Transcriptogram analysis reveals relationship between viral titer and gene sets responses during Corona-virus infection.

Rita M.C. de Almeida*[1,2,3], Gilberto L. Thomas[1], and James A. Glazier[4]

[1]*Instituto de Física and* [4]*Instituto Nacional de Ciência e Tecnologia: Sistemas Complexos, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil*

[3]*Programa de Pós-Graduação em Bioinformática, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil*

[4]*Biocomplexity Institute and Department of Intelligent Systems Engineering, Indiana University, Bloomington, Indiana, United States of America*

Abstract

To understand the difference between benign and severe outcomes after Coronavirus infection, we urgently need ways to clarify and quantify the time course of tissue and immune responses. Here we re-analyze 72-hour time-series microarrays generated in 2013 by Sims and collaborators for SARS-CoV-1 *in vitro* infection of a human lung epithelial cell line. Using a Transcriptogram-based top-down approach, we identified three major, differentially-expressed gene sets comprising 219 mainly immune-response-related genes. We identified timescales for alterations in mitochondrial activity, signaling and transcription regulation of the innate and adaptive immune systems and their relationship to viral titer. At the individual-gene level, EGR3 was significantly upregulated in infected cells. Similar activation in T-cells and fibroblasts in infected lung could explain the T-cell anergy and eventual fibrosis seen in SARS-CoV-1 infection. The methods can be applied to RNA data sets for SARS-CoV-2 to investigate the origin of differential responses in different tissue types, or due to immune or preexisting conditions or to compare cell culture, organoid culture, animal models and human-derived samples.

Introduction

Severe respiratory syndromes during the two previous major outbreaks of lethal Coronavirus, SARS-CoV-1 in 2003 [1] and Middle-Eastern Respiratory Syndrome (*MERS*) in 2012 (for a Review, see [2] and references therein), as well as the current SARS-CoV-2 pandemic, often result from dysfunctional immune responses triggered by the interaction of the host immune system with the virus [3,4]. While strong immune responses are essential to contain and clear viral infection, excessive inflammation may damage tissues, delay tissue healing after viral clearance, and lead to acute inflammatory responses and/or sepsis. In the case of SARS-CoV-2, the degree and severity of immune-response pathologies differs greatly between individuals: while 81% of infected individuals show either mild or no symptoms, 14% patients develop severe pneumonia and 5% develop acute respiratory critical conditions associated with multi-organ failure that require intensive medical care and ventilation and may lead to life-threatening sequels or death [5]. Mortality is higher in men than in women, strongly increases after 60 years of age and

increases with pre-existing co-morbidities including diabetes, obesity and cardio-vascular diseases (https://globalhealth5050.org/covid19/). Because of the complexity of the many patterns of response to SARS-CoV-2 we critically need ways to identify important biological mechanisms which act at different phases of infection and allow us to reliably identify differences in pathway and gene activity between individual patients, tissues within patients, individuals with pre-exiting conditions, sex and ethnic differences and age. The immune system is complex, sensitive and dynamic, with a delicate balance of triggers, high-gain feed-back loops, and complex interactions between its many agents, complicating interpretation of experimental measurements of immune-response components and the origins of their variation between individuals. In this case, for diagnostic, prognostic, and therapeutic purposes, detailed mathematical models of patient-specific immune responses might help us understand the range of possible immune responses, and how they depend on patient-specific variables, ranging from initial exposure level and coinfections, to age, sex, preexisting conditions and medications, *etc*. Furthermore, besides those directly related to immune dysfunction, in serious cases COVID-19 symptoms may also include blood and vascular disruption, meaning that the co-activation of other pathways with deleterious effects may play an important role in disease outcomes [6].

Both constructing mathematical models of a complex system like the human immune response and validating such models sufficiently for use to propose therapies or assist with diagnoses or prognoses requires integration of extensive data from *in vitro*, organoid and animal experiments with the more limited clinical observations in humans. Acute inflammatory responses lead to dramatic and rapid changes in expression of large numbers of genes, requiring extensive transcriptome analyses to interpret. For construction and validation of immune-response models, qualitative information is insufficient; we also need specific quantitative information on the time course of immune response and its relationship to viral titer.

Statistical analyses of transcriptome data are generally classified as either *bottom-up*, starting by identifying differentially-expressed genes, clustering them into differentially-expressed pathways and then describing the biological functions these pathways alter, or *top-down*, starting by identifying altered biological functions, then refining the analysis to hierarchically discover the relevant differently-expressed pathways and then genes. In cases of immune-system response to viral infection, where changes in gene expression are genome-wide, top-down approaches may be more practical, since the large number of differentially-expressed genes can be overwhelming to analyze and understand using bottom-up techniques.

RNA-Seq or microarray transcriptomes are affected by many sources of variability, including differences in experimental techniques, biological differences between apparently similar samples, and other confounding variables within samples, like the effect of cell-cycle phase. Our Transcriptogram method to quantify whole-genome-level expression changes reduces noise and enhances signal-to-noise ratio in transcriptome analyses, increasing the power of statistical tests to identify significantly-affected pathways and timescales [7]. Transcriptograms provide a high-level visualization of significant changes in gene expression and have proved useful in identifying relationships between pathways in fungi [8,9], plants [10,11], and humans [12,13,7]. The Transcriptogramer software tool is freely available for download at https://lief.if.ufrgs.br/pub/biosoftwarestranscriptogramer/ and has a Bioconductor application [14].

Here, as a pattern for future Transcriptogram analyses of SARS-CoV-2 data and to illustrate the power of the method in quantifying the detailed and complex temporal pattern of immune response to viral infection in cell culture, we present Transcriptogram analyses for SARS-CoV-1 time-series data sets of Sims *et al.* [15]. Sims *et al.* [15] infected cultures of a clonal population of Calu3 2B4 cells, a lung adenocarcinoma cell line isolated from the pleural effusion of a 25-year-old Caucasian male, sorted for high expression of the enzyme ACE-2, the cellular receptor for SARS-CoV-1 (and SARS-CoV-2). They inoculated cultures with either a wild type SARS-CoV-1 virus (*WT* samples) or a mutant SARS-CoV-1 strain (*DORF6* samples) that does not express the accessory protein ORF6 at high concentration (a multiplicity of infection *MOI* of 5), so that the probability of cell contamination in the culture approached 1. As controls, they also inoculated cultures with a sterile solution (Mock samples). After inoculation, they incubated the cultures at 37°C for 40 min, then changed their medium. They then harvested samples for microarray assays in triplicate at times they labeled $0\ h, 3\ h, 7\ h, 12\ h, 24\ h, 30\ h, 36\ h, 48\ h, 54\ h, 60\ h,$ and $72\ h$. Because they did not report the time for the medium change or the time between inoculation and initial harvest, their data lack a consistent time-0 data set and all time labels refer to the time after the first RNA harvest. As a result, even at $0\ h$, expression in the infected and control cultures differs (see below). We analyzed these data because of the short time intervals between samples at early times, which are critical to understanding the rapid changes occurring in tissue response to viral infection, and the relatively long duration of the experiment. The experiments also have matched-time controls in triplicate at all time points. Sims *et al.* [15] made their data available through Gene Expression Omnibus (*GEO*) under accession number GSE33267 (http://www.ncbi.nlm.nih.gov/geo) and we used these data for our analyses.

Sims *et al.* focused their analyses on the role of ORF6 in the immune response, examining the differences between the WT and DORF6 time series [15]. Here, we focus on large-scale and single-gene transcriptomic changes caused by the WT virus *w.r.t.* the control. Our analyses confirm that gene expression changes massively within $24\ h$, but we also identified relevant responses before $7\ h$ and complex temporal changes in expression throughout the time course of the experiment. Our analyses identify specific additional significant changes in expression in different pathways and individual immune-related genes at $12\ h$, $36\ h$ and $54\ h$. We identified 219 genes with differential expression at some point of the time sequence and, to illustrate the potential of our method, we selected 4 genes with large expression differences *w.r.t.* to control for further scrutiny, EGR3, TWIST1, JUN, and TNFAIP3, all related to immune response. To validate our findings, we also examined a pair of genes, HSD11B1 and HSD11B2, with known associated effector action on Cortisol/Cortisone balance.

Statistical Methods

Overview and analyses pipeline

We performed a Transcriptogram-based [9,16] top-down analysis of whole-genome transcriptome time-series for human epithelial cells cultures, comparing cultures inoculated with either a mock or a SARS-CoV-1-containing (*WT*) solution. We first assessed patterns of expression change shared by large numbers of genes, then considered smaller gene sets which had strongly covariant temporal signatures and finally examined single genes whose variance was statistically significant withing these sets. At each stage, we filtered the data based on the statistical significance of the subseries variability. The next section briefly discusses the Transcriptogram method and relevant parameters. We then present the data and Transcriptograms for the Mock and Wild Type (*WT*) virus strain time courses, indicate the gene sets we focus on and provide a genome-wide visualization of the main patterns of time evolution for the covariance-clustered gene data, comparing the infected and non-infected samples. We then consider the time evolution of 219 genes whose time courses show fold-changes larger than two compared to their pair-matched Mock samples at at least one time point. We clustered the genes by time-course similarity and identified 6 clusters. We determined the mean time evolution for each cluster. Several of these clusters show complex non-monotonic time courses, which we compare to the viral titer. From the changes in behaviors presented by the gene clusters, we infer the typical patterns and timescales for Calu3 2B4 epithelial-cell immune response to SARS-CoV-1 infection, correlate these timescales with viral titer and identify single genes that may possible targets for therapy development.

