS panels (individual queries) and four verification gene signatures, GSE47962-derived icSARSvsmock, GSE37827-derived icSARSvsmock, GSE48142-derived icSARSvsmock, and GSE33267-derived icSARSvsmock, (Table 2), derived from independent datasets (individual references). We found significant similarity between positive and negative icSARS panels and GSE47962-derived icSARSvsmock (NES=3.51, Figure 3A, and NES=-3.05, Figure 3B, for positive and negative icSARS panel, both GSEA p-value<0.001). To determine how likely the NES achieved for icSARS panels would be achieved by random chance, we generated 1000 randomly selected 233- or 114-gene panels from the GSE47960-derived icSARSvsmock gene signature to match the size and potential composition of our icSARS panels. We then repeated GSEA using these randomly generated gene panels (individual queries) and the GSE47962-derived icSARSvsmock (reference) to generate a null distribution of NES achieved via random chance. From this, we found random NES ranged from 1.47 to -1.66 and 1.56 to -1.95 for the 233- and 114-gene panels, respectively (Figure 3C), illustrating that NES achieved by our icSARS panels are non-random (null distribution p-value<0.001). For GSE37827-derived icSARSvsmock, we again observed similarity between positive and negative icSARS panels and GSE37827-derived icSARSvsmock (NES=2.96, Figure 3D, and NES=-2.16, Figure 3E, for the positive and negative icSARS panels, both GSEA p-value<0.001). Random NES ranged from 1.45 to -1.60 and 1.55 to -1.72 for the 233- and 114-gene panels, respectively (Figure 3F), showing it was non-random (both null p-value<0.001). For GSE48142-derived icSARSvsmock, we again observed similarity between positive and negative icSARS panels and GSE48142-derived icSARSvsmock (NES=3.16, Figure 3G, and NES=-2.15, Figure 3H, for the positive and negative icSARS panels, both GSEA p-value<0.001). Random NES ranged from 1.34 to -1.52 and 1.56 to –1.61 for the 233- and 114-gene panels, respectively (Figure 3I), showing it was non-random (both null p-value<0.001). For GSE33267-derived icSARSvsmock, we observed similarity between positive and negative icSARS panels and GSE33267-derived icSARSvsmock (NES=3.55, Figure 3J, and NES=-2.30, Figure 3K, for the positive and negative icSARS panels, both GSEA p-value<0.001). Random NES ranged from 1.46 to -1.71 and 1.61 to -1.83 for the 233- and 114-gene panels, respectively (Figure 3L), showing it was non-random (both null p-value<0.001). Taken together, these results demonstrate that the enrichment achieved from our icSARS panels was true.

To determine which of our icSARS panel genes were verified across all signatures, we examined leading-edge genes identified by GSEA for each verification signature. Leading-edge genes for GSE47962-derived icSARSvsmock, GSE37827-derived icSARSvsmock, GSE48142-derived icSARSvsmock, and GSE33267-derived icSARSvsmock are listed in Supplemental Material STable 2, respectively. We observed 51 genes from the positive icSARS panel and 22 genes from the negative icSARS panel were shared across all verification signatures. Of the 51 positive icSARS panel genes, we noted 10 genes had reported associations with icSARS infection via single-gene analysis with minimum fold change of 2.0 and maximum false discovery rate-corrected p-value of 0.05 between mock and icSARS time-matched samples (Mitchell et al., 2013): nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (Entrez ID 4792), zinc finger CCCH-type containing 12A (ID 80149), cysteine-serine-rich nuclear protein 1 (ID 64651), zinc finger protein 433 (ID 163059), hairy and enhancer of split 1 from Drosophila (ID 3280), transformer 2 beta homolog (ID 6434), V-rel reticuloendotheliosis viral oncogene homolog B (ID 5971), pim-3 oncogene (ID 415116), chemokine (C-X-C motif) ligand 2 (ID 2920), and C-X-C motif chemokine ligand 10 (ID 3627). These data together verify our shared leading-edge genes are associated with

icSARS infection in human lung cultures and support the hypothesis that identified genes without previously reported associations are also associated with icSARS infection in human lung cultures.

3.3 Genes Associated with icSARS Infection in Mice Revealed through Enrichment between Positive icSARS Gene Panel and Mouse-derived icSARS Gene Signature

To determine if our icSARS gene panels are also associated with response to an icSARS infection in a mouse model, we examined the GSE50000-derived icSARSvsmock gene signature (Table 2), representing differential gene expression associated with icSARS infection in a mouse model. We used GSEA to calculate enrichment between our icSARS panels (individual queries) and the GSE50000-derived icSARSvsmock signature (reference) and observed significant enrichment with the positive icSARS panel (NES=1.54, GSEA p-value<0.001, Figure 4A) but not the negative icSARS panel (NES=0.81, p-value=0.848, Figure 4B). Achieved NES were non-random for the positive icSARS panel (NES range: -1.65 to 1.57, null distribution p-value=0.002, Figure 4C), but not for the negative icSARS panel (NES range: -1.72 to 1.64, p-value=0.601). These results suggest that genes in the positive icSARS panel are associated with an icSARS infection in both in human lung cultures and mouse lung samples, but the same cannot be concluded for the negative icSARS panel. Therefore, we do not consider leading-edge genes identified by GSEA from the negative icSARS panel.

When examining leading-edge genes identified by GSEA from the positive icSARS panel (Supplemental Material STable 3), we found nine of the 51 genes verified across human lung cultures (Figure 4D): interferon-induced protein with tetratricopeptide repeats 3 (human Entrez ID: 3437, mouse Entrez ID: 15959, gene symbol: IFIT3), C-X-C motif chemokine ligand 10 (human: 3627, mouse: 15945, symbol: CXCL10), XIAP associated factor 1 (human: 54739, mouse: 327959, symbol: XAF1), 2'-5'-oligoadenylate synthetase 3 100kDa (human: 4940, mouse: 246727, symbol: OAS3), zinc finger CCHC domain containing 2 (human: 54877, mouse: 227449, symbol: ZCCHC2), interferon induced protein 35 (human: 3430, mouse: 70110, symbol: IFI35), yrdC domain containing (human: 79693, mouse: 230734, symbol: YRDC), poliovirus receptor-related 4 (human: 81607, mouse: 71740, symbol: PVRL4), and cyclin L1 (human: 57018, mouse: 56706, symbol: CCNL1). We also noted that several icSARS panel genes were not included in this dataset's platform (Supplemental Material STable 4) with 11 of these genes contained in leading-edges for all icSARS gene signatures derived from human cell cultures, so we did not exclude them from further consideration. A heat map of differential gene expression (*i.e.*, T-scores) for these 20 genes across all six icSARS gene signatures revealed that only XAF1 was statistically significant (Welch's two-sampled, two-sided T-test p-value<0.05) in all icSARS gene signatures (Figure 3E), highlighting the advantage of using a GSEA-based rather than single-gene (*e.g.*, T-score only) meta-analysis, which would have likely missed the other genes identified here because of borderline significance in at least one signature.

3.4 Positive icSARS Gene Panel is Significantly Enriched Across Other SARS Strains Regardless of Strain Virulence Levels and Mechanisms

After identifying and verifying genes associated with an icSARS infection, we set out to find differential gene expression similarities between icSARS infection and infections from other SARS stains (Urbani, MA15, ΔORF6, BAT, ΔNSP16, and ExoNI) with varying levels of virulence caused by established strain mutation variations. Identified gene similarities represent genes associated with a SARS infection regardless of, and we hypothesize these shared genes may be useful targets clinically to preclude or overcome SARS infection. To identify genes associated with SARS infection across strains, we first compared icSARS gene panels to 18 gene signatures derived from samples of

This is a provisional file, not the final typeset article

human lung cultures or mouse lungs mock- and SARS-infected with one of six SARS strains (Urbani, MA15, ΔORF6, BAT-SRBD, ΔNSP16, and ExoNI; Table 2) using GSEA to calculate enrichment. We found the positive icSARS panel significantly enriched (NES>1.66, GSEA p-value<0.01) across all 18 gene signatures (Figure 5A) while the negative icSARS panel was significantly enriched across all signatures (NES<-1.52, p-value<0.01) except two BATvsmock signatures (Figure 5A), the GSE47961-derived signature from human cultures (NES=-0.95, p-value=0.569) and the GSE50000-derived signature from mouse samples (NES=-0.81, p-value=0.846). These data support our previous findings that enrichment of the negative icSARS panel was inconsistent in icSARS signatures (Figure 3B). We confirmed that achieved significant positive NES were non-random (null distribution p-value<0.002, Figure 5B and Figure 5C) via random modelling as previously described for all signatures except for the GSE33266-derived MA15(10^2)vsmock signature, which is the lowest infectious dose examined in this study. Due to this finding, we remove the MA15(10^2)vsmock signature from further analysis.

3.5 Meta-analysis of Leading-edge Genes from Positive icSARS Panel GSEA Revealed Five Top Gene Candidates

To determine which positive icSARS panel genes may make good therapeutic targets for SARS infections caused by specific strains with varying virulence levels and mechanisms, we compared inclusion of leading-edge genes identified by GSEA across 22 gene signatures derived from five SARS strains (Supplemental Material STables 5-13). We began by examining leading-edge gene inclusion in each SARS strain specifically (Figure 6A). Genes identified through leading-edge intersections represent genes associated with infection of that specific SARS strain. We identified several genes with known associations to infection with specific SARS strains, such as *fos* and *jun* that were found in leading-edges from signatures derived from dORF6 (Sims et al., 2013) and Urbani (Yoshikawa et al., 2010) infections in human lung cell cultures. We also found several genes with no previously reported associations to that specific SARS infection. Taken together, these findings are akin to findings from our meta-analysis of icSARS infection earlier that demonstrate the ability of our meta-analysis approach to identify genes associated with specific SARS infections and predict new genes associated with infections of specific SARS strains.

To determine which positive icSARS panel genes may be best therapeutic targets to preclude or overcome SARS infections regardless of strain, we analyzed the intersection of common leading-edge genes across all six SARS strains examined in this study. We found five positive icSARS panel genes, interferon-induced protein with C-X-C motif chemokine ligand 10 (CXCL10), 2'-5'-oligoadenylate synthetase 3 (OAS3), 2'-5'-oligoadenylate synthetase-like (OASL), tetratricopeptide repeats 3 (IFIT3), and XIAP associated factor 1 (XAF1), in all 22 positive icSARS panel leading-edges (Figure 6B). Differential gene expression (*i.e.*, T-scores) heat maps illustrated the strong consistency and extent of expression changes observed across gene signatures (Figures 4C and 6C), further supporting the conclusion that these five genes are associated with SARS infection regardless of strain.

4 Discussion

This paper used a GSEA-based meta-analysis approach incorporating 29 different gene signatures into identifying gene expression changes associated with SARS-CoV infection in human lung cell cultures and mouse lung samples and prioritize generated gene candidate lists to select five top gene candidates, CXCL10, OAS3, OASL, IFIT3, and XAF1, for future experimental examination to develop novel treatment options for SARS-CoV and potentially other β-coronavirus infections. Some

of our identified top gene candidates had previously reported associations to SARS infections. For example, CXCL10 is already well established with infections of several SARS strains, specifically icSARS (Mitchell et al., 2013) and Urbani (Yoshikawa et al., 2010) in human lung cell cultures and MA15 in mouse lung samples (Gralinski et al., 2018). The two OAS genes (OAS3 and OASL) both had reported associations with Urbani infections in human lung cell cultures (Yoshikawa et al., 2010). While IFIT3 was not specifically named in prior reports examining these datasets, IFIT1 had reported associations with MA15 infections in mouse lung samples (Totura et al., 2015; Gralinski et al., 2018), so it is not surprising to see IFIT3 included among identified top gene candidates. It was surprising to find XAF1, an apoptotic gene whose reduced or absent expression in tumor samples and cell lines leads to poorer survival in gastric adenocarcinomas (Xia et al., 2006; Shibata et al., 2008), among identified top gene candidates. This paper is the first study to report an association between XAF1 and SARS infections for any strain, to the best of our knowledge, though a recent report examining differential gene expression associated with SARS-CoV-2 infection identified increased expression of CXCL10, OAS3, IFIT3, and XAF1 in human epithelial lung cells and lung samples from *Cynomolgus maca* (cynomolgus monkey) and mice (Hachim et al., 2020). We speculate that applying our GSEA-based meta-analysis approach to examine gene expression datasets across β-coronavirus infections may produce some interesting and novel insights into molecular changes associated with all β-coronavirus infections which potentially could contribute to the identification of therapeutic targets.

4.1 Dataset Platform and Signature Considerations

Defining a gene signature with the most unique genes possible is critical in optimizing achievable results from GSEA (Subramanian et al., 2005). There are several factors that impact how a gene signature is defined and thus impact the overall outcome of the GSEA-based meta-analysis approach. Such factors include gene presence in a platform and gene annotation. For example, if a gene is not included in a dataset's platform (*e.g.*, XAF1 in GSE33266-derived signatures, Figure 6C), it cannot be analyzed. We also notice a substantial loss of platform genes (Table 1) due to incomplete or incompatible genome annotation (*i.e.*, probe has no gene symbol and/or Entrez ID). This can become more problematic when analyzing genes across different species since not all human genes have homologs in other species. While both issues can be avoided if all datasets selected for analysis are profiled on the same platform, these issues must be considered when comparing signatures across platforms and species as done here. We could not improve initial gene inclusion in a platform or gene annotation to recover lost genes, but we included icSARS panel genes that were not represented in each signature specifically alongside identified leading-edge genes while considering genes associated with infection with a specific SARS strain to ensure no genes were lost during our meta-analysis. As a result, not all signatures have expression data for our final gene candidates (Figures 4 and 6).

