Construction of a Transcriptome-Driven Network at the Early Stage of Infection with Influenza A H1N1 in Human Lung Alveolar Epithelial Cells

Myungguen Chung†, Soo Young Cho† and Young Seek Lee†*

1Division of Molecular and Life Sciences, Hanyang University, Ansan 15588, 2National Cancer Center, Goyang 10408, Republic of Korea

Abstract
We aimed to understand the molecular changes in host cells that accompany infection by the seasonal influenza A H1N1 virus because the initial response rapidly changes owing to the fact that the virus has a robust initial propagation phase. Human epithelial alveolar A549 cells were infected and total RNA was extracted at 30 min, 1 h, 2 h, 4 h, 8 h, 24 h, and 48 h post infection (h.p.i.). The differentially expressed host genes were clustered into two distinct sets of genes as the infection progressed over time. The patterns of expression were significantly different at the early stages of infection. One of the responses showed roles similar to those associated with the enrichment gene sets to known ‘gp120 pathway in HIV.’ This gene set contains genes known to play roles in preventing the progress of apoptosis, which infected cells undergo as a response to viral infection. The other gene set showed enrichment of ‘Drug Metabolism Enzymes (DMEs).’ The identification of two distinct gene sets indicates that the virus regulates the cell’s mechanisms to create a favorable environment for its stable replication and protection of gene metabolites within 8 h.

Key Words: A549 Cells, Influenza A virus H1N1 subtype, High-throughput nucleotide sequencing, Sequence analysis RNA, Apoptosis, Gene expression regulation

INTRODUCTION
Influenza A H1N1 Virus (INV) is a RNA virus belonging to the family Orthomyxoviridae. Commonly referred to as the flu virus, H1N1 is specific to birds and mammals. The latest worldwide flu pandemic occurred in April 2009 and accounted for an estimated 284,500 global deaths (Dawood et al., 2012).

In general, a variety of host responses during viral infection have been identified, including activation of numerous cell death and survival pathways. These pathways include: 1) programmed cell death I (apoptosis), 2) programmed cell death II (autophagy), and 3) endoplasmic reticulum stress with subsequent unfolded protein response (UPR). There has been extensive research on the regulatory roles of these pathways during the influenza virus life cycle (Yeganeh et al., 2013). Pathogen sensory pathways (e.g., RIG-I) displayed long-lasting associations with cytokine/chemokine signaling through day 8 (Dimitrakopoulou et al., 2014). Transcriptome research for H1N1 infection revealed that host cell innate immunity is induced at 3 to 5 days post infection (Zou et al., 2013). Another report described that the influenza induced an early immune response 3 to 7 days after infection (Park et al., 2015). However, influenza symptoms typically begin two days after exposure to the virus (Nair et al., 2011). In order to understand host cell pathway hijacking in H1N1 infection within two days, transcriptome changes during the first 48 h post infection could provide important insights into the mechanisms underlying the pathogenicity of influenza A viruses. The host cell transcriptome profile was rapidly changed in the early stages of infection (within 8 h of infection). Several mediators of apoptosis were activated in this stage. After 8 h, the most significant change was found to be that in expression of genes encoding drug-metabolizing enzymes.
MATERIALS AND METHODS

Host cell preparation

A549 cells are adenocarcinoma human alveolar basal epithelial cells. The A549 cell line is widely used as an in vitro model as a type II pulmonary epithelial cell model for studies of drug metabolism and as a transfection host. A549 cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM). All media contained 10% fetal bovine serum and antibiotics (penicillin streptomycin).

Viral infection and total RNA extraction

A549 cells were infected with influenza A virus at a multiplicity of infection (MOI) of 3 for 30 min at 37°C. The virus was propagated at 37°C in 11-day-old chicken embryos. After being washed with phosphate-buffered saline (PBS), cells were infected with influenza virus (IV) at an MOI of 3 for 30 min at 37°C. Infection media consisted of 0.1% glucose, 0.05% vitamin solution, and 0.5 mg/mL L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK)-treated trypsin in DMEM. Infected cells were then washed and incubated in medium without serum for varying lengths of time. All infected cells were inoculated at least than 3 C (cycle of threshold) for RT-PCR validation experiments. Total RNA was extracted from A549 cells with the RNeasy Mini Kit (74104, QIAGEN, MD, USA).

Sample QC & RNA quantification

RNA integrity was assessed using an Agilent Technologies 2100 Bioanalyzer with an RNA Integrity Number (RIN) value greater than or equal to 7. Cell-to-cell variability of influenza viral infection was quantified by subjecting the cell supernatant to a plaque assay to determine the load of infectious viral particles and then performing real-time reverse transcription quantitative PCR (RT–qPCR) on cell lysates to quantify the level intracellular viral RNA (vRNA) of individual genome segments. The size of PCR-enriched fragments was verified by assessing the template size distributions on an Agilent Technologies 2100 Bioanalyzer using a DNA 1000 chip (Agilent, CA, USA).

