



OPEN

## Novel interferon-sensitive genes unveiled by correlation-driven gene selection and systems biology

Cristina Cheroni<sup>1,5,7</sup>, Lara Manganaro<sup>1,7</sup>, Lorena Donnici<sup>1</sup>, Valeria Bevilacqua<sup>1</sup>, Raoul J. P. Bonnal<sup>2,5,6</sup>, Riccardo L. Rossi<sup>3,8</sup>✉ & Raffaele De Francesco<sup>1,4,8</sup>✉

Interferons (IFNs) are key cytokines involved in alerting the immune system to viral infection. After IFN stimulation, cellular transcriptional profile critically changes, leading to the expression of several IFN stimulated genes (ISGs) that exert a wide variety of antiviral activities. Despite many ISGs have been already identified, a comprehensive network of coding and non-coding genes with a central role in IFN-response still needs to be elucidated. We performed a global RNA-Seq transcriptome profile of the HCV permissive human hepatoma cell line Huh7.5 and its parental cell line Huh7, upon IFN treatment, to define a network of genes whose coordinated modulation plays a central role in IFN-response. Our study adds molecular actors, coding and non-coding genes, to the complex molecular network underlying IFN-response and shows how systems biology approaches, such as correlation networks, network's topology and gene ontology analyses can be leveraged to this aim.

The hepatitis C virus (HCV) is a positive-sense single-stranded RNA virus belonging to the family of *Flaviviridae*. HCV is the causative agent of hepatitis C and the major cause of hepatocellular carcinoma. HCV infection potently stimulates the production of endogenous IFNs, which leads to ISG up-regulation in infected hepatocytes<sup>1</sup>.

The study of HCV in the laboratory has been particularly challenging over the years because of the narrow host range (it infects only humans and chimpanzees) and due to the paucity of cell lines permissive to HCV replication. Most of the cell systems tested over the years, in fact, did not support viral replication. A Huh7 derived clone termed Huh7.5 was found to be susceptible to HCV infection and highly permissive for HCV replication. This cell line was rapidly established as a robust and widely used in vitro infection model for HCV. Subsequent studies revealed that Huh7.5 permissiveness to HCV replication could possibly be a consequence of a single point mutation in the dsRNA sensor retinoic acid-inducible gene-I (RIG-I)<sup>2</sup>. RIG-I is a pattern recognition receptor that senses non-self viral nucleic acids. Specifically RIG-I recognizes pathogen-associated molecular patterns (PAMPs) characterized by a short 5' triphosphate uncapped double stranded or single stranded RNA. After binding to its PAMP, RIG-I starts a signaling cascade that leads to the activation of transcription factors IRF3 and NF- $\kappa$ B. These two factors activate the transcription of type I and III interferons as well as other antiviral genes that induce a general antiviral state of the infected cells<sup>3</sup>. Notwithstanding the possible molecular defects in the RIG-I sensor, however, treatment of Huh7.5 with type I IFNs potently inhibit replication of sub genomic RNA HCV replicons (reviewed in<sup>4</sup>).

Interferon is a class of primordial cytokines that plays a crucial role in the cellular response to the presence of viral pathogens. There are three classes of IFN, type I, II and III and each class comprises different members. Type I IFNs are constituted by IFN- $\beta$ , IFN- $\epsilon$ , IFN- $\omega$  and 13 different subtypes of IFN- $\alpha$ , type II IFN group is composed by only one member named IFN- $\gamma$  and type III IFN family comprises IFN- $\lambda$ 1, IFN- $\lambda$ 2 IFN- $\lambda$ 3 and IFN- $\lambda$ 4. All IFNs signal through the JAK-STAT pathway, type I and type III IFN share similar antiviral functions, while the type II IFN functions as a pro-inflammatory and immunomodulatory molecule. Type I and III IFN

<sup>1</sup>Virology, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", 20122 Milan, Italy. <sup>2</sup>Integrative Biology, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", 20122 Milan, Italy. <sup>3</sup>Bioinformatics, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", 20122 Milan, Italy. <sup>4</sup>Department of Pharmacological and Biomolecular Sciences (DiSFeB), University of Milan, Milan, Italy. <sup>5</sup>Present address: High Definition Disease Modelling Lab, Stem Cell and Organoid Epigenetics, Department of Oncology and Hemato-Oncology, IEO, European Institute of Oncology, IRCCS, University of Milan, Milan, Italy. <sup>6</sup>Present address: FIRC Institute of Molecular Oncology (IFOM), 20139 Milan, Italy. <sup>7</sup>These authors contributed equally: Cristina Cheroni and Lara Manganaro. <sup>8</sup>These authors jointly supervised this work: Riccardo L. Rossi and Raffaele De Francesco. ✉email: rossi@ingm.org; defrancesco@ingm.org

engage two different heterodimeric receptors: the type I IFN receptor (IFNAR) is composed of subunits IFNAR1 and IFNAR2 while type III IFNs receptor (IFNLR), comprises of IFNLR1 (also termed IL28R $\alpha$ ) and IL10R $\beta$ . Upon interaction with its receptor, both type I and III IFN activate the JAK-STAT signaling pathway, leading to the phosphorylation of STAT1 and STAT2 and interaction with IRF9 protein to form a trimeric complex named transcription factor complex ISGF3<sup>5</sup>. The complex migrates into the nucleus and binds to the IFN-stimulated response elements (ISREs) in the promoter regions of a set of genes, overall defined as interferon-stimulated genes (ISGs) that are responsible for the antiviral state of the cell. The binding of ISGF3 complex to ISREs elements induces the transcription of ISGs and initiates the cellular antiviral response.

Despite binding to different receptors, type I and III IFN trigger the same cascade of intracellular events, therefore having a similar effect on cell physiology<sup>5</sup>.

However, the potency and the kinetics of the antiviral response are different. Indeed, type III IFN response is delayed and less potent compared to type I IFN induced, but more sustained over time. In addition, the expression pattern of IFNAR and IFNLR is distinct, with IFNAR being expressed on the surface of all cell types, while IFNLR has been observed to be expressed mainly by epithelial cells and neutrophils. IFNLR has also been shown to be expressed by human hepatocytes and that HCV seems to preferentially stimulate IFN- $\lambda$  production rather than type I IFNs. Of note, IFNL locus polymorphisms are associated with different clinical outcomes during hepatitis C virus (HCV) infection suggesting the importance of this molecule in the response and pathogenesis related to HCV infection<sup>6–8</sup>.

Long non-coding RNAs are a class of RNA molecules longer than 200 nucleotides that are transcribed, but lack the ability to code for a protein product. Given their increasingly described role in the transcriptional and post-transcriptional regulation of gene expression, long non-coding RNAs have been emerging as a new layer of biological complexity<sup>9</sup>. While a canonical set of protein-coding genes acting as interferon stimulated genes is well-known<sup>10</sup>, a comprehensive view on the network of protein coding and long non-coding genes with a central role in IFN response still needs to be elucidated.

Although recently developed therapy regimens have strongly increased the cure rate, infection by hepatitis C is still a major cause of chronic liver disease<sup>11</sup>. In hepatocytes, the infection by HCV induces an interferon response that is aimed at suppressing viral replication. In fact, in cellular models, treatment with both class I and class III interferon molecules is able to suppress HCV replication and several protein-coding genes have been reported to mediate this phenomenon, thus acting as HCV restriction factors<sup>12</sup>. We chose to use the Huh7.5 cell line along with the parental Huh7 because of the higher permissiveness to HCV infection of the first, and in order to focus also on downstream RIG-1 mediated signaling. So we performed RNAseq on Huh7.5 in the presence or in the absence of IFN treatment including also the parental clone Huh7 in the same conditions.

Each gene is estimated to interact with at least four other genes in the actualization of regulatory networks<sup>13</sup> and to have functional relevance in many biological functions<sup>14</sup>. Uncovering correlations among genes is instrumental in deciphering the regulatory rationales of signal transduction pathways. Adopting higher level approaches able to analyze correlations as a whole can bypass the intrinsic limitations of the classical differential expression relying on arbitrary significance or expression cut-offs<sup>15</sup>. Weighted Gene Correlation Network Analysis (WGCNA) provides a system perspective of gene expression by grouping genes into modules selected according to correlation with a relevant phenotypic or molecular trait; this allows to find and integrate transcriptional patterns without using arbitrary selection cutoffs<sup>16</sup>. Using WGCNA it is possible to correlate traditional gene expression with any external measurement obtained from the same samples set: this cross-domain data integration leverages the discovery of otherwise hidden modules of highly correlated genes that can be used as new candidates or to unveil subtle molecular or biochemical mechanisms<sup>17,18</sup>.