Signature skew can substantially affect obtainable results from GSEA. In strongly skewed signatures (Table 2), such as the icSARSvsmock validation signature derived from GSE33267 (T-score range: 12.2 to -12.2 with a midpoint around 1996 of 19751 genes), GSEA was unable to calculate enrichment scores. To overcome this problem but maintain gene ranking, we adjusted all T-scores in the strongly skewed signature by the T-score of the gene at that signature's actual midpoint. We did not expect this adjustment to affect the overall results of this study since there were similarities in leading-edge gene membership between adjusted and non-adjusted signatures. We recommend caution when interpreting leading-edge gene membership from skewed signatures as membership does not necessarily reflect the true expression relationship (*e.g.*, membership in a positive leading-edge may arise from a raw negative T-score in that signature).

This is a provisional file, not the final typeset article

Another important consideration is signature inclusion in our meta-analysis. We included MA15-infected mouse signatures over a range of inoculation doses (Figure 6) to mimic the range of infection severity that would be encountered clinically (i.e., asymptomatic to severe patient presentation). Including these signatures identified genes across inoculation doses with the best therapeutic potential regardless of infection severity. We also removed the GSE33266 MA15(10^2)vsmock signature from inclusion in our meta-analysis because its observed enrichment was not statistically different from NES achieved randomly (Figure 6). However, we noted that IFIT3 was the only top candidate in the leading-edge of the MA15(10^2)vsmock signature and it was not statistically significant individually (Welch's T-test p-value=0.160). These findings suggest there is a lower inoculation dose limit to our approach's ability to detect relevant genes, which should be considered when applying our meta-analysis to future research.

4.2 icSARS Panel Generation and Timepoint Selection Considerations

There are several factors to consider with respect to how this study was conducted, particularly icSARS panel generation (*i.e.*, query gene set size and signature selection) and time point selection. To determine how query gene set sizes might impact our overall results, we ran GSEA between GSE47961-derived icSARSvsmock (reference) and GSE47960-derived icSARSvsmock gene sets ranging in size (queries) and found significant enrichment (p-value<0.05) with gene set sizes down to 100 genes (Supplemental Material SFig 1), again showing the high degree of similarity between GSE47960-derived icSARSvsmock and GSE47961-derived icSARSvsmock gene signatures. However, we noted that four of the five gene candidates identified here were only included in query gene sets sized 400 genes or above and CXCL10 was only included in the 500 gene query set size used in this study. Since 500 is the largest recommended query set size for accurate GSEA calculations (Subramanian et al., 2005), this discussion supports our decision that 500 was an optimal query gene set size to use in this study.

Further, we noticed that enrichment did not substantially change if positive and negative tail query gene sets were generated from GSE47961-derived icSARSvsmock and compared to GSE47960-derived icSARSvsmock as reference (Supplemental Material SFig 1). This highlights the similarity between these signatures. When examining identified positive leading-edge genes, we noted that the size of the positive icSARS panel did not substantially change from 233 genes (Table 1) to 240 genes (Supplemental Material STable 14) and that this subtle change did not affect observed enrichment across 27 signatures (Supplemental Material SFig 2) or the inclusion of our five gene candidates across leading-edges. We also noted that the size of the negative icSARS panel did change significantly (114 to 331 genes), but enrichment for the 331 gene icSARS panel was inconsistent across signatures which was similar to what we originally observed with the 113 gene icSARS panel. From this we conclude that reversing which signature is used as reference and to derive query gene sets does not change the overall result (*i.e.*, identification of five top gene candidates) of our study, and assume the same overall result would be observed if we derived query gene sets from any other signature used in this study due to use of a meta-analysis approach.

Since gene expression changes throughout a time course (Sims et al., 2013), we repeated the process of defining icSARS gene panels over a range of time points (24, 72, and 96hrs) and found that enrichment was consistently significant across the 72 and 96hr time points for all query gene set sizes regardless of signature was selected for deriving query gene sets, as previously observed for the 48hr time point (Supplemental Material SFig 1). However, significant enrichment is inconsistent for the 24hr timepoint, suggesting gene expression changes generated by an icSARS infection become more predictable later in the growth curve. Based on these data, we used the 48hr time point since

identified leading-edge genes represent over- and under-expressed genes associated with icSARS infection at the earliest time point with consistent detectable gene signature similarities. We further noted all five gene candidates identified here were included in leading-edges identified by the 500 gene query set size at the 72hr and 96hr time points, suggesting that these gene candidates likely would have been identified if later time points were selected for use, thus not altering our overall result.

4.3 Considerations on Pathway Enrichment and Potential Therapeutic Options

To expand on our analysis, we examined the cellular roles that genes in the icSARS panels are involved in. We used DAVID to calculate enrichment between the icSARS panels and pathways in commonly used knowledgebases. We noted that Gene Ontology (GO) Biological Processes (BP) database returned the most identified significantly enriched pathways compared to other databases (data not shown), so we focused this discussion on GO-BP data to avoid confusion from overlapping pathway and gene inclusion variations across different knowledgebases. DAVID identified 49 significant GO-BP pathways (EASE score p-value<0.05) from the positive icSARS panel and 10 significant pathways from the negative icSARS panel (Supplemental Material STable 15). Most significantly enriched pathways have experimentally established associations with icSARS and other β-coronavirus infections, such as up-regulation of NF-kappaB processes (transcription factor activity: GO:0051092, p-value=0.001; signaling: GO:0043123, p-value=0.043), p38MAPK cascade (GO:1900745, p-value=0.015) and apoptotic process (GO:0006915, p-value=0.018) and down-regulation of cilium morphogenesis (GO: 0060271, p-value=0.038) and assembly (GO: 0042384, p-value=0.0.030), demonstrating our gene signature approach's ability to detect pathways associated with response to an icSARS infections (Mizutani et al., 2004; Fung and Liu, 2019; Wehbe et al., 2020). We also find several pathways, such as SMAD protein signal transduction (GO:0060395, p-value=0.037), with no prior associations to icSARS infection though they have been identified in mouse infections with the dORF strain (Sims et al., 2013). Therefore, we speculate that pathways without prior association to icSARS infections identified here are also involved in response to an icSARS infection.

Finally, we consider how our findings can be translated into possible therapeutic options using the Comparative Toxicogenomics Database (Mattingly et al., 2003; Davis et al., 2019) to identify chemicals that interact with each gene candidate identified here. We then compared across chemical lists from each gene candidates and found eight chemicals that interact with all gene candidates: benzo(a)pyrene, bisphenol A, cyclosporine, estradiol, (+)-JQ1 compound, nickel, tobacco Smoke Pollution, and tributyltin. Of these, (+)-JQ1 compound and tributyltin were reported to decrease expression of all gene candidates which is an important consideration since our results show the gene candidates are usually over-expressed during SARS infection. Tributyltin is toxic to mammals (Kotake, 2012), so it cannot be considered further. JQ1 is a thienotriazolodiazepine that strongly inhibits the BET family of bromodomain proteins including BRD2, BRD3, and BRD4 (Shi et al., 2018; Lambert et al., 2019). We noted that BRD1 is among our positive icSARS panel genes (Table 1) though none of the BRD genes showed consistent changes in expression across signatures (Supplemental Material SFig 3). Several studies have examined the use of BET inhibitors structurally similar to JQ1 to treat a variety of cancers (Da Costa et al., 2013; Lochrin et al., 2014; McCleland et al., 2016; Jiang et al., 2020), and are reported to promote viral re-activation (Banerjee et al., 2012; Chen et al., 2017). We also considered how drugs that are frequently discussed for treating COVID-19 might interact with our gene candidates (Group et al., 2020; Pooladanda et al., 2020; White et al., 2020). We noted that acetaminophen reportedly decreased expression of XAF1 and OAS3 in human liver cells (Elferink et al., 2011; Jiang et al., 2015) but had mixed reports in how it affected CXCL10

This is a provisional file, not the final typeset article

and IFIT3 expression. Hydroquinone was reported to decrease XAF1 and IFIT3 expression but increase OASL expression in human epidermal melanocytes (Pyo et al., 2019). Interestingly, hydroxychloroquine was reported to increase expression of CXCL10 and IFIT3 in human lung carcinoma cells and mouse macrophages (Wang et al., 2015), the opposite of the desired effect proposed from our computational findings. This finding is surprising since other compound affect expression of some of our gene candidates in a manner consistent with our computational findings, and the efficacy of hydroxychloroquine in treating COVID-19 is debated (Agarwal et al., 2020; Badgujar et al., 2020; Meo et al., 2020; Rosenke et al., 2020). Further work is needed to assess the viability of these therapeutic options and interpret findings presented here for the advancement therapeutic options for SARS and other β-coronavirus infections.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

L.H. lead manuscript preparation. A.P. generated the supplemental materials and time course data. All authors contributed to the interpretation of results and manuscript review.

7 Funding

No funding support was provided for this work.

8 Acknowledgments

The authors would like to thank Jared Geller for the pathway analysis and Terri Pulice for the graphic design assistance.

References

Addinsoft (2019). XLSTAT statistical and data analysis solution.

Agarwal, M., Ranjan, P., Mittal, A., and Baitha, U. (2020). Use of hydroxychloroquine for pre-exposure prophylaxis in COVID 19: debate and suggested future course. *Expert Rev Anti Infect Ther***,** 1-5. doi: 10.1080/14787210.2021.1828062.

Badgujar, K.C., Badgujar, A.B., Patil, V.P., and Dhangar, D.V. (2020). Hydroxychloroquine for COVID-19: A review and a debate based on available clinical trials/case studies. *Journal of Drug Delivery and Therapeutics* 10(3)**,** 304-311.

Bahadur, S., Long, W., and Shuaib, M. (2020). Human coronaviruses with emphasis on the COVID-19 outbreak. *Virusdisease***,** 1-5. doi: 10.1007/s13337-020-00594-y.

Banerjee, C., Archin, N., Michaels, D., Belkina, A.C., Denis, G.V., Bradner, J., et al. (2012). BET bromodomain inhibition as a novel strategy for reactivation of HIV-1. *J Leukoc Biol* 92(6)**,** 1147-1154. doi: 10.1189/jlb.0312165.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., et al. (2011). NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* 39(Database issue)**,** D1005-1010. doi: 10.1093/nar/gkq1184.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41(Database issue)**,** D991-995. doi: 10.1093/nar/gks1193.

Becker, M.M., Graham, R.L., Donaldson, E.F., Rockx, B., Sims, A.C., Sheahan, T., et al. (2008). Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A* 105(50)**,** 19944-19949. doi: 10.1073/pnas.0808116105.

Bharat Siva Varma, P., Adimulam, Y.B., and Kodukula, S. (2015). In silico functional annotation of a hypothetical protein from Staphylococcus aureus. *J Infect Public Health* 8(6)**,** 526-532. doi: 10.1016/j.jiph.2015.03.007.

Boulos, M.N. (2004). Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. *Int J Health Geogr* 3(1)**,** 2. doi: 10.1186/1476-072X-3-2.

Chen, H.S., De Leo, A., Wang, Z., Kerekovic, A., Hills, R., and Lieberman, P.M. (2017). BET-Inhibitors Disrupt Rad21-Dependent Conformational Control of KSHV Latency. *PLoS Pathog* 13(1)**,** e1006100. doi: 10.1371/journal.ppat.1006100.

Da Costa, D., Agathanggelou, A., Perry, T., Weston, V., Petermann, E., Zlatanou, A., et al. (2013). BET inhibition as a single or combined therapeutic approach in primary paediatric B-precursor acute lymphoblastic leukaemia. *Blood Cancer J* 3**,** e126. doi: 10.1038/bcj.2013.24.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., McMorran, R., Wiegers, J., et al. (2019). The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 47(D1)**,** D948-D954. doi: 10.1093/nar/gky868.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1)**,** 207-210.

Elferink, M.G., Olinga, P., van Leeuwen, E.M., Bauerschmidt, S., Polman, J., Schoonen, W.G., et al. (2011). Gene expression analysis of precision-cut human liver slices indicates stable expression of ADME-Tox related genes. *Toxicol Appl Pharmacol* 253(1)**,** 57-69. doi: 10.1016/j.taap.2011.03.010.

Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* 73**,** 529-557. doi: 10.1146/annurev-micro-020518-115759.

Gardinassi, L.G., Souza, C.O.S., Sales-Campos, H., and Fonseca, S.G. (2020). Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. *Front Immunol* 11**,** 1636. doi: 10.3389/fimmu.2020.01636.

Goad, B., and Harris, L.K. (2018). Identification and prioritization of macrolide resistance genes with hypothetical annotation in Streptococcus pneumoniae. *Bioinformation* 14(9)**,** 488-498.

Gralinski, L.E., Bankhead, A., 3rd, Jeng, S., Menachery, V.D., Proll, S., Belisle, S.E., et al. (2013). Mechanisms of severe acute respiratory syndrome coronavirus-induced acute lung injury. *mBio* 4(4). doi: 10.1128/mBio.00271-13.

Gralinski, L.E., Sheahan, T.P., Morrison, T.E., Menachery, V.D., Jensen, K., Leist, S.R., et al. (2018). Complement Activation Contributes to Severe Acute Respiratory Syndrome Coronavirus Pathogenesis. *mBio* 9(5). doi: 10.1128/mBio.01753-18.

Group, R.C., Horby, P., Mafham, M., Linsell, L., Bell, J.L., Staplin, N., et al. (2020). Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19. *N Engl J Med*. doi: 10.1056/NEJMoa2022926.

This is a provisional file, not the final typeset article

Hachim, M.Y., Al Heialy, S., Hachim, I.Y., Halwani, R., Senok, A.C., Maghazachi, A.A., et al. (2020). Interferon-Induced Transmembrane Protein (IFITM3) Is Upregulated Explicitly in SARS-CoV-2 Infected Lung Epithelial Cells. *Front Immunol* 11**,** 1372. doi: 10.3389/fimmu.2020.01372.

Ijaq, J., Chandrasekharan, M., Poddar, R., Bethi, N., and Sundararajan, V.S. (2015). Annotation and curation of uncharacterized proteins- challenges. *Front Genet* 6**,** 119. doi: 10.3389/fgene.2015.00119.