Next generation sequencing

The sequencing library was prepared by random fragmentation of the DNA or cDNA sample, followed by 5’ and 3’ adapter ligation. The Illumina Hiseq 4000 generated raw images utilizing HiSeq Control Software (HCS, Illumina, CA, USA) v3.3 for system control and base calling through the integrated primary analysis software Real Time Analysis (RTA, Illumina) v2.5.2. The binary BCL (base calls) value was converted into FASTQ utilizing Illumina package bcl2fastq (V2.16.10, Illumina).

Sequence alignment

The alignment software application “New Tuxedo Protocol” was used for transcript assembly (Pertea et al., 2016). In the alignment step, HISAT2 was used to generate the alignment data (Kim et al., 2015a), which is a fast alignment program for mapping next-generation sequencing to read segments of the human genome (Siren et al., 2014).

Transcript assembly

In the transcript assembly, StringTie (Pertea et al., 2015) is known as a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts. It uses a novel network flow algorithm and an optional de novo assembly step to assemble and quantitate full-length transcripts representing multiple splice variants for each gene locus. StringTie estimates gene-level measurements by appropriately combining fragments per kilobase of exon per million reads (FPKMs) from the transcripts of each gene.

Time-series expression analysis

Ballgown (Frazee et al., 2015) is a bioconductor-based suite that hosts a comprehensive set of bioconductor communities. The “stattest” function of Ballgown is available with a gene feature that prevents data skewing in the time series and was used for statistical tests of time-series expression. Gene-level measurements were performed by appropriately combining FPKMs from the transcripts comprising the gene (Li and Dewey, 2011) and it proposed a directed graph model combined with an expectation-maximization algorithm to estimate abundance. Ballgown generates a reference index by preparing reference transcript data and calculated RSEM values by inputting RNA-Seq data. A natural splines model with
default settings was used to fit time profiles and six degrees of freedom (df) were used to fit the splines. “Timecourse=TRUE” parameters were assigned for time profiling using the natural splines function (Supplementary Table 1).

### Functional annotation

DAVID bioinformatics resources consist of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists (Huang et al., 2009).

### Gene set variation analysis (GSVA)

GSVA is a gene set enrichment method which estimates variation of pathway activity over a sample population in an unsupervised manner. Program operation was employed using basic parameters in ‘RNA-seq’ mode and ‘kernel=TRUE’ using ‘gsva’ function (release version 3.5). The known-gene sets database was updated with the ‘Hallmark class’ and ‘c7 class’ by extending the default class (Hänzelmann et al., 2013).

### Weighted correlation network analysis (WGCNA)

Time-series differentially expressed genes (DEGs) were categorized by WGCNA. Categorized data were divided into 300-1000 gene categories with respect to their biological annotation-interpretation (Supplemental Table 2). Constructing a weighted gene network entails calculation of the soft thresholding power β, to which co-expression similarity is increased to calculate adjacency. We did this with a power of 16, which is the lowest power for which the scale-free topology fit index curve flattens upon reaching a high value (in this case, approximately 0.95), a relatively large minimum module size of 30, and a moderate sensitivity (Supplementary Table 3).

### Biological metabolic network in-silico validation

Elsevier’s Pathway Studio Mammalian, ChemEffect, DiseaseFX analysis: The set of biological networks in this study was created using Pathway Studio 11.0 (Ariadne) software (https://mammalcedfx.pathwaystudio.com/). Elsevier’s Pathway Studio® enables users to explore molecular interactions and cause-and-effect relationships associated with biological processes by integrating a vast knowledge base of biological relationships with analytical and visualization tools.

Ingenuity Pathway Analysis (IPA) metabolomics analysis: DEGs were analyzed using QIAGEN’s IPA software. The canonical pathways and functional processes with the most significant biological importance were identified using the list of DEGs identified with RNA-seq and the Ingenuity Pathways Knowledge Base. Pathway enrichment p-values (Fisher’s exact test) and activation z-scores were calculated using IPA.

### RESULTS

The biological functions are divided into two groups approximately 8 h after infection

In hierarchical clustering, all samples were divided into two groups (Supplementary Fig. 1). Early infection and late infection state samples were clustered into separate groups (Supplementary Fig. 2). Known biological function pathways were calculated using GSVA and all samples were divided into two groups by pathway activity score (Fig. 1). GSVA provides
increased power to detect subtle pathway activity changes with the ‘Hallmark class’ and ‘c7 class’: ‘Apoptosis (p<0.002)’, ‘G2M checkpoint (p<0.002)’, ‘Interferon alpha response (p<0.002)’ and other biological functions were highly enriched in samples collected from cells infected for less than 8 h. ‘Ara-chidonic acid metabolism (p<0.001)’, ‘glutathione metabolism (p<0.001)’ and others were highly enriched in samples collected from cells infected for more than 8 h. The clustering result for whole transcriptome profile and pathway activity showed the same pattern, indicating that 8 h is a critical point for functional changes in host cells.