Transcriptograms

Transcriptograms are expression profiles, obtained by running a window average for expression levels of multiple genes, previously organized in an ordered gene list, representing the whole human genome. Here we consider windows of radius 30, that is, intervals around a given gene position including 30 genes to its left and 30 genes to its right in the ordered gene list. Averaging over these intervals for each gene in the list produces a smoothed mean expression profile. We generate the ordered gene list by first filtering gene products that share at least one association as inferred from the STRING Protein-Protein Interaction database with confidence scores of 800 or better [12]. The gene-list ordering clusters genes by their biological function as defined in the STRING database. The ordered gene list we use for this analysis is available as Table ST1 in supplementary information online. Ref. [12] explains the construction of the gene list in detail. We then apply this ordered list of genes to analyze gene expression data from micro-arrays or RNA-Seq experiments. Because the list clusters genes by attributed function, the running window averages expression levels over genes believed to participate in the same or similar biological functions.

One major problem in detecting differential gene expression in microarray or RNA-Seq experiments is with the high variance of the data, that can result from measurement noise or confounding variables that were not explicitly controlled for. Ref. [7] shows that Transcriptograms can reduce the variance of gene expression measurements and enhance the power of statistical tests when comparing gene expression levels between gene samples. We characterize the ordered genes by projecting onto the gene list selected biological Gene

Ontology (*GO*) terms or KEGG pathways, to associate regions of the list with key biological mechanisms.

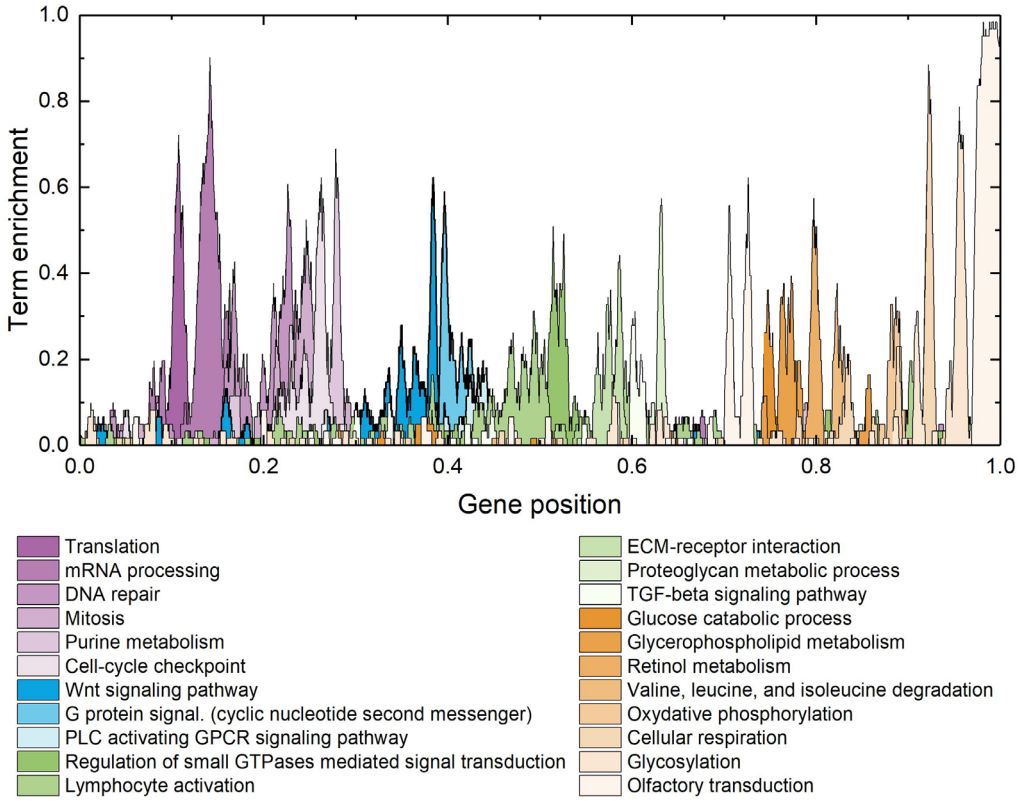

Fig. 1 Gene list and enrichment of terms related to critical biological functions as a function of position in the list. From left to right, in shades of purple, the list is enriched with genes associated with translation and mRNA processing then pathways linked to the cell cycle. Next, in shades of blue, genes associated with cell differentiation and, in shades of green, genes associated with immune response, cytokine production and interaction with the extra-cellular matrix (ECM). Finally, shades of orange denote genes associated with energy metabolism.

Fig. 1 shows term-enrichment profiles projected on the ordered list, obtained for selected KEGG pathways and Gene Ontology: Biological function terms (*GO:BP*). The gene list we use comprises 9684 genes, representing those genes whose products participate in at least one Protein-Protein Interaction (with a score of 800 or better) as listed in STRING. The horizontal axis (intrinsically numbered by gene position from 1 to 9684) has been rescaled to fit the interval [0,1]. At each position in the gene list, represented by the horizontal axis, we plot the fraction of genes within a window of radius 30 genes around that position associated with a specific term or pathway. A profile value near 1 means that almost all genes in that interval link to the term. Moving from left to right, we observe successive enrichment of terms associated with specific biological functions: at the far left, we see enrichment linked to RNA processing and metabolism, then enrichment related to the cell cycle, followed by cell differentiation, the actin cytoskeleton and immune systems. Further to the right, we see enrichment for signaling pathways associated with secretion, ECM receptors and finally, energy metabolism. Consequently, a running window

average of expression data over this ordering, averages the expression of genes linked to the same or similar biological functions.

Normalization check using Transcriptograms

Transcriptograms provide a powerful test for sample normalization by revealing undesired variable offsets in expression levels between samples. We ensure sample normalization we as follows: we plotted the Transcriptograms of radius 30 for the normalized data available in the GEO database (shown in Fig. S1 in the supplementary information online) and verified that each sample set shows offsets in its mean in relation to the other sample sets. We then re-normalized the expressions levels for each sample data set to set its mean expression to 1. Fig.S1 shows the resulting renormalized, single-sample Transcriptograms.

Relative Transcriptograms

We obtain *relative Transcriptograms* by dividing the Transcriptogram profile values at each point in the ordered gene list by the Transcriptogram profile value for a control sample at the same position in the list.

Differential Transcriptograms

We obtain *differential Transcriptograms* between **two time-series** by obtaining the relative Transcriptogram for the *WT* samples at a given time point w.r.t. time-matched Mock samples. In a time-series for differential Transcriptograms the control sample is different for each time point.

Term Enrichment

We determined term enrichment for the gene sets consisting of the genes in a given interval of the ordered gene list using the Term Enrichment Panther Service, on the Amigo 2 home page (http://amigo.geneontology.org/amigo) [17,18,19].

Covariance matrix

For each gene i from a gene set with N elements ($i = 1, \dots, N$) we define the *differential expression* $e_i(t)$ as:

$$e_i(t) = \frac{w_i(t)}{m_i(t)} ,$$

where t represents a time in the time series and $w_i(t)$ and $m_i(t)$ are the averages over replicates for the gene-expression values from, respectively, the WT or Mock transcriptomes' normalized datasets.

We define the *covariance matrix* C_{ij} as:

$$C_{ij} = \frac{1}{\sigma_i \sigma_j} \frac{1}{N_t} \sum_t (e_i(t) - \langle e_i \rangle)\big(e_j(t) - \langle e_j \rangle\big),$$

Where N_t is the number of the experiment time points (here $N_t = 11$) and $\langle e_i \rangle$ is the time average of the differential expression of the i-th gene,

$$\langle e_i \rangle = \frac{1}{N_t} \sum_t e_i(t),$$

and σ_i is the standard deviation calculated as:

$$\sigma_i = \sqrt{\frac{1}{N_t} \sum_t (e_i(t) - \langle e_i \rangle)^2} \quad .$$

Results

We start our analysis of expression profiles for WT and Mock samples at different times by generating relative Transcriptograms as described in the Statistical Methods section, taking as the control the Transcriptogram for the Mock sample at time $0\ h$ (which is the time of the first RNA harvest, at least 40 minutes after inoculation). Fig. 2 shows the relative Transcriptograms at different times for Mock (blue lines) and WT (red lines) samples. The relative Transcriptogram for the control expression levels (Mock samples at $0\ h$) appears as a black horizontal line. We also plot the relative Transcriptogram standard errors (due to the variance among replicates) for each point of the ordering: these errors are represented by gray, light red, and cyan shading around, respectively, the black, red and blue lines. The Transcriptogram's window average reduces the variance between replicates, so the error bars are barely visible.