Islam, M.S., Shahik, S.M., Sohel, M., Patwary, N.I., and Hasan, M.A. (2015). In Silico Structural and Functional Annotation of Hypothetical Proteins of Vibrio cholerae O139. *Genomics Inform* 13(2)**,** 53-59. doi: 10.5808/GI.2015.13.2.53.

Jiang, G., Deng, W., Liu, Y., and Wang, C. (2020). General mechanism of JQ1 in inhibiting various types of cancer. *Mol Med Rep* 21(3)**,** 1021-1034. doi: 10.3892/mmr.2020.10927.

Jiang, J., Briede, J.J., Jennen, D.G., Van Summeren, A., Saritas-Brauers, K., Schaart, G., et al. (2015). Increased mitochondrial ROS formation by acetaminophen in human hepatic cells is associated with gene expression changes suggesting disruption of the mitochondrial electron transport chain. *Toxicol Lett* 234(2)**,** 139-150. doi: 10.1016/j.toxlet.2015.02.012.

Kolker, E., Makarova, K.S., Shabalina, S., Picone, A.F., Purvine, S., Holzman, T., et al. (2004). Identification and functional analysis of 'hypothetical' genes expressed in Haemophilus influenzae. *Nucleic Acids Res* 32(8)**,** 2353-2361. doi: 10.1093/nar/gkh555.

Kotake, Y. (2012). Molecular mechanisms of environmental organotin toxicity in mammals. *Biol Pharm Bull* 35(11)**,** 1876-1880. doi: 10.1248/bpb.b212017.

Lai, S.T. (2005). Treatment of severe acute respiratory syndrome. *Eur J Clin Microbiol Infect Dis* 24(9)**,** 583-591. doi: 10.1007/s10096-005-0004-z.

Lambert, J.P., Picaud, S., Fujisawa, T., Hou, H., Savitsky, P., Uuskula-Reimand, L., et al. (2019). Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. *Mol Cell* 73(3)**,** 621-638 e617. doi: 10.1016/j.molcel.2018.11.006.

Li, X., Luk, H.K.H., Lau, S.K.P., and Woo, P.C.Y. (2019). Human Coronaviruses: General Features. *Reference Module in Biomedical Sciences***,** B978-970-912-801238-801233.895704-801230. doi: 10.1016/B978-0-12-801238-3.95704-0.

Lochrin, S.E., Price, D.K., and Figg, W.D. (2014). BET bromodomain inhibitors--a novel epigenetic approach in castration-resistant prostate cancer. *Cancer Biol Ther* 15(12)**,** 1583-1585. doi: 10.4161/15384047.2014.962297.

Mattingly, C.J., Colby, G.T., Forrest, J.N., and Boyer, J.L. (2003). The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect* 111(6)**,** 793-795. doi: 10.1289/ehp.6028.

McCleland, M.L., Mesh, K., Lorenzana, E., Chopra, V.S., Segal, E., Watanabe, C., et al. (2016). CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest* 126(2)**,** 639-652. doi: 10.1172/JCI83265.

Menachery, V.D., Gralinski, L.E., Mitchell, H.D., Dinnon, K.H., 3rd, Leist, S.R., Yount, B.L., Jr., et al. (2018). Combination Attenuation Offers Strategy for Live Attenuated Coronavirus Vaccines. *J Virol* 92(17). doi: 10.1128/JVI.00710-18.

Meo, S.A., Klonoff, D.C., and Akram, J. (2020). Efficacy of chloroquine and hydroxychloroquine in the treatment of COVID-19. *Eur Rev Med Pharmacol Sci* 24(8)**,** 4539-4547. doi: 10.26355/eurrev_202004_21038.

Mitchell, H.D., Eisfeld, A.J., Sims, A.C., McDermott, J.E., Matzke, M.M., Webb-Robertson, B.J., et al. (2013). A network integration approach to predict conserved regulators related to pathogenicity of influenza and SARS-CoV respiratory viruses. *PLoS One* 8(7)**,** e69374. doi: 10.1371/journal.pone.0069374.

Mizutani, T., Fukushi, S., Saijo, M., Kurane, I., and Morikawa, S. (2004). Phosphorylation of p38 MAPK and its downstream targets in SARS coronavirus-infected cells. *Biochem Biophys Res Commun* 319(4)**,** 1228-1234. doi: 10.1016/j.bbrc.2004.05.107.

Modjarrad, K. (2016). Treatment strategies for Middle East respiratory syndrome coronavirus. *J Virus Erad* 2(1)**,** 1-4.

Mohan, R., and Venugopal, S. (2012). Computational structural and functional analysis of hypothetical proteins of Staphylococcus aureus. *Bioinformation* 8(15)**,** 722-728. doi: 10.6026/97320630008722.

Omeershffudin, U.N.M., and Kumar, S. (2019). In silico approach for mining of potential drug targets from hypothetical proteins of bacterial proteome. *Int J Mol Biol Open Access* 4(4)**,** 145-152.

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., and Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect* 26(6)**,** 729-734. doi: 10.1016/j.cmi.2020.03.026.

Pooladanda, V., Thatikonda, S., and Godugu, C. (2020). The current understanding and potential therapeutic options to combat COVID-19. *Life Sci* 254**,** 117765. doi: 10.1016/j.lfs.2020.117765.

Pyo, J.J., Ahn, S., Jin, S.H., An, S., Lee, E., Choi, J., et al. (2019). Keratinocyte-derived IL-36gamma plays a role in hydroquinone-induced chemical leukoderma through inhibition of melanogenesis in human epidermal melanocytes. *Arch Toxicol* 93(8)**,** 2307-2320. doi: 10.1007/s00204-019-02506-6.

Rosenke, K., Jarvis, M.A., Feldmann, F., Schwarz, B., Okumura, A., Lovaglio, J., et al. (2020). Hydroxychloroquine Proves Ineffective in Hamsters and Macaques Infected with SARS-CoV-2. *bioRxiv*. doi: 10.1101/2020.06.10.145144.

School, K., Marklevitz, J., W, K.S., and L, K.H. (2016). Predictive characterization of hypothetical proteins in Staphylococcus aureus NCTC 8325. *Bioinformation* 12(3)**,** 209-220. doi: 10.6026/97320630012209.

Shi, X., Liu, C., Liu, B., Chen, J., Wu, X., and Gong, W. (2018). JQ1: a novel potential therapeutic target. *Pharmazie* 73(9)**,** 491-493. doi: 10.1691/ph.2018.8480.

Shibata, T., Noguchi, T., Takeno, S., Gabbert, H.E., Ramp, U., and Kawahara, K. (2008). Disturbed XIAP and XAF1 expression balance is an independent prognostic factor in gastric adenocarcinomas. *Ann Surg Oncol* 15(12)**,** 3579-3587. doi: 10.1245/s10434-008-0062-4.

Sims, A.C., Tilton, S.C., Menachery, V.D., Gralinski, L.E., Schafer, A., Matzke, M.M., et al. (2013). Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. *J Virol* 87(7)**,** 3885-3902. doi: 10.1128/JVI.02520-12.

Sivashankari, S., and Shanmughavel, P. (2006). Functional annotation of hypothetical proteins - A review. *Bioinformation* 1(8)**,** 335-338. doi: 10.6026/97320630001335.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide

This is a provisional file, not the final typeset article

expression profiles. *Proc Natl Acad Sci U S A* 102(43)**,** 15545-15550. doi: 10.1073/pnas.0506580102.

Totura, A.L., Whitmore, A., Agnihothram, S., Schafer, A., Katze, M.G., Heise, M.T., et al. (2015). Toll-Like Receptor 3 Signaling via TRIF Contributes to a Protective Innate Immune Response to Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio* 6(3)**,** e00638-00615. doi: 10.1128/mBio.00638-15.

Wang, L.F., Lin, Y.S., Huang, N.C., Yu, C.Y., Tsai, W.L., Chen, J.J., et al. (2015). Hydroxychloroquine-inhibited dengue virus is associated with host defense machinery. *J Interferon Cytokine Res* 35(3)**,** 143-156. doi: 10.1089/jir.2014.0038.

Wehbe, Z., Hammoud, S., Soudani, N., Zaraket, H., El-Yazbi, A., and Eid, A.H. (2020). Molecular Insights Into SARS COV-2 Interaction With Cardiovascular Disease: Role of RAAS and MAPK Signaling. *Frontiers in Pharmacology* 11(836). doi: 10.3389/fphar.2020.00836.

White, N.J., Watson, J.A., Hoglund, R.M., Chan, X.H.S., Cheah, P.Y., and Tarning, J. (2020). COVID-19 prevention and treatment: A critical analysis of chloroquine and hydroxychloroquine clinical pharmacology. *PLoS Med* 17(9)**,** e1003252. doi: 10.1371/journal.pmed.1003252.

Xia, Y., Novak, R., Lewis, J., Duckett, C.S., and Phillips, A.C. (2006). Xaf1 can cooperate with TNFalpha in the induction of apoptosis, independently of interaction with XIAP. *Mol Cell Biochem* 286(1-2)**,** 67-76. doi: 10.1007/s11010-005-9094-2.

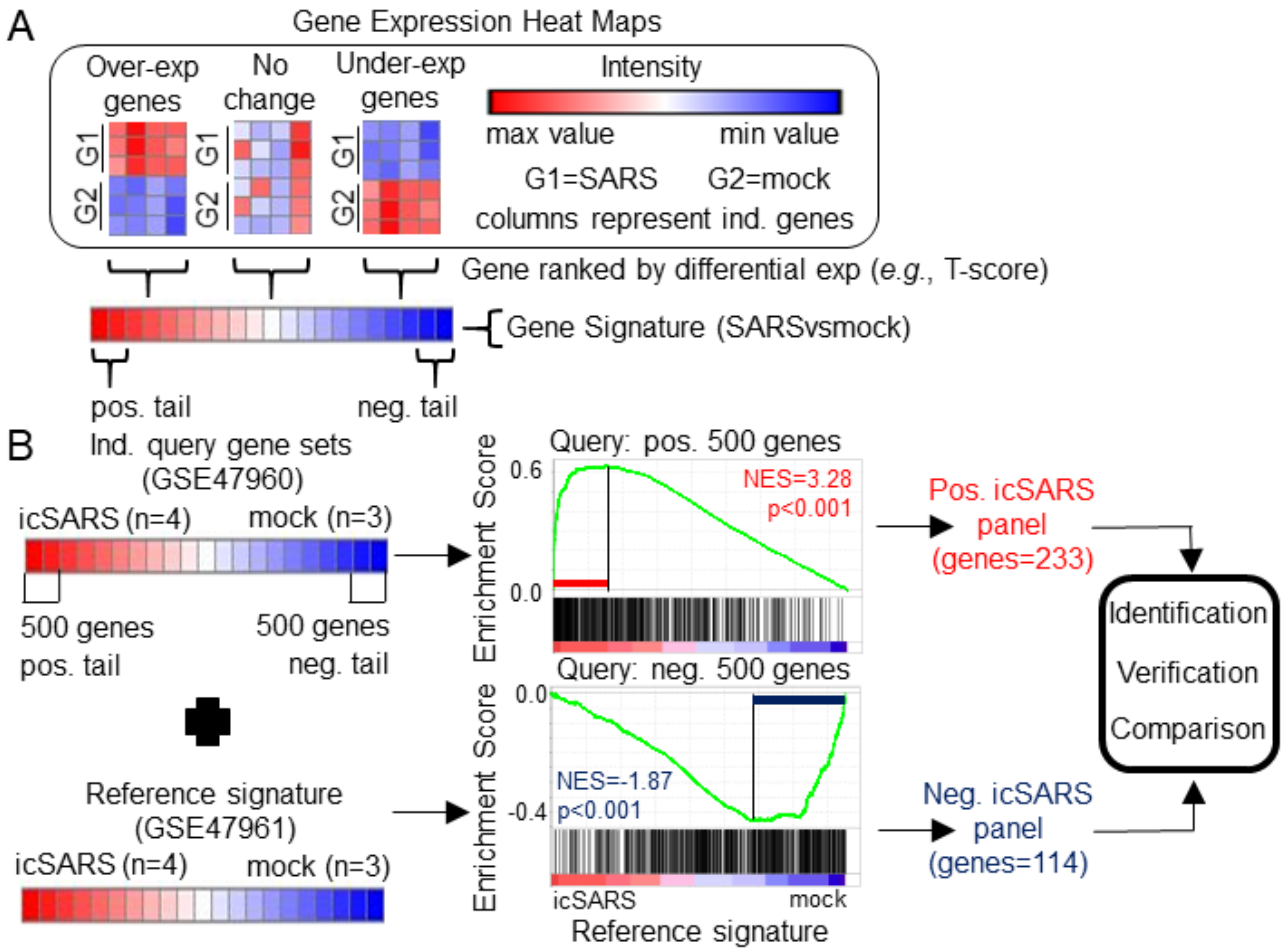
XLSTAT, A. (2013). Data analysis and statistics software for Microsoft Excel. *New York, NY: Addinsoft.*

Yi, Y., Lagniton, P.N.P., Ye, S., Li, E., and Xu, R.H. (2020). COVID-19: what has been learned and to be learned about the novel coronavirus disease. *Int J Biol Sci* 16(10)**,** 1753-1766. doi: 10.7150/ijbs.45134.

Yoshikawa, T., Hill, T.E., Yoshikawa, N., Popov, V.L., Galindo, C.L., Garner, H.R., et al. (2010). Dynamic innate immune responses of human bronchial epithelial cells to severe acute respiratory syndrome-associated coronavirus infection. *PLoS One* 5(1)**,** e8729. doi: 10.1371/journal.pone.0008729.

Yount, B., Roberts, R.S., Sims, A.C., Deming, D., Frieman, M.B., Sparks, J., et al. (2005). Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. *J Virol* 79(23)**,** 14909-14922. doi: 10.1128/JVI.79.23.14909-14922.2005.