In order to obtain a genome-wide view of the network of coding and non-coding RNA genes in IFN treated hepatoma cells and to identify possible novel effectors of IFN antiviral activity we applied to our set of RNA sequencing data a systems biology approach modeled on independent phenotypical pattern of the two well-known IFN sensitive genes<sup>19</sup> individually measured in the populations of IFN-treated and untreated hepatoma cell lines. After the identification of differentially expressed (DE) genes of IFN treated and untreated Huh7 and Huh7.5 cells, we constructed a weighted gene co-expression network including both coding and non-coding molecules, gene modules and sub-networks were identified as significantly related to IFN-treatment as a whole. Integrating topological network metrics and gene ontology terms we identified 42 novel hub genes, including 7 long noncoding RNAs (lncRNAs) and 7 pseudogenes, in the modules associated with IFN-response. Notably, putative binding sites for transcription factors involved in the IFN-response (such as IRF9 and STAT1) were also predicted for most of the genes included in the selected modules.

## Results

**IFN treated cells show a robust over expression of hundreds of genes.** In order to decipher the genome-wide landscape of the network of protein coding and long-non coding genes mediating interferon antiviral activity in hepatoma cell lines, we profiled by RNAseq the transcriptome of Huh7 and Huh7.5 cells treated with interferon alpha (IFN $\alpha$ ), beta (IFN $\beta$ ) or lambda (IFN $\lambda$ ) for 8 h. Each treatment was compared to respective mock-treated cells in 6 pairwise comparisons, for a total of 43 samples classified in 12 experimental conditions (Table 1). After gene expression quantification in read counts, dimensionality reduction by principal component analysis (PCA) was applied on the complete dataset: samples nicely segregated according to cell line identity and treatment with no outliers, as shown by PCA whose first two components are able to explain more than 90% of the entire dataset variability (Supplementary Figure 1).

Gene expression was robustly modulated by IFN $\alpha$  and IFN $\beta$  treatments, with IFN $\lambda$  inducing a milder perturbation (Fig. 1a) as expected. Differential expression analysis using 0.05 false discovery rate (FDR) and a twofold change (FC) cut-offs showed hundreds of differentially expressed genes, mostly up-regulated by the treatment

Cell line	Treatment (# of samples)	Control (# of samples)
Huh7	IFN alpha (4)	Mock (3)
	IFN beta (3)	Mock (3)
	IFN gamma (4)	Mock (4)
Huh7.5	IFN alpha (4)	Mock (4)
	IFN beta (3)	Mock (3)
	IFN gamma (4)	Mock (4)

**Table 1.** Experimental design. The 6 biological comparisons (3 treatments in each cell line) are described with the number of biological replicates (from 3 to 4) for each condition in parenthesis.

(Fig. 1a and Supplementary Figure 1). Specifically, the following number of genes were retrieved as significantly differentially expressed (DE) in treated compared to mock-treated cells: 320 for Huh7-IFN $\alpha$ ; 263 for Huh7-IFN $\beta$ ; 79 for Huh7-IFN $\lambda$ ; 252 for Huh7.5-IFN $\alpha$ ; 302 for Huh7.5-IFN $\beta$ ; 78 for Huh7.5-IFN $\lambda$  (Fig. 1a).

Comparing results (numbers of DE genes) of treatments with different IFNs, the fraction of DE genes shared by all contrasts (i.e. observed in all cases) was about 20% (23 and 21.9 in Huh7 and Huh7.5 cells respectively) with IFN $\lambda$ -modulated genes completely included in the overlapping region for both Huh7 and Huh7.5 (Fig. 1b,c). A group of 185 genes was consistently modulated by type I IFNs and in both cell lines (Fig. 1d): these genes are all redundantly involved in IFN-related functions, as confirmed by a scan of the biological processes with Gene Ontology, where mostly high level terms such as “immune system process”, “response to stimulus”, “defense to response”, “innate immune response” and “cytokine production” were overrepresented (not shown).

Taking into account the differentially expressed genes after IFN treatments (Fig. 1a), we assessed the functional relevance of cell lines specific subsets of these genes. Comparing genes perturbed by the IFN treatment in Huh7 and Huh7.5, we found that genes that were differentially expressed in both Huh7 and Huh7.5 cell lines (the Venn overlaps and middle panels in Supplementary Figure 2a and b) were the only ones triggering relevant ontologies and pathway enrichments, for both IFN $\alpha$  and IFN $\beta$ . Genes that were perturbed exclusively in Huh7 cells enriched a few general terms, while genes perturbed exclusively in Huh7.5 cells did not trigger any significant enrichment. This is an indication that the two cell lines appeared functionally homogeneous according to ontologies and pathways’ enrichment analyses after IFN treatments.

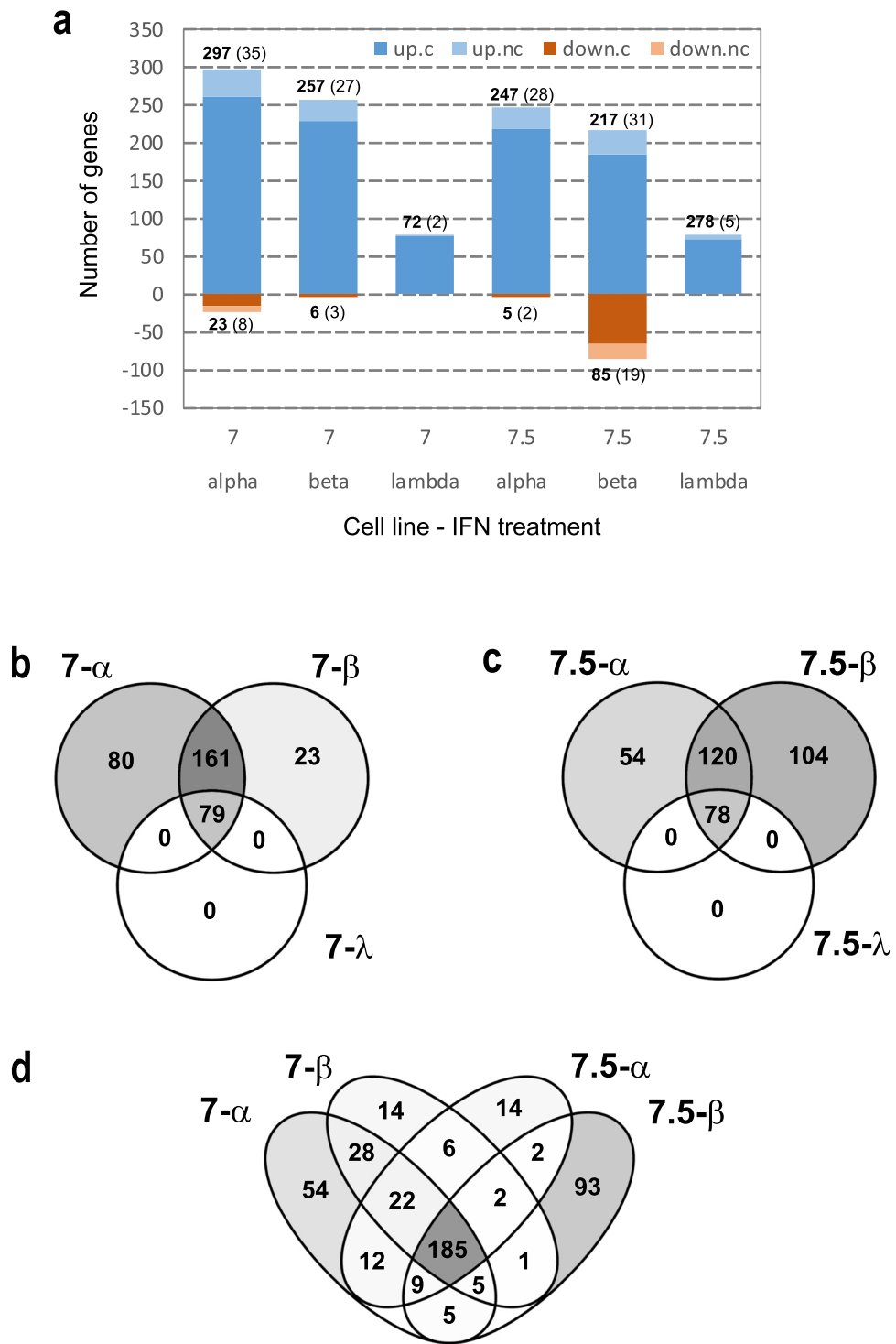
Most DE genes were protein-coding, but also some non-coding RNAs (lncRNA) were identified (Fig. 1a and Supplementary Figure 1). Specifically, lncRNA represented the 13.4% of all IFN $\alpha$  induced genes in Huh7 and the 11.9% in Huh7.5. Analysis of IFN $\beta$  stimulation dataset revealed that the lncRNA constituted the 11.4% and 16.5% of induced genes in Huh7 and Huh7.5 respectively.

On the other end, IFN $\lambda$  stimulated lncRNA represented only the 2.5% and 6.4% of induced genes in Huh7 and Huh7.5 respectively. All differentially expressed gene lists at the set cut-off values can be found in Supplementary File 1.