Fig. 2 presents the relative Transcriptogram profiles at $0\ h, 24\ h$, and $48\ h$ after first RNA harvest. The top panel shows that at $t = 0\ h$ (approximately 40 minutes after the viral inoculation and washout at the start of the experiment) the Mock and WT samples are similar, although in some regions the errors do not overlap (around gene positions 0.41 or 0.53, for example), indicating that cells are already responding to viral infection at the time of first RNA sampling. Fig. 2 shows that at $24\ h$ and $48\ h$, the relative Transcriptogram profiles of both Mock and WT samples differ significantly from the Mock sample mean profile at $0\ h$. The direction of deviations at a given gene location at $48\ h$ is typically the same as that at $24\ h$, but of greater amplitude. To identify intervals in the relative Transcriptograms with significant expression

variations, we define and consecutively label contiguous regions along the horizontal axis with values larger than 9/5-fold changes in the WT Relative Transcriptogram values at $48\ h$ (labels A1, A2,...,A11). From the many terms enriching each region (see Methods section) we select a representative term as a label, based on the number of genes associated to that term in that region.

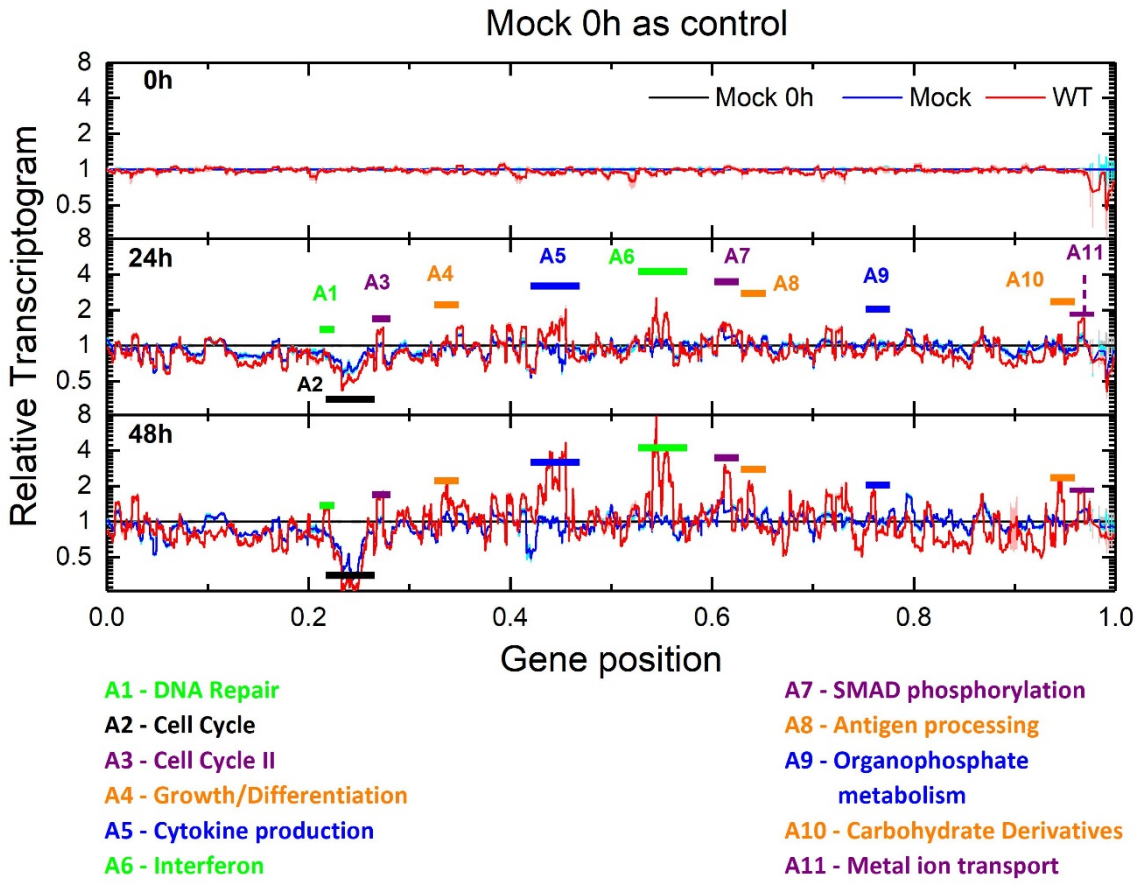

Fig. 2. Relative Transcriptograms of radius 30 for Mock and WT samples, using the Mock sample's expression at $t = 0\ h$ as the control. The labeled time is the experimental time after the first RNA harvest (over 40 minutes after inoculation). Vertical axes are on log_2 scale. Black horizontal lines represent the control sample (Mock) expression. Red and Blue lines represent the relative Transcriptograms for, respectively, WT and Mock samples. Gray, Light red and cyan shading indicate the standard errors of the respective relative Transcriptograms. We identify 11 intervals, indicated by the horizontal color bars, where the red line differs from the control by more than 9/5 at 48h.

Fig. S2 in the supplementary information online shows equivalent panels for all time points ($0\ h, 3\ h, 7\ h, 12\ h, 24\ h, 30\ h, 36\ h, 48\ h, 54\ h, 60\ h, 72\ h$). While our method does not seek to identify genes related to immune response, most of the bands of significant variance correspond to regions enriched with genes participating in pathways linked to the immune response. Genes linked to the cell-cycle I region (interval A2) are depressed in both Mock samples and WT samples, probably reflecting contact inhibition of proliferation in confluent *in vitro* cultures. However, Fig. S1 in the supplementary information online also shows that at $54\ h$ WT samples show some expression recovery of genes linked to the cell cycle. This recovery may reflect the onset of cell cycle after the death of some infected cells, which reduces contact

inhibition of proliferation, or another tissue-recover mechanism. Changes in expression across multiple functional bands of the relative Transcriptogram appear at $24\ h$. These bands stay fixed in width but increase in amplitude until $54\ h$, after which they slowly decrease in amplitude. For more details on the changes in each band, refer to Fig. S2 and movies SM1 and SM2 in the supplementary information online.

The important messages in Fig. 2 and S1 and the animation of the time changes of the relative Transcriptograms in movies SM1 and SM2 are that: i) major changes in band expression start after $12\ h$; ii) the bands of expression change in amplitude but not in width, reflecting their correspondence with changes in activity of specific biological mechanisms; and iii) gene expression in the control samples also changes in time, because cell state changes in culture conditions, even in the absence of infection.

To distinguish cell-culture effects, which affect both WT and control cultures, from the effects of infection, we considered time-matched mock expression profiles as controls for the WT expression profiles. We define the *differential* Transcriptogram profile as the ratio of the WT transcriptogram value at a given time and gene position to the matched-time Mock transcriptogram value at the same gene position. Using a time-matched control helps reduce the signal from tissue-culture effects common to both WT and control samples and accentuates specific differential infection effects. Differential profiles do not show changes in expression of cell-cycle-related genes, for example, since both control and WT expression change in the same way in time. Figure 3 shows differential profiles as violet lines, with the light violet shading showing the standard error. The horizontal black line shows the control differential expression profile for the Mock sample at the corresponding time. Fig. 3 presents the differential profiles (WT(t)/Mock(t)) at three time points, Fig. S3 in the supplementary information online presents the differential profiles for all time points and SM3 animates these time changes into a movie.

Differential Transcriptogram profiles show noticeable alterations after $12\ h$. As before, we to associate the most altered bands to biological functions and labeled used Panther to label them accordingly. As in Fig. 2, the altered bands remain constant in width but change in amplitude.

We next consider the time evolution of the differential expression (WT(t)/Mock(t)) of the 590 individual genes that participate in the 17 bands identified as significantly differentially expressed in the differential Transcriptograms in Fig. 3 (where we define a significant change as either larger than 9/5 or less than 5/9). Table ST2 in the supplementary information online presents these results as an Excel file containing the gene names and plots for each gene's relative and differential expression evolution, with brief information on each gene. Among these

590 genes, we found 219 for which the expression $w_i(t)$ at one or more time points t differed from the time-matched control $m_i(t)$ more than two fold (*i.e.* $w_i(t) > 2m_i(t)$ or $w_i(t) < 0.5\, m_i(t)$). Our significance limit is higher because here we are considering single-gene differential expression, rather than the differential Transcriptogram values, which are averages over the expression of neighboring genes in the list.

