RESEARCH PAPER

RNA Biology 11:7, 875–890; July 2014; © 2014 Landes Bioscience

Annotation of long non-coding RNAs expressed in Collaborative Cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts

Laurence Josset[1,2], Nicolas Tchitchek[1,2], Lisa E Gralinski[2,3], Martin T Ferris[2,4], Amie J Eisfeld[5], Richard R Green[1,2], Matthew J Thomas[1,2], Jennifer Tisoncik-Go[1,2], Gary P Schroth[6], Yoshihiro Kawaoka[5], Fernando Pardo-Manuel de Villena[4], Ralph S Baric[2,3], Mark T Heise[2,4], Xinxia Peng[1,2], and Michael G Katze[1,2*]

[1]Department of Microbiology; School of Medicine; University of Washington: Seattle, WA USA; [2]Pacific Northwest Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research: Portland, OR USA; [3]Department of Epidemiology; University of North Carolina-Chapel Hill; Chapel Hill, NC USA; [4]Department of Genetics; University of North Carolina-Chapel Hill; Chapel Hill, NC USA; [5]Department of Pathobiological Sciences; Influenza Research Institute; University of Wisconsin-Madison; Madison, WI USA; [6]Illumina, Inc.; San Diego, CA USA

Keywords: long non-coding rna, influenza virus, sars-cov, rna-seq, interferon, collaborative cross.

The outcome of respiratory virus infection is determined by a complex interplay of viral and host factors. Some potentially important host factors for the antiviral response, whose functions remain largely unexplored, are long non-coding RNAs (lncRNAs). Here we systematically inferred the regulatory functions of host lncRNAs in response to influenza A virus and severe acute respiratory syndrome coronavirus (SARS-CoV) based on their similarity in expression with genes of known function. We performed total RNA-Seq on viral-infected lungs from eight mouse strains, yielding a large data set of transcriptional responses. Overall 5,329 lncRNAs were differentially expressed after infection. Most of the lncRNAs were co-expressed with coding genes in modules enriched in genes associated with lung homeostasis pathways or immune response processes. Each lncRNA was further individually annotated using a rank-based method, enabling us to associate 5,295 lncRNAs to at least one gene set and to predict their potential *cis* effects. We validated the lncRNAs predicted to be interferon-stimulated by profiling mouse responses after interferon-α treatment. Altogether, these results provide a broad categorization of potential lncRNA functions and identify subsets of lncRNAs with likely key roles in respiratory virus pathogenesis. These data are fully accessible through the **MO**use **NO**n-**C**ode **L**ung interactive database (MONOCLdb).

Introduction

Influenza A virus (IAV) and severe acute respiratory syndrome coronavirus (SARS-CoV) are two respiratory pathogens that belong to independent viral families yet can cause similar acute lung disease.[1] In 2012–2013, the emergence of a novel IAV, the avian H7N9 virus, and of a novel human CoV, the Middle East respiratory syndrome coronavirus (MERS-CoV), has raised pandemic concerns and highlights the importance of deciphering general mechanisms of respiratory virus pathogenesis. Respiratory virus infection outcome is determined by a complex game between the virus and the host, the rules of which are not fully understood, but where the host-response can be more deleterious than the virus itself for inducing lung disease.[2] High-throughput methods have been used to globally characterize the host response to IAV and SARS-CoV infections and have revealed that the dynamics and magnitude of the innate immune response to infection, as well as immune cell infiltration, are crucial aspects of pathogenesis.[3-5]

Some potentially important host factors for the antiviral response, whose functions remain largely unexplored, are non-protein-coding RNAs (ncRNAs). There is an increasing number of different classes of these regulatory ncRNAs: small interfering RNA (siRNA), microRNA (miRNA), Piwi-interacting RNA (piRNA), promoter-associated small RNA (PASRs), small nucleolar RNA (snoRNA) and long non-coding RNAs (lncRNAs). lncRNAs are endogenous cellular RNAs that are mRNA-like in length (> 200 nt) but which lack any positive-strand open-reading frames longer than 30 amino acids. A recent review estimates the number of total lncRNAs is in the range of ~20,000 transcripts, but only about 200 lncRNAs have been characterized to date.[6] Known lncRNAs are involved in many complex

*Correspondence to: Michael G Katze; Email: honey@uw.edu
Submitted: 03/26/2014; Revised: 05/28/2014; Accepted: 06/03/2014; Published Online: 06/12/2014
http://dx.doi.org/10.4161/rna.29442

human diseases and regulate key cellular processes by a variety of molecular mechanisms. Among the most well studied lncRNAs, *Xist* and *Air* have been shown to epigenetically silence transcription by targeting chromatin-modifying complexes to particular genes in *trans* and *cis*, respectively.[7,8] Other lncRNAs act at the post-transcriptional level, such as *H19* lncRNA, which serves as the precursor for miR-675 to moderate cell growth,[9] and *Malat1*, which forms a molecular scaffold for several proteins present in nuclear speckles and which regulates pre-mRNA alternative splicing.[10]

Recently, several studies have identified lncRNAs as major players in the host-response to pathogens. Differential expression of lncRNA is observed in response to viral infection[11] and in immune cells after stimulation or differentiation.[12] In particular, we previously observed that 500 annotated and 1,000 novel lncRNAs are differentially expressed in mice after SARS-CoV infection.[13] About 40% of these changes were similarly observed in mice and mouse embryonic fibroblasts (MEF) infected with influenza virus A/PR/8/34 and in response to interferon (IFN) treatment.[13] A few lncRNAs have been functionally studied for their role in viral pathogenesis. For example, *Tmevpg1* (also known as *NeST*), is an antisense transcript distal to *IFNG* that is involved in Theiler's virus persistence and decreased *Salmonella enterica* pathogenesis, and it enhances *IFNG* gene expression by binding to the histone methyltransferase complex and altering histone 3 methylation at the IFN-γ locus.[14,15] *Neat1* is one of many lncRNAs induced by HIV-1 infection, and it serves as a scaffold for the nuclear paraspeckle substructure that can sequester some mRNAs in the nucleus.[16] Importantly, *Neat1* deficiency enhances HIV-1 replication.[17]

Identifying the role of all lncRNAs involved in the host-response to infection is especially challenging because of their large number and variety of functions. It has been hypothesized that lncRNAs function through their secondary structure rather than through their primary sequence.[6] However, there are currently no computational methods to reliably predict a single secondary structure for a single sequence of long RNA,[18] which could in turn be used to predict lncRNA function. In addition, minimal lncRNA expression, localization and interactome data are available, which also limits our understanding of lncRNA function. With the large amount of transcriptome data generated by high-throughput technologies, predicting gene function on the basis of expression is an attractive strategy for the characterization of novel or unannotated transcripts.[19] One approach for predicting the function of unknown genes is the 'guilt by association' approach, according to which genes with similar expression profiles are functionally associated. This strategy was successfully applied to 340 mouse lncRNAs after re-annotation of the Affymetrix Mouse Array using 34 data sets derived from diverse mouse tissues.[20]

Here, we expanded this approach by using total RNA-Seq to profile pulmonary transcriptomic responses in mice infected with either highly pathogenic IAV or SARS-CoV. Eight mouse strains with large genetic diversity and that constitute the Collaborative Cross (CC) founder strains[21] were infected with either mouse-adapted H1N1 influenza virus or with recombinant mouse-adapted SARS-CoV, providing a wide range of host transcriptional responses to two different respiratory viruses. We found that lncRNAs accounted for about 40% of total genes differentially expressed (DE) upon infection. Of these DE lncRNAs, 5,295 were functionally annotated using module-based and rank-based enrichment methods, with universal and ad hoc gene sets. To validate the lncRNAs predicted to be IFN-stimulated genes (ISGs) in the context of respiratory disease, we profiled mouse pulmonary transcriptomic responses after IFNα treatment by an independent total RNA-Seq experiment. We anticipate that our lncRNA annotation, entirely available through a user-friendly web interface, MONOCLdb (www.monocldb.org), will accelerate mechanistic characterization of lncRNA function(s) that are of general interest to the infectious disease and immunology fields.

Results

CC founder strains have a wide range of susceptibility to PR8 and MA15 infection

To systematically characterize lncRNAs involved in mouse pulmonary responses to respiratory virus infection, eight different strains of mice were infected intranasally with sublethal doses of highly pathogenic mouse-adapted IAV (PR8) or SARS-CoV (MA15) and the lungs used for transcriptome sequencing. These eight mouse strains – A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILt, CAST/EiJ, PWK/PhJ, and WSB/EiJ – represent the founder strains for the CC mouse resource project.[21] They were chosen for the CC resource because of their large genetic diversity and they have also been previously shown to have a wide range of susceptibility to PR8 infection.[22] Weight loss was monitored daily over the course of infection and we found that the CC founders had a wide range of morbidity after either PR8 or MA15 infection (**Fig S2A**). We also noted that the strains most susceptible to PR8 infection (C57BL/6J and A/J) were not the most susceptible to infection with MA15. While C57BL/6J and A/J mice lost the most weight at four days post infection [DPI] when infected with PR8, these two strains were regaining weight between three and four DPI when infected with MA15. The two mouse strains most susceptible to MA15 were PWK/PhJ and CAST/EiJ, but these strains had intermediate to low susceptibility to PR8 infection. In addition to the wide range of weight loss, the CC founders also supported PR8 and MA15 viral replication to different levels (**Fig S2B**). Overall, viral replication was not significantly correlated with weight loss after either MA15 or PR8 infection (**Fig S2C**). However, when considering samples at four DPI only, there was a significant correlation between weight loss and viral replication (p-value < 0.01), especially after MA15 infection (**Fig S2C**).

Global changes in lncRNA expression are as discriminative as changes in protein-coding gene expression

Whole-transcriptome analysis of the pulmonary response of all eight CC founder strains at two and four DPI was performed by total RNA-Seq to a high depth of sequencing (median: 50.3 million (M) total reads per sample) (**Fig S3**). After normalization and expression-based filtering, the distribution of $\log_2$ scaled

counts showed that the 12,211 lncRNAs that passed our criteria (see Methods) were generally expressed to a lower level than the 15,355 coding genes, with a median of 8.2 and 5.7 counts (in $\log_2$) per coding and non-coding genes, respectively (**Fig S4**). However, multidimensional scaling (MDS) representation of samples based on lncRNA expression (**Fig. 1A**) or on coding-gene expression (**Fig. 1B**) showed that lncRNA expression levels differentiated infection conditions as well as coding gene expression. In addition to clustering by infection condition, samples were clustered based on their genetic background, with three main clusters that were representative of mouse phylogenic origin: *M.m. domesticus* (WSB/EiJ, NOD/ShiLtJ, NZO/HILt, C57BL/6J, 129S1/SvImJ and A/J), *M.m. castaneus* (CAST/EiJ) and *M.m. musculus* (PWK/PhJ) (**Fig. 1A**). Notably, this clustering was less striking when based on coding-gene expression, with CAST/EiJ samples being closer to *M.m. domesticus* samples, indicating that lncRNA basal levels might be more strain specific than coding gene expression (**Fig. 1B**). In addition, the dynamic range of lncRNA expression following either MA15 or PR8 infection was as large as the coding gene expression range, though lncRNA expression levels were more downregulated while coding gene expression levels were more upregulated after infection (**Fig. 1C**). Finally, we found a large number of genes were DE after either MA15 or PR8 infection, with differences in the magnitude of response that depended on the mouse strain and virus (**Fig. 1D**). For example, PWK/PhJ mice, which were highly susceptible to MA15 infection, had up to 5,098 DE genes at four DPI but only 869 DE genes after PR8 infection. In contrast, C57BL/6J, WSB/EiJ and 129S1/SvImJ mice had more DE genes after PR8 infection compared with MA15 infection at four DPI, with, for example, 5,986 DE genes after PR8 infection of 129S1/SvImJ mice but only 926 genes after MA15 infection.

