Genome-Wide Network Analysis Reveals the Global Properties of IFN-β Immediate Transcriptional Effects in Humans[1,2]

Guy Haskin Fernald,* Simon Knott,† Andrew Pachner,‡ Stacy J. Caillier,* Kavitha Narayan,‡ Jorge R. Oksenberg,* Parvin Mousavi,† and Sergio E. Baranzini[3]*

IFN-β effectively controls clinical exacerbations and magnetic resonance imaging activity in most multiple sclerosis patients. However, its mechanism of action has not been yet fully elucidated. In this study we used DNA microarrays to analyze the longitudinal transcriptional profile of blood cells within a week of IFN-β administration. Using differential expression and gene ontology analyses we found evidence of a general decrease in the cellular activity of T lymphocytes resembling the endogenous antiviral response of IFNs. In contrast, most of the differentially expressed genes (DEGs) from untreated individuals were involved in cellular physiological processes. We then used mutual information (MI) to build networks of coregulated genes in both treated and untreated individuals. Interestingly, the connectivity distribution (k) of networks generated with high MI values displayed scale-free properties. Conversely, the observed k for networks generated with suboptimal MI values approximated a Poisson distribution, suggesting that MI captures biologically relevant interactions. Gene networks from individuals treated with IFN-β revealed a tight core of immune- and apoptosis-related genes associated with higher values of MI. In contrast, networks obtained from untreated individuals primarily reflected cellular housekeeping functions. Finally, we trained a neural network to reverse engineer the directionality of the main interactions observed at the biological process level. This is the first study that incorporates network analysis to investigate gene regulation in response to a therapeutic drug in humans. Implications of this method in the creation of personalized models of response to therapy are discussed. *The Journal of Immunology,* **2007, 178: 5076–5085.**

Interferon β is a member of the type I IFN family and, as a recombinant reagent, is currently the most prescribed therapeutic agent for controlling exacerbations in relapsing-remitting multiple sclerosis. IFN-β has been an FDA-approved drug for >10 years and many hundred of thousands of individuals have been treated worldwide but, remarkably, its precise mechanism of action remains unknown (1).

At the molecular level, IFN-β is recognized by a heterodimeric receptor complex triggering a downstream signaling cascade that involves the phosphorylation of several intracellular mediators including Jak1, Tyk2, STAT1, STAT2, p38, and NF-κB. The interaction of these proteins results in the intranuclear formation of the IFN-stimulated gene factor 3 and the γ IFN activation factor, which, after binding to their DNA responsive elements (IFN-stimulated response element and γ IFN-activating sequence, respectively), drive the transcription of a multitude of genes (see Ref. 2 for a review on the IFN-β signaling pathway). At the cellular and organismal level, however, the mechanism by which IFN-β exerts its immunomodulatory effect is not well established (3). Although it is generally accepted that IFN-β counteracts the effects of IFN-γ (4, 5) and inhibits the production of matrix metalloproteases (6), its precise role in modulating the expression of adhesion molecules (7, 8), inducing apoptosis (9–13), and regulating the Th1/Th2 balance (14, 15) remains controversial.

Recently, a number of high throughput analyses have been completed in an attempt to characterize the global transcriptional effects of IFN-β in vitro and ex vivo (14, 16–20). These studies provided important insights into the intricate regulatory effects of IFN-β in different experimental settings. However, with the exception of a single longitudinal study (19), these reports focused on the identification of DEGs before and after treatment, thereby capturing primarily a static picture of a complex dynamical process. The advent of high throughput methodologies and comprehensive computational algorithms has facilitated the analysis of data from time series. Such analyses may provide a deeper understanding of the regulatory networks at play (21).

The systems biology paradigm states that the behavior of cells and organisms are emergent properties of their underlying molecular and cellular regulatory networks. Relevant knowledge about these processes can be obtained by studying the properties of such networks, including their topology and dynamics (22–28). Networks of coexpressed genes can, for example, be obtained by recursively computing their pairwise correlation. Genes with correlation coefficients exceeding a predetermined threshold are then considered coexpressed and graphically linked by a segment (edge). As an alternative to correlation, the pairwise mutual information (MI)[4] can be computed. The advantage of MI is that, unlike

*School of Medicine, University of California, San Francisco, CA 94143; †School of Computing, Queen's University, Kingston, Ontario, Canada; and ‡University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, NJ 07103

Received for publication September 22, 2006. Accepted for publication February 2, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was funded by National Institutes of Health Grant 1R01 AI42911 and National Multiple Sclerosis Society Grant NMSS CA 1035-A-7.