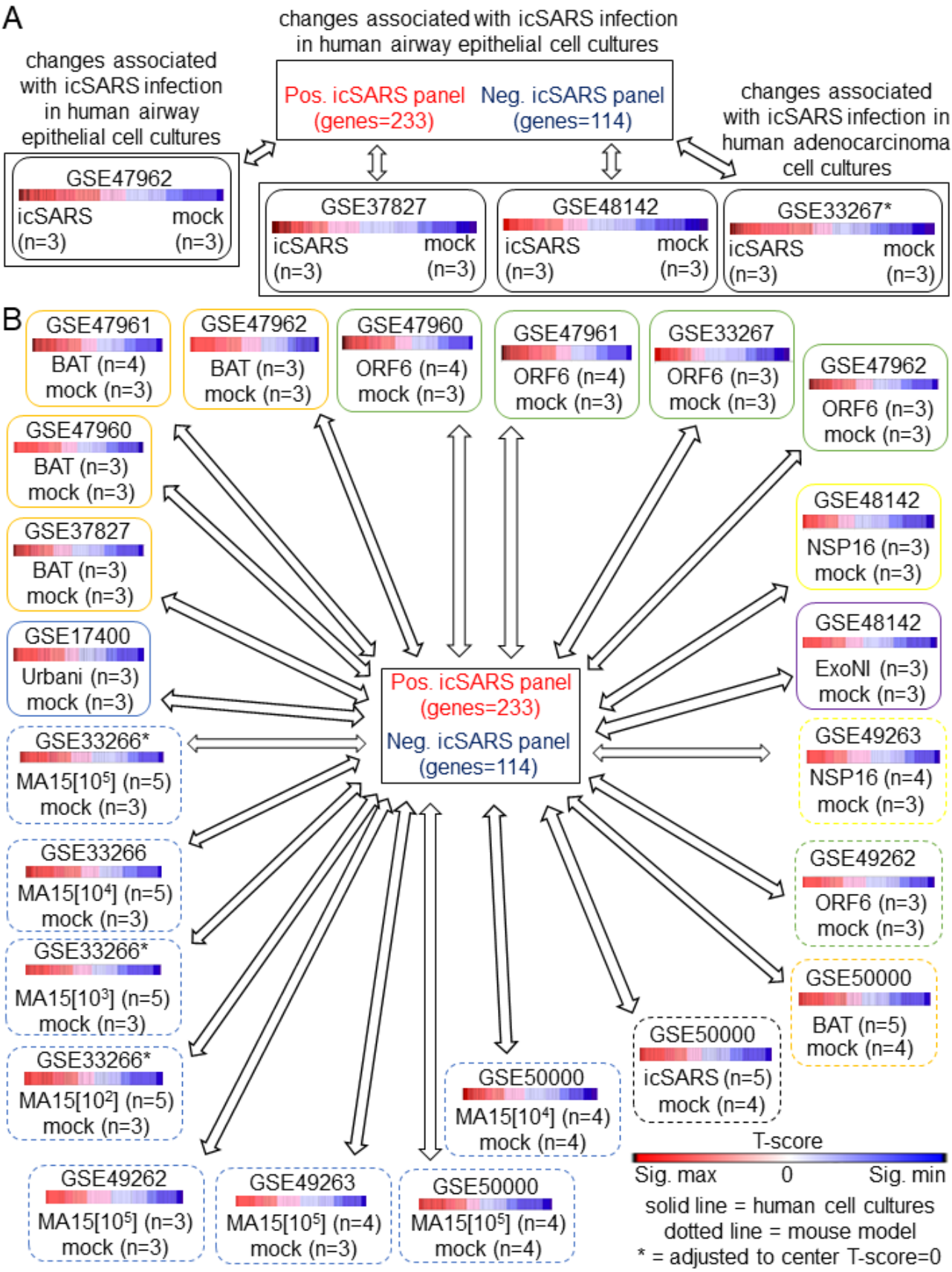
Figure 1. Gene Signature Definition and Generation icSARS Gene Panels



A) Schematic definition of a gene signature. Differences in gene expression between two groups, such as SARS- and mock-infected lung cells, are measured by Welch's two-sample T□test score. Gene signatures are ranked lists of genes from high (red) to low (blue) differential mRNA expression between groups. B) Generation of icSARS gene panels for use in this study. To identify differentially expressed genes associated with icSARS infection in human airway epithelial cell cultures, query gene sets containing either the 500 most over- or under-expressed genes from positive or negative tails of the gene signature generated from the Gene Expression Omnibus (GEO) accession number GSE47960 mRNA expression dataset. The positive and negative tail query sets were compared individually to the gene signature generated from the GEO GSE47961 dataset, which was used as reference for Gene Set Enrichment Analysis (GSEA). From this GSEA computed two enrichment plots, one for each query set, and their associated normalized enrichment score (NES) and p-value which represent the extent of enrichment between query set and reference signature. GSEA also identified leading-edge genes, which are genes that contribute most to achieving maximum enrichment. Two gene panels were defined from leading-edge genes identified in each query set. These gene panels were used in this study for three purposes: 1) identification of gene expression changes associated with icSARS infection in human airway epithelial cell cultures, 2) verification of identified findings in independent datasets, and 3) comparison to other gene signatures representing changes in gene expression associated with other SARS infections.

Figure 2. Schematic of icSARS Panel Verification and Comparison Across SARS Infections

This is a provisional file, not the final typeset article

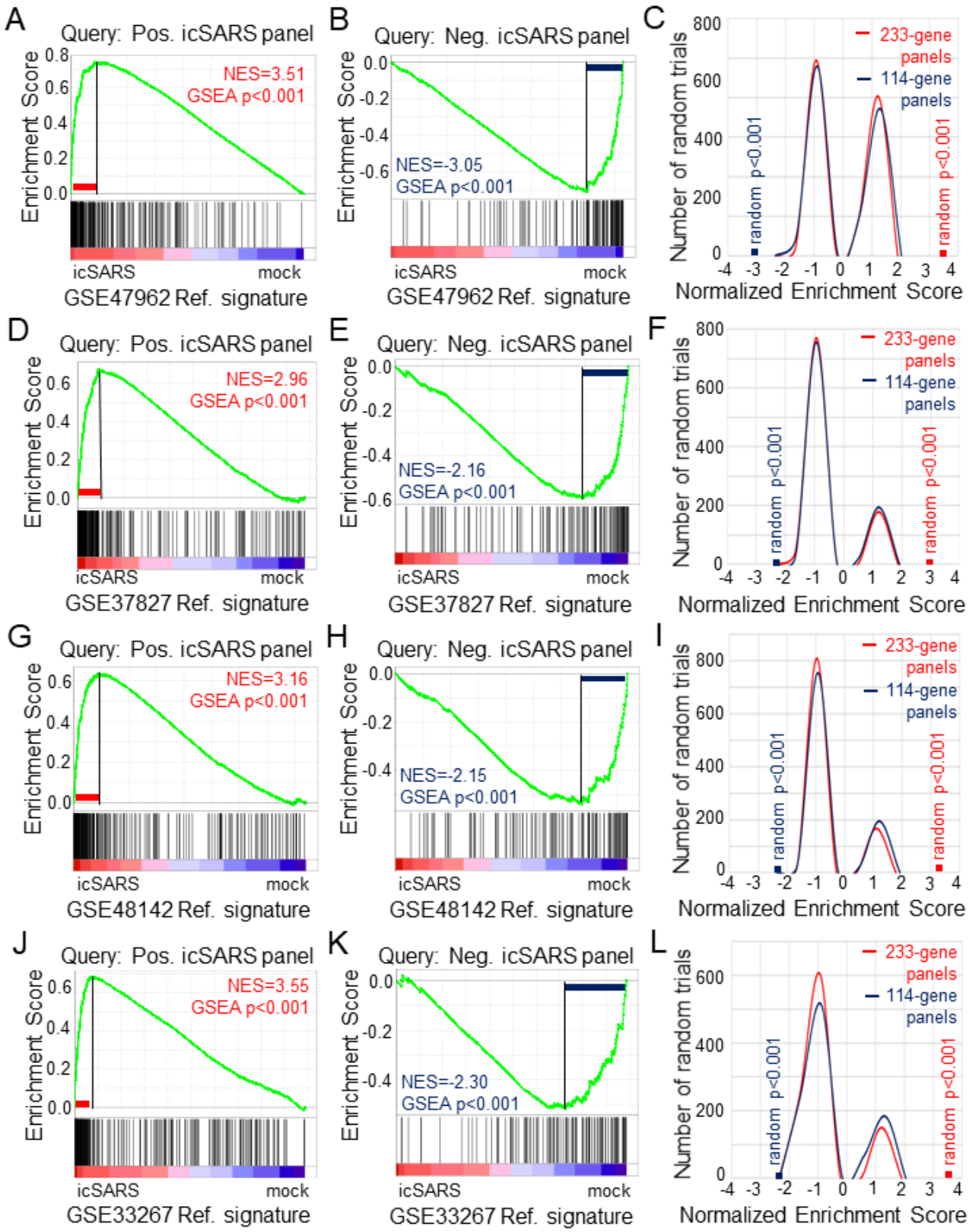

A
changes associated with icSARS infection in human airway epithelial cell cultures
changes associated with icSARS infection in human airway epithelial cell cultures
changes associated with icSARS infection in human adenocarcinoma cell cultures
Pos. icSARS panel (genes=233)
Neg. icSARS panel (genes=114)
GSE47962
icSARS (n=3)
mock (n=3)
GSE37827
icSARS (n=3)
mock (n=3)
GSE48142
icSARS (n=3)
mock (n=3)
GSE33267*
icSARS (n=3)
mock (n=3)
B
GSE47961
BAT (n=4)
mock (n=3)
GSE47962
BAT (n=3)
mock (n=3)
GSE47960
ORF6 (n=4)
mock (n=3)
GSE47961
ORF6 (n=4)
mock (n=3)
GSE33267
ORF6 (n=3)
mock (n=3)
GSE47962
ORF6 (n=3)
mock (n=3)
GSE47960
BAT (n=3)
mock (n=3)
GSE37827
BAT (n=3)
mock (n=3)
GSE17400
Urbani (n=3)
mock (n=3)
GSE48142
NSP16 (n=3)
mock (n=3)
GSE48142
ExoNI (n=3)
mock (n=3)
Pos. icSARS panel (genes=233)
Neg. icSARS panel (genes=114)
GSE33266*
MA15[10^5] (n=5)
mock (n=3)
GSE49263
NSP16 (n=4)
mock (n=3)
GSE33266
MA15[10^4] (n=5)
mock (n=3)
GSE49262
ORF6 (n=3)
mock (n=3)
GSE33266*
MA15[10^3] (n=5)
mock (n=3)
GSE50000
BAT (n=5)
mock (n=4)
GSE33266*
MA15[10^2] (n=5)
mock (n=3)
GSE50000
icSARS (n=5)
mock (n=4)
GSE50000
MA15[10^4] (n=4)
mock (n=4)
T-score
Sig. max
0
Sig. min
GSE49262
MA15[10^5] (n=3)
mock (n=3)
GSE49263
MA15[10^5] (n=4)
mock (n=3)
GSE50000
MA15[10^5] (n=4)
mock (n=4)
solid line = human cell cultures
dotted line = mouse model
* = adjusted to center T-score=0

754

A) Gene signatures used for verification of icSARS panels. Four verification signatures were defined from datasets containing cultures of mock- and icSARS-infected human airway epithelial cells (GSE47962) or lung adenocarcinoma cells (GSE36727, GSE48142, and GSE33267). (B) Gene signatures used for comparison to icSARS panels. Twenty-three comparison signatures were defined from SARS strains varying in virulence (icSARS, Urbani, MA15, BAT-SRBD, ORF6, NSP16, and ExoNI) with 11 comparison signatures derived from human lung adenocarcinoma cultures (solid ovals) and 12 comparison signatures derived from mouse lung samples (dashed ovals).