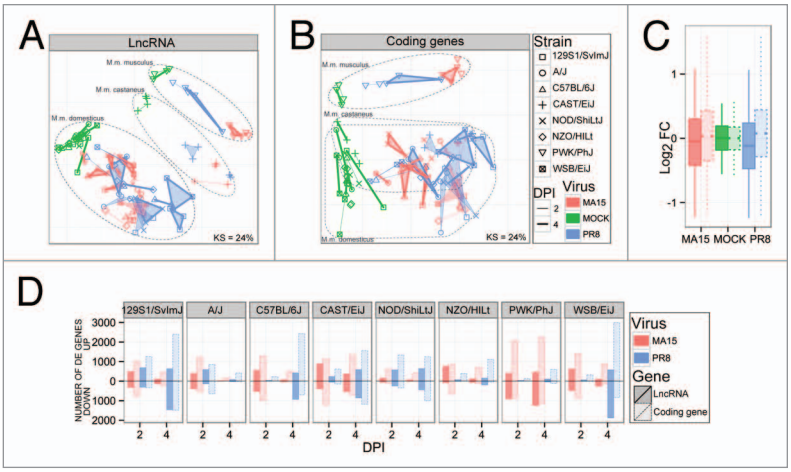

Figure 1. Characterization of lncRNA pulmonary expression in mice after infection with either IAV PR8 or SARS-CoV MA15. (**A-B**) Similarities in lncRNA (**A**) or coding gene (**B**) expression profiles are depicted using non-parametric multidimensional scaling (MDS). Each RNA sample is represented as a single point colored by viral treatment (green for mock-, salmon for MA15- and blue for PR8-infected samples), and with a different shape according to mouse strain. Convex hulls link samples belonging to the same condition, with different line width depicting the DPI. Euclidian distance was calculated using the normalized counts data for lncRNA passing QC (**A**) or coding genes passing QC (**B**), such that proximity indicates similarity, while distance indicates dissimilarity of gene-expression profiles. Kruskal's stress (KS) quantifies the quality of the representations as a fraction of the information lost during the dimensionality reduction procedure. (**C**) Dynamic range of expression after infection for lncRNA compared with coding genes. Boxplots represent the 5% and 95% quantile (lower and upper extreme whiskers), 25% and 75% (lower and upper hinges) and the median of gene expression changes after infection in $\log_2$FC, considering data for two and four DPI and for all eight mice strains together. (**D**) Number of differentially expressed (**DE**) lncRNA and coding genes after infection at each DPI and for each mouse strain (FDR = 1%). Dark colors represent lncRNAs and the light colors represent coding genes.

Importantly, lncRNAs accounted for about 40% of the total number of DE genes. In total, there were 8,270 coding DE genes and 5,329 non-coding DE genes in at least one condition. Notably, DE lncRNA were as strongly correlated with viral replication and morbidity as DE coding genes (**Fig S5**). Many DE coding and non-coding genes were highly correlated with viral replication, while the association with mouse weight loss was weaker. However, 62% of DE lncRNAs were negatively correlated with viral replication while only 42% of coding genes were positively correlated with viral replication (**Fig S5**), which was consistent with DE lncRNAs being more downregulated after infection compared with the coding genes. Altogether, these results show that while lncRNAs were on average slightly less expressed than the coding genes, their differential expression and dynamic range after infection and association with viral replication was just as strong.

DE lncRNAs are tightly co-expressed with DE coding genes

Genes sharing similar functions tend to be co-expressed.[20] To computationally characterize functions of DE lncRNAs, we determined whether they were co-expressed with DE genes of known functions. We first evaluated several parametric and non-parametric methods, including Pearson, Spearman, Kendall, maximal information coefficient (MIC), Hoeffding, distance correlation (dcor) and biweight midcorrelation (bicor) to determine the optimal method for detecting co-expressed coding genes sharing similar function (Supplemental Materials and **Fig S6**). The signed bicor metric outperformed other methods, especially for associating coding genes belonging to similar reactome pathways (**Fig S6**). We then computed pairwise correlation between DE genes using the signed bicor. We compared the distribution of bicor coefficient between pairs of lncRNAs or pairs of coding genes, or mixed pairs of coding and non-coding genes (**Fig S7**). The median bicor coefficient was similar between pairs of lncRNAs and coding genes, and coding genes were more likely to be strongly correlated together than were pairs of lncRNAs. On the other hand, a higher number of mixed pairs of coding and non-coding genes were highly negatively correlated than "pure" pairs of coding genes or of lncRNAs (**Fig S7**), consistent with their different trend of regulation after infection. Based on these pairwise correlations, a complete weighted network was inferred and 11 modules comprised of tightly co-expressed coding and non-coding genes were detected. These modules were classified arbitrarily by color names. **Figure 2A** shows that within each

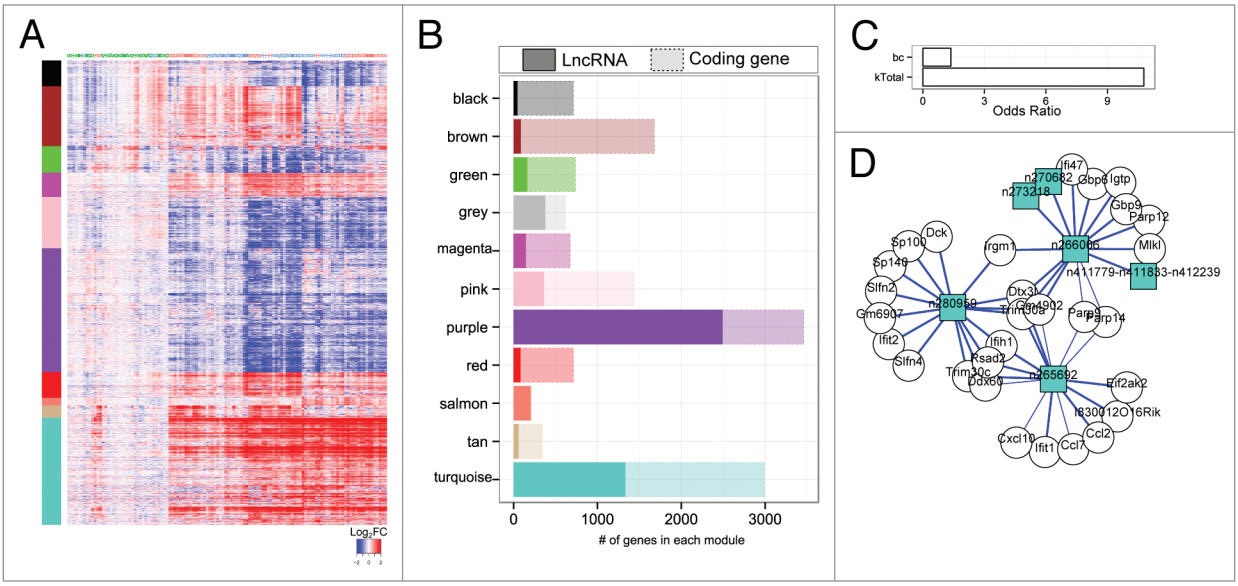

Figure 2. Modular annotation of lncRNA. (**A**) Heatmap depicting expression values for DE coding and non-coding genes. Samples were clustered by hierarchical clustering and represented by symbols similar to the ones used in **Figure 1A and B**. Genes were grouped into modules (co-expressed sets of transcripts), which were arbitrary labeled and depicted by different colors. (**B**) Number of coding and non-coding genes comprising each module. (**C**) Odds ratio of being a key point in the network given the gene is coding compared with non-coding. Key points are defined as bottlenecks: top 5% genes with highest betweeness centrality (bc); and hubs: top 5% genes with highest degree in the whole network (kTotal). (**D**) Example of lncRNA hubs within the turquoise module: n266006, n265692, and n280959. The turquoise module is enriched in ISGs (**Table 1**). For clarity, only the top 15 most correlated genes for each hub lncRNA are shown. lncRNAs are colored based on their MONOCLdb module membership and represented by square symbols, while coding genes are depicted as open circles, but please note that all genes in panel D belong to the turquoise module. This representation was generated using **MONOCLdb.**

module, gene expression levels changed very similarly after infection. Whereas the brown and salmon modules included 95% of coding genes or 94% of lncRNAs (respectively), the other modules included both coding genes and lncRNAs that were strongly co-expressed (**Fig. 2B**).

Module-based annotation provides a first level of annotation for lncRNAs and identifies lncRNAs with a central position in each module

To determine whether modules of co-expressed genes were associated with specific biological functions, we performed a functional enrichment analysis using several gene-sets from seven categories: three categories universally used in biology (GO Biological Process, Reactome pathways and TF binding motifs) and four categories relevant to respiratory virus pathogenesis (Immgen, GeneAtlas, ISGs and QTL determining MA15 and PR8 susceptibility) (**Table 1**, **Table S1**). The rationale for using Immgen and GeneAtlas gene-sets is that immune cells infiltrating the lungs contribute to respiratory virus pathogenesis and account for a large part of the pulmonary transcriptomic response observed after infection.[3] Therefore, we specifically determined whether co-expressed genes in immune cells (Immgen) or genes predominantly expressed in immune cells compared with lung epithelial cells (GeneAtlas) were enriched in each module. In addition, modules were also correlated with weight loss data and viral replication to determine their relevance during infection (**Table 1**). Enrichment for each module is described in Suppl Text. Some modules had specific expression patterns, depending on the mouse strain and infecting virus. For example, the green module, which was enriched in cytoskeleton and epithelial cilium functions, was specifically correlated with PR8 viral replication and was downregulated after PR8 infection in all founders except for NZO/HILt mice, which were resistant to PR8 infection (**Fig S8**). The turquoise module was the largest upregulated module, with 1,331 lncRNAs and 1,664 coding genes highly upregulated to different extents in all eight founders infected with either PR8 or MA15 (**Fig. 2A-B**). This module was highly enriched in ISGs and inflammatory/IFN related pathways, and enriched in genes with promoters containing the ISGF3 binding motif. The turquoise module was also the most highly correlated module with viral replication following either MA15 or PR8 infection.

Module functional enrichment allowed us to describe the global host-response network to either PR8 or MA15 infection. This also provided a primary level of annotation for lncRNAs belonging to each module. Moreover, an advantage of module definition was that we were able to determine which lncRNAs were highly connected in each module (intra-modular hubs) and which might regulate the module. Considering the whole network, we found that coding genes were more likely to be key points (hubs or bottlenecks) of the network than were lncRNAs (**Fig. 2C**). However, considering centrality within each module, we found that some lncRNAs were among the top intramodular hubs. For example, n280959, n266006 and n265692 were the most highly connected lncRNAs within the turquoise module (hub percentile ranks > 98%) and may have key roles in regulating the IFN response against viral infection (**Fig. 2D**).