[2] The data presented in this article have been submitted in the Gene Expression Omnibus under accession number GSE5678.

[3] Address correspondence and reprint requests to Dr. Sergio E. Baranzini, 513 Parnassus Avenue S-256, San Francisco, CA 94143. E-mail address: sebaran@cgl.ucsf.edu

[4] Abbreviations used in this paper: MI, mutual information; DEG, differentially expressed gene; FDR, false discovery rate; GO, gene ontology; SA, sensitivity analysis; O, observed; E, expected; O/E, observed to expected (ratio); SSE, sum squared error; TMI, threshold of MI.

Copyright © 2007 by The American Association of Immunologists, Inc. 0022-1767/07/$2.00

Table I. *Samples and blood draw schedule*

Sample[a]	IFN	Relnet[b]	Time Point (hours)								
			1	2	3	4	5	6	7	8	9
T1	Yes	P	0	3.5	6.25	9.5	11.5	16.5	25	49	
T2	Yes	P			7.25	10.25	13.0	15.5	33		156
U1	No	C1	0	3.5	6.25	9.5	11.5	16.5	25	49	156
U2	No	C1			7.25	10.25	13.0	15.5	33		156
U3	No	C2	0	3.5	6.25	9.5	11.5	16.5	25	49	156
U4	No	C2			7.25	10.25	13.0	15.5	33		156

[a] T, Treated; U, untreated.
[b] P, Patient; C, control.

correlation, it can also capture nonlinear relationships. Relnet (relevance networks), a method for generating statistically significant MI networks, seems ideally suited to the analysis of the higher order properties of coordinated gene expression (29). We generated longitudinal transcriptional profiles of closely spaced blood draws of IFN-β-treated and untreated individuals. We then applied Relnet to the obtained gene expression matrix and report the topological properties of the coexpression networks triggered in vivo by IFN-β. In addition, we used sensitivity analysis (SA) of trained neural networks to reverse engineer the causal relationships among processes elicited by IFN-β administration. This knowledge could lead to a better understanding of the mode(s) of action of this drug as well as the pathogenesis underlying multiple sclerosis.

Materials and Methods

Subjects and sample collection

Repeated blood draws were conducted at University of Medicine and Dentistry of New Jersey (Newark, NJ) into PAXgene tubes (Qiagen) in two individuals after the administration of a therapeutic dose of IFN-β1a (Avonex, Biogen) in accordance with institutional guidelines. In one individual (T1), samples were obtained at baseline and at 3.5, 6.25, 9.5, 11.5, 16.5, 25, 49, and 156 h. In the other (T2), blood was collected at 7.25, 10.25, 13, 15.5, and 33 h. Samples from four healthy volunteers matched for age and gender were also collected. Because samples from the two IFN-β-treated individuals were collected at slightly different times, we matched the collection times of two controls to follow each time series.

RNA purification, fluorescent labeling, and microarray hybridization

Total RNA from all samples was obtained by using the PAXgene blood RNA kit (Qiagen). To obtain sufficient amounts for microarray hybridization, RNA was amplified essentially as described (30). Amplified RNA samples were labeled with Cy-3 (Amersham Biosciences), mixed in equimolar quantities with a Cy5-labeled pool of human RNA (common reference), and hybridized in triplicate onto microarrays containing over 22,000 spotted 70-mer oligonucleotide probes (National Institutes of Health/National Institute of Neurological Disorders and Stroke Microarray Consortium core facility). The data discussed in this article have been deposited in National Center for Biotechnology Information Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE5678.

Data analysis and network generation

Raw data were imported into Biometrics Research Branch array tools National Institutes of Health), log transformed, and normalized. Replicate arrays were kept only if their overall pairwise correlation was >0.8. To identify DEGs, F tests adjusted for a false discovery rate (FDR) were run on the complete time series for each individual. To create the input expression matrix for the MI networks, the union of each FDR-adjusted gene list was computed. This resulted in a reduced dataset of 1827 genes for each individual. To ensure that a similar number of arrays were included in each matrix, we created one file with concatenated expression values from both treated individuals (T1 plus T2) and two with concatenated expression values from two untreated individuals each (controls 1, U1 plus U2; controls 2, U3 plus U4). Next, each of the resulting three expression matrices was discretized in the range 0–4. Discretized expression matrices were loaded into MATLAB (MathWorks) and Relnet (29) was run. We modified Relnet to run on a 12-processor computer cluster and to generate output files that could be imported into Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/) for subsequent analysis. To determine the significance of the MI values, we performed 30 random permutations of the columns in each of the three matrices and recorded the maximum MI value achieved with the permuted data as the threshold of mutual information (TMI). The number of permutations was determined empirically such that the maximum MI achieved did not differ significantly among the 30 runs. All MI values in the original matrices above TMI were considered significant.