Figure 3. Verification of icSARS Gene Panels in Independent Datasets

This is a provisional file, not the final typeset article

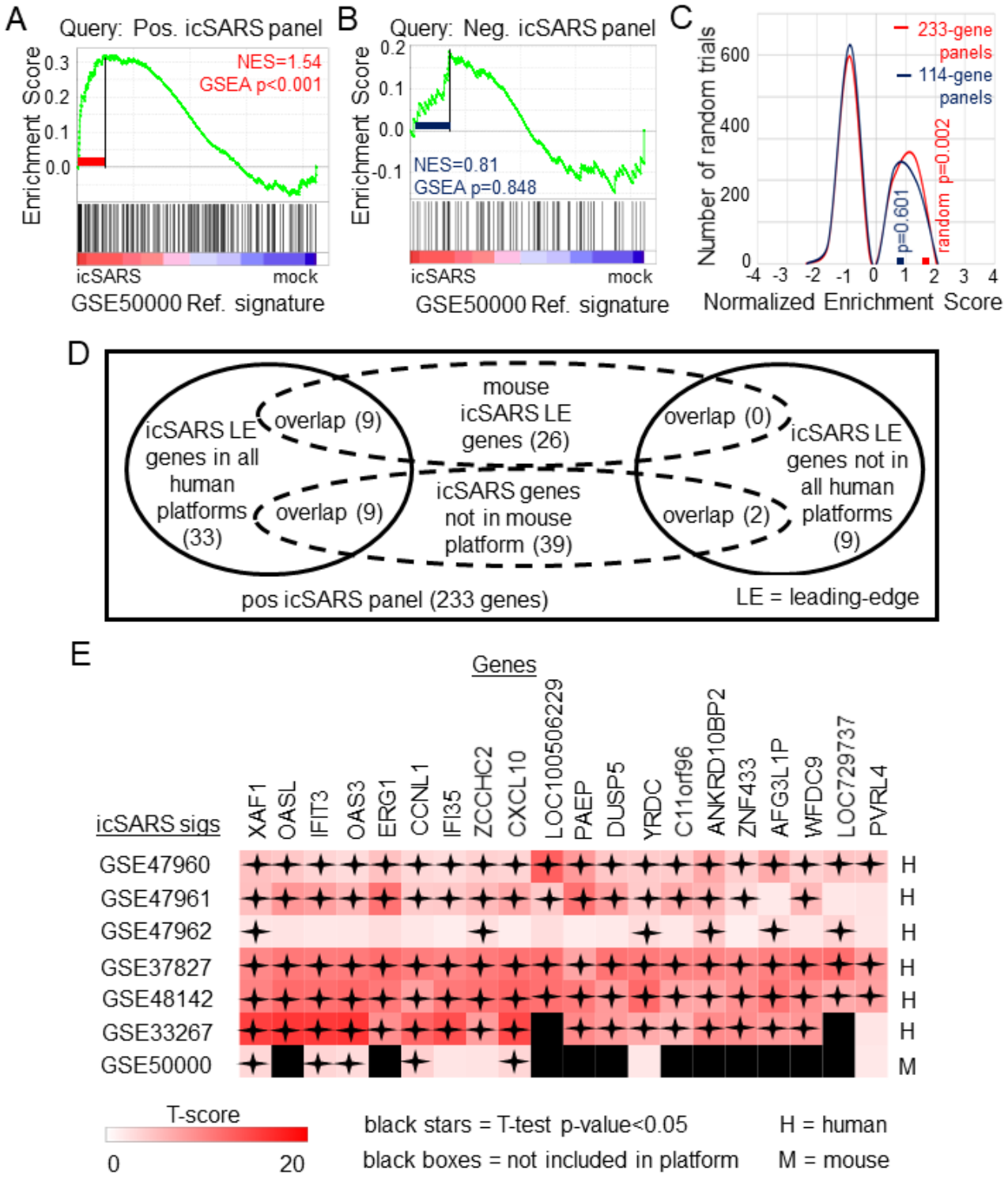


A) Gene Set Enrichment Analysis (GSEA) detected significant enrichment, as determined by normalized enrichment score (NES), between the positive icSARS gene panel and the positive tail of the GSE47962-derived gene signature. B) GSEA detected significant enrichment between the

negative icSARS panel and the negative tail of the GSE47962-derived signature. C) Distribution plot of NES from 1000 randomly generated gene panels (individual queries) compared to the GSE47962-derived signature (reference) shows that NES achieved in (A) and (B) are non-random. D) GSEA found significant enrichment between the positive icSARS panel and the positive tail of the GSE37827-derived gene signature. E) GSEA detected significant enrichment between the negative icSARS panel and the negative tail of the GSE37827-derived signature. F) Distribution plot of 1000 randomly generated NES from the GSE37827-derived signature (reference) illustrated that NES achieved in (D) and (E) are non-random. G) GSEA detected significant enrichment between the positive icSARS gene panel and the positive tail of the GSE48142-derived gene signature. H) GSEA detected significant enrichment between the negative icSARS panel and the negative tail of the GSE48142-derived signature. I) Distribution plot of 1000 randomly generated NES from the GSE48142-derived signature (reference) illustrated that NES achieved in (G) and (H) are non-random. J) GSEA detected significant enrichment between the positive icSARS panel and the positive tail of the GSE33267-derived gene signature. K) GSEA found significant enrichment between the negative icSARS panel and the negative tail of the GSE33267-derived signature. L) Distribution plot of 1000 randomly generated NES from the GSE33267-derived signature (reference) shows NES achieved in (J) and (K) are non-random.

Figure 4. Positive icSARS Panel Enrichment in icSARS Infected Mouse Model Revealed Genes Associated with icSARS Infection

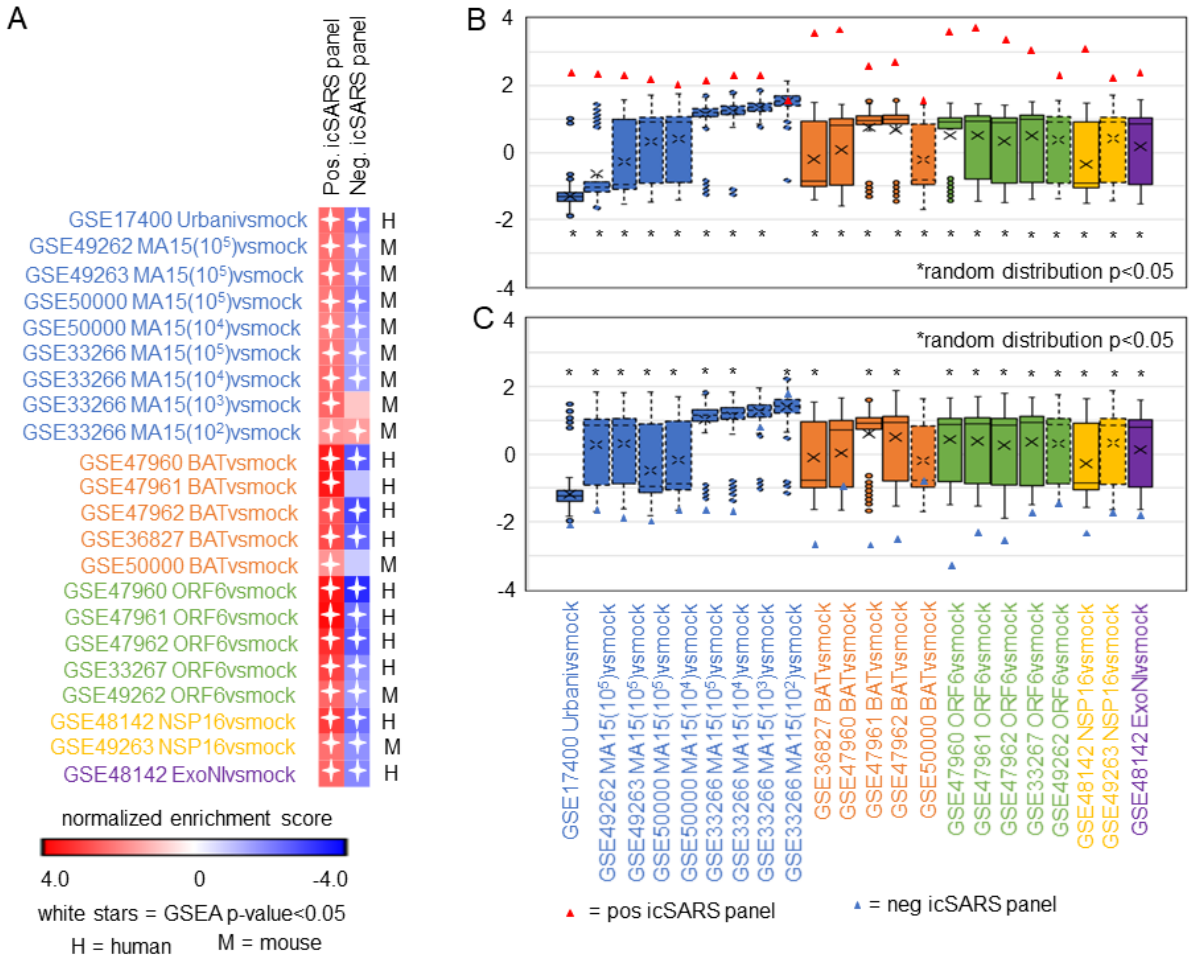
This is a provisional file, not the final typeset article



A) Gene Set Enrichment Analysis (GSEA) detected significant enrichment, as determined by normalized enrichment score (NES), between the positive icSARS gene panel and the positive tail of the GSE50000-derived gene signature. B) GSEA did not detect enrichment between the negative icSARS panel and the negative tail of the GSE50000-derived signature. C) Distribution plot of 1000 randomly generated NES from gene panels (individual queries) compared to the GSE50000-derived signature (reference) illustrated that the NES achieved by the positive icSARS panel in (A) is non-random while NES achieved by the negative icSARS panel in (B) can be randomly generated.

D) Venn diagram illustrated the overlap of positive icSARS panel genes between human genes with verified association to icSARS infection in human cultures (found across all positive icSARS leading edges in Figure 2 unless not included in platform) and mouse genes from either the GSE50000 gene signature's leading-edge in (A) or genes not included in the platform used to profile GSE50000. E) Heat map of T-scores for overlapping genes from (D) showed extent of over-expression across icSARS gene signatures.

Figure 5. icSARS Panel Enrichment Detect Differential Gene Expression Similarities Across SARS Strains with Varying Virulence

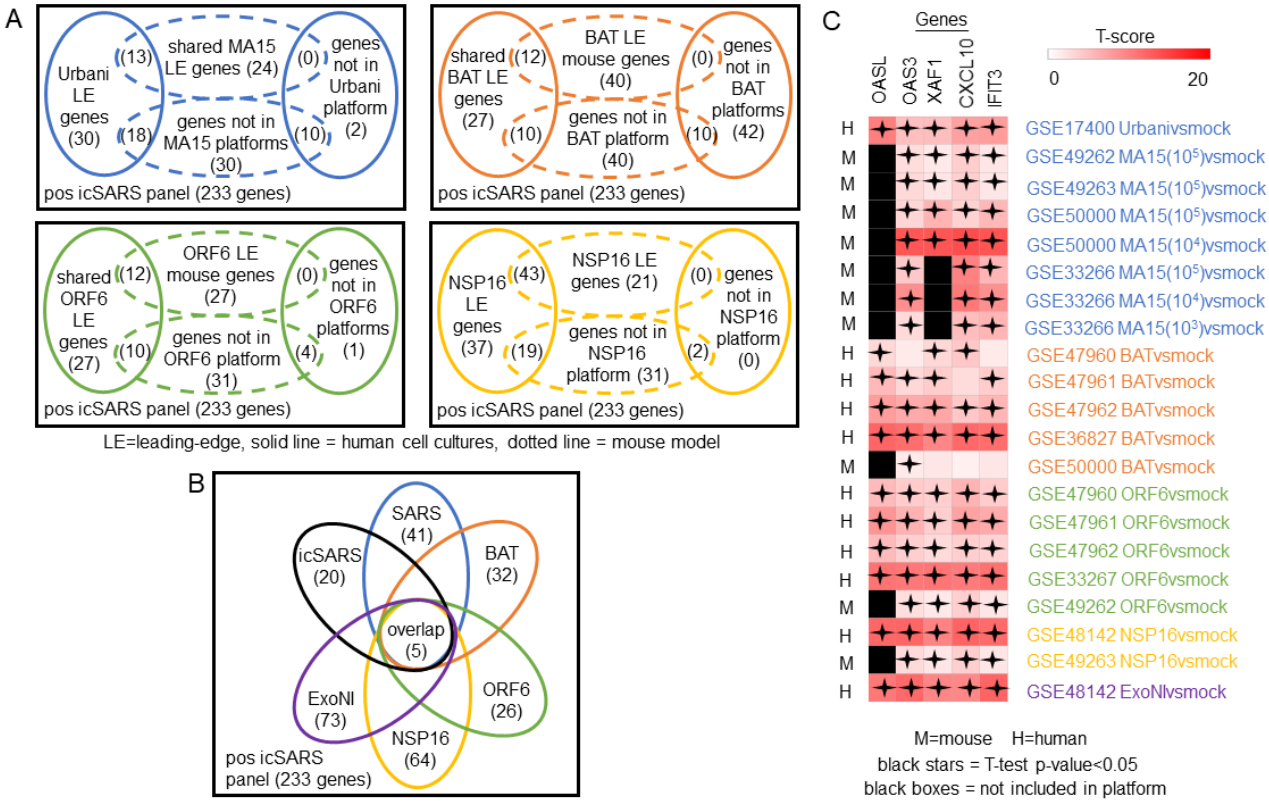


A) Heat map of Gene Set Enrichment Analysis (GSEA) calculated normalized enrichment score (NES) show consistent significant enrichment of the positive icSARS panel across SARS strains with varying levels of virulence in both human lung cultures and mouse lung samples. The negative icSARS panel is consistent across all SARS strains except BAT-SRBD and low inoculation doses of MA15. B) Box and whiskers plot of NES from 1000 randomly generated gene panels containing 233 genes (individual queries) compared to 23 comparison gene signature (individual references, Figure 1) demonstrate that the NES achieved by comparing the reference gene signature to the positive icSARS gene panel used as query in (A) is better than NES obtained randomly. C) Box and whiskers plot of NES from 1000 randomly generated gene panels containing 114 genes (individual queries) compared to 23 comparison gene signature (individual references) demonstrate that the NES

This is a provisional file, not the final typeset article

achieved by comparing the reference gene signature to the negative icSARS gene panel used as query in (A) is consistently better than NES obtained randomly for all strains except BAT-SRBD and low inoculation doses for MA15.

Figure 6. Meta-analysis Across 28 Gene Signatures Representing Seven SARS Strains Varying in Virulence Identifies Five Over-Expressed Genes Associated with SARS Infection



A) Venn diagrams illustrated the inclusion and overlap of positive icSARS panel genes in identified leading-edges and dataset platforms across human cell culture (*in vitro*) and mouse infection (*in vivo*) models using SARS strains that are 1) fully virulent (top-left), 2) isolated from bats and synthetically made to infect animal cells (top-right), 3) lacking ORF6 function resulting in less virulence (bottom left), and 4) lacking NSP function resulting in less virulence (bottom right). Overlapping positive icSARS panel leading-edge genes represent genes that change in response to a SARS infection with a specific SARS strain. B) Venn diagram showed the intersection of positive icSARS panel leading-edges from 22 verification and comparison gene signatures across seven strains (icSARS, Urbani, MA15, BAT-SRBD, dORF6, dNSP16, and ExoNI) in both human cultures and mouse samples that identified five shared genes. Overlapping positive icSARS panel leading-edge genes represent genes associated with SARS infection regardless of virulence strength and mechanism. C) Heat map of T-scores for the five positive icSARS panel leading-edge genes identified in (B) showed the extent of over-expression in each gene signature from SARS strains with varying virulence levels.