Table 1. Functional enrichment for each module of co-expressed coding and non-coding genes

MONOCLdb module	GO	Reactome	GeneAtlas	Immgen	motif	ISG	QTL	Correlation with WL and viral replication (bicor)
black	GO:0007275_ multicellular organismal development (ES = 5.91)	Metabolism (ES = 13.85)	T-cells CD4+ (ES = 1.79)	#44: "Downregulated with differentiation, except some myeloids. High in stromal"(ES = 2.84)	SP1(MA0079.2) (ES = 12.82)		QTL_SARS_ eosinophilia (ES = 1.62)	SARS_MA15_vRNA (-0.63);PR8_vRNA (-0.37)
brown	GO:0006412_ translation (ES = 14.48)	Gene Expression (ES = 40.25)	B-cells follicular (ES = 1.84)	#4: "Ribosomal proteins" (ES = 4.03)	Klf4(MA0039.2) (ES = 32.6)		QTL_FLU_ Hrl4 (ES = 7.24)	SARS_WL (-0.42);SARS_MA15_ vRNA (0.3);PR8_WL (-0.41);PR8_vRNA (0.35)
green	GO:0007018_ microtubule-based movement (ES = 9.32)	Potassium Channels (ES = 3.63)		#37: "High in stromal and blood endothelial cell" (ES = 2.9)	Rfx4_primary (UP00056_1) (ES = 15.88)			PR8_WL (0.47); PR8_vRNA (-0.72)
grey	GO:0042113_B cell activation (ES = 2.56)	GPCR downstream signaling (ES = 2.78)	Bcells common (ES = 7.81)	#33: "Early B module" (ES = 4.61)	Otx1_2325.1 (UP00229_1) (ES = 3.91)			SARS_WL (0.37)
magenta	GO:0042254_ ribosome biogenesis (ES = 8.45)	Gene Expression (ES = 24.61)	Mast cells (ES = 1.75)	#5: "Downregulated with differentiation" (ES = 10.38)	GABPA (MA0062.2) (ES = 5.12)			SARS_WL (-0.71); SARS_MA15_vRNA (0.84); PR8_WL (0.49); PR8_vRNA (0.71)
pink	GO:0008152_ metabolic process (ES = 10.29)	Metabolism (ES = 21.02)		#35: "Endothelial genes, extracellular matrix " (ES = 7.16)	SP1(MA0079.2) (ES = 11.13)			SARS_WL (0.59); SARS_MA15_vRNA (-0.92); PR8_WL (0.45); PR8_vRNA (-0.52)
purple	GO:0009404_ toxin metabolic process (ES = 1.43)	Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell (ES = 1.72)	DC lymphoid (ES = 1.77)	#36: "Fibroblasts genes, extracellular matrix " (ES = 7.38)	Hoxa10_2318.1 (UP00217_1) (ES = 15.63)		QTL_FLU_ Hrl3 (ES = 2.89)	SARS_WL (0.67); SARS_MA15_vRNA (-0.85); PR8_WL (0.52); PR8_vRNA (-0.7)
red	GO:0007165_ signal transduction (ES = 4.75)	Immune System (ES = 25.77)	T-cells foxP3+ (ES = 2.47)	#25: "Low in T cells, intermediate in B cells, high in myeloids" (ES = 3.46)	Klf4(MA0039.2) (ES = 11.09)	ISG (ES = 2.77)	QTL_FLU_ Hrl4 (ES = 1.53)	SARS_WL (-0.67); SARS_MA15_vRNA (0.8); PR8_WL (-0.52); PR8_vRNA (0.68)
salmon	GO:0031123_ RNA 3′-end processing (ES = 1.34)		B-cells marginal (ES = 2.2)	#1: "Downregulated in myeloids and stromal" (ES = 2.33)	Foxl1_ secondary (UP00061_2) (ES = 6.95)			SARS_MA15_vRNA (0.68); PR8_vRNA (0.39)
tan	GO:0051301_ cell division (ES = 48.25)	Cell Cycle (ES = 71.45)	Macrophage BM_0hr (ES = 15.16)	#11: "cell cycle genes" (ES = 75.87)	E2F2_ secondary (UP00001_2) (ES = 6.36)			SARS_WL (-0.55); SARS_MA15_vRNA (0.49); PR8_WL (-0.52); PR8_vRNA (0.39)
turquoise	GO:0006955_ immune response (ES = 23.55)	Immune System (ES = 39.88)	Macrophage common (ES = 10.6)	#52: "Interferon response" (ES = 12.59)	Isgf3g_primary (UP00074_1) (ES = 6.55)	ISG (ES = 87.9)	QTL_SARS_ eosinophilia (ES = 1.92)	SARS_WL (-0.51); SARS_MA15_vRNA (0.95); PR8_WL (-0.39); PR8_vRNA (0.76)

ES, Enrichment score (ES) defined as $-\log_{10}$ p-value calculated by exact Fisher's test.

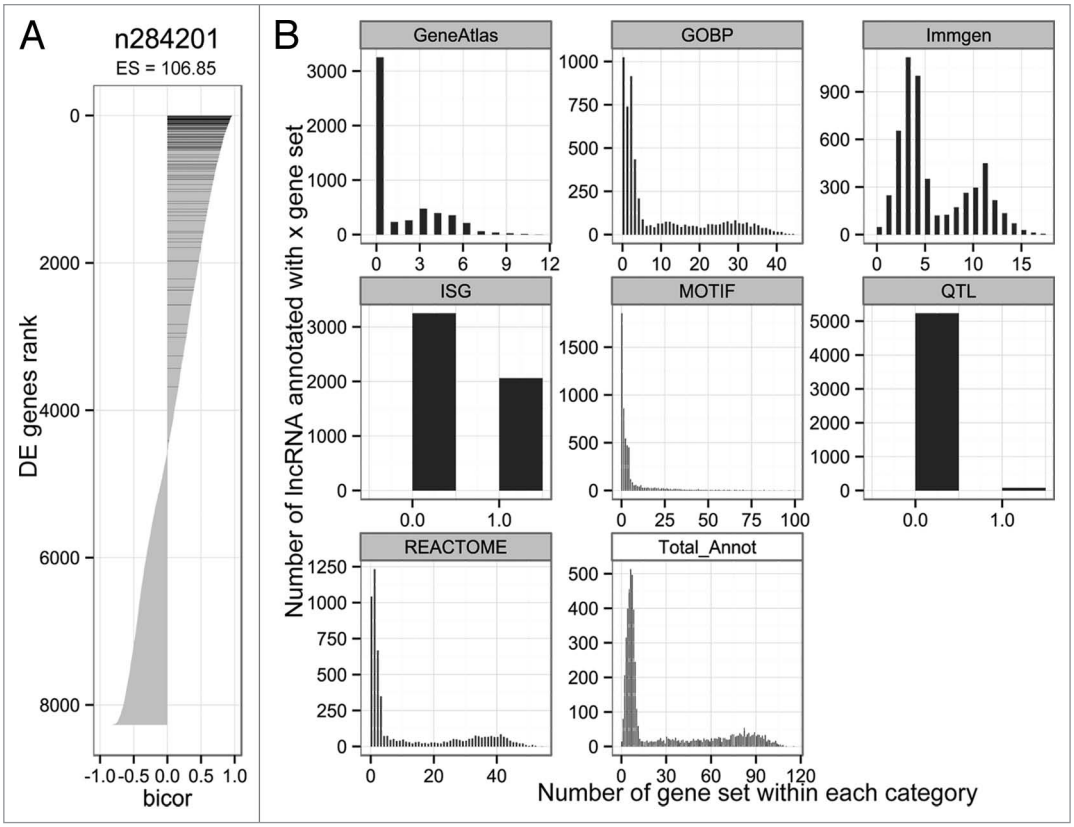

Figure 3. Individual lncRNA annotation based on ranked correlation. (**A**) Example of ranked-correlation annotation for n284201. DE genes are ranked based on their bicor coefficient with n284201 and colored in black for ISG and grey for not ISG. Functional enrichment was performed with the Wilcoxon Rank-Sum (WRS) test, which defined whether genes from one gene-set are significantly found at the top of the list. Enrichment score (ES) is defined as $-\log_{10}$ (Bonferoni adjusted p-value) for n284101 was highly significant (ES = 110) and therefore n284101 was annotated as an ISG. (**B**) Distribution of the ranked annotation in each functional category. "GeneAtlas" gene-sets were defined as genes highly expressed in immune cell populations compared with lung profiles in GeneAtlas, "GOBP" gene-sets are the Gene Ontology Biological Processes, "Immgen" gene-sets are modules of co-expressed genes across various immune cell types as defined in the Immgen project, "ISG" is a list of IFN response genes, "Motif" gene-sets are lists of genes whom promoters have TF motif binding sites, "QTL" gene-sets are QTL regions identified for susceptibility of SARS or IAV in the CC mice, and "Reactome" are reactome pathways. Finally, "Total_annot" is the sum of GeneAtlas, GOBP, Immgen and Reactome annotations.

Rank-based annotation reveals that most lncRNAs are associated with a few functions but a few lncRNAs might have pervasive functions

To more precisely predict lncRNA functions, we used a second method referred to as "rank-based annotation." The principle of this method is illustrated in **Figure 3A** for n284201. DE genes were ranked based on their bicor coefficient with n284201. We then used the Wilcoxon-Rank Sum (WRS) test to determine whether genes from a given gene-set, ISGs in our example **Figure 3A**, were significantly found in the top of the list (i.e., positively correlated with n284201). The enrichment score (ES), determined as $-\log_{10}$ of the Bonferroni adjusted p-value, was highly significant (ES = 110), therefore predicting that n284201 may be an ISG.

We performed this annotation for all DE lncRNA and for all gene-sets. The results of this annotation can be retrieved in www.monocldb.org where we provide the ES and percentile rank (PR) based on the p-value of the lncRNA for each function. It is therefore possible to know which lncRNA was found to be most highly enriched in any given gene-set (in the lowest PR). Using the Bonferroni adjusted p-value < 0.05 as the cutoff for significance, we determined how many lncRNAs were associated with one or more functions (**Fig. 3B**). About 1,000 lncRNAs were not significantly associated with any GO biological process (BP) or any Reactome pathway, but 1,232 lncRNAs (23%) were significantly enriched in one pathway and 915 lncRNAs were enriched in two GO processes. Notably, a handful of lncRNAs were associated with more than 40 BPs or pathways and could have more pervasive functions, similarly to DE coding genes. For example, *Mapk3* or *Cdk1* DE genes belong to more than 40 Reactome pathways.