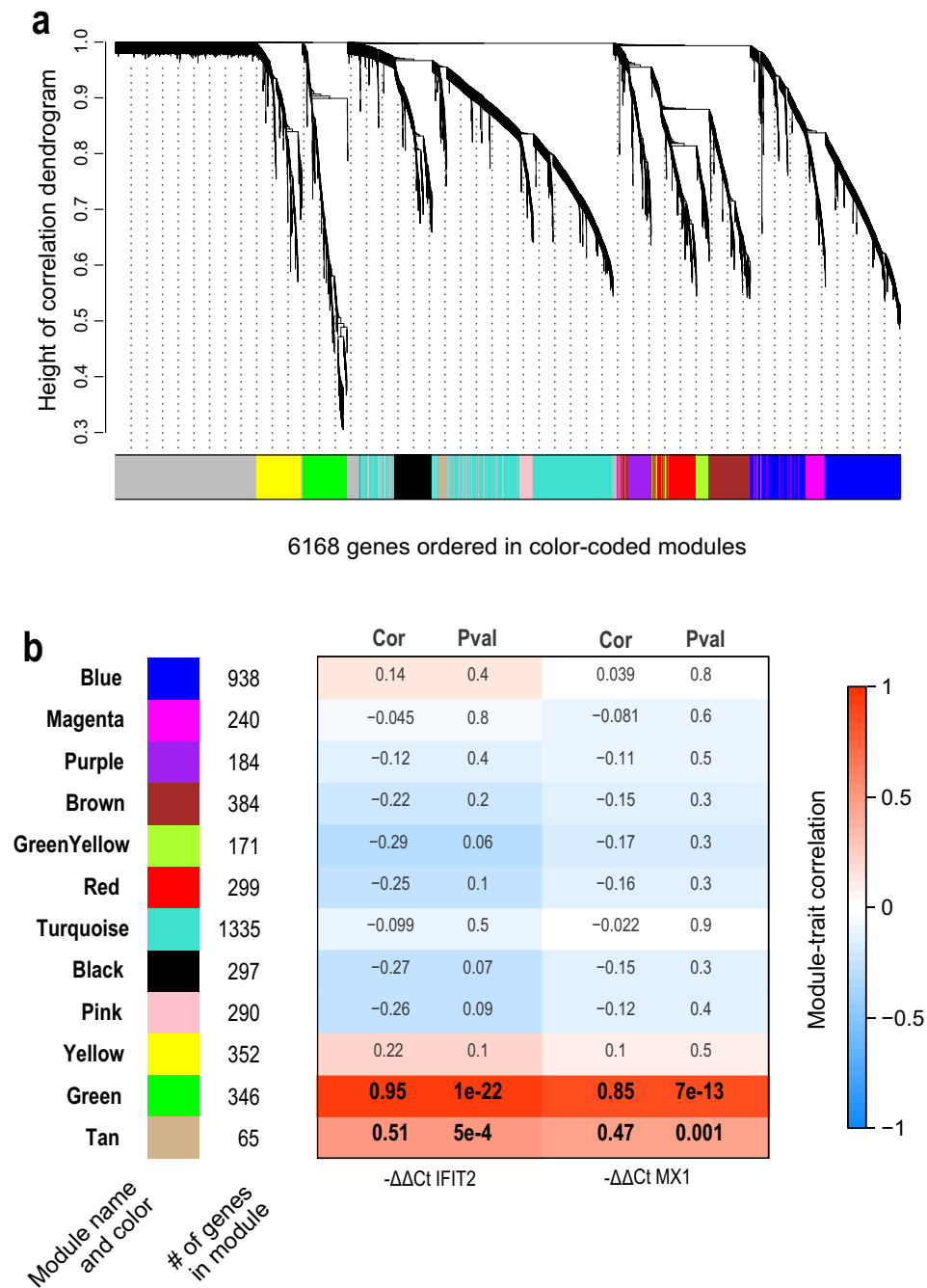
### Two gene expression modules among the IFN-related co-expression network are specifically associated with IFN response.

Conventional differential expression analysis has several shortcomings when applied to a complex experimental design composed by multiple comparisons, mostly related to experimental noise masking lower intensity signals that collectively can be biologically relevant in terms of regulation. Since we wanted to select genes associated to a continuous effect such as IFN response and at the same time overcome the limitations of a conventional pairwise comparison, we applied weighted gene co-expression network analysis (WGCNA)<sup>16</sup> that is able to group genes into modules with a similar expression pattern and then classifying modules according to their correlation to a continuous independent external variable, usually connected to the phenotypic perturbation under investigation. This method allowed us to summarize the large number of profiled genes in a small set of gene modules exemplifying the transcriptional behaviors in response to all IFN treatments. Instead of analyzing the treated–untreated contrasts separately in the two cell lines and for the three IFN treatments, as it is usually done in differential expression, we applied the WGCNA framework on normalized and appropriately transformed read count values of 6168 genes (selected on the basis of the coefficient of variation) across all 43 samples in our experimental design. By employing a gene clustering strategy on the Topological Overlap Dissimilarity Measure, we identified 12 gene modules with size ranging from 65 genes in the Tan module to 1335 genes in the Turquoise module (Fig. 2). 1267 genes remained unassigned (Fig. 2a, grey color). Multidimensional scaling confirmed a clear clustering in the three-dimensional space of the genes assigned to the same module (Supplementary Figure 3a).

In order to pinpoint the modules related to IFN response we used the expression pattern of two induced well known IFN sensitive genes (ISGs)<sup>19</sup>, MX1 and IFIT2, as independent experimental variables to scan the summarized expression profile of each module. To this aim, expression of MX1 and IFIT2 was measured by RT qPCR on the same sample set (Supplementary Figure 3b,c), and after identification of the expression profile of each module with the WGCNA summarizing metric (its first principal component or module eigengene) we related it to the induction of the two ISGs and selected the modules with a similar expression profile. Two modules (Green and Tan) out of twelve identified were selected due to their strong positive correlation to both IFN response proxies (MX1 and IFIT2 expression), with statistically significant *p*-Values (Fig. 2b).



**Figure 1.** Differential expression analysis on RNASeq transcriptome profiling. **(a)** Number of differentially expressed genes that are significantly modulated in each treatment compared to mock-treated cells after DESeq2 analysis with  $FDR < 0.05$  and absolute  $\text{Log}_2\text{FC} > 1$ . Over expressed (blue) and downregulated genes (rust red): dark colours for protein coding genes (actual numbers in bold) and light colours for non-coding genes (actual numbers in brackets). **(b–d)** Venn diagrams illustrating the overlap of modulated genes across treatments. Panel **(b)** shows the overlap between IFN $\alpha$  and IFN $\beta$  treatments in the Huh7 line, while panel **(c)** shows the same results for Huh7.5. In both lines, the genes modulated by IFN $\lambda$  are entirely included in the overlapping region of other two Interferons. Panel **(d)** shows how all genes in **(b)** and **(c)** overlap, showing a group of 185 genes modulated in all cell lines and by both type I IFN perturbations.



**Figure 2.** Generation of the gene co-expression network and selection of modules associated with IFN treatment. **(a)** Gene dendrogram generated on the Topological Overlap Dissimilarity matrix: each branch of the dendrogram corresponds to a module of highly interconnected genes, for a total of 12 modules identified. The color row below the dendrogram represents the module assignment. **(b)** Identified modules with relative colour “names”: numbers beside each colour block indicates the number of genes assigned to each of them. The heat map shows the correlation values (Pearson correlation degree “Cor” and statistical significance “Pval”) between each module eigengene and the induction of IFIT2 and MX1 transcripts measured by real-time PCR. Negative correlations are in blue, positive correlations are in red. Only Green and Tan modules show a significant positive correlation with the ISGs expression.

The Green module is composed of 346 genes, that overall shows a remarkable pattern of induction by IFN $\alpha$  and IFN $\beta$  (and weaker induction by IFN $\lambda$ ) in both Huh7 and Huh7.5 cells (Supplementary Figure 4). The Tan module groups 65 genes that are induced by IFN treatment, but also show differential baseline levels in Huh7 compared to Huh7.5 (basal higher expression in the latter) (Fig. 2b and Supplementary File 2).

The remaining modules, Black, Brown, GreenYellow, Purple, Red and Yellow do not show a clear change of gene expression pattern according to treatment or cell identity. Blue and Magenta modules group genes whose expression levels are lower in Huh7.5 compared to Huh7, while Turquoise and Pink modules are composed by genes expressed at higher levels in Huh7.5 compared to Huh7 cells (data not shown). None of them shows a statistically significant correlation with the induction of the two IFN sensitive genes used (MX1 and IFIT2) (Fig. 2b).