**Changes in gene expression were observed in the two groups at 8 h**

To identify host cell pathway hijacking, DEGs were selected by statistical testing of time-series expression data (Fig. 2). Gene expression changes were divided into two major groups. Fig. 2 shows 1,031 genes of the time-series DEGs, which were filtered by a threshold Q-value<0.05 (Supplementary Table 2). The blue module (blue line) shows increased expression after 4 h and decreases after 6 h, while the turquoise module (turquoise line) decreases at 4 h and then increases after 24 h.

Gene ontology (GO) enrichment analysis (Dennis et al., 2003) results of these two modules are displayed in Table 1. The envelop surface glycoprotein gp120 (p<0.007) was enriched in group 1. The neuroactive ligand-receptor interaction (p<1.97E−04) and metabolism of xenobiotics by cytochrome P450 (p<2.60E−04) were enriched in group 2.

**Biological features of host cell before 8 h after infection: apoptosis**

The protein gp120 is essential for viral entry into cells as it plays a vital role in attachment to specific cell surface receptors (Gram and Hansen, 1998). Viruses, such as HIV, with gp120 in their genome induce cellular apoptosis (Kapasi et al., 2002). Apoptosis is important in an infection response and is activated at the early infection stage (Fig. 1). Supplementary Fig. 3 shows that apoptosis marker genes were expressed in early infection stage (Subramanian et al., 2005). The viral gene gp120 is known to be associated with BCL-2 expression (Gaubin et al., 1999). BCL-2 and other apoptosis-related genes are down-regulated in the early stages of influenza A infection. At timepoints before 8 h post-infection, expression of apoptosis marker genes, including BCL2, was decreased (Supplementary Fig. 4A).

**Specific biological features after 8 h: DME-centric pathways**

DMEs are the set of metabolic pathways that modify the chemical structure of xenobiotics, which are compound of foreign materials such as replicated viral RNA and other metabolites (Mackenzie et al., 2017). Suppression of DMEs is necessary for viruses to maintain their presence in a cell and replicate themselves. In general, ‘endoplasmic reticulum (ER) stress response’ and ‘cytochrome P450’ are typical examples of DMEs. Both were found to be DEGs in the experimental results (Supplementary Fig. 4B, 4C).

In the Supplementary Fig. 4B, the endoplasmic reticulum (ER) stress response, also known as the unfolded protein response (UPR), is a primitive, evolutionary conserved molecular signaling cascade which has been implicated in multiple...
biological processes, including innate immunity and the pathogenesis of certain viral infections. Influenza A virus induces ER stress in a pathway-specific manner (Hassan et al., 2012). In the Supplementary Fig. 4C, various cytochrome P450 (<2.06E−04) genes were differentially expressed 8 h post infection. The major role of cytochrome P450s is in steroid and bile acid syntheses (Nelson, 2009; Lorbek et al., 2012). They are essential for the solubilization and absorption of lipids and fat-soluble vitamins in the intestine and they contribute to the major DME pathways by which cholesterol is excreted from the body (Lu et al., 2000).

DISCUSSION

The timepoint of 8 h post infection is an important inflection point for gene expression and biological functions

Kim et al. (2015b) reported that relatively few host cell genes were differentially expressed at 8 h post infection. In this study, our data are in agreement with the report by Kim et al. (2015b). This study indicates that 8 h post infection is an inflection point for host cell function and viral transition from the early to late infection stage.

In the early infection stage ‘Apoptosis’, ‘G2M checkpoint’, ‘Interferon alpha response’, and other biological functions are enriched. One study reported that apoptosis of virus-infected cells is one important host strategy used to limit viral infection (Scott and Norris, 2008). Cell cycle arrest may inhibit early cell death of infected cells, allow the cells to evade immune defenses, or help promote virus assembly (Bagga and Bouchard, 2014). Innate cytokine responses, such as alpha interferon (IFN-alpha) have roles in determining the rate of virus replication in the initial stages of infection (Price et al., 2000). These results indicate that influenza virus naturally establishes an environment for self-reproduction in early infection stage and hijacks host cell pathways, termed ‘cytopathic viral effects’ (Heaton, 2017) (Fig. 1).