The Transcriptogram analyses identify 219 genes in differentially-expressed Transcriptogram bands which are also individually differentially expressed relative to their time-matched Mock samples. This gene set comprises the genes that respond more intensely to virus inoculation. To identify their associated biological functions, we used the Over Representation test by Panther, available on the Gene Ontology-Amigo home page (http://amigo.geneontology.org/amigo) to find the Reactome Pathways that enrich this set of 219 genes. Among others, we find that 51 of these genes participate in "Cytokine signaling in immune system," 36 participate in "Innate immune system," 18 participate in "Toll-like-receptor cascade," 15 participate in "Interleukin 4 and Interleukin 13 signaling" and 8 genes are participate in "TNFR2 non-canonical NF-κB´ pathways." Table ST3 in the supplementary information online gives the complete list of over-represented Reactome Pathways for the 219 genes (P<0.05, Buonferroni corrected). Of the 219 genes, 35 have not been classified as forming a representative set for some Reactome Pathway. Thus 84% of the 219 significantly variant genes participate in pathways either directly or indirectly involved in components of the immune response.

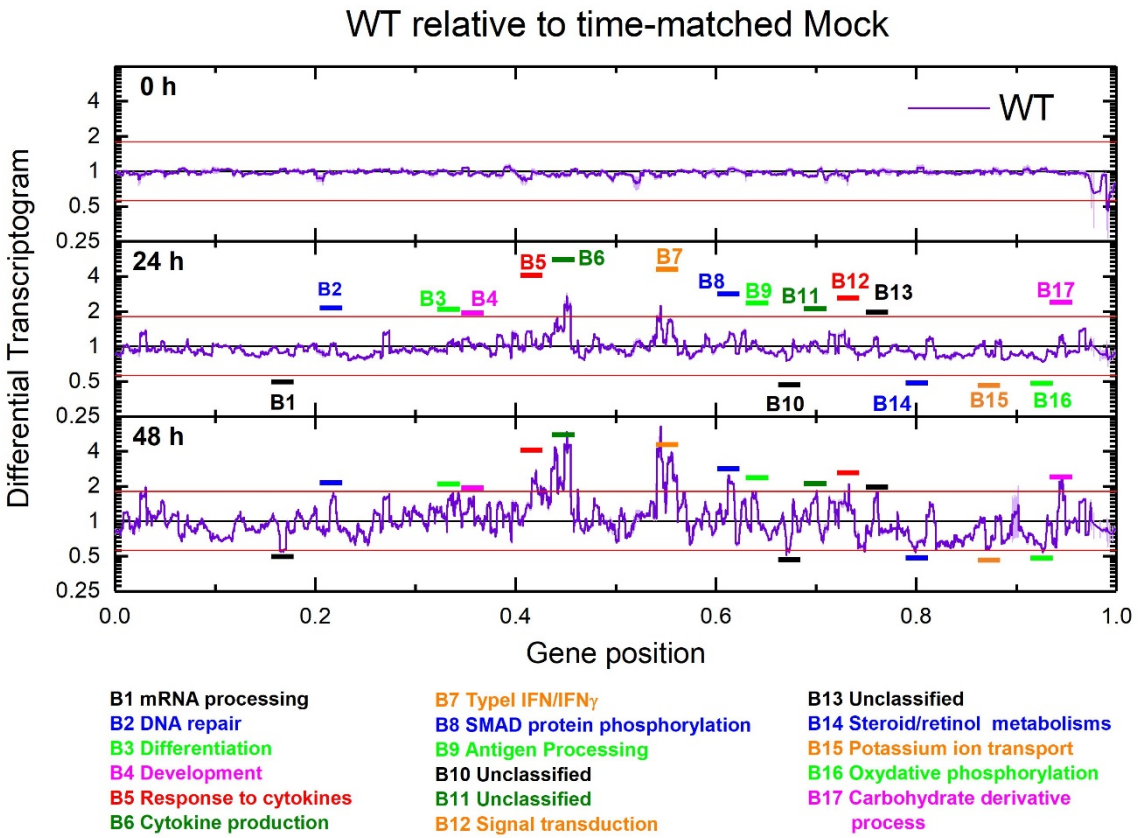

Fig. 3. . Differential Transcriptograms WT(t)/Mock(t) (radius 30). Time is time after the first RNA harvest. Vertical axes are on log_2 scale. Black horizontal lines represent the control sample Mock(t)/Mock(t). Violet lines are the differential Transcriptograms for WT(t)/Mock(t). Light violet shading indicates standard errors for WT Transcriptograms. We identify 17 bands where the violet line differs from the control more than 9/5-fold at 24 h. The horizontal red lines denote the 9/5-fold and 5/9-fold lines in all panels.

Proceeding with our top-down strategy, we look for temporal patterns in the differential expression of single genes (not Transcriptogram values) related to immune response by considering covariant gene sets for the 219 genes identified as presenting significantly different single-gene expression time series among the 590 genes that participate in the significantly different bands in the differential Transcriptograms in Fig.3. To find genes with similar patterns of temporal evolution, we first calculated the covariance matrix for the differential expression time series of the 219 genes (see Statistical methods section). When two genes have the same pattern of time change of differential expression, their temporal covariance approaches one. Using the covariance matrix, we order genes into covariance clusters (Fig. 4). We identify 3 large clusters, A, B and C. The time changes of differential expression of genes in cluster D are less strongly correlated than those in clusters A, B and C, so we have subdivided D into three more strongly covariant sub-clusters, D1, D2 and D3. As discussed previously, the ordering of the gene list is based on biological function attributed to the genes. The covariance matrix clusters genes by the similarity of their differential expression time series. The activation a pathway associated

with one biological function may lead to activation of genes associated with different biological functions, leading to temporal correlations in their differential expression patterns. Also, different genes within a pathway may activate with different time patterns, reducing their temporal covariance. Consequently, we do not expect the covariance matrix clusters to directly correspond to the differential transcriptogram bands in Fig.3. Instead, Clusters A to D3 are covariant gene sets for the genes which both Transcriptogram and single-gene analysis identified as significantly differentially expressed in the present experiment. Table ST2 in the supplementary materials online is an Excel file containing a separate worksheet for each of the 17 bands which the Transcriptogram analysis identified as significantly differentially expressed, as shown in Fig. 3. For each of these bands, Table ST2 presents plots of the relative and differential expression of the 219 genes, in different colors, depending on the time-series pattern they follow. All bands have genes which belong to all covariant clusters, showing that proximity in the ordered list does not correlate with covariant cluster identity.

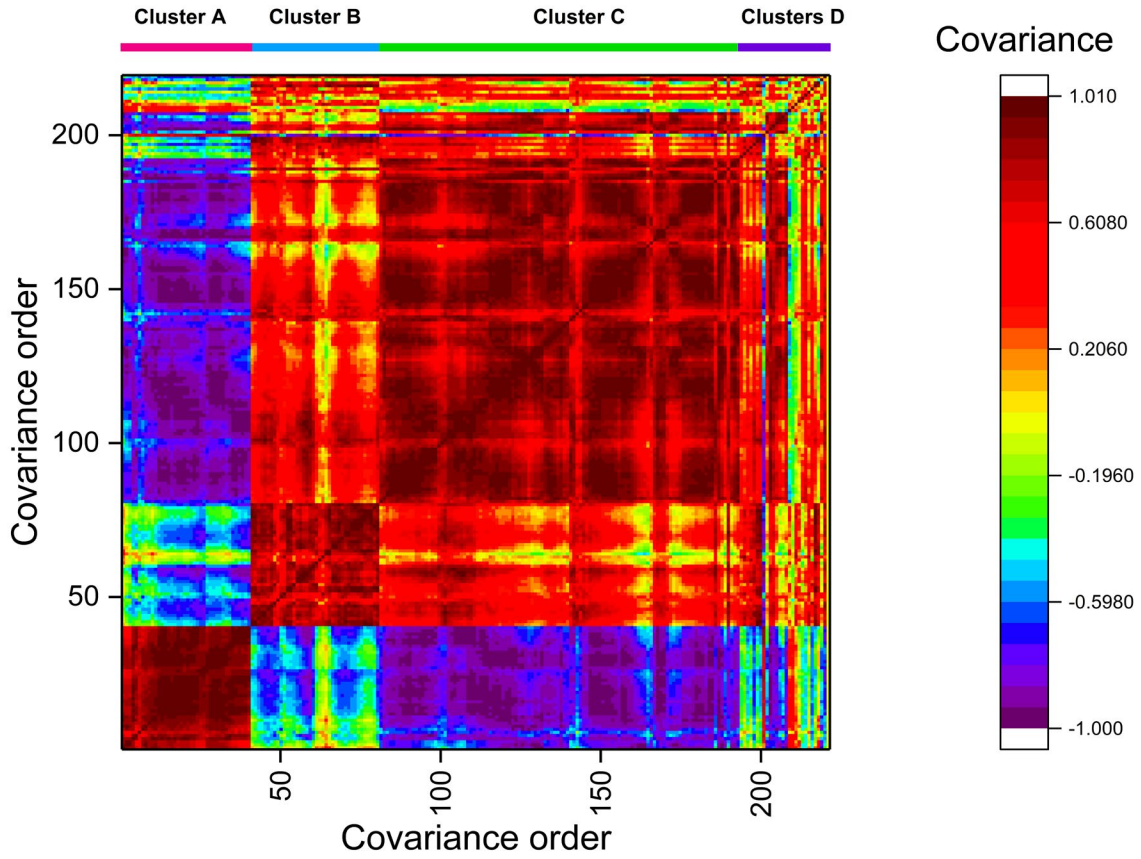

Fig. 4. Covariance matrix for the time changes of differential expression (time-matched control) for 219 genes selected from the significantly differentially expressed Transcriptogram bands identified in Fig. 3. We find 4 major covariant clusters. Table 1 lists the genes in each cluster.Fig.5 shows the time-series for the differential expression for each gene in each cluster and the averaged differential expression time-series for each cluster.