### Gene modules associated with IFN-response share functional roles and display clear regulatory motifs.

Green and Tan modules together count 411 genes, and they do contain almost all the 185 initially selected DE genes while adding many more, witnessing the added value of a selection method not strictly based on hard cutoffs imposed by differential expression. To characterize the two modules, the most informative genes belonging to them can be found using significance measures or topological properties such as high neighborhood connectivity. A functional neighborhood is composed of nodes (genes) that are highly connected (i.e. significantly co-expressed) to a given set of nodes and its analysis facilitates a guilt-by-association screening strategy to prioritize nodes for their potential ability to interact in the biological context under investigation<sup>15,17</sup>. In order to prioritize most relevant genes and to limit noise in subsequent functional analyses potentially triggered by peripheral ones<sup>20,21</sup> we applied this procedure and defined modules' subnetwork using the adjacency matrix (the matrix of co-expression measures, or weights, between all gene pairs) obtained by the WGCNA method: we imported nodes and edges of this matrix into Cytoscape and imposed a threshold of 0.1 as minimum edge weight. Sub-networks of highly connected nodes were thus selected from both the green and tan modules: 282 nodes (255 protein-coding and 27 non-coding genes) and 18,085 edges were extracted from the Green module, while 54 nodes (50 protein-coding and 4 non-coding) and 601 edges from the Tan module (Supplementary File 3).

Once the modules' sub-networks of highly-connected (i.e. highly co-expressed) genes were defined, we checked their functional composition with gene ontology (GO) enrichment analysis for biological processes (BP). For Green genes, a remarkable enrichment was observed for categories associated to IFN-response and defense to virus (Fig. 3a). A similar outcome, although less prominent, was also obtained for the Tan genes (Fig. 3b). GO enrichment analysis thus confirmed that Green and Tan sub-networks are indeed functionally enriched for genes associated with innate immunity phenomena.

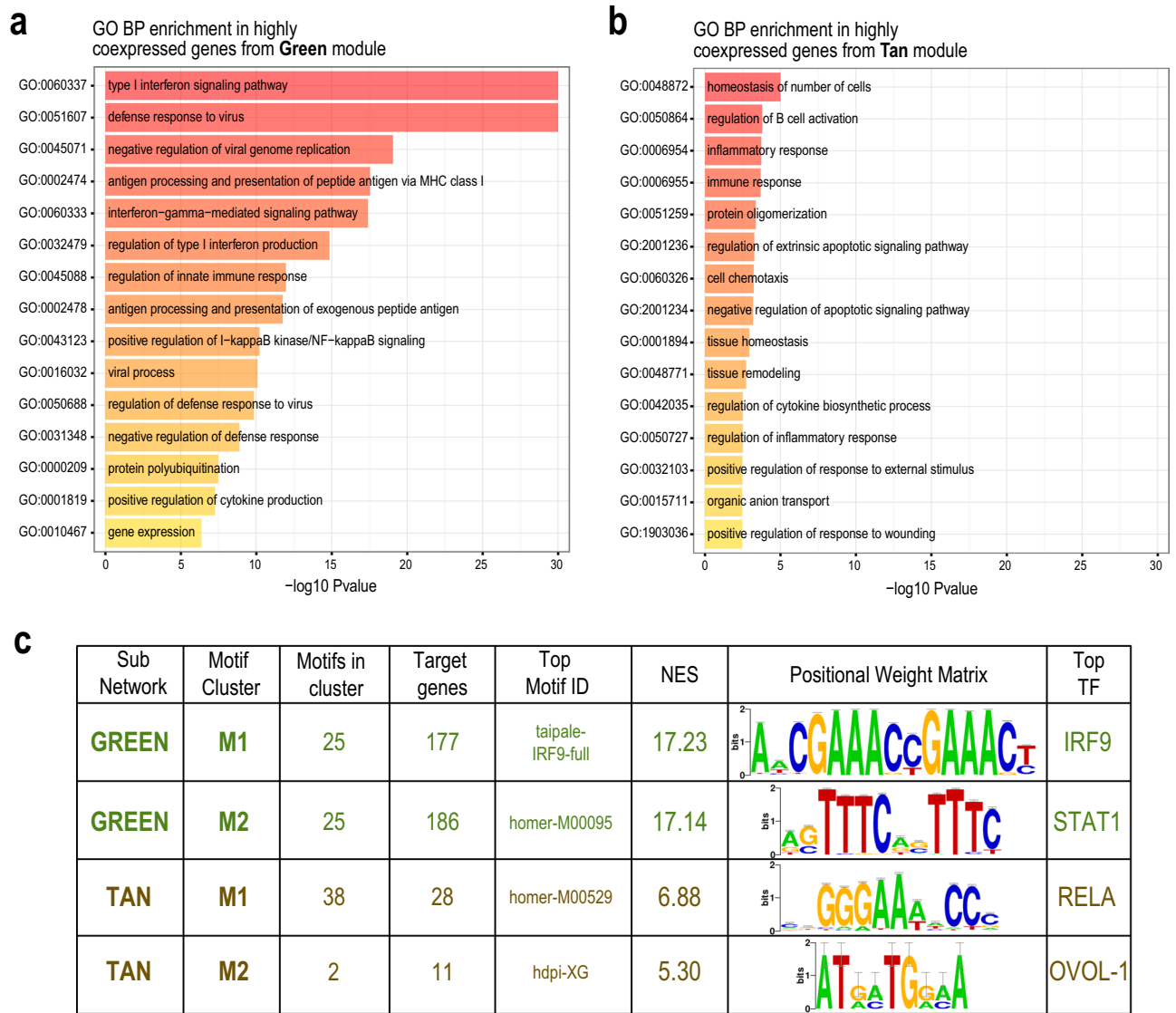
Co-expressed genes have been successfully used as input datasets for the search of transcription factors (TF) binding sites<sup>22,23</sup>, thus we went further into the functional characterization of our gene modules by inferring their upstream transcriptional regulatory network: to do so we applied the ranking-and-recovery method implemented with the iRegulon motifs collection<sup>24</sup>. Using the iRegulon Cytoscape plugin we searched the Green subnetwork (282 genes) for most probable regulons, i.e. the set of transcription factors and their targets, on the basis of presence of shared TF binding sites in the cis-regulatory control region of the genes. We found two highly enriched motif clusters, each composed by 25 regulatory sequences. On the basis of the Normalized Enrichment Score (NES), taipale-IRF9-full was selected as top motif for the first cluster (Green-M1), with IRF9 as top-ranked master regulator. Similarly, homer-M00095 and STAT1 were found respectively as top motif and top regulator for the second cluster, called Green-M2 (Fig. 3c). Among the 282 genes of the green sub-network that are putative targets, 163 are regulated by both regulators classes (Supplementary Figure 5a). Since IRF9-STAT1 axes is the key pathway regulating the IFN response<sup>25</sup>, the identification of IRF9- and STAT1-related regulons in Green sub-network further highlights its relevance in the cellular response to interferon treatment.

The same analysis was then applied on Tan genes. For this sub-network, two different motif clusters were found to be overrepresented: homer-M00529 (Tan-M1) and hdpi-XG (Tan-M2), with associated master regulators RelA, subunit of NF-kB complex required during the early response to RNA viruses, and OVOL1, respectively (Fig. 3c). A total of 30 putative target genes were identified, with 9 genes predicted to be the target of both transcription factors classes (Supplementary Figure 5b). Interestingly, the expression of OVOL1, a transcription factor involved in cell differentiation seems to be dependent on IRF6, suggesting that OVOL1 responsive genes could be activated at a later time point and play a role in tissue remodeling after a damage induced by the pathogen or the host immune response<sup>26</sup>.

Taken together these data confirm the ability of our strategy to detect canonical genes involved in IFN response through STAT1-IRF9 pathway as well as novel gene clusters involved in a less characterized IFN modulation.

### Novel potential ISGs are uncovered with network pruning and knowledge subtraction.

Network metrics and Gene ontology annotations can be exploited to identify the most central players (most connected nodes, or hub nodes) in the network, allowing topological data reduction and facilitating the discovery of not previously annotated key molecular drivers that may not emerge otherwise (Fig. 4a). We examined more in detail two measures of network centrality in the previously selected sub-network: the eigenvector centrality, which is a measure of the influence of a node in a network, and the closeness centrality, a measure of the proximity of each node to all the others, thus evaluating each node's contribution to the network structure. The two metrics showed a linear correlation for the highly connected nodes (above a certain threshold), identifying a group of peripheral nodes, for both Green and Tan sub-networks (Fig. 4b,c). The highly co-expressed (and as such, highly connected) genes are likely to play a more relevant role in molecular regulation of IFN induction and when visualized in the network graph they appear as densely connected nodes. The nodes with lower