"Motif" gene-sets were used to determine whether some lncRNA might be tightly co-regulated with genes having similar TF binding motifs in their promoter. Among the 3,454 lncRNAs positively correlated with genes sharing one or more motifs, 976 lncRNAs (28%) had one of these motifs in their promoter, including several lncRNAs with interferon regulatory factor (IRF) binding motifs (**Table S2**). This implies that lncRNA could be co-regulated with a group of coding genes by specific transcription factors. Looking at genes that were highly expressed

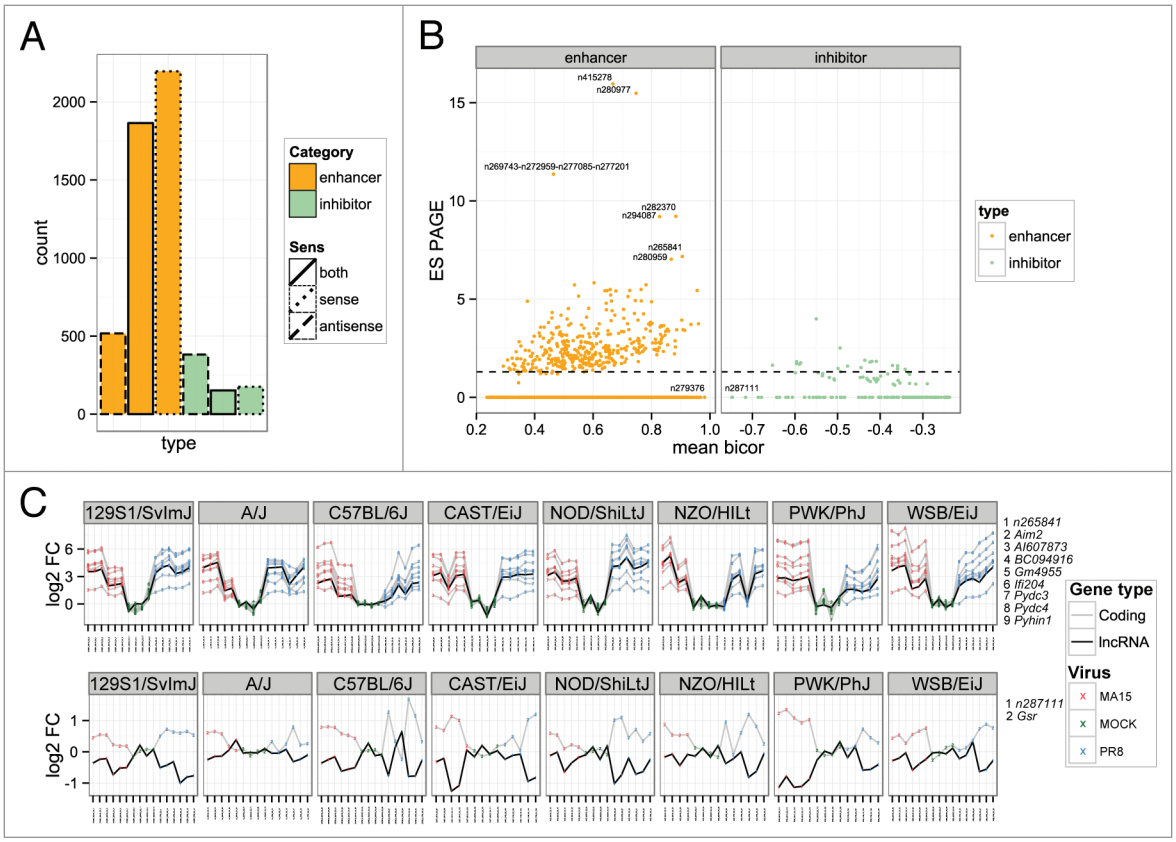

Figure 4. Prediction of potentially *cis*-acting lncRNAs. (**A**) Number of lncRNA positively (enhancer-like function) or negatively (inhibitors) correlated with neighbor coding genes (within 200 kb) considering all genes regardless of their strand (both), or only genes on the same strand as the lncRNA (sense) or on the opposite strand (antisense). (**B**) Specificity and strength of *cis* lncRNA correlation with neighbor genes, regardless of their strand. ES PAGE were defined as $-\log_{10}$ p-value calculated by PAGE test which assess whether neighbor genes were among the most positively correlated (for enhancer-like lncRNA) or negatively correlated (for inhibitor lncRNA) genes. ES PAGE was calculated only for lncRNAs with more than 3 coding neighbors; otherwise this score was set arbitrarily to 0. The x-axis represents the arithmetic mean of bicor coefficient between a given lncRNA and all its coding neighbor genes. lncRNAs with the highest specificity for correlation with coding neighbor genes, or the most correlated with their neighbor genes (mean bicor) are indicated with their names. Similar plots for lncRNA specificity for antisense or sense neighbors are depicted **Fig S11**. (**C**) Expression levels (in Log_2FC) of n265841, n287111, and their neighbor genes, across the different CC founder mice and viral conditions.

in immune cells compared with whole lung (the "GeneAtlas" category), we determined that 2,056 lncRNAs might be associated with immune cell infiltration. Most of these lncRNAs were upregulated after both PR8 and MA15 infection in all CC founder strains and belong to the turquoise module (**Fig S9**). Immgen gene-sets include co-expressed genes in immune cells as well as in fibroblasts, endothelial cells or the extracellular matrix, therefore it was not surprising that only 67 lncRNAs were not associated with one Immgen module, compared with 3,273 lncRNAs not associated with any GeneAtlas category. Finally, 2,059 DE lncRNAs (39%) were predicted to be ISGs. In total, we were able to associate 5,295 out of the 5,329 DE lncRNAs with at least one gene-set.

Potential *cis*-regulatory lncRNAs were mostly positively correlated with coding-gene neighbors

Some lncRNAs have been reported to have *cis*-regulatory effects on multiple flanking genes. To determine potential *cis*-acting lncRNAs, we analyzed the correlation of each lncRNA with its coding-gene neighbors (**Fig. 4**). We defined as potential *cis*-acting lncRNA the genes whose neighbors were all significantly positively correlated (lncRNAs with potential transcriptional "enhancer-like" function) or all significantly negatively correlated (lncRNAs with potential transcriptional "inhibitor" function), regardless of the chromosome strand or considering coding genes that were on the same strand (sense) or on the opposite strand (antisense) (**Fig. 4A**). Considering all neighbor coding genes, a large number of lncRNAs (1864; 35%) were classified as potential *cis* enhancer-like while only 152 lncRNAs (3%) were classified as potential *cis* inhibitors (**Fig. 4A**). However, enhancer-like lncRNAs were mostly on the same strand as positively correlated neighbors while inhibitor lncRNAs were mostly on the opposite strand as negatively correlated neighbors (**Fig. 4A**). Most of the *cis*-acting lncRNAs had only one DE coding gene neighbor, while most of the *trans*-acting lncRNAs had no DE coding gene neighbors (**Fig S10**).

We performed the same analysis on DE coding genes for comparison (**Fig S10**). In constrast to lncRNAs, there was no coding gene negatively correlated with all of its neighbor genes or with sense neighbor genes. However, similar to *cis* enhancer-like lncRNA, most of the *cis* enhancer-like coding genes had one

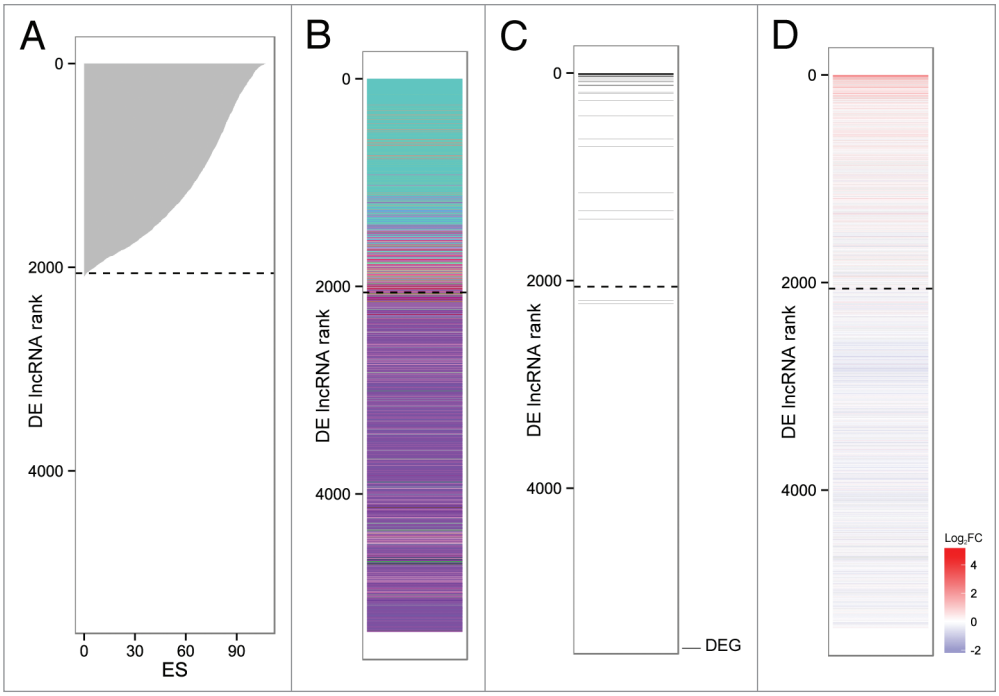

Figure 5. Validation of ISG annotation. (**A**) Enrichment score (ES) of each lncRNA for ISG annotation. Dashed line indicates the rank above which lncRNAs had a significant ES > 1.3. (**B**) Module membership for each lncRNA ranked as in panel A. Each line represent a lncRNA colored based on its MONOCLdb module membership (**C**) lncRNAs that were found DE in an additional RNA-Seq data set of mice treated with IFN-α are displayed with black lines. (**D**) Expression level for each ISG in C57BL/6J mice treated with IFN over untreated mice is depicted in a blue to red gradient. In B, C and D, lncRNA are ranked as in panel A, based on their ES for ISG annotation. Top ranked lncRNA were highly and significantly upregulated in mice treated with IFN.

Figure 6 (Opposite page). MONOCLdb. (**A**) Presentation of the MONOCLdb pipeline. Users can select lncRNAs by: noncode ID (e.g., "n424068"), GO term found significantly enriched with the rank-based annotation (e.g., "GO:0007010"), Immgen Coarse module number found significantly enriched with the rank-based annotation (e.g., "Immgen_Coarse.module_28"), Ensembl gene ID of most correlated coding-genes (e.g., "ENSMUSG00000029088"), or Ensembl gene ID of chromosomic neighbor (within 200 kb) coding-genes (e.g., "ENSMUSG00000030921"). (B-G) Examples of figures generated by MONOCLdb after query with: n424068 (*Neat1*), n424069 (*Neat1*), n177784 (*Malat1*), n424043 (*Adapt33*), and n424044 (*Adapt33*). For simplification, we have replaced the MONOCLdb lncRNA gene names by their symbol. (**B**) Expression heatmap. Expression values of lncRNAs in Log_2FC in PR8- and MA15-infected mice are displayed as a green to red gradient (saturation levels: log_2FC from -2 to 2) (mean of biological replicate). (**C**) Module-based enrichment. Module membership is depicted by a set of colored squares with functional description of each module on the top. The second set on the right displays percentile rank (PR) of intramodular degree and betweeness centrality with a yellow to blue gradient. High PRs in dark blue indicate intramodular hubs and bottlenecks. (**D**) Pathogenicity Association. Bubble plot showing the correlation between lncRNA expression and phenotypic data. The size of each bubble is relative to the absolute bicor coefficient, with green indicating anti-correlation and red positive correlation. (**E**) Genomic Co-Expression Network. Genomic network showing the top 15 most correlated genes with each queried lncRNA (|bicor| > 0.7). The position of each lncRNA in the chromosomic circle is relative to its coordinate (middle of the gene). lncRNA classified as potential *cis* lncRNA are represented in blue while *trans* lncRNA are in purple. (**F**) Rank-based Enrichment. Radial plot showing results of rank-based enrichment for Neat1 in Reactome pathways. Distance from the center to each edge is relative to the enrichment score (ES) defined as $-log_{10}$ Bonferoni corrected p-value of WRS test. (**G**) Co-expression network. Relationships between each queried lncRNA and their top 15 most correlated genes (|bicor| > 0.7) are represented as a network with yellow edges indicating negative correlation and blue edges indicating positive correlation. Coding genes are depicted as circles and non-coding genes as squares. lncRNAs are colored based on their module membership.

coding gene neighbor while most of the *trans* coding genes had two or three coding gene neighbors (sense and both strands, respectively) (**Fig S10**). The specificity of correlation with neighbor coding genes of lncRNAs with more than two neighbor coding genes was determined by PAGE (**Fig. 4B**). We found that enhancer-like *cis* lncRNAs were more specifically and strongly associated with coding neighbor genes than potential *cis* inhibitors (**Fig. 4B**, **Fig S11**). However, we did not find any *cis*-acting lncRNA specifically associated only with its neighbor coding genes, indicating that it might be difficult to untangle direct and indirect effects of *cis*-acting lncRNAs from in vivo experiments.