Table 1 List of genes in each covariant cluster (Fig. 4). The time-series of differential expression of genes in cluster D are less strongly covariant than those in Clusters A, B and C and form 3 distinct sub-clusters.

Cluster A	Cluster B	Cluster C			Cluster D1	Cluster D2	Cluster D3
ABCC3	ARNT2	ADAMTS3	IFIT5	SAMD9	CD1A	CCL5	CACHD1
AKR1C3	B3GAT1	AHR	IL18R1	SAMD9L	CXCR3	CFD	CER1
AKR7A2	BMP6	AR	IL18RAP	SERTAD1	KCNAB3	DHRS9	GLI1
ASB9	CCL17	ATF3	IL1A	SH2D1A	KCNF1	FNDC5	GNGT2
ATP6V0A1	CCL3	AXL	IL1B	SIRT1	KCNQ5	NID2	HSD17B6
ATPIF1	CCL4	AZIN1	IL8	SMAD7	KCNV2	PLEKHM2	INHA
BNIP3	CCR10	BACH2	INHBA	SMURF2	LBP	POP1	KCNA3
BNIP3L	CRYAA	BANP	IRAK2	STAG1		ZYG11A	KCNQ4
CAT	CTRC	BCL3	JUN	STON2			KL
CFTR	CYP17A1	BRIP1	KLHL20	SYT1			SRGN
COX7A2L	DNER	CCL2	KRT14	THRB			TMOD4
CPE	DUSP7	CCR1	LTB	TICAM1			TNFRSF1B
CYP2B6	EGR1	CCR7	MAN1A1	TLE4			
DEFB1	EGR3	CDK8	MAP3K1	TLR1			
ETFB	ELN	CHUK	MAP3K8	TNF			
HYAL3	ERRFI1	CNN1	MAP4K3	TNFAIP3			
KCND1	ESR1	CNTN6	MED13	TNFSF13B			
NDUFA2	FFAR2	CXCL5	MT1G	TNFSF15			
NDUFB2	FOS	CYR61	MT1H	TNFSF9			
NDUFB8	GDF6	DLL1	NEXN	TRAF1			
NEBL	GPR132	DUSP1	NFKB1	TRAF2			
NFIA	HNF4G	DUSP10	NFKBIA	TRIB1			
OBSL1	HSD11B2	DUSP16	NLRC5	TRIM21			
PAQR6	IRF8	DUSP2	NMI	TTBK2			
PNPLA4	JUNB	DUSP4	NR1D1	UBD			
RAB13	KISS1	DUSP5	NR1D2	UGCG			
RAB5B	LOX	EIF2AK3	NTNG2	UGT2B15			
RDH5	NPR1	FBXO6	PELI1	UNC13B			
ROPN1	NR0B1	FOSL1	PELI2	USP18			
S100A9	NUMBL	FOXO1	PML	USP25			
SCNN1A	OR1S2	GATA3	POU2F2	USP9X			
SPA17	OR2A7	HERC6	PPIF	ZBTB32			
SVIP	OR52K2	HLA-A	PPP1R15A				
TFF1	SERPINE1	HLA-B	PPP4R4				
TGFB1	SIRPB1	HLA-C	PRTN3				
THRA	SLC18A3	HLX	PTPRR				
TIMP1	SPN	HSD11B1	REL				
TLE2	THPO	HTR3B	RGS20				
UGT1A8	TWIST1	ID3	RIPK2				
UQCRC1	ZNF14	IFI16	RORA				

Fig. 5 presents the time series for the differential expression of each gene in each cluster, together with the cluster average of these values. Within clusters A, B, C, and D1, the genes have similar patterns of change of differential expression in time. Clusters D2 and D3 show more diverse temporal patterns of evolution. Fig. 6 summarizes the averaged temporal patterns of differential gene expression of the clusters.

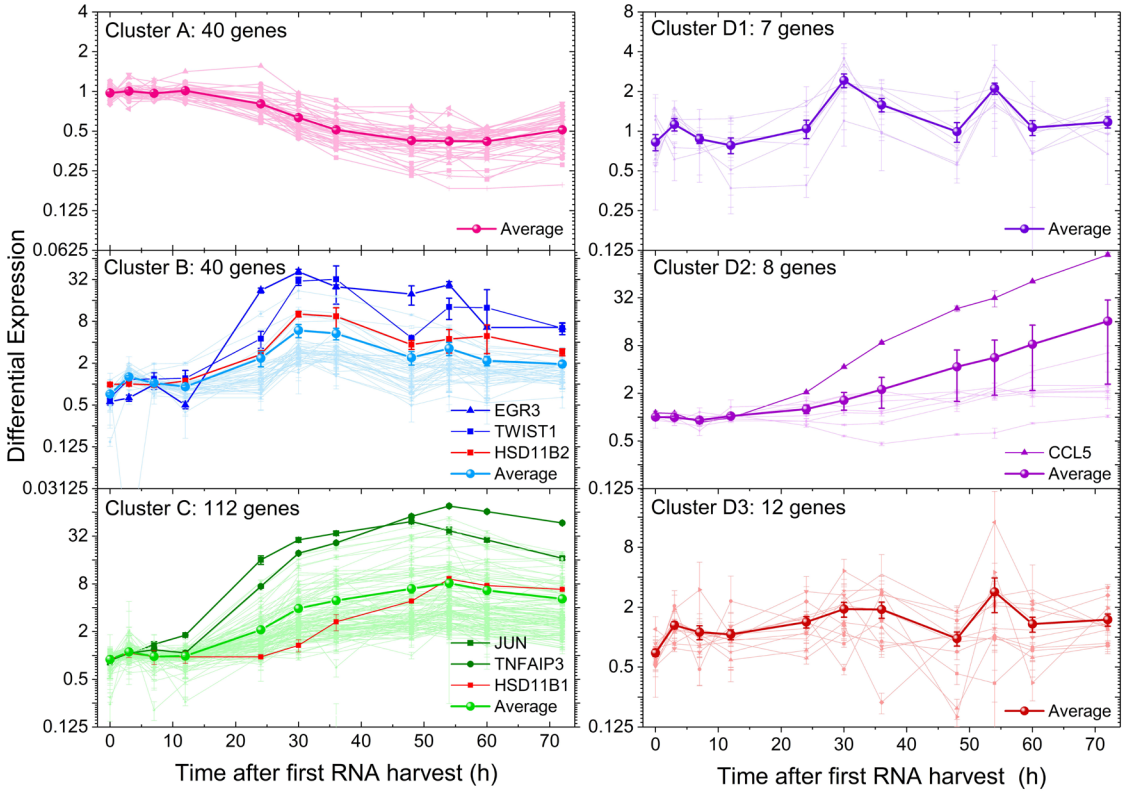

Fig. 5. Time evolution for the differential expression of the mean for all genes and for individual selected genes in covariant clusters A, B, C, D1, D2 and D3. Selected genes in each cluster are highlighted. The control for each gene at each time point is the Mock sample expression of the same gene at a matched time. Expression is for individual genes, not Transcriptogram averages over the neighbors in the ordered gene list. Clusters A (40 genes),B (40 genes), and C (112 genes) have highly covariant differential expression time series. Cluster D is subdivided in three sub-clusters: D1 (7 genes), D2 (8 genes) and D3 (12 genes). The time series in cluster D1 are more homogeneous than those in D2 and D3. The large symbols show the average differential expression vs time for each cluster. Table ST2, in the supplementary information online shows the individual relative expression time series for each gene in full detail.

Fig. 6 consolidates the results for the averaged cluster time series. Because clusters B, D1, and D3 have similar temporal dynamics, we merge them into a single cluster B' using an average weighted by their number of components. Fig. 6 shows three distinct time courses for elements of the cells´ response to virus inoculation. The vertical axes represent single gene differential expression in the WT samples w.t.r. the Mock samples. Genes in the consolidated cluster B' decrease ($P < 0.0052$) in expression relative to the control at $t = 0\ h$. After $12\ h$ changes in differential expression increase rapidly, with $36\ h$ and $54\ h$ having (local) maxima in the collective differential expression dynamics. To test the significance of the difference of the fold-change we estimated the P-value for each time-point using a two-tailed Welch test. We also

calculated P values to assess significant differences in mean differential expression at each time-point across clusters. We present these results in Table 2.