Network analysis

Using a custom MATLAB script, each relevant network obtained with Relnet was exported and loaded into Pajek, where all network measurements and graphics were produced.

Sensitivity analysis

Dimensionality reduction was achieved by grouping genes with similar expression profiles into metagenes. Genes showing correlations of >0.75 along the time series were grouped together. Gene groups with equal slope signs in between time points were also joined together into metagenes. Finally, each metagene was assigned to one gene ontology category from all those known to be active in the IFN-β canonical pathway. A recurrent Elman neural network was trained with the expression levels of all metagenes at times t-n to predict the expression levels of a target metagene at time t, where $n = 1,\ldots,n$. A "jitter" was then randomly sampled from a gene-specific noise model and added to the inputs of the network one metagene at a time. The noise model for a metagene nG_i followed a normal distribution with a mean of 0 and a SD of its expression profile $SDnG_i$, where $SDnG_i$ is the SD of the metagene's expression. The effect of the added jitter, measured as the increase in the error of the network, reflects the importance of that gene in predicting the expression levels of the target metagene. Metagenes with the smallest influence were iteratively removed from the input space and the network was retrained with the resulting subset of metagenes. This procedure was repeated, removing two metagenes on each iteration until a minimal set of metagenes with high predictive power was ascertained (a detailed description of this method will appear in a future article; S. Knott, S. E. Baranzini, and P. Mousavi, manuscript in preparation). For robustness, SA was performed 20 times, each time taking advantage of small differences in the perturbation scheme. The 20 subsets are then tallied to identify the metagenes with highest occurrence as the most likely regulator of each of the remainder target metagenes.

Results

Longitudinal RNA samples collected from two individuals after the administration of a therapeutic dose of IFN-β1a and from four untreated volunteers were hybridized in triplicates onto 70-mer spotted oligonucleotide microarrays containing >22,000 probes. Although 130 arrays were processed, 36 were discarded after rigorous quality control (see *Materials and Methods*). Because the treated samples were collected at slightly different time points, we chose two untreated controls to closely match each of the IFN-β time series, thereby obtaining one combined dataset for treated (T1 plus T2) and two for control individuals (U1 plus U2 and U3 plus U4) (Table I).