Figure 4C depicts the expression values for a *cis* enhancer-like lncRNA, n265841, and its sense and antisense coding neighbors, and an example of a *cis* inhibitor lncRNA, n287111, and its coding neighbor.

Validation of predicted IFN-stimulated lncRNAs

Figure 5 shows lncRNAs that were predicted to be ISGs by the rank-based annotation method (**Fig. 5A**) and by the module-based method (**Fig. 5B**). There was good agreement between the two methods, with lncRNAs mostly associated with ISGs belonging to the turquoise module (**Fig. 5A and B**). To validate our functional predictions, we performed an independent experiment

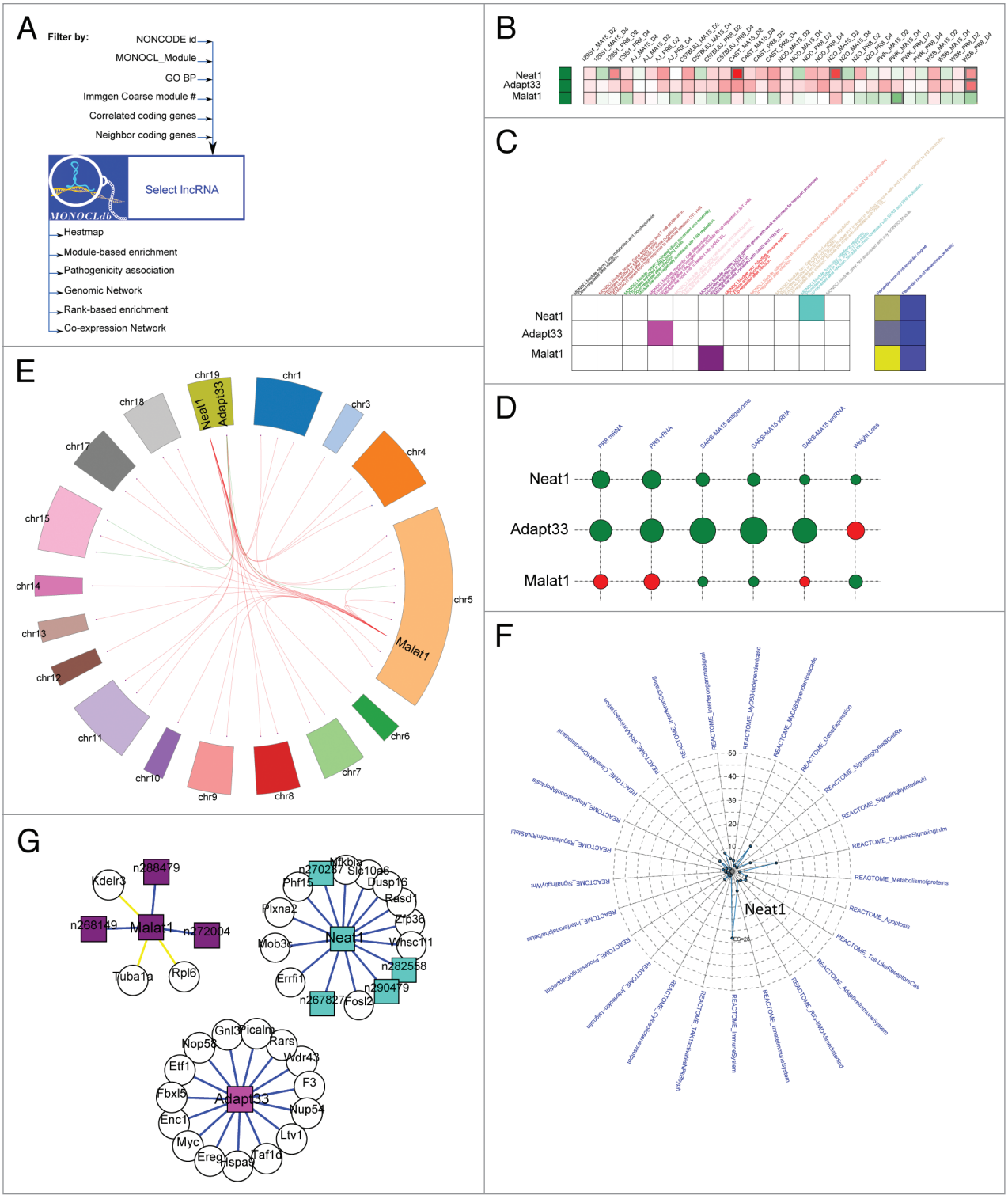

Figure 6. See opposite page for figure legend.

by treating C57BL/6J mice with IFN-α. Whole transcriptome pulmonary responses were determined at 12 h post-treatment by total RNA-Seq and statistical analysis was performed to identify DE coding and non-coding genes in IFN-α treated mice compared with mock treated mice. We did not observe any immune cell infiltration by hematoxylin and eosin staining at this time point (data not shown) and significantly induced genes were consequently defined as ISGs. In our experimental conditions at 12 h post-treatment, we found only 240 significantly upregulated genes after IFN treatment, including 187 coding genes and 53 lncRNAs. lncRNAs that were upregulated after IFN treatment, depicted in black in **Figure 5C**, were significantly enriched in the

top of the list of predicted IFN-stimulated lncRNAs. Other predicted IFN-stimulated lncRNAs that were not found DE in the lungs of IFN-treated C57BL/6J mice were mostly upregulated but did not pass the statistical threshold (**Fig. 5D**). It is possible that these lncRNAs would be more upregulated following IFN treatment of other CC founder strains. Finally, the top predicted IFN-stimulated lncRNAs (*i.e.*, lncRNA with lowest p-values of enrichment in ISG = in the lowest PR for enrichment in ISG) were the most significantly highly induced after IFN treatment (**Fig. 5D**). This indicates that p-values of enrichment (or PR) were predictive of function and that a higher confidence in functional annotation should be placed in enrichment within lower PR.

The MONOCL database

We provide an interactive database (the MOuse NOnCode Lung database – MONOCLdb, www.monocldb.org/) that allows users to query and analyze sets of lncRNA in the context of respiratory virus infection from our study (**Fig. 6**). Using this web portal, users can select lncRNAs by the following means: NONCODE identifiers, inferred associated MONOCLdb co-expression modules, inferred associated GO terms, inferred associated IMMGEN modules, or by neighbor coding genes. MONOCLdb can then be used to produce figures and raw files for expression values, module-based enrichment, rank-based enrichment, co-expression network, genomic network, and phenotypic data associations. **Figures 2 and 6** provide example illustrations of the MONOCLdb web-interface with a subset of available interfaces. All of the different produced charts and figures are interactive and user-friendly. Users can easily download the figures (svg files) as well as the raw results table (txt files).

A HTTP web service is also available on MONOCLdb allowing automatic retrieval of analysis result tables and images via a specific web URL (www.monocldb.org/content/web-service). We provide an example of R script code for automatic querying of the MONOCL database via this URL. Further details about MONOCLdb automatic querying can be found in the "About" section of the MONOCLdb web-portal (www.monocldb.org/content/about).

A Distributed Annotation System (DAS) service is also available for visualization of lncRNA annotations. DAS is a protocol for requesting and returning annotation data for genomic regions and can be integrated into a large variety of genome browsers. We provide a DAS track for the GRCm38 mouse genome that allows retrieval of lncRNAs described in the present study by using their genomic positions. For each mouse lncRNA, the DAS service provides the main annotations as well as a link redirecting to the MONOCLdb website for more specific details.

Discussion

lncRNAs are increasingly implicated in infectious disease, however, only a few have been functionally characterized for their role during viral infection. Here we quantified the expression of 20,728 mouse lncRNA genes, 5,329 of which were differentially expressed after IAV or SARS-CoV infection. Using a 'guilt-by-association' approach to annotation, 5,295 lncRNAs were characterized by at least one gene set. This greatly expands the work of Liao et al., who used similar methods to characterize the lncRNAs present on the Affymetrix Mouse 430 2.0 array and to annotate 340 mouse lncRNAs based on their expression in 34 data sets.[20] While Liao et al. included diverse tissues and biological conditions, they did not include viral infection and had only one data set derived from a bacterial infection (*Aeromonas spp.* infected intestinal cells). In the present study, we focused on characterizing lncRNAs that were involved in respiratory virus pathogenesis. In terms of methodology, we used both a module-based and a rank-based annotation. However, whereas Liao et al. only considered the top 0.05 percentile first degree of each lncRNA for their "hub-based method," we did not threshold the correlated genes but rather used the whole weighted network by performing a ranked functional enrichment for each lncRNA. We found that thresholding the first degree of each lncRNA gave enrichment results that were highly dependent on the cutoff used, and consequently we chose a method that was independent of any threshold and which was more robust.

In addition, we used several data-driven gene sets for functional enrichment that were relevant to our focused biological question, as genes co-expressed or specific to immune cells, ISGs, or QTL determining susceptibility to IAV and SARS-CoV infection. Our rationale for using immune cell or IFN-related gene sets was that it was previously shown that pulmonary transcriptional changes after IAV infection are driven mainly by the IFN response and immune cell infiltration.[3,23,5] These different levels of annotation may help characterize important lncRNAs relevant for infectious disease. Prioritization for functional characterization of lncRNAs should also consider correlation of expression with viral replication and weight loss, and potential key position (hub or bottleneck) within a network.

It was surprising that 57% of DE lncRNAs vs. 40% of DE coding genes belonged to modules mostly downregulated after infection (black, green, pink and purple modules). The four downregulated modules were enriched in genes associated with metabolism, development, transport processes, and the cytoskeleton. A higher proportion of down- vs. upregulated lncRNA was observed previously in SARS-CoV infected mice[13] and in TNFα stimulated MEFs.[24] It was also shown that lncRNAs have higher tissue specificity than coding genes.[25] Decreased expression of lung-specific lncRNAs might thus be explained by pulmonary cell death induced by infection, or by a relative decrease in the number of lung cells after immune cell infiltration. Alternatively, some of the downregulated lncRNAs might be highly expressed in normal cells to maintain homeostasis and downregulated following infection.