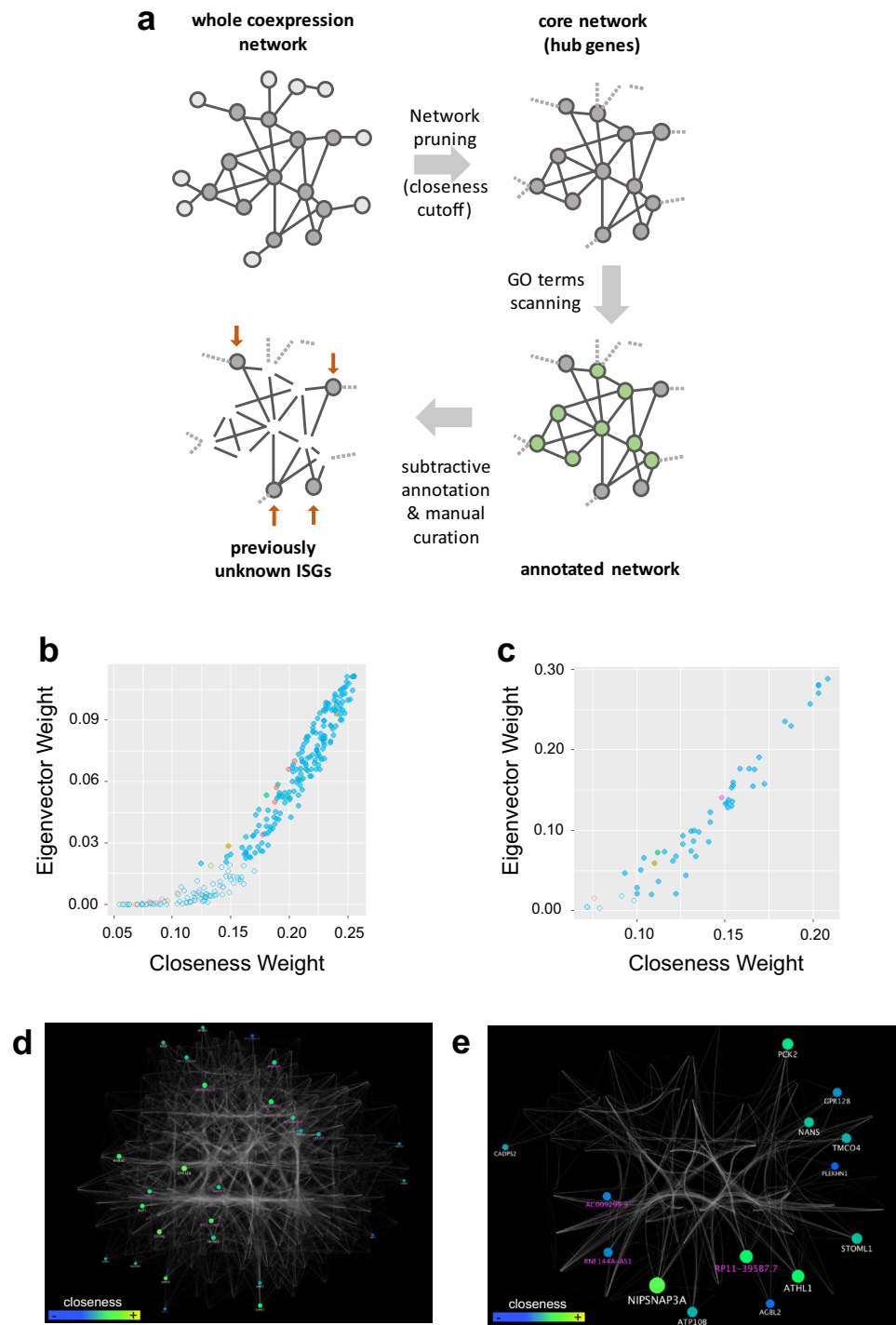


**Figure 3.** Functional characterization of Green and Tan modules. **(a,b)** Results of gene ontology enrichment analysis performed on Green **(a)** and Tan **(b)** sub-networks; bar plots depict the  $-\log p$ -values for the top-15 Biological Process terms. **(c)** Tabular results of the analysis of regulatory motifs performed on Green and Tan sub-networks by IRegulon. The motif clusters identified as significantly enriched for each module ( $NES > 5$ ) are reported. **(d,e)** Circular plots illustrating the regulatory relationship among each motif cluster and its putative target genes for Green and Tan modules. The thickness of each line and correspondent gene sector represents the number of motifs in the cluster found as putative regulators of the gene.

eigenvector (we set an arbitrary cut-off at 0.02) have lower connectedness/co-expression and indeed appeared topologically peripheral in the network graphs (Supplementary Figures 6–7). Using this cut-off, we identified as peripheral and discarded 96 and 6 nodes from the Green and Tan sub-networks respectively (Fig. 4d,e).

We used the result of the previously performed gene ontology analysis and specifically took into account the enrichment terms related to innate immunity response. We then tagged as known players all nodes associated to at least one term among the following: innate immune and inflammatory response; type I interferon production; defense response to virus and viral process; antigen processing and presentation of peptide antigen via MHC class I; I-kappaB kinase/NF-kappaB signaling.

These nodes can be considered as known players: we then subtracted them and the remaining, un-annotated nodes were selected as potential novel players not yet characterized as related to innate immunity. Moreover, to be more stringent we also complemented this analysis with a manual curated literature search, so that only genes not previously reported as IFN-related were included in our final list. Adopting this approach, we identified 26 putative new and relevant mediators of interferon antiviral activity in the green network. The same procedure was then applied to the Tan network: topological network analysis was performed on a sub-network of 54 nodes and 601 edges; 48 genes were selected as non-peripheral and subjected to ontology annotation and manual literature search. Finally, 14 protein-coding and non-coding were selected as potential new players. The resulting



**Figure 4.** Green and Tan sub networks of putative new ISGs. **(a)** The method for extraction of putative, previously unknown IFN sensitive genes, using network pruning by centrality metrics cutoffs and subtractive gene ontology annotations. **(b)** Eigenvectors and closeness of genes in the Green module. These two metrics are linearly correlated above a certain threshold: higher eigenvector values correspond to higher closeness and identify hub genes. Nodes with eigenvector < 0.02 are defined as peripheral and depicted as transparent. **(c)** Eigenvectors and closeness of genes in the Tan module. **(d)** Cytoscape-generated sub-network for the Green module. Gene ontology annotation of surviving nodes after subtractive gene ontology annotation are displayed, while nodes (genes) associated to immune-related biological processes are not, allowing focus on non-peripheral novel players. Eigenvector is proportional to node size, closeness to color gradient. **(e)** Similar cytoscape-generated sub-network for the Tan module.