Fig. 6 shows the temporal pattern of response in the infected cells triggered by virus inoculation together with the time evolution of the viral titer, reproduced from Ref. [15], measured in units of PFU/ml (plaque forming units per milliliter) for 6 samples for each time point. We can observe that:

1) Viral titer initially decreases from $0\ h$ to $3\ h$, then increase rapidly from $3\ h$ to $36\ h$, decrease between $36\ h$ and $48\ h$, then increase to a small, but statistically significant second maximum at $54\ h$, and finally decrease from $54\ h$ to the end of the experiment. The first genes to respond significantly to viral inoculation belong to cluster B´. Cluster B' is the only cluster whose mean expression is significantly different from Mock sample expression before $12\ h$ (see Table 2).
2) In Cluster A (40 genes) average differential expression does not differ significantly from the control until $12\ h$. Between $12\ h$ and $54\ h$ its average differential expression decreases, reaching a minimum at $54\ h$. After $54\ h$ average differential expression increases until the end of the experiment but always remains less than 0.5. This cluster is enriched in genes involved in mitochondrial activity. Shi and collaborators showed that the SARS-CoV-1 protein designated *opening reading frame-9b* (ORF-9B) localizes to the outer mitochondrial membrane, manipulating host-cell mitochondria, and disturbing mitochondrial anti-viral signaling [20]. This interference could explain why Cluster A's mean differential expression moves opposite to the viral titer.
3) The genes in Cluster B´ (40+7+12 genes) have the richest temporal dynamics of mean differential expression. At $0\ h$, differential expression is already depressed in the WT relative to the control, indicating that some genes change their expression very rapidly w.r.t. the control, during the 40 min incubation time before RNA harvesting $(0\ h)$. Average differential expression increases between $0\ h$ and $7\ h$ then decreases until $12\ h$, then increases again, reaching a maximum at $30\ h$. It then decreases until $48\ h$, then increases to a second, more modest maximum, at $54\ h$ and finally decreases until the end of the experiment. Although the maximum at $54\ h$ is modest in amplitude, it is statistically significant and coincides with the minimum value of the average differential gene expression in Cluster A and the second peak in viral titer, suggesting that it reflects a real change in biological function. The term enrichment analysis for Cluster B showed that 36 of the 40 genes in this cluster participate in Gene Ontology (*GO*) terms linked to "response to

stimulus," with 13 specifically tagged as "response to cytokines." The great majority of the products of these genes localize either to the extracellular matrix, indicating signaling activity, or to the nucleus, indicating response to signaling. The other two clusters forming cluster B′, Cluster D1 (7 genes) contains genes associated with terms for ion transport, potassium included, while Cluster D2 (12 genes) contains a diverse spectrum of genes that all associate with the broad GO term "Immune Response."

4) In Cluster C (112 genes), after $12\,h$ mean differential expression increases monotonically to a maximum at $54\,h$ after which it decreases until the end of the experiment. This maximum at $54\,h$ coincides with the maximum at $54\,h$ in Cluster B′, the minimum in Cluster A, and the modest second peak in viral titer. Term enrichment analysis of the genes in Cluster C shows that the great majority of these genes have GO annotations involved in "immune response and signaling," including "production and regulation of cytokines" (58 genes), ``I-kappaB kinase/Nf-kappaB signaling′′ (19 genes),``response to hormone′′ (19 genes), and ``innate and adaptive immune response′′.
5) In Cluster D2 (8 genes) mean differential expression (Fig. 6 inset) increases monotonically after $12\,h$. The individual gene differential expression time series show that this increase is due to a strong monotonic increase of CCL5 differential expression(see Fig. 5, right-middle panel), a chemoattractant for blood monocytes, memory T-helper cells and eosinophils. The other genes present significant differential expression only at a few time points and their time series show modest differential expression as compared to CCL5. Also, their differential expression time series have different patterns: some monotonically increase, others are more stable in time, and a few first, then recover after $36\,h$.

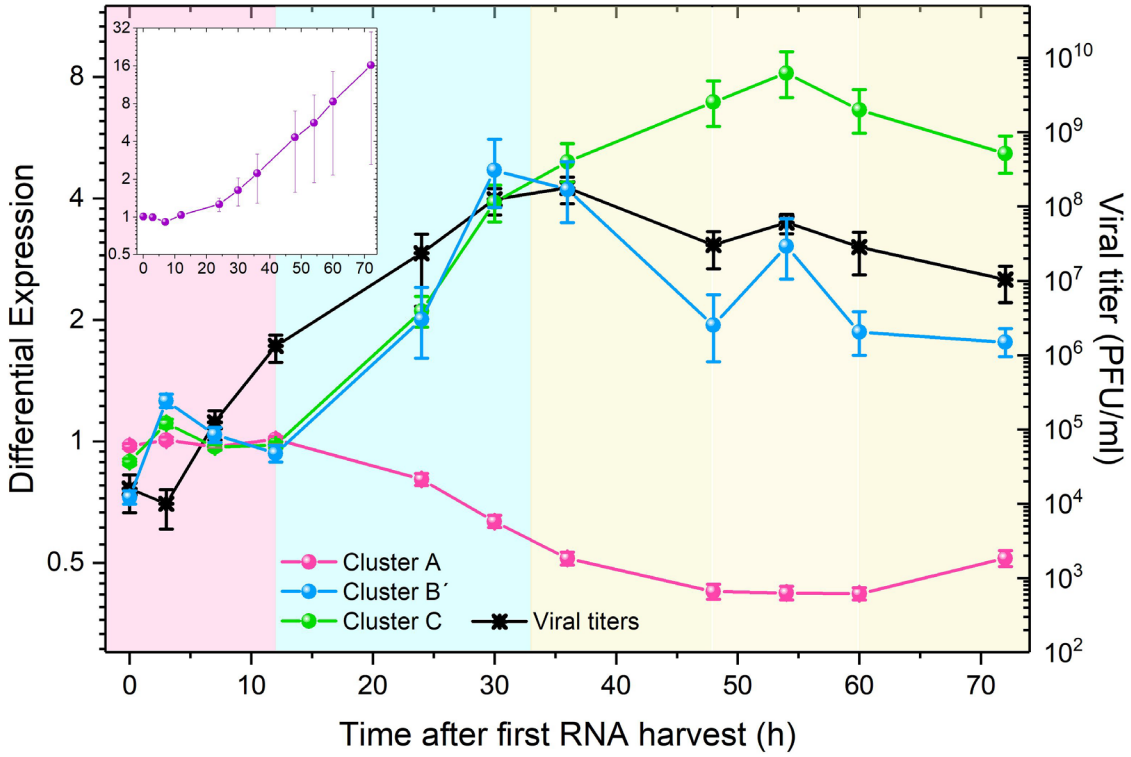

Fig. 6. Time evolution of viral titer [15] (right log_{10} axis) and average differential expression of the covariant gene clusters A, B′ and C (left log_2 axis). The inset shows the time evolution of cluster D2. The control for each time point is the time-matched Mock sample. We identify three main phases for the host-virus interaction in the cell cultures. In the first phase, denoted by a pale-pink background, clusters A and C have differential expression near 1, while cluster B′ differential expression moves opposite to the viral titer. In the second phase, denoted by a pale-blue background, average differential expression in cluster A decreases monotonically while average differential expression in both clusters B′ and C increase similarly, in parallel with increasing viral titer. In the third phase, denoted by a pale-yellow background, differential expression in clusters B′ and C diverge after the viral titer reaches its maximum, with Cluster B´ tracking viral titer evolution. After $54\ h$, however, viral titer and the average differential expression of clusters B´ and C decrease while average differential expression in cluster A increases.

Table 1. P values for the time series depicted in Fig. 6.

Time after first RNA harvest (h)	P-values					
	WT and Mock matched time samples			Between Clusters		
	Cluster A	Cluster B+D1+D3	Cluster C	Cluster A and B+D1+D3	Cluster A and Cluster C	Cluster B+D1+D3 and Cluster C
0	0.5759	0.0052	0.1272	<0.0001	<0.0001	<0.0001
3	0.0446	0.4778	0.6286	<0.0001	<0.0001	<0.0001
7	0.2360	0.0003	0.8210	<0.0001	<0.0001	<0.0001
12	0.9110	0.0654	0.1710	<0.0001	<0.0001	0.0908
24	0.0089	0.0021	0.0011	<0.0001	<0.0001	0.6809
30	<0.0001	0.0159	0.0007	<0.0001	<0.0001	0.5409
36	<0.0001	0.0045	<0.0001	<0.0001	<0.0001	0.4408
48	0.0003	0.0006	<0.0001	<0.0001	<0.0001	<0.0001
54	0.0008	<0.0001	0.0022	<0.0001	<0.0001	0.0023
60	0.0021	0.0126	0.0007	<0.0001	<0.0001	<0.0001
72	0.0004	0.0007	0.0001	<0.0001	<0.0001	<0.0001

In summary, we have identified three gene clusters (A, B´, and C) with distinct temporal evolution. Clusters B´ and C have the most distinctive patterns of temporal change, probably reflecting their specific functional roles during early infection. To illustrate the power of our method, we selected six genes from these clusters and analyzed their differential expression evolution. We then discuss their possible roles in the cellular response to virus inoculation.