Among the downregulated DE lncRNAs, only two have been previously described: *Mrhl* (n342983) and *Malat1* (n177784). They both belonged to the purple module, which is enriched in genes from Immgen_Coarse.module_36 (ES = 7.38) specific to fibroblasts and non-immune stromal cells, and they were downregulated to different levels according to mouse strain and infecting virus. *Malat1* is an abundant nuclear lncRNA localized in nuclear speckles and has been described as a regulator of

gene expression governing hallmarks of lung cancer metastasis.[26] *Malat1* depletion results in the activation of p53 and its target genes[27] and its downregulation during infection could therefore activate the p53 pathway. In our study, *Malat1* was highly negatively correlated with genes coding for 60S ribosomal protein L6 (*Rpl6*), the endoplasmic reticulum protein retention receptor (*Kdelr3*) and tubulin α1 (*Tuba1a*), while several lncRNAs were positively correlated with *Malat1* and could have similar functions in nuclear speckles.

On the other hand, 36% of DE lncRNAs belonged to modules of genes upregulated after infection (brown, magenta, red, salmon, tan, and turquoise modules). These modules were enriched in immune cell proliferation or differentiation, the IFN response and pro-inflammatory pathways. Using rank-based enrichment, we found that most of the upregulated genes were associated at different levels with the IFN response and with genes specific to immune cells. We validated this annotation by performing additional experiments using mice treated with IFN-α, and we showed that the rank of enrichment for each gene set was an important parameter of functional prediction, with lncRNAs with lowest p-values of enrichment for ISGs being significantly upregulated after treatment with IFN-α.

At the module level, we found that the turquoise module was highly enriched in ISGs. Several lncRNAs (n280950, n266006, n265692 for example, **Figure 2D**) were highly connected in this module (hubs) and could therefore have a role in controlling the IFN response. These three lncRNAs: n265692 (*AK156844*), n280959 (*AK080205*), n266006 (*AK156398*) have not been described previously to our knowledge. n265692 has a motif for ISGF3G (aka IRF9) in its promoter and rank-based enrichment revealed a significant co-regulation with genes also having a binding motif for IRF9. A total of 177 other lncRNAs were co-regulated with genes sharing IRF3, IRF4, IRF5 and/or IRF9 binding motifs and had these motifs in their promoter. IRFs are major transcription factors regulating the IFN response. This observation implies that some IFN-stimulated lncRNAs may be induced by the same pathway as protein-coding ISGs. The two other hub lncRNAs (n280959, n266006) were co-regulated with genes having an IRF binding motif, although they did not have such motifs in their promoter. However, these lncRNAs had binding motifs for other TFs implicated in regulation of inflammatory response (including *Stat3* for n266006, and *Klf4* for n280950).

Among the few known lncRNAs that were DE after infection, *Neat1* (n424068 and n424069) was significantly upregulated in PR8-infected 129S1/SvImJ and WSB/EiJ mice and MA15-infected CAST/EiJ and NZO/HILt mice (**Fig. 6**). *Neat1* is a scaffold for nuclear paraspeckles formation and is upregulated after HIV infection and can sequester some HIV mRNAs.[28] Here we found that *Neat1* belonged to the turquoise module and the rank-based annotation predicted it was among the 17% top predicted ISGs. In addition, *Neat1* was highly enriched in pathways related to defense response to virus, innate immune response, and inflammatory response. Other known lncRNAs that were DE after infection included *Adapt33* (n424043-n424044), which was slightly upregulated in PR8-infected WSB/EiJ mice and belonged to the magenta module enriched in cell differentiation genes (**Fig. 6**). We found that this transcript was negatively correlated with both IAV and SARS-CoV replication and highly correlated with several stress and cell-cycle coding genes (*Hspa9* and *Myc*). Reactome pathways that were associated with *Adapt33* with lowest PR by rank-based analysis included tRNA aminoacylation (PR = 1.5%), regulation of apoptosis (PR = 3.2%), and innate immune system (PR = 5.4%). Interestingly, *Adapt33* was previously described as a stress-inducible riboregulator correlated with the apoptosis response,[29] but it has never been described in the context of infectious disease.

It is important to note that we were able to annotate lncRNA functions in the context of respiratory infection thanks to the diverse response of the CC founder mice to SARS-CoV and IAV infection. We observed that the eight CC founder mice had a large range of phenotypic response to infection, associated with a large difference in the magnitude of the transcriptomic response. We have previously shown that NZO/HILt and PWK/PhJ resistance to PR8 infection was due to the dominant gene *Mx1*, that acts in the context of IAV infection but not SARS-CoV infection.[22] The present study sheds light on other genes that may be involved in IAV and SARS-CoV susceptibility. Specifically, 34 of 210 lncRNAs that were found in regions controlling SARS-CoV [Gralinski et al., in preparation], and 53 of 296 lncRNAs in regions controlling IAV resistance,[22] were DE. None of these lncRNAs has been previously functionally described. Among this very rich list, an interesting lncRNA to further explore is n268833 (*AK142945*), which belongs to the QTL *HrI2* (Host response to Influenza).[22] This lncNRA was significantly upregulated after PR8 infection in all CC mice except CAST/EiJ, NZO/HILt and PWK/PhJ, which were the three strains the most resistant to PR8 infection. The expression of n268833 was highly correlated with PR8 replication, belonged to the turquoise module, and was strongly positively correlated with *IL-18*. Among the list of DE lncRNAs present in QTL controlling SARS-CoV resistance, n276032 (*AK047596*), n290720 (*AK017435*) and n292484 (*AK132900*) were specifically upregulated in CAST/EiJ mice infected by MA15, which had the highest viral replication and the most weight loss. n276032 is in the QTL associated with SARS-CoV titer [Gralinski et al., in preparation], was annotated in the turquoise module and was enriched in innate immune pathways by the rank-based method. These results suggest that some lncRNAs might control mouse genetic susceptibility to respiratory viruses, and highlight the richness of this data set to mine from different angles to further hypothesis generation and an understanding of respiratory virus pathogenesis and lncRNA functions.

To conclude, we have greatly expanded the available annotation of lncRNAs and described the significant regulation of 5,329 lncRNAs (most of which have not been described previously) after infection of mice with IAV or SARS-CoV. We provide the scientific community with a database (MONOCLdb) to easily retrieve expression values and annotation of any given lncRNA. In addition, we generated a large RNA-Seq data set, with gene-expression profiles from 120 CC founder mice. This represents a valuable resource for mouse genomic studies and for the Collaborative Cross. We expect that this work will help to

design the experimental characterization of important lncRNAs and will accelerate general knowledge about lncRNA functions. In particular, mechanistic characterization of lncRNAs predicted to belong to the IFN response would have a broad impact for immunology and infectious disease fields.

Materials And Methods

Animals

Eight-to-16-wk-old female animals from the eight CC founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILt, CAST/EiJ, PWK/PhJ, and WSB/EiJ) originally from the Jackson Laboratory (jax.org) were bred at UNC Chapel Hill under specific pathogen free conditions. All experiments were approved by the UNC Chapel Hill Institutional Animal Care and Use Committee.

Virus and cell lines

The mouse-adapted influenza A strain A/PR/8/34 (H1N1) [PR8] or recombinant mouse-adapted SARS-CoV (MA15) were used for infection studies. PR8 virus was grown in 10-d-old embryonated chicken eggs and titered on MDCK cells, as previously described.[30] SARS-CoV MA15 was propagated and titered on Vero E6 cells.[31]

Infections

Animals were anesthetized via inhalation of isoflurane (Piramal, Bethlehem, Pa) and subsequently infected intranasally with 5 × 10^2 pfu of PR8 or 10^4 PFU of MA15 in 50 μL of phosphate buffered saline (PBS), while mock infected animals received 50 μL of PBS. Animals were assayed and scored daily for morbidity (determined as percent weight loss), mortality and clinical disease. At two or four days post infection [DPI], animals (n = 2–3 for infected conditions, n = 2 for mocks) were euthanized via isoflurane overdose and cardiac puncture and lungs were harvested and used for total RNA-Seq and viral titration.

IFN treatment of MEF cells and mice

Mouse embryonic fibroblast (MEF) cells derived from the eight CC founder strains were treated individually with either mouse recombinant IFN-α4 (50 U/ml; PBL InteferonSource 12110–1), or IFN-β (100 U/ml; PBL InterferonSource 12400–1). After 16 h, MEF cells were washed once with 1X Dulbecco's phosphate buffered saline (D-PBS) and cell lysates collected in 500 μL of QIAzol Lysis Reagent for total RNA extraction. Gene expression was measured using 4X44K Mouse Whole Genome Gene Expression Microarrays (Agilent Technologies).

Six-week-old female C57BL/6J mice were intranasally treated with 10,000 units of recombinant IFN-α (Universal Type I IFN, Recombinant Human IFN-α A/D [BglII], R&D Systems) dissolved in endotoxin-free phosphate-buffered saline (EF-PBS), or with EF-PBS alone. Four IFN-treated mice and 3 EF-PBS treated mice were euthanized at 12 h post-treatment and lungs were preserved in RNA-Later before transcriptome profiling by total RNA-Seq (**Supplemental Materials**).

RNA extraction

Total RNA was extracted from MEF cell lysates and lung tissue homogenates using the miRNeasy mini kit (Qiagen). RNA sample concentrations were quantified on an ND-2000c UVVis spectophotometer (Nanodrop, Wilmington, DE) and controlled for integrity and purity on a capillary electrophoresis system (Agilent 2100 Bioanalyzer; Agilent Technologies, Santa Clara, CA).

Stranded whole transcriptome library preparation and sequencing

Whole transcriptome libraries were constructed using TruSeq Stranded Total RNA with Ribo-Zero Gold (Illumina, San Diego, CA) according to the manufacturer's guide. Libraries were quality controlled and quantitated using the BioAnalzyer 2100 system and qPCR (Kapa Biosystems, Woburn, MA). The resulting libraries were then sequenced initially on a HiSeq 2000 using HiSeq v3 sequencing reagents, with additional sequencing on a Genome Analyzer IIx using GA v5 sequencing reagents, both of which generated paired-end reads of 100 nucleotides (nt). The GAIIx was used to ensure samples had 30 million reads or more. The libraries were clonally amplified on a cluster generation station using Illumina HiSeq version 3 and GA version four cluster generation reagents to achieve a target density of approximately 700,000 (700K)/mm2 in a single channel of a flow cell. Image analysis, base calling, and error estimation were performed using Illumina Analysis Pipeline (version 2.8).

lncRNA annotation

We downloaded the non-coding annotation from the NONCODEv3 database http://www.noncode.org/NONCODERv3/datadownload/lncRNA_mouse.zip, which included most of the published mouse lncRNAs sequences and lncRNAs annotated in a number of well-known databases before 2012.[32] Out of the 37,049 mouse non-coding sequences, we selected 36,073 non-coding sequences that included the term 'lncRNA' in their type. In addition, 209 lncRNA sequences were added from Gutmann et al.[33] As multiple isoforms of lncRNAs were present in NONCODEv3, we defined a gene level by aggregating transcripts with overlapping exons (> 50% sequence overlap) using intersectBed (bedtools-2.17.0)[34] and MM9 coordinates. A translation table between transcript and gene ID is available in www.monocldb.org. lncRNA features overlapping with exons of protein-coding genes on the same strand were subsequently filtered out for each of the CC founder genome, as described below, resulting in 25,891 lncRNA transcripts (21,839 lncRNA genes).