Table 1. SARS Infection Datasets Utilized for this Study

Dataset	Description[1]	Platform	Probes	IDs	Genes
GSE47960	Unlogged cultures of human primary tracheobroncial epithelium cells with mock, infectious clone SARS-CoV (icSARS) derived from Urbani strain, SARS-dORF6, or SARS-BatSRBD infection	GPL6480	32067	25666	16479
GSE47961	Unlogged cultures of human primary tracheobroncial epithelium cells with mock, icSARS, SARS-dORF6, or SARS-BatSRBD infection	GPL6480	33421	26500	16978
GSE47962	Unlogged cultures of human primary tracheobroncial epithelium cells with mock, icSARS, SARS-dORF6, or SARS-BatSRBD infection	GPL6480	32388	25897	16603
GSE37827	Unlogged cultures of 2B4 cells (a clonal derivative of human lung adenocarcinoma Calu-3 cells) with mock, icSARS, or SARS-BatSRBD infection	GPL6480	41000	30936	19596
GSE48142	Unlogged cultures of human lung adenocarcinoma 2B4 cells with mock, icSARS, icSARS-ExoNI or icSARS-dNSP16 infection	GPL6480	33321	26212	16794
GSE33267	Unlogged cultures of human lung adenocarcinoma Calu3cells with mock, icSARS, or SARS-dORF6 infection	GPL4133	45015	32696	19751
GSE17400	Unlogged cultures of human lung adenocarcinoma 2B4 cells with mock or Urbani strain of SARS-CoV (SARS) infection	GPL570	54675	44134	21182
GSE50000	Unlogged lung samples from 20-week-old C57BL6 mice infected with mock, icSARS, either 10^4 or 10^5 plaque-forming unit wild-type mouse-adapted SARS MA15 or SARS-BatSRBD	GPL7202	41174	34907	21200
GSE33266	Unlogged lung samples from 20-week-old C57BL6 mice infected with mock or wild-type mouse-adapted SARS MA15 at 10^2,10^3, 10^4, or 10^5 plaque-forming units over a time course	GPL4134	45057	36933	21608
GSE49262	Log2 normalized lung samples from 20-week-old C57BL6 mice infected with mock,10^5 plaque-forming unit wild-type SARS MA15, or SARS-dORF6	GPL7202	41174	34907	21200
GSE49263	Log2 normalized lung samples from 20-week-old C57BL6 mice infected with mock,10^5 plaque-forming unit wild-type SARS MA15, or SARS-dNSP16	GPL7202	41174	34907	21200

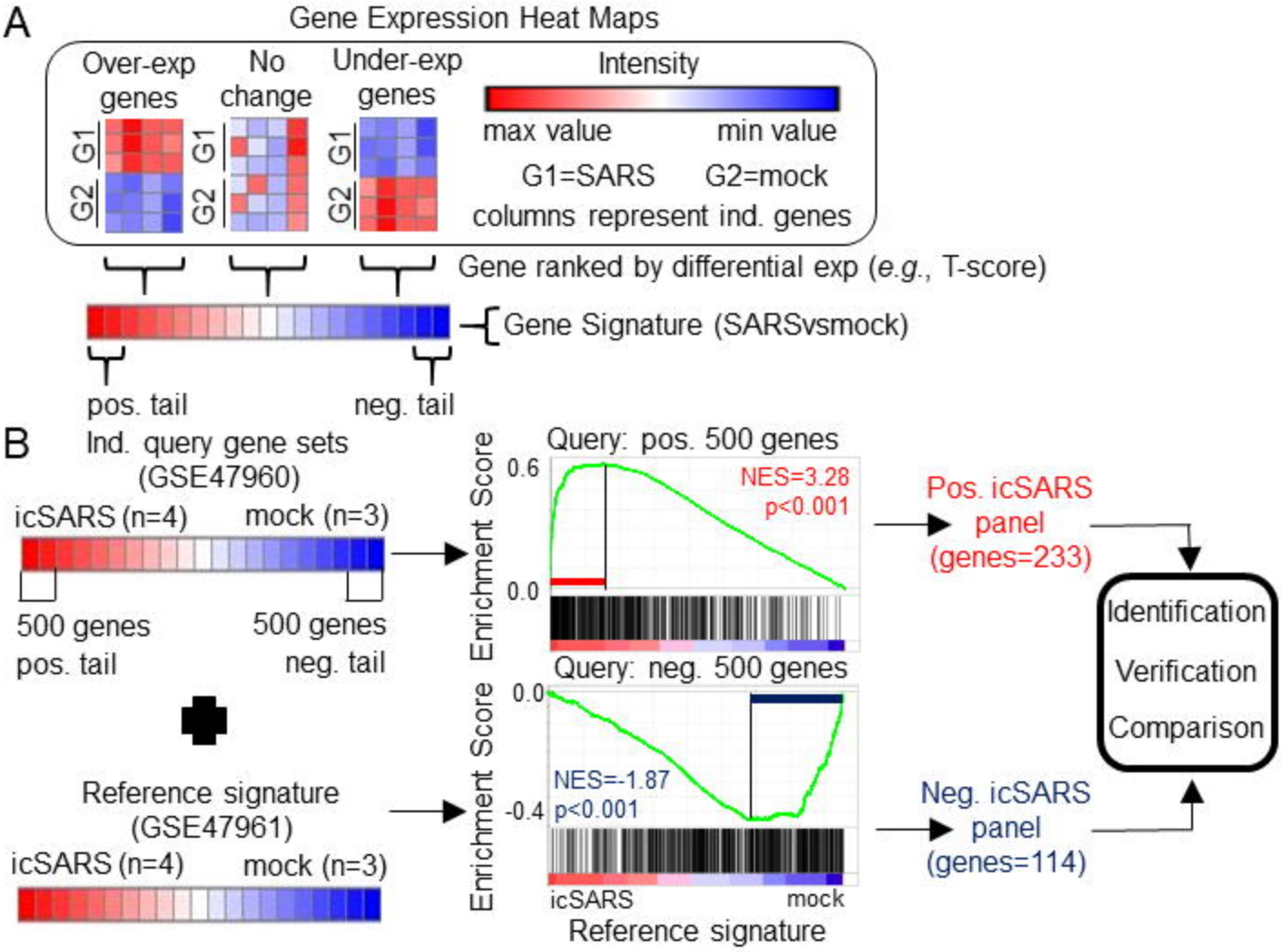
[1]All datasets collected samples over a time course. 48hr samples were used for this study unless otherwise stated. Unlogged samples were z-scored across all dataset samples regardless of time or treatment for normalization before use.

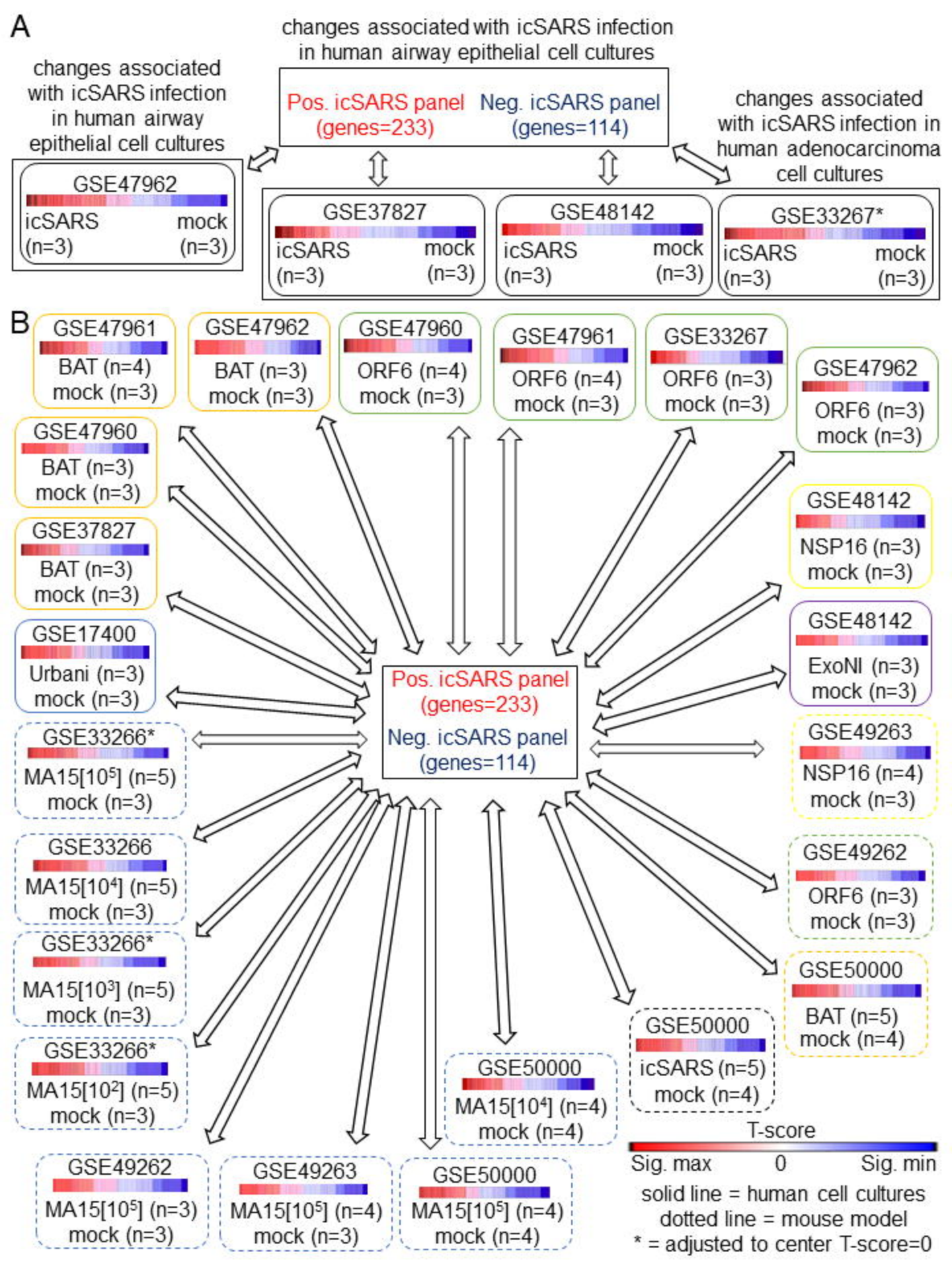
Platform, Gene Expression Omnibus platform identifiers. Probes, number of platform probes with expression data in dataset. IDs, number of probes with Gene Expression Omnibus provided Entrez identifiers. Genes, number of unique Entrez identifiers used for gene signature definition with representatives for duplicate Entrez identifiers being selected by highest coefficient of variation. Values provided in the Genes column reflect the number of genes in gene signatures defined for this study.

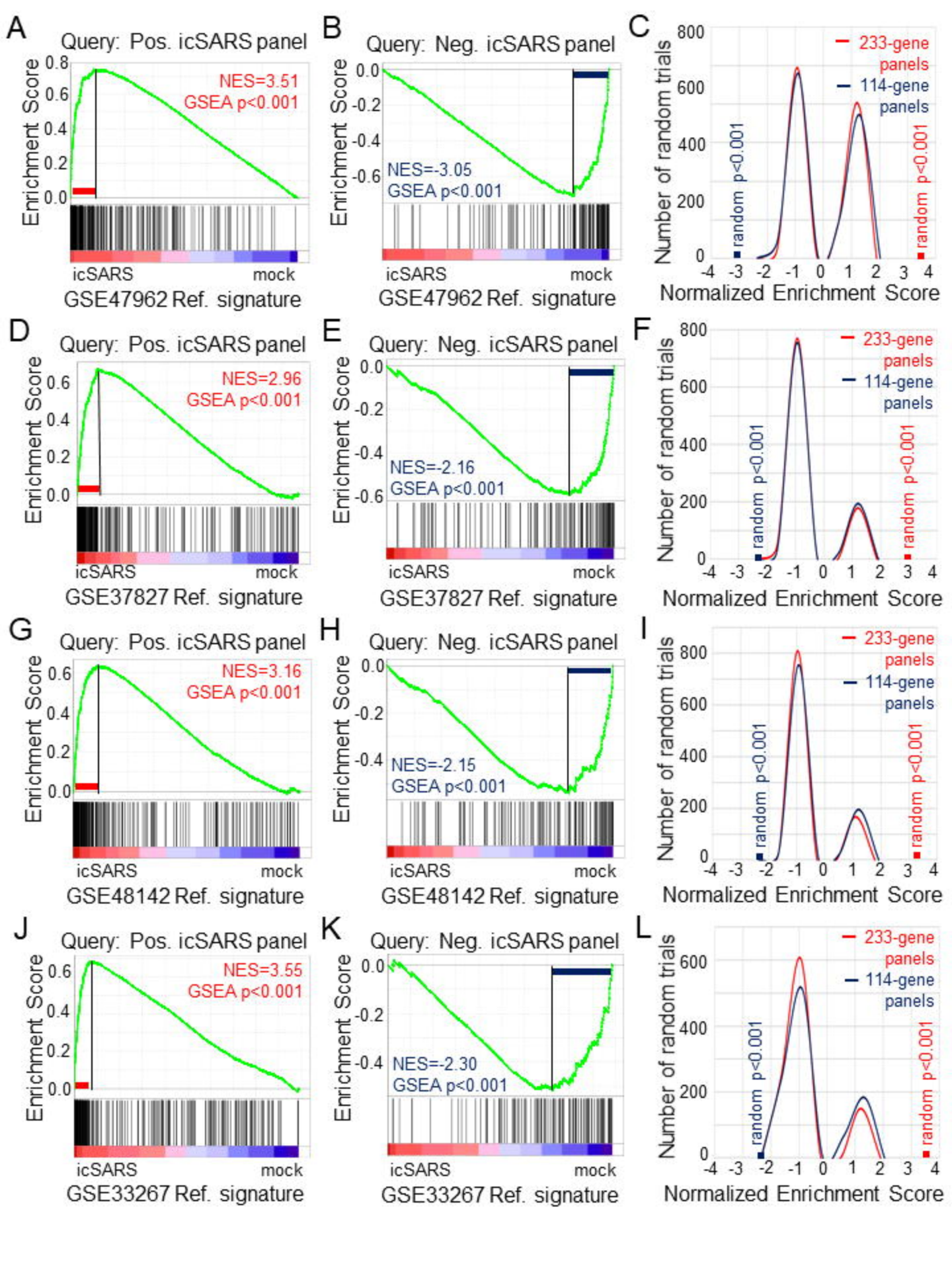
Table 2. Gene Signatures Defined in this Study