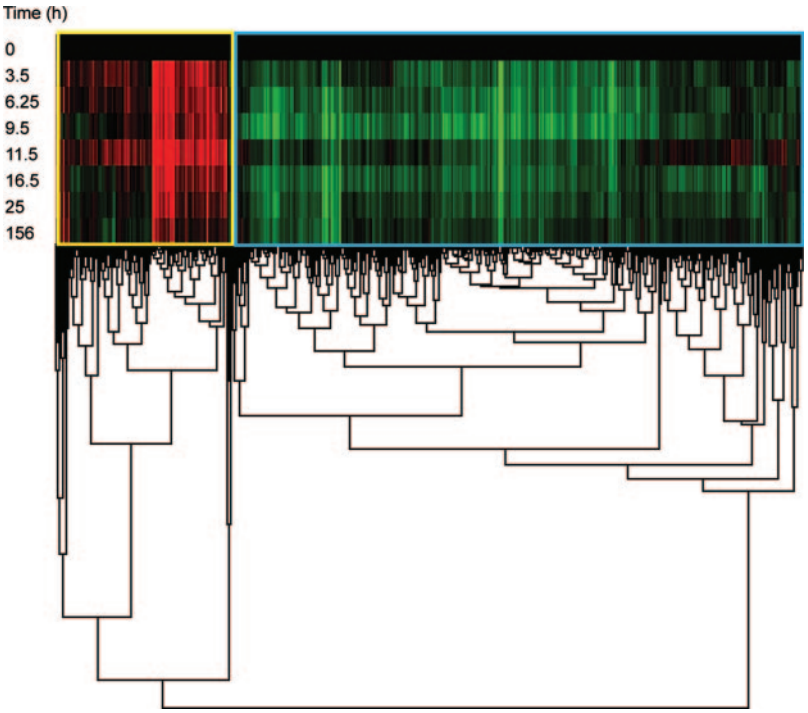

FIGURE 1. Longitudinal changes in gene expression after IFN-β administration. Average expression of 438 genes (columns) along time (rows) in PBMC from two individuals at baseline and soon after IFN-β administration. Time is shown in hours. High expression is indicated in red, medium in black, and low in blue. A significant fraction of the regulated genes showed a rapid increase from baseline in response to the drug (yellow box). However, most of the significant changes were toward down-regulation as shown by a rapid decrease in expression from baseline (cyan box). Different profiles can be seen later on for both groups of genes.

Differential expression analysis

Initially we sought to identify the DEGs along time in each individual by performing a F test statistic corrected by the FDR. Between 300 and 1000 genes were identified as differentially expressed in time within each individual. To obtain an estimate of the quality and magnitude of the detectable response to IFN-β by peripheral blood cells, we combined and averaged comparable time points in both of the treated individuals and applied filtering to select only those genes showing differential expression along time. Fig. 1 shows a hierarchical clustering of the 438 genes that were identified as differentially expressed in time in the IFN-β-treated individuals but not in the controls. When taking baseline (T0) as a reference, a clear immediate effect of IFN-β can be observed. Among the expected up-regulated transcripts we detected MX1, an influenza virus-resistant gene, and OAS1 and OAS3, members of the 2-5A synthetase family, essential proteins involved in the innate immune response to viral infection. In addition, *ISG15* (*G1P2*), *G1P3*, *IFIT*-1, *IFIT*-4, *IFI44L*, *IRF7*, and *IFI27*, known for their immediate responsiveness to type I IFNs, were also found elevated in the treated individuals. Tables II and III show a gene ontology (GO) classification for biological process category of the DEGs. The up-regulated genes (Table II) revealed statistically significant deviation toward the response to stress (observed (O) to expected (E) ratio (O/E): 4.23, $p = 3.88\ 10^{-7}$), immune response

Table II. *Up-regulated genes*

Biological Process[a]	Observed	Expected	Ratio	p Value	Immune Response	Response to Stress	Response to Wounding	JAK/STAT Cascade
Physiological process	48	41.04	1.17	5.36E-04	FCGR1A	DNAJA1	IL1RN	JAK2
Response to stimulus	23	8.41	2.73	1.21E-06	G1P3	G1P3	IRF7	NMI
Response to biotic stimulus	19	3.61	5.26	5.20E-10	GBP5	GADD45B	KRT1	STAT1
Organismal physiological process	19	7.84	2.42	9.78E-05	IFI35	IFI35	NMI	
Immune response	18	3.02	5.96	2.42E-10	IFIT1	IL1RN	PLAUR	
Defense response	18	3.44	5.23	1.99E-09	IGHM	IRF7	SERPING1	
Response to stress	16	3.78	4.23	3.88E-07	IL1RN	KRT1	TNFAIP6	
Response to pest: pathogen or parasite	12	1.97	6.09	3.55E-07	IRF7	LTF	ZNF148	
Response to other organism	12	2.1	5.71	7.28E-07	KRT1	MX1		
Response to wounding	8	1.41	5.67	7.00E-05	LTF	NMI		
Response to external stimulus	8	1.83	4.37	4.19E-04	MX1	PLAUR		
Response to virus	4	0.27	14.81	1.39E-04	NFIL3	SERPING1		
Inflammatory response	4	0.78	5.13	7.47E-03	NMI	STAT1		
JAK-STAT cascade	3	0.11	27.27	1.94E-04	PSMB8	TNFAIP6		
Blood coagulation	3	0.3	10	3.43E-03	SERPING1	TREX1		
Coagulation	3	0.31	9.68	3.68E-03	TNFAIP6	ZNF148		
Hemostasis	3	0.33	9.09	4.21E-03	TNFSF13B			
Wound healing	3	0.33	9.09	4.35E-03	ZNF148			
Regulation of body fluids	3	0.37	8.11	5.89E-03				
Membrane fusion	2	0.09	22.22	3.69E-03				
Complement activation	2	0.14	14.29	8.19E-03				

[a] Gene ontology categories differ significantly between treated and untreated individuals. Genes corresponding to four selected ontologies (category names set in boldface) are listed in the four far-right columns under the corresponding category name.