Alignments of reads to CC founder strain transcriptomes

To infer the function of conserved lncRNAs across mouse strains, we focused our analysis on genes with conserved sequence across the eight CC founders (80% of exonic GRCm38.70 or NONCODEv3 reference sequence conserved). For this reason, we aligned RNA-Seq reads to each CC founder transcriptome, as described below. Accuracy of gene quantification following this pipeline was checked for three C57BL/6J samples by comparing gene counts after alignment to the C57BL/6J transcriptome and gene counts quantified after alignment to the *Mus musculus* reference genome (GRCm38.70) (**Supplemental Materials** and **Fig S1**).

The eight CC founder strain genomes were downloaded from the UNC Systems Genetics website (version 2012–11–08). To retrieve the specific coding and non-coding sequences for each

CC founder strain, 36,282 lncRNA transcript sequences and 74,418 protein-coding cDNA sequences from Ensembl release 70 (selecting "gene_biotype:protein_coding") were aligned against each founder genome using BLAT.[35] The best alignment per query (transcript sequence) was kept and alignments for which less than 80% of the query sequence was aligned were filtered out. Overlapping introns, and those introns < 4 bp were removed using gffread with the following options: -E -T -Z.[36] lncRNA features that overlapped with protein-coding sequences on the same strand were removed using intersectBed.[34] In total, the sequence of 74,182 protein-coding transcripts (22,521 coding genes) and 25,891 lncRNA transcripts (21,839 lncRNA genes) passed our criteria in all eight CC strains. To check this pipeline, we aligned the sequences of 25 randomly selected coding transcripts from the eight CC transcriptomes by multiple alignments (using BLASTN) and we verified that known SNPs and indels were correctly retrieved (data not shown).

Raw reads were trimmed using fastq_quality_trimmer from the FASTX-toolkit with the following options: -Q33 -l 25 -t 20. The order of paired-end reads in the two fastq files were subsequently fixed using Picard tools (picard.sourceforge.net). Reads that mapped directly with no gaps to MM9 ribosome sequence using Bowtie[37] were filtered out. Read alignments against PR8 and MA15 viral sequences are described in **Supplemental Materials**. Remaining reads were mapped against specific CC founder strain transcriptomes with SOAPaligner/soap2.[38] For each read, a maximum number of two mismatches were allowed, and repeat hits were kept. The insert window for paired-end reads was set between 20 and 500 nt. To determine fragment count on the gene level from the SOAP output, we used a custom script in Java reproducing HTSeq paired and strand-specific union mode.[39] Out of 44,360 genes common in all CC founder transcriptomes, 40,566 genes, including 19,838 coding and 20,728 non-coding genes, were quantified with at least one read count across the experiment.

Data normalization and differential expression analysis

Technical replicates were in strong agreement with each other (Pearson correlation coefficient of their $\log_2$ raw gene counts $r^2 > 0.9$) and they were summed as recommended in the DESeq package.[39] Three samples were further excluded based on their low raw gene counts distribution (GEO accession numbers: GSM1265573, GSM1265541, GSM1265528). This resulted in a total of 120 samples that were used for subsequent analysis (**Table S3**). We filtered out the genes that were not consistently expressed by keeping only genes that had at least 10 raw read counts in 75% of the samples of a single biological condition, defined based on mouse strain and viral infection condition.[40] The expression-based filtering resulted in 15,355 coding-genes and 12,211 non-coding genes that passed inspection. Data normalization was performed using a scaling method, as implemented in the DESeq bioconductor package.[39] Individual Log_2 fold change (FC) were calculated after offsetting the normalized data by 1 and by subtracting individual $\log_2$ values by the mean of $\log_2$ expression values from mouse strain-matched mock samples. To determine differentially expressed (DE) genes in response to infection, samples from each mouse strain infected with MA15 or PR8 at each DPI were compared with the pool of strain-matched mock-infected mice. Differential expression was assessed using the negative binomial model implemented in DESeq,[39] with genes with a false discovery rate (FDR) of < 1% defined as DE. Five samples from infected mice with very low viral read counts and showing a similar response to mocks based on multidimensional scaling (MDS) were excluded from the differential expression analysis ("NZO_PR8_D2_39," "NZO_PR8_D4_45," "C57BL6J_PR8_D4_95," "C57BL6J_PR8_D2_89," "CAST_MA15_D4_103"). In total, 8,270 coding genes and 5,329 non-coding genes were determined to be DE in at least one infection condition.

Co-expression network inference

Co-expression between all pairs of DE genes using $\log_2$FC expression values was determined using the biweight midcorrelation (bicor) method implemented in WGCNA R package.[41] This method was chosen after benchmarking several parametric and non-parametric methods (**Supplemental Materials** and **Fig S6**). A complete signed weighted co-expression network was built following the WGCNA method.[42] Briefly, the adjacency matrix was computed using $[(1 + A)/2]^{\beta}$ where A is the adjacency matrix of biweight midcorrelations and the soft-thresholding power β was fixed to 12 based on the scale free topology criterion as previously described.[42] Bottlenecks of the weighted network were determined by estimating the number of shortest paths going through each node (betweenness centrality, bc) with a maximum path length of 20 using igraph R package.[43] Central genes of the networks that are heavily connected nodes, or hubs, were determined by calculating weighted degree for each gene considering the whole network (kTotal) or only genes belonging to the same module (kWithin).

Gene-sets used for functional enrichment

Gene Ontology (GO) Biological Process (BP) gene-sets were retrieved from Ensembl using the Biomart interface.[44] Reactome pathway gene-sets were retrieved from the reactome website.[45] Co-expressed modules of genes in immune cells were downloaded from the Immgen website (www.immgen.org/ModsRegs/modules). In addition, genes highly expressed in immune cells compared with lung were defined as genes expressed 20-fold more in each immune cell subset than in lung based on microarray analysis from GeneAtlas V3 (GSE10246) and that were expressed only in that cell subset. IFN-stimulated genes (ISGs) were defined as the union of genes significantly upregulated 12 and 24 h after treatment of BALB/c mice with IFN-α [46] and genes significantly upregulated in at least one of the eight CC founder-strain-derived MEF cells treated with IFN-α or IFN-β ($\text{Log}_2\text{FC} > 2$ and adjusted Student's p-value < 0.01). For transcription factor (TF) binding motifs, we scanned promoters (defined as –450 to +50 nt from cDNA start using GRCm38.70 sequences) for the presence of mouse TF motifs contained in the JASPAR CORE[47] and UniPROBE[48] databases using FIMO software[49] from the MEME suite.[50] The presence of a motif in each gene promoter was defined as having a p-value < 10^-4. Finally, the last category of gene-sets was genes present in QTL regions determining PR8[22] or MA15 responses [Gralinski et al., in preparation].

Module-based annotation

A coarse annotation of lncRNA was provided by the annotation of the modules to which they belong. Module definition was performed using the dynamicTreeCut R package[42] based on the topological overlap matrix calculated in WGCNA.[42] The minimal module size was set to 150 genes determined as the number giving highest module enrichment scores in GO BP. Modules were given color names arbitrarily and genes that did not belong to any module were assigned the color grey. Association between each module and phenotypic data (weight loss and viral replication) was calculated by computing the biweight midcorrelation between phenotypic data and each module representative expression profile ("module eigengene") using the WGCNA package. In addition, each module was characterized functionally by calculating enrichment scores in each of the gene-sets described above as $-\log_{10}$(p-value) determined by one-sided Fisher's exact test with background set as all genes passing our expression-based filtering.

Rank-based annotation

An individual and finer annotation of lncRNA was obtained by using a rank-based method. For each lncRNA, the list of DE coding and non-coding genes was ranked based on the signed biweight midcorrelation coefficient. Enrichment in each gene-set was computed using the Wilcoxon rank–sum (WRS) test implemented in the Piano Bioconductor Package.[51] Significance was estimated from the normal distribution and p-values were adjusted with the Bonferroni method. We used the up distinct-directional p-value, which assesses whether genes belonging to a given gene-set are significantly enriched in the top of the ranked list (i.e., highly positively correlated with the lncRNA). We chose to consider only positively correlated genes (and not both highly positively and negatively correlated genes) because we found that positive signed correlation outperformed unsigned correlation to associate genes with similar functions (**Fig S6**). Adjusted p-value < 0.05 were considered as significant. For each gene-set, we further determined which lncRNA were the most associated with the gene-set by computing the percentile ranked (PR) on significant p-values.

We checked our functional prediction by using another rank-based enrichment method implemented in the Piano package: parametric analysis of gene-set enrichment (PAGE).[51] PAGE results were similar to WRS results, but PAGE was too sensitive with many lncRNAs that were enriched in some gene sets with similar highly significant p-value < 10^-100, and therefore it was not possible to rank them for their association with these gene sets.

***Cis/trans* annotation**

We considered correlation with chromosomal neighboring genes to determine whether lncRNA could regulate transcription in a *cis* manner. Neighbor genes were defined as genes within 200 kb from the middle of the lncRNA gene, using Grm38.70 coordinates. The middle of each gene was calculated as the arithmetic mean of the middle of its transcripts (defined as the difference between stop and start coordinates). A given lncRNA was defined as *cis* enhancer-like if it was found significantly positively correlated with all its coding neighbors, regardless of the chromosomal strand, or only considering neighbors on the same strand (sense) or on the opposite strand (antisense). Inversely, a given lncRNA was defined as a potential *cis* inhibitor if it was found significantly negatively correlated with all its coding neighbors. Significance of biweight midcorrelation was defined as two sided Student p-value < 0.01. lncRNAs with no significantly correlated neighboring gene or with both positively and negatively correlated neighbors were classified as potential *trans* lncRNAs. Specificity of potential *cis* lncRNA effect on coding neighbors was computed using PAGE analysis[51] for *cis* lncRNA with more than two coding neighbors. Up distinct-directional p-values were used for *enhancer-like cis* lncRNA to assess specific *positive* correlation with coding genes, and down distinct-directional p-values were used for *inhibitor cis* lncRNA to assess specific *negative* correlation with coding genes.

Design of the database, web portal, and automatic querying

The MONOCLdb web portal was created using Drupal (http://drupal.org/), a free and open-source content management framework. The different visualization interfaces of MONOCLdb, as well as the automatically querying web-service, were created using a collection of PHP, SQL, R, and JavaScript scripts. MySQL (http://www.mysql.com/) was used as the database engine for MONOCLdb. The JavaScript Data-Driven Documents (http://d3js.org/) library was used to create the different interactive figures.

Distributed Annotation System service

The Distributed Annotation System (DAS) service was set up using ProServer.[52] ProServer is a Perl DAS server, developed by the Wellcome Trust Sanger Institute. The DAS provides annotation information of genomics data into a large variety of Genome Browsers (e.g., Ensembl, NCBI, and UCSC). Further information regarding DAS can be found at http://www.dasregistry.org and http://www.biodas.org. The DAS track that we provide has been set up for the Ensembl Grm38 and NCBI MM9 coordinates systems. Please use www.monocldb.org:9000/das as the DAS entry point for the MONOCL database.

Data accession number

NCBI Gene Expression Omnibus (GEO), GSE52405, GSE55480, and GSE53057. GSE52405 ("RNA-Seq based characterization of long non-coding RNA involved in respiratory viruses pathogenesis") contains 123 total RNA-Seq samples from the eight CC founders mice infected with PR8, MA15 or mock-infected. Please note that mice strains were abbreviated as follow for sample names in GSE52405: A/J [AJ], C57BL/6J [C57BL6J], 129S1/SvImJ [129S1], NOD/ShiLtJ [NOD], NZO/HILt [NZO], CAST/EiJ [CAST], PWK/PhJ [PWK], and WSB/EiJ [WSB]. GSE55480 ("RNA-seq based characterization of long non-coding RNA involved in respiratory viruses pathogenesis") contains 12 total RNA-Seq samples from C57BL/6J mice treated with IFN-α or PBS. Finally, GSE53057 ("Transcriptomic Profiling of Collaborative Cross Founder Mouse Embryonic Fibroblasts stimulated with Type I, II and III Interferons") contains 71 microarray samples from the eight CC founders mice stimulated with either IFN-α or IFN-β.