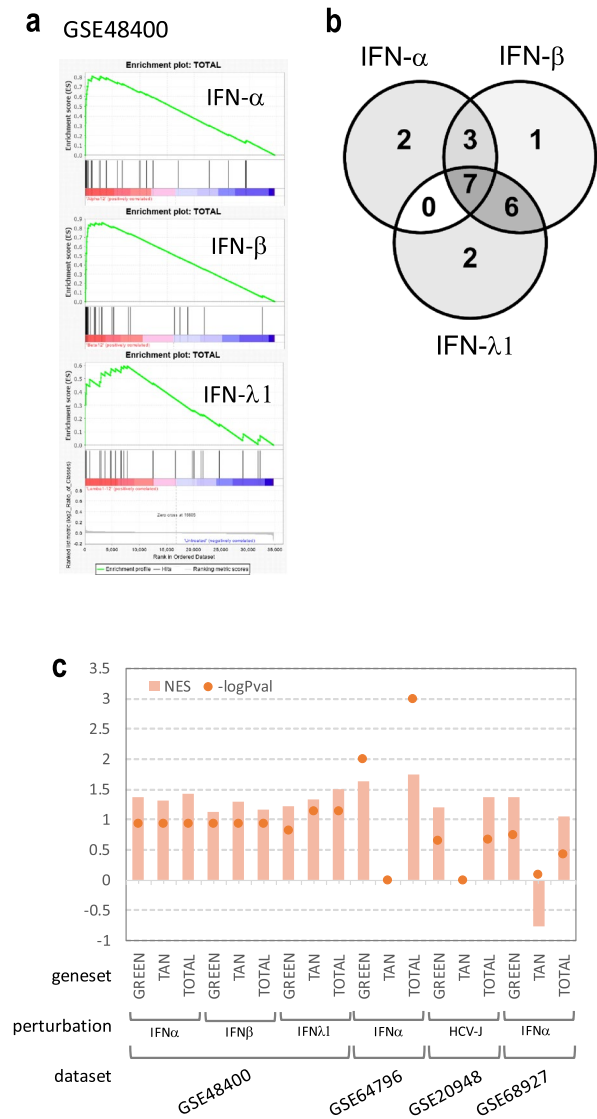


Source module	Gene name	Ensembl Gene ID	Biotype	Description (Biomart)
Green	SIDT1	ENSG00000072858	protein_coding	SID1 transmembrane family member 1
Green	CYTH1	ENSG00000108669	protein_coding	cytohesin 1
Green	BLZF1	ENSG00000117475	protein_coding	basic leucine zipper nuclear factor 1
Green	DNPEP	ENSG00000123992	protein_coding	aspartyl aminopeptidase
Green	RARRES3	ENSG00000133321	protein_coding	phospholipase A and acyltransferase 4
Green	MOB3C	ENSG00000142961	protein_coding	MOB kinase activator 3C
Green	NME7	ENSG00000143156	protein_coding	NME/NM23 family member 7
Green	GRIP2	ENSG00000144596	protein_coding	glutamate receptor interacting protein 2
Green	GPR158	ENSG00000151025	protein_coding	G protein-coupled receptor 158
Green	MFSD12	ENSG00000161091	protein_coding	major facilitator superfamily domain containing 12
Green	KCNT2	ENSG00000162687	protein_coding	potassium sodium-activated channel subfamily T member 2
Green	PLEKHA7	ENSG00000166689	protein_coding	pleckstrin homology domain containing A7
Green	GRINA	ENSG00000178719	protein_coding	glutamate ionotropic receptor NMDA type subunit associated protein 1
Green	ZNF107	ENSG00000196247	protein_coding	zinc finger protein 107
Green	LAP3P2	ENSG00000213500	processed_pseudogene	leucine aminopeptidase 3 pseudogene 2
Green	SMTNL1	ENSG00000214872	protein_coding	smoothelin like 1
Green	AC009948.5	ENSG00000223960	antisense	cholesterol induced regulator of metabolism RNA
Green	PSME2P2	ENSG00000225131	processed_pseudogene	proteasome activator subunit 2 pseudogene 2
Green	AC009950.2	ENSG00000225963	antisense	novel transcript, antisense to SP110
Green	AP001610.5	ENSG00000228318	antisense	novel transcript
Green	PNPT1P1	ENSG00000229241	processed_pseudogene	polyribonucleotide nucleotidyltransferase 1 pseudogene 1
Green	PATL2	ENSG00000229474	protein_coding	PAT1 homolog 2
Green	NAMPTL	ENSG00000229644	processed_pseudogene	nicotinamide phosphoribosyltransferase pseudogene 1
Green	HLA-K	ENSG00000230795	unprocessed_pseudogene	major histocompatibility complex, class I, K (pseudogene)
Green	RP11-274E7.2	ENSG00000238000	processed_pseudogene	proteasome activator subunit 2 pseudogene 1
Green	RP11-54O7.17	ENSG00000272512	lincRNA	novel transcript
Tan	STOML1	ENSG00000067221	protein_coding	stomatin like 1
Tan	NANS	ENSG00000095380	protein_coding	N-acetylneuraminate synthase
Tan	ATP10B	ENSG00000118322	protein_coding	ATPase phospholipid transporting 10B (putative)
Tan	PRRG4	ENSG00000135378	protein_coding	proline rich and Gla domain 4
Tan	NIPSNAP3A	ENSG00000136783	protein_coding	nipsnap homolog 3A
Tan	ATHL1	ENSG00000142102	protein_coding	protein-glucosylgalactosylhydroxylysine glucosidase
Tan	GPR128	ENSG00000144820	protein_coding	adhesion G protein-coupled receptor G7
Tan	TMCO4	ENSG00000162542	protein_coding	transmembrane and coiled-coil domains 4
Tan	AGBL2	ENSG00000165923	protein_coding	ATP/GTP binding protein like 2
Tan	PLEKHN1	ENSG00000187583	protein_coding	pleckstrin homology domain containing N1
Tan	AC009299.3	ENSG00000227403	lincRNA	long intergenic non-protein coding RNA 1806
Tan	RNF144A-AS1	ENSG00000228203	processed_transcript	glycosaminoglycan regulatory associated long non-coding RNA
Tan	RP11-395B7.7	ENSG00000260336	Sense_overlapping	<i>na, deprecated id</i>
Tan	CADPS2	ENSG00000081803	protein_coding	calcium dependent secretion activator 2

**Table 2.** Novel IFN sensitive genes. The 40 novel putative IFN sensitive genes (ISGs) annotated with the source module they were extracted from (Green or Tan modules), gene names, Ensembl gene IDs, biotype and description as retrieved from Ensembl Biomart.

combined list of putative novel mediators of interferon antiviral activity is made of 40 genes from the two selected gene modules, among which 13 non-coding RNAs (Table 2 and “novel” classification in Supplementary File 3).

**The newly selected ISGs are enriched in IFN perturbed or HCV infected datasets.** In order to validate our new IFN-stimulated signature in datasets from independent works we searched the literature and selected published datasets obtained on human samples treated with IFN. We selected and downloaded gene expression datasets of four works<sup>27–30</sup> (Supplementary Table 1) and for each work we obtained the ranked gene



**Figure 5.** New IFN-stimulated signature is confirmed as ISGs with GSEA. (a) GSEA running sums showing the enrichment of custom geneset obtained by the combination of the Green and Tan modules in the Boolean et al. 2014 datasets (GSE48400) after 12 h of induction with IFN alpha, beta and lambda1. (b) Venn diagram of overlapping genes belonging to the leading edges of the three GSEA analyses in the previous panel. Seven genes (DNPEP, REC8, GRINA, CYTH1, PCK2, NME7, SIDT1) have been consistently found upregulated after different IFNs induction. (c) Normalized enrichment scores (NES) and relative GSEA p-Values for all sets of performed GSEAs (for datasets with more time points the 12 h point was selected and plotted). Dashed horizontal line is the standard GSEA significant cutoff: enrichments whose p-Value is above the dashed line are thus significant. Green, Tan and Total refer to genesets from Green module, Tan module and the combination of the two, respectively.

lists according to the differential expression metrics used (mainly log fold change) to compare treatments versus control samples. We then verified if our newly selected ISGs were positioned at the extremes of the gene expression datasets (i.e. if they were significantly over-represented among down or up regulated genes). To do so we used the 40 unique putative mediators of interferon antiviral activity identified by our analysis as elements of three custom gene sets (Green, Tan and the combination of the two) to scan all prepared ranked datasets with the GSEA algorithm<sup>31</sup>. Strong upregulation of the combined geneset (genes from both the green and tan modules) was indeed observed in data obtained after induction of different interferons (Fig. 5a), even if the genes directly responsible for the observed enrichment were not always the same with a recurrent subset of 7 genes (Fig. 5b) among the annotated genes in the signature (DNPEP, REC8, GRINA, CYTH1, PCK2, NME7, SIDT1). With minimal differences due to time and method of induction, our newly selected gene signature was always detected among the up-regulated genes after induction or interferon related perturbation: in all analyzed published datasets taken into account the custom Green, Tan and total genesets showed a significant enrichment in most cases, with the Tan geneset being the weakest with a p-Value below the cutoff in three out of four datasets and the combined geneset just below the cutoff in GSE68927 data (Fig. 5c). Individual GSEA enrichment plots are shown

in Supplementary Figures 8–13. These validation analyses confirmed that integrating WGCNA analysis with subtractive gene ontology annotation was an effective way of finding mediators of interferon antiviral activity.

## Discussion

Here we report a systems biology approach which leverages the possibility to integrate datasets at different levels: transcription, membership to co-expression modules, correlation with phenotypic traits, knowledge organized as bio-ontologies, transcription factors motifs prediction and networks centrality metrics. These aspects were coherently integrated in our workflow with the aim to identify novel molecular drivers of IFN response.

Gene expression is often victim of confounding effects and it is not always possible to pinpoint all the subtle regulations and perturbations by simple, cut-off dependent, differential expression (DE) methods. A multivariate approach to gene selection allows to dive deeper into hidden relationships between molecular actors in a complex system.

By applying such an approach, we were able to find a group of genes that are specifically associated to the continuous IFN mediated modulation that were not previously described as such, and that were, at least partly, undetected in the conventional differential expression approach.

When using differential expression methods, the user has to set cut-offs and findings are necessarily dependent upon those: the more stringent the cut-offs the fewer the results. This can be harmful in the selection process favoring the more extreme perturbations over the subtler ones, thus yielding a tip-of-the-iceberg effect. Using WGCNA and the independent external measure of the IFN sensitive genes (a measure acting as a proxy to the phenotype under investigation), almost all differentially expressed genes were also selected, plus many more, without the need of setting the hard cut-offs of canonical differential expression analyses. Almost all (180 out of 185) differentially expressed genes shared by the analyzed contrasts belong to one of the two identified gene modules, while only 44% (180 out of 411) of Tan or Green modules genes are also selected as differentially expressed by canonical analysis. This means that 56% of modules genes (231 genes) we found are not selected as DE genes with the used cut-offs (0.05 FDR and twofold changes). In a less strict way, taking into account the sum (without repetition) of all DE genes, i.e. the genes differentially expressed in all six (2 cell lines and 3 IFN types) analyzed contrasts, the fraction of modules genes not selected as DE genes drops at 19%, yet maintaining a relevant number of genes (111) exclusively selected by the multivariate approach.