In the late infection stage, ‘Arachidonic acid metabolism’ and ‘glutathione metabolism’ were found to be enriched. Lu et al. (2012) reported results from metabolic profiling and biochemical detection which indicated significant metabolic changes in the arachidonic acid metabolic pathway. Glutathione is partially depleted in influenza A virus infection (Alsuwaidi et al., 2013). These result suggest that late expressed genes create circumstances necessary to hijack host cell pathways by inhibiting the inflammatory response and the oxidation required for metabolic breakdown of xenobiotic materials by living organisms, termed the ‘Drug metabolism pathway’ (Commandeur et al., 1995).

However, driver modules in host cell function hijacking was still unclear. To identify driver function in the early infection stage, WGCNA used co-expression to predict functional modules. We predicted an early infection module (blue module) and found enriched ‘GP120 pathway in HIV’ (Fig. 3). The MAVS gene (IPS-1, VISA or Cardif) is critical for host defenses to viral infection by inducing type-1 interferons (IFN-I), though its role in virus-induced apoptotic responses has not been elucidated (Lei et al., 2009). Like gp120 in HIV, the PB1-F2 gene has been shown to induce BCL2 by changing the level of MAVS gene expression (Varga et al., 2011). Unlike HIV, influ-
enz A requires a mediat or such as gp120. The PB1-F2 viral gene in the influenza genome is a candidate to be a mediator of gene sets regulat ed by gp120 (Fig. 3). The PB1-F2 viral gene is known to induce cell death (Chen et al., 2001) and apoptosis (Nicholson et al., 2001), similar to gp120. In the WGNC A results for late stages of infection, we predicted a late module (turquoise module) and found the ‘Drug metabolism enzyme pathway (DMEs)’ to be enriched (Supplementary Fig. 5). DMEs were additionally found to be enriched in the late stage of infection by DAVID: ‘metabolism of xenobiotics by cytochrome 450 (p<0.004)’ and ‘drug metabolism (p<0.0003)’ (Supplementary Table 3). P450 converts exotic materials to chemically reactive metabolites, which, if not detoxified, may lead to various forms of hepatic and extrav hepatic toxicity, including cellular necrosis (Park et al., 1995). In Supplementary Table 4, we show that CYP2C9 (P-value=7.46E-08), CYP2C18 (P-value=9.41E-06), CYP2F1 (P-value=0.0012), and CYP3A7 (P-value=9.51E-05) are involved in drug and steroid metabolism (Nelson, 2009).

### The hub gene of the transcriptome-driven reconstruction network is a candidate drug targeting for influenza A

A therapeutic strategy that might escape viral resistance is to target host cellular mechanisms involved in viral replication and pathogenesis. The DGIgb integrates drug-gene interactions from 15 different sources (Wagner et al., 2016) (Table 2). Reconstruction of biological networks has been vital to understanding complex biological activities (Yu et al., 2013). In the gp120-regulated pathway of the early stages of infection, two genes (BCL-2, TERT) that surpassed the score thresholds (Sum score of DGIgb>2) were found. BCL2 was enriched in Paclitaxel, Oblimersen, Docetaxel, and Obatoclax treatments, while the TERT gene was only enriched in GV1001-treated cells.

Brown et al. reported that Docetaxel should be further studied for its use in influenza vaccine production (Chen et al., 2012; Brown et al., 2015). Obatoclax is a novel inhibitor of endosomal acidification which prevents viral fusion and could be pursued as a potential broad-spectrum antiviral candidate (Varghese et al., 2017). The TERT gene decreased rapidly at 8 h post infection (ENSG00000164362, P-value=3.34E−05). TERT is the major catalytic subunit of telomerase. TERT has additionally reported to be associated with apoptosis (Fu et al., 1999).

The final goal of studying host-microbe interactions is to inform in silico models in order to identify crucial pathways, hubs, and bottlenecks, which will lead to development of new targets and strategies for prevention, diagnosis, risk assessment, and treatment of severe influenza infection in humans (Kollmus et al., 2014). This study adopted functional transcriptome analysis with transcriptome sequencing and identification of genes after focusing on key time-series processes, providing several gene set enrichment analysis methods to identify genes related to seasonal influenza strains. In the future, a series of functional validation experiments need to be performed to compare more virulent influenza strains.

### CONFLICT OF INTEREST

The authors declare that they have no competing interests.

### ACKNOWLEDGMENTS

We thank Dr. Hueeman Kim for his advice regarding experimental design. This research was funded by the post-genome multi-ministerial genome project (NRF-2014M3C9A3064815) of the Ministry of Science, ICT; and Future Planning, Republic of Korea. This research was supported by National Cancer Center Grant (NCC-1711290).

### REFERENCES

Alsuwaidi, A. R., Almarzooqi, S., Albawardy, A., Benedict, S., Kochiyil,