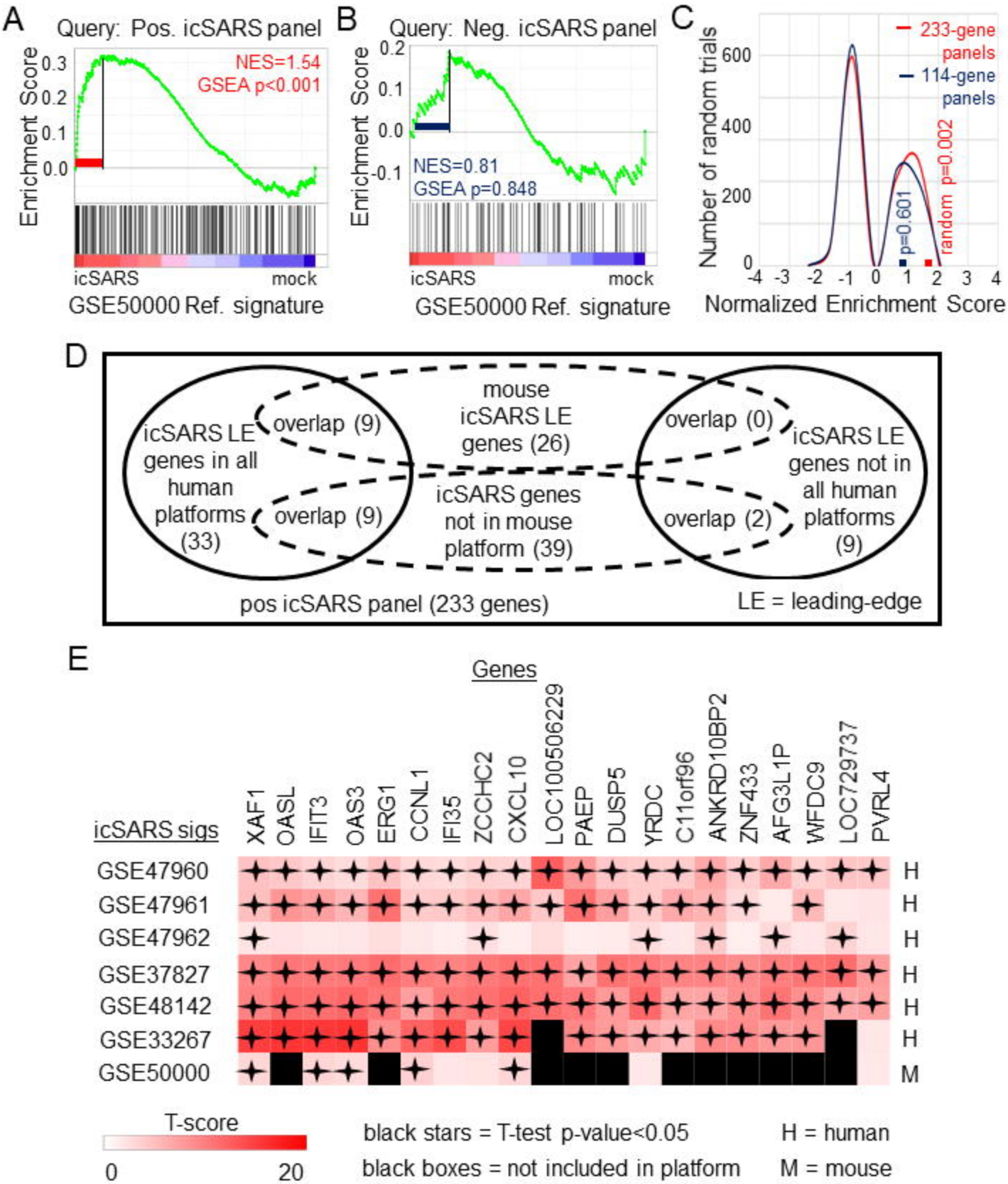
Dataset	Group 1 (N)	Group 2 (N)[1]	Gene signature	Use	High[2]	Low[2]	Cross[3]
GSE47960	icSARS (4)	mock (3)	icSARSvsmock	I	12.5	-13.2	7245
	dORF6 (4)	mock (3)	ORF6vsmock	C	12.5	-13.2	8738
	BatSRBD (3)	mock (3)	BATSRBDvsmock	C	11.6	-12.1	8076
GSE47961	icSARS (4)	mock (3)	icSARSvsmock	I	12.0	-13.1	8285
	dORF6 (4)	mock (3)	ORF6vsmock	C	13.1	-12.1	9026
	BatSRBD (4)	mock (3)	BATSRBDvsmock	C	12.1	-6.7	8421
GSE47962	icSARS (3)	mock (3)	icSARSvsmock	V	12.1	-12.1	8703
	dORF6 (3)	mock (3)	ORF6vsmock	C	12.2	-12.2	8225
	BatSRBD (3)	mock (3)	BATSRBDvsmock	C	12.2	-12.2	9157
GSE37827	icSARS (3)	mock (3)	icSARSvsmock	V	12.2	-12.2	8261
	BatSRBD (3)	mock (3)	BATSRBDvsmock	C	12.2	-11.8	8843
GSE48142	icSARS (3)	mock (3)	icSARSvsmock	V	12.2	-12.2	6918
	ExoNI (3)	mock (3)	ExoNIvsmock	C	12.2	-9.9	7570
	dNSP16 (3)	mock (3)	NSP16vsmock	C	12.2	-12.2	6638
GSE33267	icSARS (3)	mock (3)	icSARSvsmock[4]	V	15.9	-8.6	9874
	dORF6 (3)	mock (3)	ORF6vsmock	C	12.2	-12.2	7378
GSE17400	SARS (3)	mock (3)	SARSvsmock	C	12.2	-12.2	10571
GSE50000	icSARS (5)	mock (4)	icSARSvsmock	C	6.8	-7.7	10007
	$MA10^5$ (4)	mock (4)	$MA10^5$vsmock	C	9.4	-10.2	9882
	$MA10^4$ (4)	mock (4)	$MA10^4$vsmock	C	14.1	-14.1	9759
	BatSRBD (5)	mock (4)	BATSRBDvsmock	C	9.9	-9.5	10185
GSE33266	$MA10^5$ (5)	mock (3)	$MA10^5$vsmock[4]	C	13.6	-14.6	10803
	$MA10^4$ (5)	mock (3)	$MA10^4$vsmock	C	14.0	-14.1	11654
	$MA10^3$ (5)	mock (3)	$MA10^3$vsmock[4]	C	12.8	-6.6	10803
	$MA10^2$ (5)	mock (3)	$MA10^2$vsmock[4]	C	7.3	-6.4	10803
GSE49262	$MA10^5$ (3)	mock (3)	$MA10^5$vsmock	C	4.5	-2.5	10073
	dORF6 (3)	mock (3)	ORF6vsmock	C	4.4	-2.5	10211
GSE49263	$MA10^5$ (4)	mock (3)	$MA10^5$vsmock	C	4.8	-2.5	10384
	dNSP16 (4)	mock (3)	NSP16vsmock	C	5.0	-2.6	10188

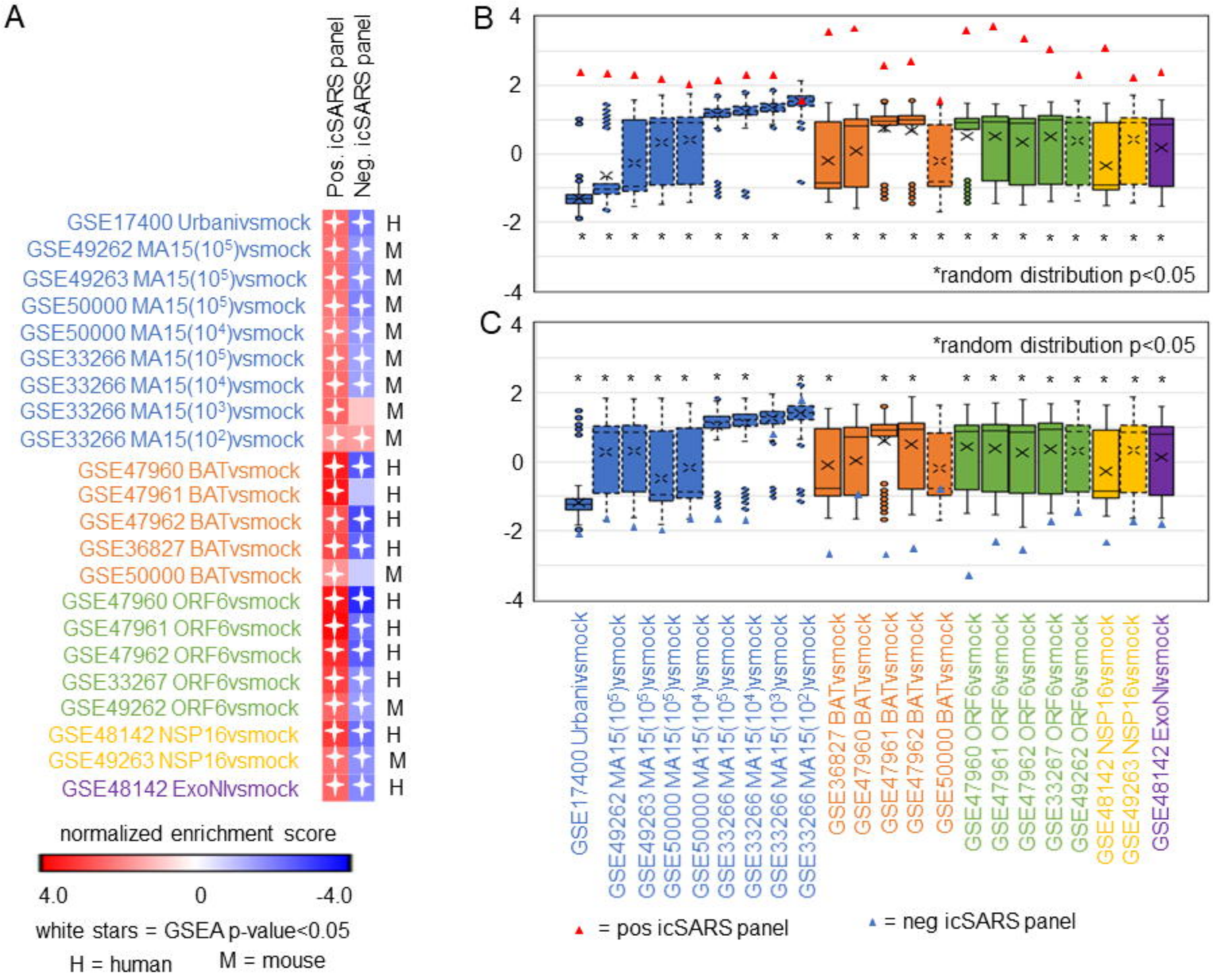
[1]Same mock samples used across signatures in same dataset.
[2]These values define the range of Welch's two-sample T-scores that were used to rank genes in the signature.
[3]Values in the Cross column reflect where (*i.e.*, which gene) in the signature the T-score crosses 0.
[4]T-scores in gene signature adjusted so Cross (T-score=0) occurs near center of the signature.
C, comparison; I, identification; N, number of samples; V, verification.