Table III. *Down-regulated genes*
Gene ontology categories significantly different between treated and untreated individuals

Biological Process[a]	Observed	Expected	Ratio	*p* Value	Immune Response	Protein Biosynthesis	Generation of Precursor Metabolites and Energy	Apoptosis	Protein Folding	Mitosis
Metabolism	112	95.49	1.17	5.24E-03	AOAH	ALG5	ATP5L	ARHGEF6	CCT4	ANAPC11
Cellular metabolism	105	89.31	1.18	8.29E-03	CCR2	COPS5	COX5B	BIRC2	CCT6A	BUB3
Macromolecule metabolism	71	53.74	1.32	3.01E-03	CD3D	EEF1G	COX6A1	BNIP2	CKAP1	CHFR
Protein metabolism	55	39.85	1.38	4.65E-03	CLC	EIF2B1	CS	GPR65	DNAJB1	CORO1A
Biosynthesis	34	17.37	1.96	9.12E-05	CSF1R	KARS	ENO1	GSTP1	FKBP1A	PIN1
Cellular biosynthesis	31	15.66	1.98	1.62E-04	CST7	MRPL20	NDUFA1	GZMB	KIAA0674	RAE1
Response to biotic stimulus	28	12.47	2.25	4.15E-05	FCER1G	MRPS18B	NDUFA2	IL2RB	PFDN5	RAN
Defense response	26	11.9	2.18	1.27E-04	FCGRT	PUM1	NDUFB2	LGALS1	PIN1	
Immune response	24	10.45	2.3	8.00E-05	FPR1	RPL18	NDUFB9	MAL	PPIA	
Protein biosynthesis	22	10.05	2.19	4.32E-04	FYB	RPL19	ATP5G3	NFKBIA	TTC1	
Macromolecule biosynthesis	22	10.96	2.01	1.35E-03	GNLY	RPL22	ATP5I	SLC25A6		
Generation of precursor metabolites and energy	17	7.76	2.19	1.96E-03	GPR65	RPL24	PGAM1	SNCA		
Response to other organism	16	7.27	2.2	2.56E-03	HLA-DPB1	RPL26	PGK1	TNFRSF1B		
Response to external stimulus	14	6.34	2.21	4.56E-03	HLA-DRB5	RPL27	PPP1CC	TRADD		
Apoptosis	14	6.73	2.08	7.65E-03	IL16	RPL27A	TBXAS1			
Programmed cell death	14	6.76	2.07	7.90E-03	LILRA2	RPL3	UQCRFS1			
Protein kinase cascade	12	3.5	3.43	2.09E-04	LILRB3	RPL36AL	COX5A			
Protein folding	10	3.1	3.23	1.11E-03	LTB	RPS19				
I-κ B kinase/NF-κ B cascade	7	1.4	5	5.13E-04	NCF2	RPS27				
Mitosis	7	1.96	3.57	3.49E-03	PSME1	RPS28				
M phase of mitotic cell cycle	7	1.98	3.54	3.76E-03	S100A12	RPS7				
Purine ribonucleotide biosynthesis	5	1.05	4.76	3.99E-03	S100A9	TLR1				
Oxidative phosphorylation	5	1.06	4.72	4.21E-03	SERPINA1					
Purine ribonucleotide metabolism	5	1.09	4.59	4.67E-03	TLR1					
Purine nucleotide biosynthesis	5	1.1	4.55	4.92E-03						
Ribonucleotide biosynthesis	5	1.14	4.39	5.70E-03						
Purine nucleotide metabolism	5	1.14	4.39	5.70E-03						
Ribonucleotide metabolism	5	1.21	4.13	7.21E-03						
Detection of biotic stimulus	3	0.25	12	1.84E-03						
Ras protein signal transduction	3	0.43	6.98	9.08E-03						
Spermidine biosynthesis	2	0.04	50	5.09E-04						
Spermidine metabolism	2	0.05	40	1.01E-03						
Negative regulation of Ras protein signal transduction	2	0.08	25	2.48E-03						
Negative regulation of small GTPase-mediated signal transduction	2	0.09	22.22	3.44E-03						
Hemocyte development	2	0.09	22.22	3.44E-03						
Polyamine biosynthesis	2	0.09	22.22	3.44E-03						
Regulation of Ras protein signal transduction	2	0.11	18.18	4.55E-03						
Hemocyte differentiation (sensu *Arthropoda*)	2	0.11	18.18	4.55E-03						
Purine nucleoside metabolism	2	0.11	18.18	4.55E-03						
Purine ribonucleoside metabolism	2	0.11	18.18	4.55E-03						
Polyamine metabolism	2	0.13	15.38	7.19E-03						
Ribonucleoside metabolism	2	0.13	15.38	7.19E-03						
Activation of NF-κ B-inducing kinase	2	0.14	14.29	8.71E-03						
Defense response to fungi	2	0.14	14.29	8.71E-03						