Differential expression is per se limited to pairwise comparisons (i.e. treated versus untreated), and each result is valid in the context of the considered pair. Taking into account only the differentially expressed genes in all comparisons under study, as it was done for the 185 perturbed in all contrasts analyzed, underestimates the biological complexity of a given set of comparisons, de facto flattening the differences. DE analysis remains the election method to select culprits in carefully designed and biologically strong contrasts, but it could be not the right choice when analyzing complex designs and looking for regulatory relationships. The fact that all genes are tested for differential expression one by one with completely independent tests is even more underestimated, since this approach once again does not take into account the natural interdependence of genes collaborating in biochemical pathways.

On the contrary, topology (centrality) metrics are a way of putting into play the relationships among genes, thus the approach is much more prone to let emerge the cooperative role of genes more than the antagonistic phenotypes. Among the genes selected in this way there could be genes that would have been pruned by an arbitrary differential expression cut-off, thus a network driven systems biology approach is more sensitive and more fit for regulatory network discovery.

For this reason, we took the advantage of having two co-expression modules well associated with the phenotype of interest to leverage them to uncover possible enrichment of transcriptional factor motifs. The findings of STAT1 and IRF9 motifs enrichment also in genes not previously described as IFN related strengthen their involvement in this process and internally validate our findings. Both IFN- $\alpha$  and IFN- $\lambda$  induce phosphorylation of STAT1 inducing, even if with different kinetics, the expression of similar sets of ISGs, with IFN- $\alpha$  antiviral action being more potent than that of IFN- $\lambda$ , and STAT1-mediated differential inhibition of HCV seems to be essential for IFN- $\lambda$  but not for IFN- $\alpha$ <sup>32</sup>. We indeed observed an overall duality between IFN- $\alpha/\beta$  and IFN- $\lambda$  gene expression patterns, but the multivariate approach that considers expression patterns across the whole samples set should act independently from such differences. The fact that motifs for IRF9 and STAT1 transcription factors are evenly distributed across the green module genes (Supplementary Figure 5a) could account for the above duality in order to conserve the possibility to adapt to environmental situations responding to both classes of treatment. A similar distribution is not observed for RELA and OVOL-1 motifs across genes in the Tan module (Supplementary Figure 5b); to this extent, the Tan module is much more a RELA related than an OVOL-1 related regulon. In addition, we also identified lncRNA specifically modulated by Type I and Type III IFN treatment. Several gene expression studies have shown that IFN profoundly changes the lncRNA cellular landscape and many lncRNAs are considered ISGs. Indeed, lncRNAs specifically regulate IFN response<sup>33</sup>. Interestingly, we identified miR-146a as an upregulated lncRNA after IFN- $\alpha$  treatment. miR-146a plays an important role in modulating type I IFN response and inflammation by targeting TRAF6<sup>34</sup>. Increase in miR-146a levels suppresses type I interferon and facilitates infection of several viruses like EV71, dengue virus and HIV<sup>34–36</sup>. On the other hand, inhibition of miR-146a restores type I interferon production and blocks enterovirus-induced cell death<sup>37</sup>. Finally, we validated our datasets of new IFN-stimulated genes, in independent works suggesting that our method is suitable to interrogate different datasets for the discovery of new mediators of interferon pathways.

## Methods

**Cells and IFN perturbation.** Huh7 cell line (reference number JCRB0403) was purchased from Japanese Collection of Research Bioresources (JCRB) Cell Bank, Osaka, Japan. Huh7.5 cell line was a gift of Prof. Charles M. Rice, Laboratory of Virology and Infectious Disease, The Rockefeller University, NY, USA. Huh7 and Huh7.5 cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin–streptomycin and 2 mM L-glutamine<sup>38</sup>. Cells were treated for 8 h with 10 ng/ml of Interferon- $\alpha$ 2A (cod.H6041, SIGMA), 5000 U/ml of Interferon- $\beta$ 1A (cod.I4151, SIGMA) or 100 ng/ml of interferon- $\lambda$ 3 (cod.E92028Hu, Uscn Life Science Inc) prior RNA extraction and library preparation.

**RNA extraction and library preparation.** RNA-Seq experiments were performed on Huh7 and Huh7.5 treated with Interferon- $\alpha$ 2A, Interferon- $\beta$ 1A (cod.I4151, SIGMA) or Interferon- $\lambda$ 3 as described above. Mock-treated cells were used as control for each treatment. At least 3 biological replicates were analysed for each biological condition. Total RNA was extracted with mirVana miRNA isolation kit and DNase digested (Ambion) according to manufacturer's instruction. RNA samples were submitted to quality control by RNA integrity number evaluation. Libraries for Illumina sequencing were constructed by polyA selection with the Illumina Kit. A paired end (2  $\times$  100) run was performed.

**RNAseq gene expression and differential expression analysis.** After initial quality check reads were filtered to remove low quality calls by Trimmomatic (version 0.XX) (24,695,404) using default parameters. Processed reads were then aligned to human genome assembly GRCh38 (Ensembl release 77) with STAR mapping software (version 2.3.0e)<sup>39</sup>. HTSeq-count algorithm (version 0.6.1)<sup>40</sup> option `-str=no`, gene annotation release 77 from Ensembl) was employed to produce read counts at the gene level for each sample. To estimate differential expression, the matrix of gene counts produced by HTSeq was analysed by DESeq2<sup>41</sup>. Differential expression analysis was performed to compare each treatment with the respective mock-treated cells (6 comparisons, 43 samples overall). Read counts were normalized by calculating size factors, while Variance Stabilizing Transformed (vst) values were employed for principal component analysis, as implemented in DESeq2.

Independent filtering procedure was then applied, setting the threshold to the 76 percentile (5.74 read counts); 15,421 genes were subsequently tested for differential expression. Significantly modulated genes were selected by considering an absolute value of log<sub>2</sub> fold change (Log<sub>2</sub>FC) higher than 1 and a false discovery rate (FDR) lower than 5%.

**Weighted gene co-expression network generation and module identification.** Variance Stabilizing Transformed values (as calculated by DESeq2) were used as input data for network analysis. Starting from the set of 15,421 genes identified by independent filtering, a gene selection strategy was applied by calculating for each gene the coefficient of variation (CV) across the 43 biological samples. A 60-percentile threshold was then imposed, thus selecting the 40% of genes showing the highest variability (highest CV values, 6168 genes). Sample average-linkage hierarchical cluster analysis was performed as a quality control step on this gene subset. A signed hybrid gene co-expression network was generated relying on the WGCNA R package (version 1.43–10)<sup>16</sup>. The correlation matrix was calculated based on the Pearson metrics and then transformed into an adjacency matrix by raising it to the power of  $\beta = 12$ , chosen on the basis of the scale-free topology criterion. Topological Overlap Measure was calculated from the adjacency matrix and the relative dissimilarity matrix was used as input for average-linkage hierarchical clustering and gene dendrogram generation. Network modules were detected as branches of the dendrogram by using the DynamicTree Cut algorithm (deepSplit = 1, minimum cluster size = 40)<sup>42</sup>. Multi-dimensional scaling was calculated on the topological overlap dissimilarity matrix.

**Module-trait correlation.** As a phenotypic trait for module-trait correlation, the transcripts of the two well-known ISGs MX1 and IFIT2 were quantified by real-time PCR in the 43 samples.

For real-time PCR, TaqMan gene expression assay (Thermo Fisher scientific) was performed according to manufacturer's instruction.  $\Delta$ Ct versus the gene normalizer GAPDH and then  $\Delta\Delta$ Ct versus the mean  $\Delta$ Ct value in corresponding non-treated cells were calculated for each of the six treatments. Minus  $\Delta\Delta$ Ct was used for downstream analysis, so that an induction of ISGs was quantified by a high  $-\Delta\Delta$ Ct value. For each module, Pearson correlation between the module eigengene and the  $-\Delta\Delta$ Ct value was calculated. For the selected modules (Green and Tan) sample clustering and heat map illustrating the gene expression profiles were generated starting from Variance Stabilizing Transformed values.

**Sub-network generation and topological network analysis in cytoscape.** Node and edge information for the selected modules (Green and Tan) were exported from the adjacency matrix relying on the appropriate WGCNA function and imposing a threshold of 0.1 as minimum edge weight. 282 nodes and 18,085 edges for the Green module and 54 nodes and 601 edges for the Tan module were therefore imported in Cytoscape for further analysis<sup>43</sup>. Topological network analysis for Green and Tan sub-networks was performed in Cytoscape relying on CytoNCA app using the “with weight” analysis<sup>44</sup>. Peripheral genes were identified as those showing an eigenvector value lower than 0.02.

**Gene ontology enrichment analysis.** Gene ontology enrichment analysis for the Biological Process domain was performed on the 282 genes belonging to the green sub-network, using the list of 6168 genes selected for network generation as a custom reference set. Analysis was performed by TopGO (version 2.18.0)<sup>45</sup> relying on Fisher test and Weight01 method to take into account ontology hierarchy; node size was set at 20. A

0.01 *p*-value cut off was applied to select significantly enriched GO terms. The same procedure was performed on the 54 nodes of the Tan sub-network. Overall screening gene ontology analysis contemplating more databases (ontologies, pathways, motifs, proteins and phenotypes) on cell-line specific differentially expressed genes were also performed using g:Profiler<sup>46</sup>.