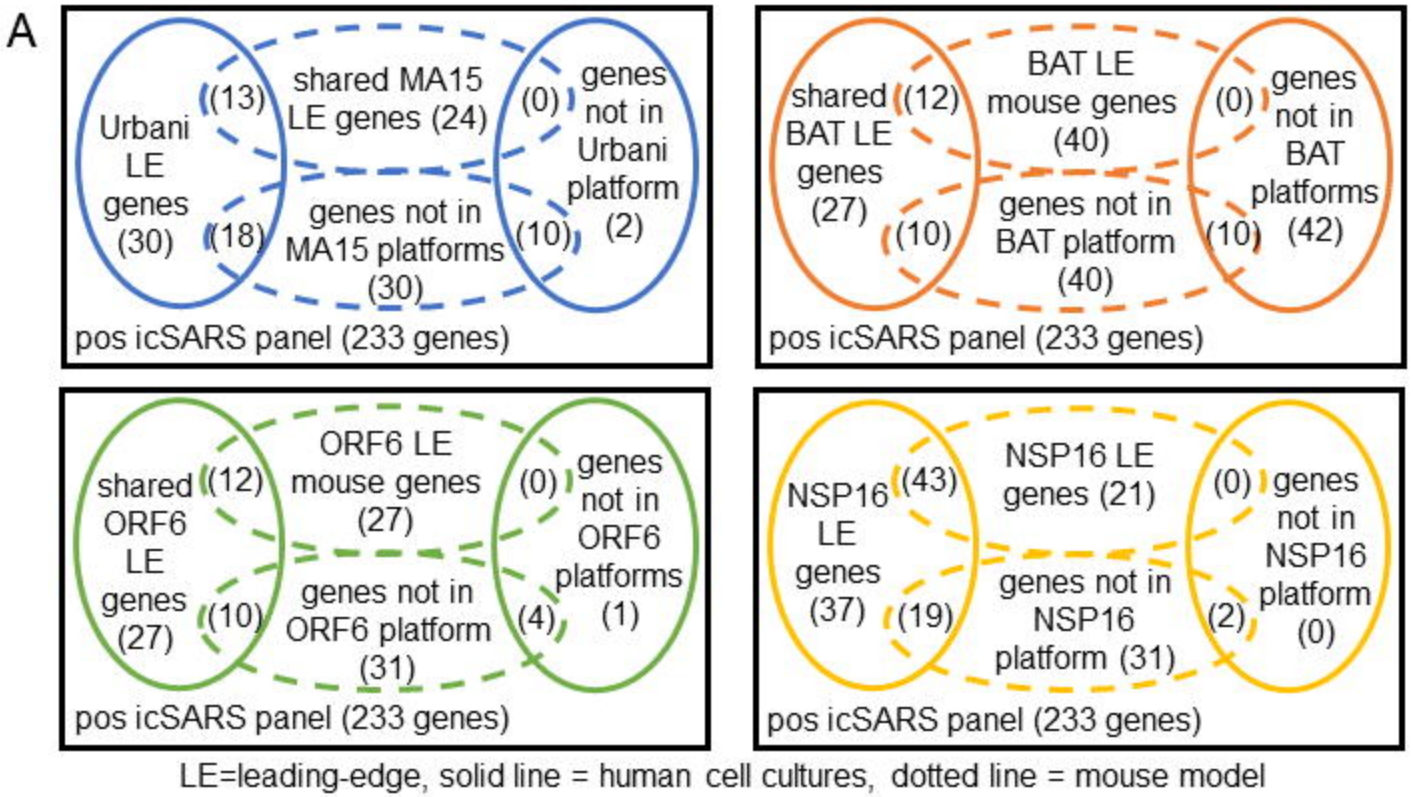

A
Gene Expression Heat Maps
Over-exp genes
No change
Under-exp genes
Intensity
max value
min value
G1=SARS
G2=mock
columns represent ind. genes
G1
G2
Gene ranked by differential exp (e.g., T-score)
Gene Signature (SARSvsmock)
pos. tail
neg. tail
B
Ind. query gene sets (GSE47960)
icSARS (n=4)
mock (n=3)
500 genes pos. tail
500 genes neg. tail
Reference signature (GSE47961)
icSARS (n=4)
mock (n=3)
Query: pos. 500 genes
Enrichment Score
0.6
0.0
NES=3.28
p<0.001
Query: neg. 500 genes
0.0
-0.4
NES=-1.87
p<0.001
icSARS
mock
Reference signature
Pos. icSARS panel (genes=233)
Neg. icSARS panel (genes=114)
Identification
Verification
Comparison


A
changes associated with icSARS infection in human airway epithelial cell cultures
Pos. icSARS panel (genes=233)
Neg. icSARS panel (genes=114)
changes associated with icSARS infection in human airway epithelial cell cultures
GSE47962
icSARS (n=3)
mock (n=3)
changes associated with icSARS infection in human adenocarcinoma cell cultures
GSE37827
icSARS (n=3)
mock (n=3)
GSE48142
icSARS (n=3)
mock (n=3)
GSE33267*
icSARS (n=3)
mock (n=3)
B
GSE47961
BAT (n=4)
mock (n=3)
GSE47962
BAT (n=3)
mock (n=3)
GSE47960
ORF6 (n=4)
mock (n=3)
GSE47961
ORF6 (n=4)
mock (n=3)
GSE33267
ORF6 (n=3)
mock (n=3)
GSE47962
ORF6 (n=3)
mock (n=3)
GSE47960
BAT (n=3)
mock (n=3)
GSE37827
BAT (n=3)
mock (n=3)
GSE48142
NSP16 (n=3)
mock (n=3)
GSE17400
Urbani (n=3)
mock (n=3)
GSE48142
ExoNI (n=3)
mock (n=3)
Pos. icSARS panel (genes=233)
Neg. icSARS panel (genes=114)
GSE33266*
MA15[10^5] (n=5)
mock (n=3)
GSE49263
NSP16 (n=4)
mock (n=3)
GSE33266
MA15[10^4] (n=5)
mock (n=3)
GSE49262
ORF6 (n=3)
mock (n=3)
GSE33266*
MA15[10^3] (n=5)
mock (n=3)
GSE50000
BAT (n=5)
mock (n=4)
GSE33266*
MA15[10^2] (n=5)
mock (n=3)
GSE50000
icSARS (n=5)
mock (n=4)
GSE50000
MA15[10^4] (n=4)
mock (n=4)
GSE49262
MA15[10^5] (n=3)
mock (n=3)
GSE49263
MA15[10^5] (n=4)
mock (n=3)
GSE50000
MA15[10^5] (n=4)
mock (n=4)
T-score
Sig. max
0
Sig. min
solid line = human cell cultures
dotted line = mouse model
* = adjusted to center T-score=0

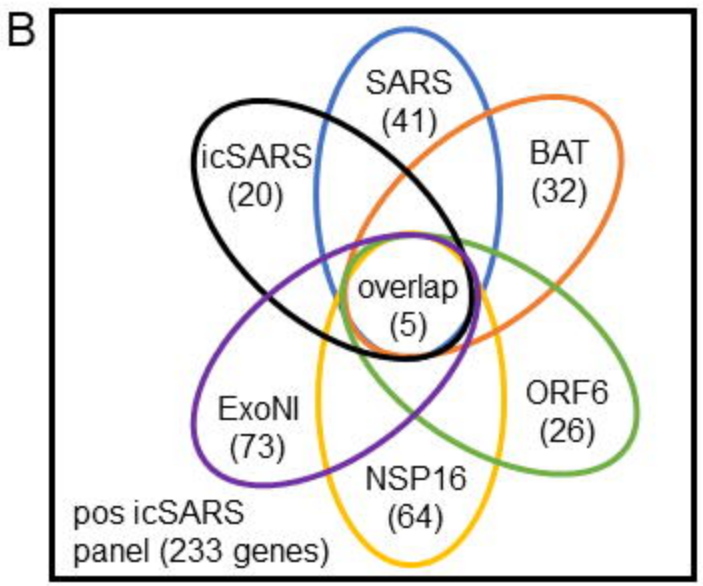

A
Query: Pos. icSARS panel
NES=3.51
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE47962 Ref. signature
B
Query: Neg. icSARS panel
NES=-3.05
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE47962 Ref. signature
C
233-gene panels
114-gene panels
random p<0.001
random p<0.001
Number of random trials
Normalized Enrichment Score
D
Query: Pos. icSARS panel
NES=2.96
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE37827 Ref. signature
E
Query: Neg. icSARS panel
NES=-2.16
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE37827 Ref. signature
F
233-gene panels
114-gene panels
random p<0.001
random p<0.001
Number of random trials
Normalized Enrichment Score
G
Query: Pos. icSARS panel
NES=3.16
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE48142 Ref. signature
H
Query: Neg. icSARS panel
NES=-2.15
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE48142 Ref. signature
I
233-gene panels
114-gene panels
random p<0.001
random p<0.001
Number of random trials
Normalized Enrichment Score
J
Query: Pos. icSARS panel
NES=3.55
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE33267 Ref. signature
K
Query: Neg. icSARS panel
NES=-2.30
GSEA p<0.001
Enrichment Score
icSARS
mock
GSE33267 Ref. signature
L
233-gene panels
114-gene panels
random p<0.001
random p<0.001
Number of random trials
Normalized Enrichment Score

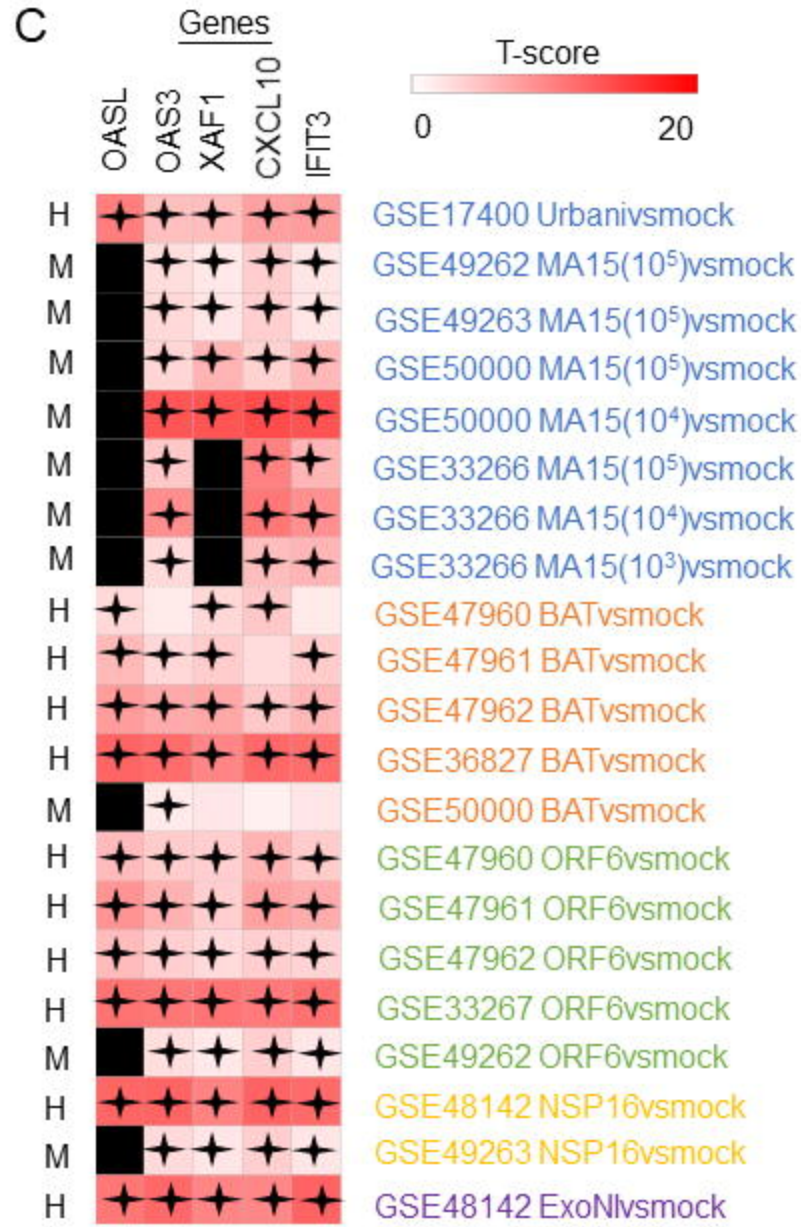

A
Query: Pos. icSARS panel
Enrichment Score
NES=1.54
GSEA p<0.001
icSARS
mock
GSE50000 Ref. signature
B
Query: Neg. icSARS panel
Enrichment Score
NES=0.81
GSEA p=0.848
icSARS
mock
GSE50000 Ref. signature
C
233-gene panels
114-gene panels
Number of random trials
p=0.601
random p=0.002
Normalized Enrichment Score
D
icSARS LE genes in all human platforms (33)
overlap (9)
overlap (9)
mouse icSARS LE genes (26)
icSARS genes not in mouse platform (39)
overlap (0)
overlap (2)
icSARS LE genes not in all human platforms (9)
pos icSARS panel (233 genes)
LE = leading-edge
E
Genes
icSARS sigs
XAF1
OASL
IFIT3
OAS3
ERG1
CCNL1
IFI35
ZCCHC2
CXCL10
LOC100506229
PAEP
DUSP5
YRDC
C11orf96
ANKRD10BP2
ZNF433
AFG3L1P
WFDC9
LOC729737
PVRL4
GSE47960
GSE47961
GSE47962
GSE37827
GSE48142
GSE33267
GSE50000
H
H
H
H
H
H
M
T-score
0
20
black stars = T-test p-value<0.05
black boxes = not included in platform
H = human
M = mouse


A
Pos. icSARS panel
Neg. icSARS panel
GSE17400 Urbanivsmock H
GSE49262 MA15(10^5)vsmock M
GSE49263 MA15(10^5)vsmock M
GSE50000 MA15(10^5)vsmock M
GSE50000 MA15(10^4)vsmock M
GSE33266 MA15(10^5)vsmock M
GSE33266 MA15(10^4)vsmock M
GSE33266 MA15(10^3)vsmock M
GSE33266 MA15(10^2)vsmock M
GSE47960 BATvsmock H
GSE47961 BATvsmock H
GSE47962 BATvsmock H
GSE36827 BATvsmock H
GSE50000 BATvsmock M
GSE47960 ORF6vsmock H
GSE47961 ORF6vsmock H
GSE47962 ORF6vsmock H
GSE33267 ORF6vsmock H
GSE49262 ORF6vsmock M
GSE48142 NSP16vsmock H
GSE49263 NSP16vsmock M
GSE48142 ExoNIvsmock H
normalized enrichment score
4.0
0
-4.0
white stars = GSEA p-value<0.05
H = human
M = mouse
B
*random distribution p<0.05
C
*random distribution p<0.05
GSE17400 Urbanivsmock
GSE49262 MA15(10^5)vsmock
GSE49263 MA15(10^5)vsmock
GSE50000 MA15(10^5)vsmock
GSE50000 MA15(10^4)vsmock
GSE33266 MA15(10^5)vsmock
GSE33266 MA15(10^4)vsmock
GSE33266 MA15(10^3)vsmock
GSE33266 MA15(10^2)vsmock
GSE36827 BATvsmock
GSE47960 BATvsmock
GSE47961 BATvsmock
GSE47962 BATvsmock
GSE50000 BATvsmock
GSE47960 ORF6vsmock
GSE47961 ORF6vsmock
GSE47962 ORF6vsmock
GSE33267 ORF6vsmock
GSE49262 ORF6vsmock
GSE48142 NSP16vsmock
GSE49263 NSP16vsmock
GSE48142 ExoNIvsmock
= pos icSARS panel
= neg icSARS panel



LE=leading-edge, solid line = human cell cultures, dotted line = mouse model





M=mouse H=human

black stars = T-test p-value<0.05

black boxes = not included in platform