We begin by considering the gene pair HSD11B1 (from cluster C) and HSD11B2 (from cluster B´). Both genes are linked to Cortisone-Cortisol balance. Cortisol is anti-inflammatory, is secreted by the adrenal gland, is present in plasma, and can be converted to inactive Cortisone by the enzyme 11-beta-hidroxysteroid dehydrogenase type 2 (HSD-2), the product of the HSD11B2 gene (For a review, see [21]). The time series for HSD11B2 presented in the middle-left panel of Fig. 5 (red line), shows that its differential expression in infected cells begins to increase after $12\ h$ up to $36\ h$ reaching a 32-fold change w.r.t. control, after which it gradually decreases to an 8-fold change . The reduction in anti-inflammatory Cortisol signaling between $12\ h$ and $36\ h$ probably enhances pro-inflammatory signaling in response to infection. HSD11B1 differential expression also starts increasing after $12\ h$, but peaks later, at about $54\ h$, as shown in the bottom-left panel of Fig. 5 (red line). 11-beta-hidroxysteroid dehydrogenase type 1 (HSD-1), the product of the gene HSD11B1, converts Cortisone back into Cortisol, possibly restoring anti-inflammatory signaling, after the initial pro-inflammatory response. The data shows the time evolution for the well-known interplay between Cortisone-Cortisol.

Fig. 5, bottom-left panel (olive green) highlights the differential expression evolution of JUN and TNFAIP3 from cluster B. JUN encodes c-Jun, a protein that participates in the transcription-factor complex "Activator Protein-1" (AP-1) that has complex context-dependent behaviors [22]. In epithelial cells AP-1 components (containing c-Jun) may participate in apoptosis or cell proliferation. JUN differential expression evolution seems to increase with viral titer with a few hours of delay. Cell cycle mean expression is depressed in both Mock and WT samples compared to Mock sample expression at $0\ h$ (the interval marked A2 in Fig. 2 and Fig. S1 in the supplemental information online), probably due to contact inhibition in the culture. However, Fig. S1 shows that cell-cycle gene expression recovers after $60\ h$. Both apoptosis and proliferation may occur in the infected culture. The observed time series for JUN differential expression may relate to these differences in cell cycle-related expression between WT and Mock samples.

TNFAIP3 encodes the protein A20, a negative regulator of the NF-κB protein complex. TNFAIP3 is thus a negative regulator of inflammation and is known to be rapidly induced after Toll-like receptors interact with a pathogen or respond to TNF-α or IL-1 cytokines [22]. The bottom-left panel in Fig. 5, shows a peak for differential expression of TNFAIP3 at $48\ h$, followed by monotonic decrease until the end of the experiment. Comparing the time-series for the differential expression of TNFAIP3 shown in the bottom-left panel in Fig. 5 with the viral titer evolution in Fig. 6, we may infer that TNFAIP3 follows the viral titer with about a $4\ h$ delay. This temporal relationship suggests that the anti-inflammatory response due to TNFAIP3 in the WT sample gradually decreases as the viral titer decreases.

Fig. 5, middle-left panel shows the dynamics of differential expression of individual genes in cluster B. We have highlighted in navy blue the expression of EGR3 and TWIST1, the two genes whose differential expression presents the largest fold changes in cluster B. TWIST1 negatively regulates the NF-κB protein complex. TWIST1 is thus anti-inflammatory [23]. The variation in the TWIST1 time series in the middle-left panel of Fig. 5 generally follows the viral titer evolution in Fig. 6. EGR3 is a zinc-finger transcription factor of the Early Growth Transcription family (EGR) that responds early to environmental stimuli to induce cell proliferation, differentiation, and immune responses [24]. In resting epithelial cells, EGR3 is usually weakly expressed, but a wide variety of extracellular signals such as cytokines and T-cell receptor (TCR) activation can promote EGR3 expression [24]. Fig. 5 middle-left panel shows that EGR3 differential expression increases after $12\ h$ and remains high, varying between 16- and 32-fold change from $20\ h$ to $54\ h$, then decreasing when the viral titer begins to decrease after $54\ h$ (Fig. 6). This correspondence suggests that the virus may activate EGR3 in epithelial cells. If the virus also promotes EGR3 in T cells, fibroblasts, and endothelial cells it could explain T-cell

anergy in SARS-CoV-1 and SARS-CoV-2 infection in T cells, since the co-activation of T-cell receptors by antigen and EGR3 may lead to T cell anergy [25,26]. Furthermore, EGR3 regulates fibrogenic responses in fibroblasts [27], and EGR3 may cause vascular disruption when active in vascular endothelial cells [28]. The infected lung contains epithelial cells, T-cells, fibroblasts, and endothelial cells, all of which express the ACE-2 receptor for SARS-CoV-1 and SARS-CoV-2 [29]. Both T-cell depletion due exhaustion or anergy, and fibrotic sequels have been reported in SARS-CoV-1 [30] and SARS-CoV-2 [31] patients. We wonder whether these effects on T-cells and fibroblasts may correlate with the activation of EGR3 by the virus. Also, since EGR3 activates VEGF in endothelial cells [28], its activation in infected cells may link to the endothelialitis, thrombosis, and angiogenesis reported in COVID-19 [6].

Conclusions and Perspectives

Transcriptogram analysis of microarray time series by Sims *et al.* [15] for SARS-CoV-1 infection of Calu3-2B4 cells, a human epithelial cell line, selected for ACE-2 expression [9] identifies three main gene sets with well-defined dynamics, summarized in Fig. 6. Differential expression profiles indicate that some cell responses (Cluster B´) begin very soon after inoculation, and that mitochondrial activity decreases until $54\ h$, then partially recovers. Considering that clusters B´ and C consist mostly of genes associated with immune response, our results show that the dynamics of these genes in response to viral inoculation follows two different time-evolution patterns. While Cluster B´ consists mainly of genes related to innate immune response, Cluster C comprises genes related to both innate and adaptive immune responses. Because mathematical models usually consider variables that aggregate the effect of multiple genes into broad representations of classes of biological mechanisms or pathways, these mean differential expression time series can serve as direct validation data for mathematical models of epithelial-cell responses to SARS-CoV-1 infection. Beyond pro/anti-inflammatory signaling via, for example, the negative regulation of NF-κB complex by TNFAIP3 and the interplay between HSD11B1 and HSDB11B2 differential expression, the genes in each cluster suggest that the response to viral inoculation also includes regulation of apoptosis and proliferation, via JUN, and has secondary effects on cell differentiation (with different possible outcomes, depending on the cell type), via EGR3. These effects follow the temporal patterns of either Cluster B´ or Cluster C suggesting coordinated patters of cellular responses. Because the Transcriptogram analysis selects genes most functionally relevant to the specific behavioral changes in a particular experiment, we could identify the correlated responses of genes annotated to different pathways or GO terms (which would be hard to identify if we conducted

correlation analyses of the temporal expression changes of all genes at once). The differential Transcriptogram, by identifying differential expression bands of functionally related genes, greatly reduces the number of "genes of interest" making their detailed temporal analysis practical.

Because gene expression changes in control samples in cell culture as well as in infected samples, using time-matched gene expression controls is critical to distinguish cell-culture effects from infection effects.

Our analysis identified six genes which we analyzed in more detail. We chose four because they had time series with the largest fold-changes in their clusters: JUN and TNFAIP3 from cluster C and TWIST1 and EGR3 from Cluster B´. We discussed their roles in the anti-viral immune response in detail in the last section.

EGR3 activation relates to the virus using ACE-2 to invade the cell. EGR3 activation may explain a number of symptoms in patients with severe responses to SARS-CoV-1 and SARS-CoV-2 infection. Other cell types in the infected lung may be infected by the virus: in particular, if EGR3 activation also occurs in infected T-cells, it could explain T-cell anergy (against viral antigens), in infected fibroblasts, it could link to observed fibrosis, and in infected endothelial cells, it could explain the endothelialitis, thrombosis, and angiogenesis reported in COVID-19.