[a] Gene ontology categories differ significantly between treated and untreated individuals. Genes corresponding to six selected ontologies (category names set in boldface) are listed in the six far-right columns under the corresponding category name.

(O/E ratio: 5.96, $p = 2.42\ 10^{-10}$), and JAK/STAT cascade (O/E ratio: 22.22, $p = 3.69\ 10^{-3}$). These findings are consistent with the canonical pathway described for IFN-β signaling and function. Although one-third of the DEGs seem to be rapidly up-regulated in response to IFN-β (Fig. 1, yellow box), most genes show a marked and sustained down-regulation (Fig. 1, cyan box). Among these, lymphocyte marker *CD3* and ribosomal genes clearly stand out. In total, the statistically significant down-regulation of 13 ribosomal genes was observed in both individuals treated with IFN-β (nine from the large subunit and four from the small subunit) (Table III). In addition, genes involved in the generation of precursor metabolites and energy, as well as in protein folding and mitosis, were strongly down-regulated. GO analysis of down-regulated genes included mitosis (O/E ratio: 3.57, $p = 3.49\ 10^{-3}$), protein folding (O/E ratio: 3.23, $p = 1.11\ 10^{-3}$), immune response (O/E ratio: 2.39, $p = 4.06\ 10^{-5}$), protein biosynthesis (O/E ratio: 2.19, $p = 4.32\ 10^{-4}$), and apoptosis (O/E ratio: 2.08, 7.65 10^{-3}). Further inspection revealed that the majority of the apoptosis-related transcripts in this list are those involved its prevention or blocking. Overall, this pattern strongly suggests that the cellular processes in

T lymphocytes such protein synthesis and folding are put on hold at the same time that antiapoptotic pathways are inactivated, thus unleashing programmed cell death.

Relevance networks

The union of all DEGs from each individual was computed, yielding 1826 annotated genes that were selected to carry all subsequent computations, and a matrix was built using the expression of these 1826 genes in all samples from both of the treated individuals (31 arrays). Two more matrices were generated, each by using concatenated expression data from two untreated individuals (31 and 32 arrays, respectively), thus totaling three expression matrices. We then computed the pairwise MI for all genes in each matrix using a modified version of the software Relnet (see *Materials and Methods*). MI is a measure of how much one variable reveals about another and can thus be used to identify groups of genes that are expressed in a coordinated fashion. We calculated the MI of all possible pairs of genes in the data and also the global maximum MI after 30 random permutations of each original dataset. These maximums were recorded as the TMI for each dataset. Of the >3.3 million MI values calculated for each original matrix, only those exceeding the global maximum obtained after the 30 permutations (TMI) were considered significant. Fig. 2 shows the distribution of MI values for each of the three matrices and the TMI indicating the proportion of significant MI values for each matrix. By connecting the pairs of genes (nodes) with the MI exceeding the TMI we were able to generate one network for the treated individuals and two for the controls. The first network comprised 713 nodes while those of the controls included 476 and 373 genes, respectively. A possible explanation for this size difference is that the larger number of nodes in the treated network is due to the many genes whose expression is modified by IFN-β, whereas the natural fluctuation of gene activity allowed us to detect only the most robust and coordinated changes in the untreated control individuals.

One of the main parameters characterizing large networks is their connectivity profile, that is, the number of nodes with a given number of links. Although random networks usually display a connectivity pattern resembling a Poisson distribution (each node has equal probability of being connected to any other node), connectivity in complex functional networks often seems to follow a power law distribution (whereas most nodes are loosely connected, only a few nodes show very high connectivity) (31). We therefore analyzed the connectivity profiles of all three networks by computing the number of links (edges) from each node and plotting them in a double log chart (Fig. 3). We observed that the connectivity distribution plot of the IFN-induced network (treated individuals) that resulted from considering only those MI values above the threshold (zone 1) closely followed a power law, consistent with what has been described for most large networks. This observation suggests that a large number of genes are activated or repressed in a coordinated fashion following drug administration. To evaluate whether this pattern was restricted to genes linked by high MI values, we plotted the connectivity distributions of a comparable number of nodes linked with smaller MI values (zones 2–5 in Fig. 2). Remarkably, as MI values decrease the shapes of the connectivity distribution plots deviate from a power law and instead approximate a Poisson distribution. This is further evidence that values above TMI represent statistically and possibly biologically significant relationships. Interestingly, we detected a similar connectivity distribution profile in the control networks as well. It is likely that these transcripts belong to the core of genes in charge of maintaining cellular homeostasis (see below).