**Regulatory motif analysis.** Cytoscape app (version 1.3)<sup>24</sup> was used to identify enriched regulatory regions in Green and Tan sub-networks. Analysis was performed on 5000 bp upstream the TSS, using 10 K motif collection and retrieving motifs with a Normalized Enrichment score higher than 5; maximum FDR on motif similarity was set at 0.005. Circular plots for result visualization were produced in R by the circlize package (version 0.2.4)<sup>47</sup>.

**Functional annotation.** Starting from the results of the gene ontology analysis, the terms found as enriched considering a threshold of *p*-Value < 0.01 were then classified as related or unrelated to innate immunity phenomena. For the identification of genes related to innate immune response, descendent terms of the following categories were considered: GO:0045087—innate immune response; GO:0006954—inflammatory response; GO:0032606—type I interferon production; GO:0051607—defense response to virus; GO:0016032—viral process; GO:0002474—antigen processing and presentation of peptide antigen via MHC class I; GO:0007249—I-kappaB kinase/NF-kappaB signalling. Genes associated with at least one of these groups of terms were therefore defined as ‘known players’, while those not associated with any were defined “IFN-unrelated”.

**Gene set enrichment analyses.** Gene set enrichment analysis (GSEA) was used to validate the selected gene signature of putative ISGs as actually over represented in other published datasets of IFN induction studies. GSEA was performed as described in the original paper<sup>31</sup> using the Broad Institute official java implementation, with custom genesets. Normalized gene expression data of the selected literature were downloaded from the GEO Omnibus repository and used as GSEA input datasets; ranked gene lists of differential expression analyses performed were obtained using Signal to Noise or Log2 Ratios metrics. Custom genesets of 28, 14 and 42 genes were built using the candidate ISGs from the Green module, Tan module and both combined, respectively, after harmonization of gene names annotation across the gene expression platforms used in the four different works. GSEA was then performed with default parameters and cutoffs. GSEA running plots depicts enrichment scores (ES), while in the summary plot of all GSEA results were used normalized enrichment scores (NES) with cognate *p*-Values.

## Data availability

The sequencing data generated during the current study have been deposited in the ArrayExpress database at EMBL-EBI under accession number E-MTAB-10469. Processed datasets are also available in table format as Supplementary Information files.

Received: 6 May 2021; Accepted: 20 August 2021

Published online: 10 September 2021

## References

1. Heim, M. H. & Thimme, R. Innate and adaptive immune responses in HCV infections. *J. Hepatol.* **61**, S14–25 (2014).
2. Bartenschlager, R. & Pietschmann, T. Efficient hepatitis C virus cell culture system: What a difference the host cell makes. *Proc. Natl. Acad. Sci. USA* **102**, 9739–9740 (2005).
3. Kell, A. M. & Gale, M. RIG-I in RNA virus recognition. *Virology* **479–480**, 110–121 (2015).
4. Bartenschlager, R., Kaul, A. & Sparacio, S. Replication of the hepatitis C virus in cell culture. *Antiviral Res.* **60**, 91–102 (2003).
5. Stanifer, M. L., Pervolaraki, K. & Boulant, S. Differential regulation of type I and type III interferon signaling. *Int. J. Mol. Sci.* **20** (2019).
6. Lazear, H. M., Schoggins, J. W. & Diamond, M. S. Shared and distinct functions of type I and type III interferons. *Immunity* **50**, 907–923 (2019).
7. Mesev, E. V., LeDesma, R. A. & Ploss, A. Decoding type I and III interferon signalling during viral infection. *Nat. Microbiol.* **4**, 914–924 (2019).
8. Lazear, H. M., Nice, T. J. & Diamond, M. S. Interferon-λ: Immune functions at barrier surfaces and beyond. *Immunity* **43**, 15–28 (2015).
9. Mathy, N. W. & Chen, X.-M. Long non-coding RNAs (lncRNAs) and their transcriptional control of inflammatory responses. *J. Biol. Chem.* **292**, 12375–12382 (2017).
10. Schoggins, J. W. Interferon-stimulated genes: What do they all do?. *Annu. Rev. Virol.* **6**, 567–584 (2019).
11. Maucort-Boulch, D., de Martel, C., Franceschi, S. & Plummer, M. Fraction and incidence of liver cancer attributable to hepatitis B and C viruses worldwide. *Int. J. Cancer* **142**, 2471–2477 (2018).
12. Marcello, T. *et al.* Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology* **131**, 1887–1898 (2006).
13. Arnone, M. I. & Davidson, E. H. The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124**, 1851–1864 (1997).
14. Miklos, G. L. & Rubin, G. M. The role of the genome project in determining gene function: Insights from model organisms. *Cell* **86**, 521–529 (1996).
15. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
16. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
17. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
18. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).

19. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* **32**, 513–545 (2014).
20. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* **8**, 22 (2007).
21. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
22. Auerbach, R. K., Chen, B. & Butte, A. J. Relating genes to function: Identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. *Bioinformatics* **29**, 1922–1924 (2013).
23. Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801–813 (2013).
24. Janky, R. *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* **10**, e1003731 (2014).
25. Stark, G. R. & Darnell, J. E. The JAK-STAT pathway at twenty. *Immunity* **36**, 503–514 (2012).
26. Kwa, M. Q. *et al.* Receptor-interacting protein kinase 4 and interferon regulatory factor 6 function as a signaling axis to regulate keratinocyte differentiation. *J. Biol. Chem.* **289**, 31077–31087 (2014).
27. Bolen, C. R., Ding, S., Robek, M. D. & Kleinstein, S. H. Dynamic expression profiling of type I and type III interferon-stimulated hepatocytes reveals a stable hierarchy of gene expression. *Hepatology* **59**, 1262–1272 (2014).
28. Carnero, E. *et al.* Type I interferon regulates the expression of long non-coding RNAs. *Front. Immunol.* **5**, 548 (2014).
29. Blackham, S. *et al.* Gene expression profiling indicates the roles of host oxidative stress, apoptosis, lipid metabolism, and intracellular transport genes in the replication of hepatitis C virus. *J. Virol.* **84**, 5404–5414 (2010).
30. Grünvogel, O. *et al.* DDX60L Is an interferon-stimulated gene product restricting hepatitis C virus replication in cell culture. *J. Virol.* **89**, 10548–10568 (2015).
31. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
32. Yamauchi, S. *et al.* STAT1 is essential for the inhibition of hepatitis C virus replication by interferon- $\lambda$  but not by interferon- $\alpha$ . *Sci. Rep.* **6**, 38336 (2016).
33. Suarez, B., Prats-Mari, L., Unfried, J. P. & Fortes, P. LncRNAs in the type I interferon antiviral response. *Int. J. Mol. Sci.* **21** (2020).
34. Wu, S. *et al.* miR-146a facilitates replication of dengue virus by dampening interferon induction by targeting TRAF6. *J. Infect.* **67**, 329–341 (2013).
35. Fu, Y. *et al.* Exosome-mediated miR-146a transfer suppresses type I interferon response and facilitates EV71 infection. *PLoS Pathog.* **13**, e1006611 (2017).
36. Teng, Y. *et al.* CRISPR/Cas9-mediated deletion of miR-146a enhances antiviral response in HIV-1 infected cells. *Genes Immun.* **20**, 327–337 (2019).
37. Ho, B.-C. *et al.* Inhibition of miR-146a prevents enterovirus-induced death by restoring the production of type I interferon. *Nat. Commun.* **5**, 3344 (2014).
38. Blight, K. J., McKeating, J. A. & Rice, C. M. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* **76**, 13001–13014 (2002).
39. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
40. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
41. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
42. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
43. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
44. Kijima, S. & Kijima, H. Statistical analysis of channel current from a membrane patch. II. A stochastic theory of a multi-channel system in the steady-state. *J. Theor. Biol.* **128**, 435–455 (1987).
45. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).
46. Raudvere, U. *et al.* g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucl. Acids Res.* **47**, W191–W198 (2019).
47. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

## Acknowledgements

We would like to thank Valeria Zanoni for technical laboratory support.

## Author contributions

R.D.F. conceived the study. C.C., L.M., R.D.F., and R.L.R. designed experiments and analytical workflows. L.D. and V.B. performed experiments. R.J.P.B. generated RNAseq raw counts data. C.C. wrote code, processed and analyzed RNAseq, network and functional enrichment data. R.L.R. analyzed functional enrichment data, GSEA validations and integrated whole data. C.C., L.M., L.D., V.B., R.L.R. and R.D.F. interpreted the result. C.C., L.M., R.L.R. wrote the manuscript. C.C. and R.L.R. made figures. L.M. contributed virological expertise and finalized the manuscript. R.L.R. and R.D.F. jointly supervised the work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97258-8>.

**Correspondence** and requests for materials should be addressed to R.L.R. or R.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2021