Our analysis also identified as differentially expressed two genes from a well-known feed-back loop which regulates Cortisol-Cortisone balance. Infection perturbs the resting Cortisone-Cortisol homeostasis and we would expect that each gene would follow a different time course in response to infection. We find that differential expression of the proinflammatory HSD11B2 follows Cluster B´ and peaks with viral titer, while the anti-inflammatory HSD11B1 follows Cluster C and peaks later. We will examine the remaining 213 genes iidentified as signifiant by our Transcriptogram analsysis in future work

Our Transcriptogram method hierarchically prioritizes groups of differentially-expressed genes by first considering those bands in the ordered list with the most altered expression w.r.t. to the control. This filtering reduces the number of genes of interest to a tractable set and suggests shared mechanistic functions for the observed gene expression patterns. These gene sets are defined by the data directly, not by reference to previously-defined pathways or biological functions. We could apply the same methodology to identify functional differences between cell-culture responses to SARS-CoV-1 infection between male-derived and female derived cells or between adult-derived and juvenile-derived cells (to identify sex-linked and age-linked changes in response pattern). The same methods could identify critical differences in cell responses to SARS-CoV-1 and SARS-CoV-2 infection or among responses to infection by other respiratory viruses. We could also study differences in response between cells

derived from different possible loci of infection (nasal, throat, bronchial, alveolar, heart, kidney), or to compare infection responses between classical cell culture and organoids, between organoids derived from different donors, or between different initial infection intensities.

Acknowledgement

This work has received support from Brazilian agencies CNPq and CAPES. JAG acknowledges support from the Falk Medical Research Trust Catalyst Program and the US National Institutes of Health, grants U01 GM111243, R01 GM076692 and R01 GM077138.

Authors Contributions

RdA and GLT ran the analyses. RdA, GLT, and JAG designed and wrote the paper.

Corresponding Author

Rita de Almeida, rita@if.ufrgs.br.

References

1 Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science. 2003;302:276-278.

2 Arabi YM, Balkhy HH, Hayden FG, Boucharma A, Luke T, Baillie JK, Al-Ornari A, Hajeer AH, Senga M, Denison MR, et al. Middle East Respiratory Syndrome. The New England Journal of Medicine. 2017;376(6):584-594.

3 Kim JM, Chung YS, Jo HJ, Nam-Joo L, Kim MS, Woo SH, Park S, Kim JW, Kim HM, Han MG. Identification of Coronavirus Isolated from a Patient in Korea. 2020:3-7.

4 Manjili RH, Zarei M, Habibi M, Manjili MH. COVID-19 as an Acute Inflammatory Disease. The Journal of Immunology. 2020 May 25:2000413.

5 Wu Z, McGoorgan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. Journal of the American Medical Association. 2020:1239-1242.

6 Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, Vanstapel A, Werlein C, Stark H, Tzankov A, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. The new england journal of medicine. 2020.

7 da Silva SRM, Perrone GC, Dinis JM, de Almeida MC. Reproducibility enhancement and differential expression of non predefined functional gene sets in human genome. BMC Genomics. 2014:1181.

8 Ferrareze PAG, Streit RSA, dos Santos PR, dos Santos FM, de Almeida RMC, Schrank A, Kmetzsch L, Vainstein MH, Staats CC. Transcriptional Analysis Allows Genome

Reannotation and Reveals that Cryptococcus gatii CGII undergoes Nutrient Restriction During Infection. Microorganisms. 2017:49.

9 Rybarczyk-Filho L, Castro MAA, Dalmolin RJ, Moreira JCF, Brunnet LG, de Almeida RMC. Towards a genome-wide transcriptogram: the Saccharomyces cerevisiae case. Nucleic Acids Research. 2011:3005-3016.

10 Miotto YE, da Costa T, de Oliveira BH, Guzman F, Margis R, de Almeida RMC, Offringa R, Maraschin FdS. Identification of root transcriptional responses to shoot illumination in Arabidopsis thaliana. Plant molecular biology. 2019;101:487-498.

11 Cadavid IC, Guzman F, Luisa AdOB, Margis R. Transxriptomic ando post-transcriptional analyses of two soybean cultivars under salt stress. Molecular Biology Reports. 2020;accepted for publication.

12 de Almeida RMC, Clendenon SG, Richards WG, Boedigheimer M, Damore M, Rossetti S, Harris PC, Herbert BS, Xu WM, Wandiger-Ness A, et al. Transcriptome analysis reveals manifold mechanisms of cyst development in ADPKD. Human Genomics. 2016:37.

13 Reis CF, de Souza ID, Morais DAA, Oliveira RAC, Imparato DO, de Almeida RMC, Dalmolin RJS. Systems Biology-Based Analysis Indicates Global Transcriptional Impairment in Lead-Treated Human Neural Progenitor Cells. Frontiers in Genetics. 2019;10:791.

14 Morais DAA, de Almeida RMC, Dalmolin RJS. Transcriptogramer: an R/bioconductor package for transcriptional analysis based on canonical protein-protein interaction data. Bioinformatics. 2019:2875-2876.

15 Sims AC, Tilton SC, Menachery VD, Gralinsky E, Schäfer A, Matzke MM, Webb-Robertson BJM, Chang J, Luna ML, Long CE, et al. Release of Severe Acute Respiratory Syndrome Coronavirus Nuclear Import Block Enhances Host Transcription in Human Lung Cells. Journal of Virology. 2013;87(7):3885-3902.

16 Bacallao R, Clendenon SG, de Almeida MC, Glazier JA, inventors. TARGETING cGMP-RELATED PHOSPHODIESTERASES TO REDUCE CYST FORMATION IN CYSTIC KIDNEY DISEASE, AND RELATED MATERIALS AND METHODS. 2018 March 22. US2018/0078559 A1.

17 Ashburner M, Ball CA, Blake JA, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics. 2000;25(1):25-29.

18 The Gene Intology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research. 2019;47(D1):D330-D338.

19 Mi H, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research. 2019;47(D1):D419-D426.

20 Shi CS, Qi HY, Boularan C, Huang NN, Abu-Asab M, Shelhamer JH, Kehrl H. SARS-CoV ORF-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. Journal of Immunology. 2014:3080-3089.

21 Miller WL, Auchus RJ. The Molecular Biology, Biochemistry, and Physiologyof Human Steroidogenesis and Its Disorders. Endocrine Reviews. 2011:81-151.

22 Momtazi G, Lambrecht BN, Naranjo JR, Schock BC. Regulators of A20 (TNFAIP3): new drug-able targets in inflammation. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2019:L456-L469.

23 Šošić D, Richardson JA, Yu K, Ornitz DM, Olson EN. Twist Regulates Cytokine Gene Expression Through a Negative Feedback Loop That Represses NF-kappaB Activity. Cell. 2003:169-180.

24 Kumbrink J, Kirsch KH, Johnson JP. EGR1, EGR2, and EGR3 activate the expression of their coregulator NAB2 establishing a negative feedback loop in cells of neuroectodermal and epithelial origin. Journal of Cell Biochemistry. 2010:207-211.

25 Safford M, Collins S, Lutz MA, Allen A, Huang CT, Kowalski J, Blackford A, Horton MR, Drake C, Schwartz RH, et al. Egr-2 and Egr-3 Are Negative Regulators of T Cell Activation. Nature Immunology. 2005:472-80.

26 Li S, Tizong M, Sebastian M, Ghaffari E, Liu M, Symonds ALJ, Wang P. The Transcription Factors Egr2 and Egr3 Are Essential for the Control of Inflammation and Antigen-Induced Proliferation of B and T Cells. Immunity. 2012:685-696.

27 Fang F, Shangguan AJ, Kelly K, Wei J, Gruner K, Ye B, Wang W, Bhattachayya S, Hinchcliff ME, Tourtellote WG, et al. Early Growth Response 3 (Egr-3) Is Induced by Transforming Growth Factor-β and Regulates Fibrogenic Responses. The American Journal of Pathology. 2013:1197-1208.

28 Suehiro Ji, Hamakubo T, Kodama T, Aird WC, Minami T. Vascular Endothelial Growth Factor Activation of Endothelial Cells Is Mediated by Early Growth response-3. Blood. 2010 March 25:2520-2532.

29 Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, Li T, Chen Q. High expression of ACE2 receptor of 2019-nCoV on the. International Journal of Oral Science. 2020:8.

30 Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, Chen L, Li M, Liu Y, Wang G, et al. Reduction and Functional Exhaustion Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). Frontiers in Immunology. 2020:827.

31 Liang Y, Wang ML, Chien CS, Yarmishyn A, Yang YP, Lai WY, Luo YH, Lin YT, Chen YJ, Chang PC, et al. Highlight of Immune Pathogenic Response and Hematopathologic Effect in SARS-CoV, MERS-CoV, and SARS-Cov-2 Infection. Fontiers in Immunology. 2020:1022.

32 Richardson S, Hirsch JS, Narashimhan M, Crawford JM, McGinn T, Davidson KW, Northwell Covid-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020.

33 Singh S, Chowdhry M, Chatarjee A, Khan A. Gender-Based Disparities in COVI D-19 P atient Outcomes: A Propensity-matched Analy. medarXiv [Internet]. 2020 April. Available from: https://doi.org/10.1101/2020.04.24.20079046.