We next evaluated the topology of the gene networks obtained after selecting only those significant nodes present in treated but not in any of the controls ($n = 438$) and those present in any of the controls but not in treated individuals ($n = 361$). Genes are represented as nodes whose sizes are proportional to the number of connections. The network induced by IFN-β displays larger nodes and more connections, suggesting a wider and coordinated transcriptional activity. Also, a noticeable difference was observed in the number of connections between the two networks, especially those with high MI values (Fig. 4). MI values of >1.8 could be only seen in the treated network as evidenced by the red and orange colored edges in Fig. 4. Further, a higher proportion of yellow and green colored edges is also evident in the patients' network. This suggests that although coordinated gene activity due to normal cellular functionality can be detected in the network, it is of a

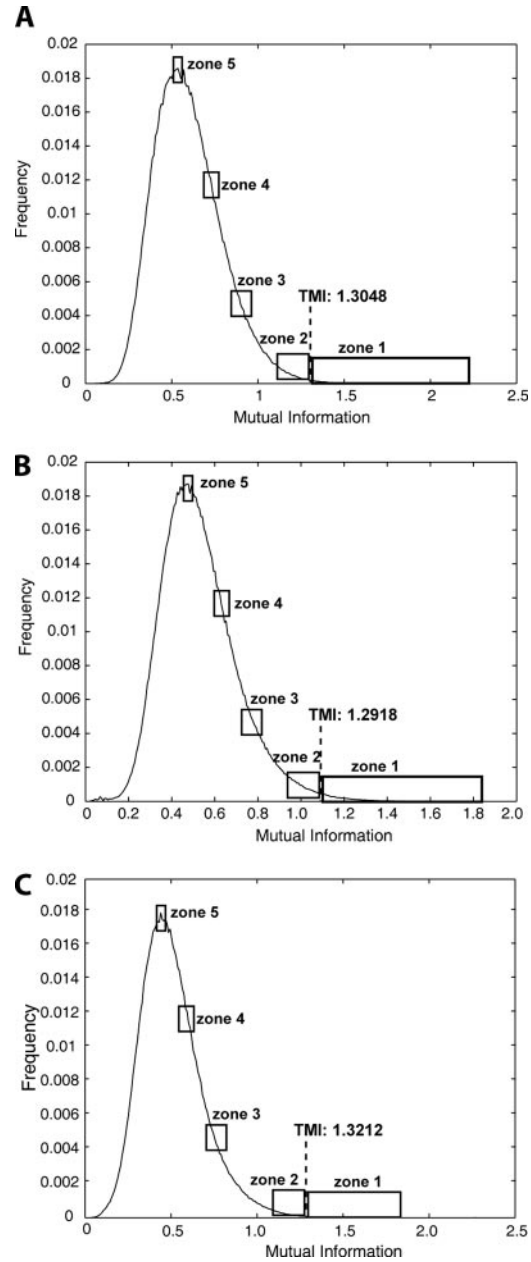

FIGURE 2. MI frequency distribution. All pairwise mutual information values for each matrix were sorted and their frequency was calculated. Relevant networks (zone 1) were defined by considering the nodes exceeding TMI (broken line). Networks of similar size were created using successively lower MI ranges (zone 2–5). *A*, Treated individuals (T1 and T2). *B*, Controls 1 (U1 and U2). *C*, Controls 2 (U3 and U4).