Acknowledgment

The authors thank Marcus Korth for valuable feedback on the manuscript. We also acknowledge Illumina for providing library and sequencing reagents, David Threadgill who kindly provided the primary MEF cells isolated from the eight founder strains, and Elizabeth Rosenzweig who performed the IFN treatments of MEF cells and corresponding microarray experiments. Research in the Author's laboratory is supported by Public Health Service grants U19 AI100625–02 and U54AI081680, and with Federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200800060C.

Supplemental Materials

Supplemental Materials may be found here:
www.landesbioscience.com/journals/rnabiology/article/29442

References

1. Ng WF, To KF, Lam WW, Ng TK, Lee KC. The comparative pathology of severe acute respiratory syndrome and avian influenza A subtype H5N1--a review. Hum Pathol 2006; 37:381-90; PMID:16564911; http://dx.doi.org/10.1016/j.humpath.2006.01.015
2. Josset L, Tisoncik-Go J, Katze MG. Moving H5N1 studies into the era of systems biology. Virus Res 2013; 178:151-67; PMID:23499671; http://dx.doi.org/10.1016/j.virusres.2013.02.011
3. Josset L, Belser JA, Pantin-Jackwood MJ, Chang JH, Chang ST, Belisle SE, Tumpey TM, Katze MG. Implication of inflammatory macrophages, nuclear receptors, and interferon regulatory factors in increased virulence of pandemic 2009 H1N1 influenza A virus after host adaptation. J Virol 2012; 86:7192-206; PMID:22532695; http://dx.doi.org/10.1128/JVI.00563-12
4. Tchitchek N, Eisfeld AJ, Tisoncik-Go J, Josset L, Gralinski LE, Bécavin C, Tilton SC, Webb-Robertson BJ, Ferris MT, Totura AL, et al. Specific mutations in H5N1 mainly impact the magnitude and velocity of the host response in mice. BMC Syst Biol 2013; 7:69; PMID:23895213; http://dx.doi.org/10.1186/1752-0509-7-69
5. Brandes M, Klauschen F, Kuchen S, Germain RN. A systems analysis identifies a feedforward inflammatory circuit leading to lethal influenza infection. Cell 2013; 154:197-212; PMID:23827683; http://dx.doi.org/10.1016/j.cell.2013.06.013
6. Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res 2012; 40:6391-400; PMID:22492512; http://dx.doi.org/10.1093/nar/gks296
7. Jeon Y, Lee JT. YY1 tethers Xist RNA to the inactive X nucleation center. Cell 2011; 146:119-33; PMID:21729784; http://dx.doi.org/10.1016/j.cell.2011.06.026
8. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 2008; 322:1717-20; PMID:18988810; http://dx.doi.org/10.1126/science.1163802
9. Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol 2012; 14:659-65; PMID:22684254; http://dx.doi.org/10.1038/ncb2521
10. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 2010; 39:925-38; PMID:20797886; http://dx.doi.org/10.1016/j.molcel.2010.08.011
11. Yin Z, Guan D, Fan Q, Su J, Zheng W, Ma W, Ke C. lncRNA expression signatures in response to enterovirus 71 infection. Biochem Biophys Res Commun 2013; 430:629-33; PMID:23220233; http://dx.doi.org/10.1016/j.bbrc.2012.11.101
12. Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, Mattick JS. Genome-wide identification of long noncoding RNAs in CD8+ T cells. J Immunol 2009; 182:7738-48; PMID:19494298; http://dx.doi.org/10.4049/jimmunol.0900603
13. Peng X, Gralinski L, Armour CD, Ferris MT, Thomas MJ, Proll S, Bradel-Tretheway BG, Korth MJ, Castle JC, Biery MC, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. MBio 2010; 1:e00206-10; PMID:20978541; http://dx.doi.org/10.1128/mBio.00206-10
14. Collier SP, Collins PL, Williams CL, Boothby MR, Aune TM. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J Immunol 2012; 189:2084-8; PMID:22851706; http://dx.doi.org/10.4049/jimmunol.1200774
15. Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, Chang HY, Brahic M, Kirkegaard K. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. Cell 2013; 152:743-54; PMID:23415224; http://dx.doi.org/10.1016/j.cell.2013.01.015
16. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 2009; 33:717-26; PMID:19217333; http://dx.doi.org/10.1016/j.molcel.2009.01.026
17. Zhang Q, Chen CY, Yedavalli VS, Jeang KT. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. MBio 2013; 4:e00596-12; PMID:23362321; http://dx.doi.org/10.1128/mBio.00596-12
18. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. Nat Rev Genet 2011; 12:641-55; PMID:21850044; http://dx.doi.org/10.1038/nrg3049
19. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ. Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat Genet 2002; 31:255-65; PMID:12089522; http://dx.doi.org/10.1038/ng906
20. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 2011; 39:3864-78; PMID:21247874; http://dx.doi.org/10.1093/nar/gkq1348
21. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, et al.; Complex Trait Consortium. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 2004; 36:1133-7; PMID:15514660; http://dx.doi.org/10.1038/ng1104-1133
22. Ferris MT, Aylor DL, Bottomly D, Whitmore AC, Aicher LD, Bell TA, Bradel-Tretheway B, Bryan JT, Buus RJ, Gralinski LE, et al. Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. PLoS Pathog 2013; 9:e1003196; PMID:23468633; http://dx.doi.org/10.1371/journal.ppat.1003196
23. Perrone LA, Plowden JK, García-Sastre A, Katz JM, Tumpey TM. H5N1 and 1918 pandemic influenza virus infection results in early and excessive infiltration of macrophages and neutrophils in the lungs of mice. PLoS Pathog 2008; 4:e1000115; PMID:18670648; http://dx.doi.org/10.1371/journal.ppat.1000115
24. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. Elife 2013; 2:e00762; PMID:23898399; http://dx.doi.org/10.7554/eLife.00762
25. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009; 458:223-7; PMID:19182780; http://dx.doi.org/10.1038/nature07672
26. Gutschner T, Hämmerle M, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Gross M, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res 2013; 73:1180-9; PMID:23243023; http://dx.doi.org/10.1158/0008-5472.CAN-12-2850
27. Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, Zhang Y, Gorospe M, Prasanth SG, Lal A, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet 2013; 9:e1003368; PMID:23555285; http://dx.doi.org/10.1371/journal.pgen.1003368
28. Zhang Q, Chen C-Y, Yedavalli VSRK, Jeang K-T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. MBio 2013; 4:e00596-12; PMID:23362321; http://dx.doi.org/10.1128/mBio.00596-12
29. Wang Y, Davies KJA, Melendez JA, Crawford DR. Characterization of adapt33, a stress-inducible riboregulator. Gene Expr 2003; 11:85-94; PMID:12837039; http://dx.doi.org/10.3727/000000003108748982
30. Sheahan T, Whitmore A, Long K, Ferris M, Rockx B, Funkhouser W, Donaldson E, Gralinski L, Collier M, Heise M, et al. Successful vaccination strategies that protect aged mice from lethal challenge from influenza virus and heterologous severe acute respiratory syndrome coronavirus. J Virol 2011; 85:217-30; PMID:20980507; http://dx.doi.org/10.1128/JVI.01805-10
31. Gralinski LE, Bankhead A 3rd, Jeng S, Menachery VD, Proll S, Belisle SE, Matzke M, Webb-Robertson BJ, Luna ML, Shukla AK, et al. Mechanisms of severe acute respiratory syndrome coronavirus-induced acute lung injury. MBio 2013; 4:4; PMID:23919993; http://dx.doi.org/10.1128/mBio.00271-13

32. Bu D, Yu K, Sun S, Xie C, Skogerbø G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res 2012; 40:D210-5; PMID:22135294; http://dx.doi.org/10.1093/nar/gkr1175
33. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 2011; 477:295-300; PMID:21874018; http://dx.doi.org/10.1038/nature10398
34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26:841-2; PMID:20110278; http://dx.doi.org/10.1093/bioinformatics/btq033
35. Bhagwat M, Young L, Robison RR. Using BLAT to find sequence similarity in closely related genomes. Curr Protoc Bioinformatics 2012; Chapter 10:Unit10 8.
36. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7:562-78; PMID:22383036; http://dx.doi.org/10.1038/nprot.2012.016
37. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10:R25; PMID:19261174; http://dx.doi.org/10.1186/gb-2009-10-3-r25
38. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009; 25:1966-7; PMID:19497933; http://dx.doi.org/10.1093/bioinformatics/btp336
39. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010; 11:R106; PMID:20979621; http://dx.doi.org/10.1186/gb-2010-11-10-r106
40. Chang ST, Thomas MJ, Sova P, Green RR, Palermo RE, Katze MG. Next-generation sequencing of small RNAs from HIV-infected cells identifies phased microrna expression patterns and candidate novel microRNAs differentially expressed upon infection. MBio 2013; 4:e00549-12; PMID:23386435; http://dx.doi.org/10.1128/mBio.00549-12
41. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw 2012; 46:46; PMID:23050260
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9:559; PMID:19114008; http://dx.doi.org/10.1186/1471-2105-9-559
43. Mueller LA, Kugler KG, Graber A, Emmert-Streib F, Dehmer M. Structural measures for network biology using QuACN. BMC Bioinformatics 2011; 12:492; PMID:22195644; http://dx.doi.org/10.1186/1471-2105-12-492
44. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. Ensembl 2013. Nucleic Acids Res 2013; 41:D48-55; PMID:23203987; http://dx.doi.org/10.1093/nar/gks1236
45. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014; 42:D472-7; PMID:24243840; http://dx.doi.org/10.1093/nar/gkt1102
46. Cilloniz C, Pantin-Jackwood MJ, Ni C, Carter VS, Korth MJ, Swayne DE, Tumpey TM, Katze MG. Molecular signatures associated with Mx1-mediated resistance to highly pathogenic influenza virus infection: mechanisms of survival. J Virol 2012; 86:2437-46; PMID:22190720; http://dx.doi.org/10.1128/JVI.06156-11
47. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 2010; 38:D105-10; PMID:19906716; http://dx.doi.org/10.1093/nar/gkp950
48. Robasky K, Bulyk ML. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res 2011; 39:D124-8; PMID:21037262; http://dx.doi.org/10.1093/nar/gkq992
49. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics 2011; 27:1017-8; PMID:21330290; http://dx.doi.org/10.1093/bioinformatics/btr064
50. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 2009; 37:W202-8; PMID:19458158; http://dx.doi.org/10.1093/nar/gkp335
51. Väremo L, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res 2013; 41:4378-91; PMID:23444143; http://dx.doi.org/10.1093/nar/gkt111
52. Finn RD, Stalker JW, Jackson DK, Kulesha E, Clements J, Pettett R. ProServer: a simple, extensible Perl DAS server. Bioinformatics 2007; 23:1568-70; PMID:17237073; http://dx.doi.org/10.1093/bioinformatics/btl650