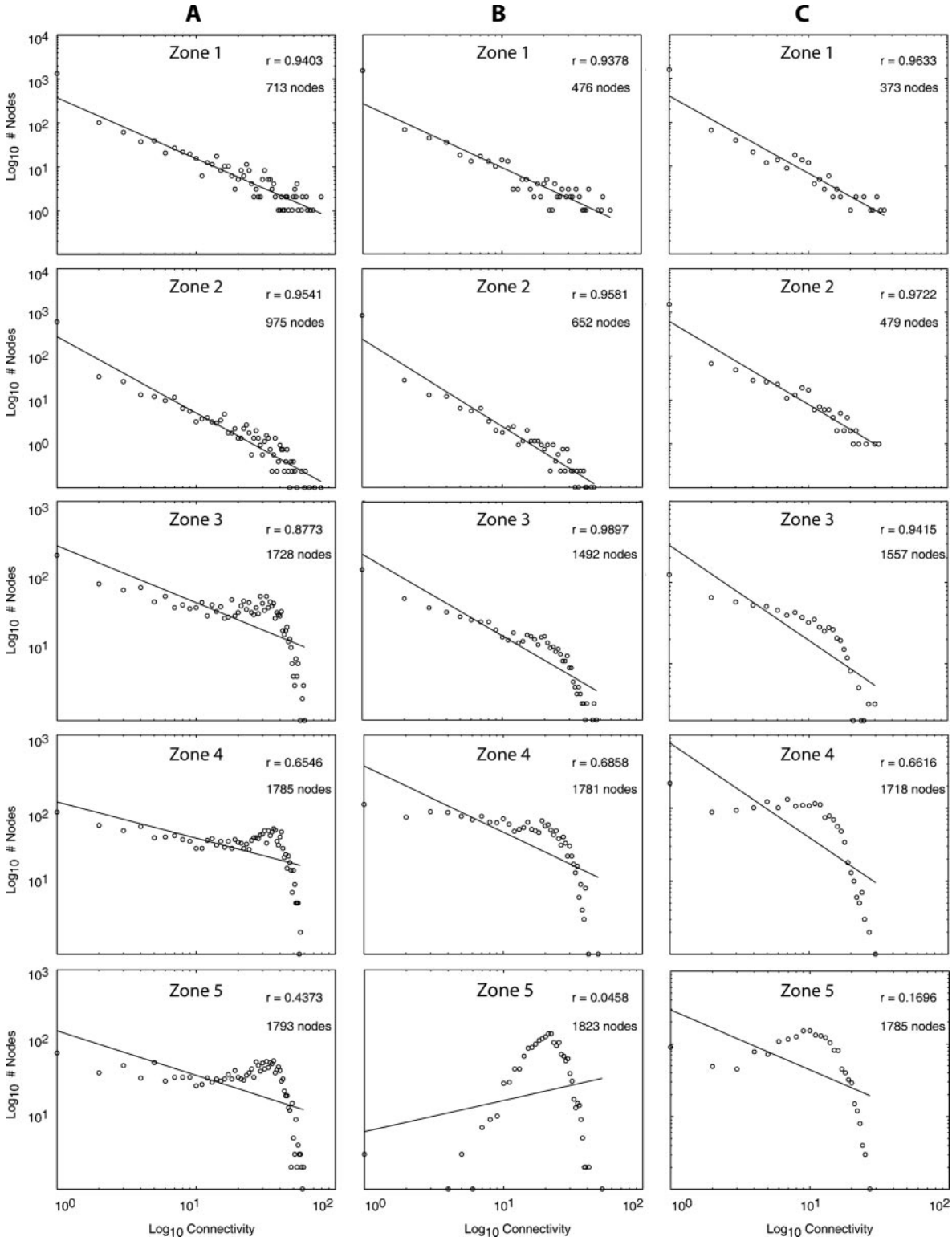

FIGURE 3. Connectivity distribution plots. Connectivity plots were generated for networks built using different mutual information values from Fig. 2 and a fitter to a power law. For each network the MI range, number of nodes, goodness of fit coefficient, and γ exponent is indicated. *A*, Treated individuals (T1 and T2). *B*, Controls 1 (U1 and U2). *C*, Controls 2 (U3 and U4).

smaller magnitude than that elicited by IFN-β administration. The gene ontologies from the genes in each network were retrieved and those categories showing statistically significant enrichment between networks were encircled (Fig. 4). Interestingly, categories like death, cell adhesion, response to stress, and response to biotic stimulus appear as a signature of the IFN-β-elicited network.

To formally address the hypothesis that genes in the control networks reflect normal gene activity, we conducted a gene ontology search on the 361 genes present in controls but not in the treated individuals. We observed that almost half of these genes either belonged to the category physiological process or macromolecule catabolism. For example, 155 genes were observed in a physiological

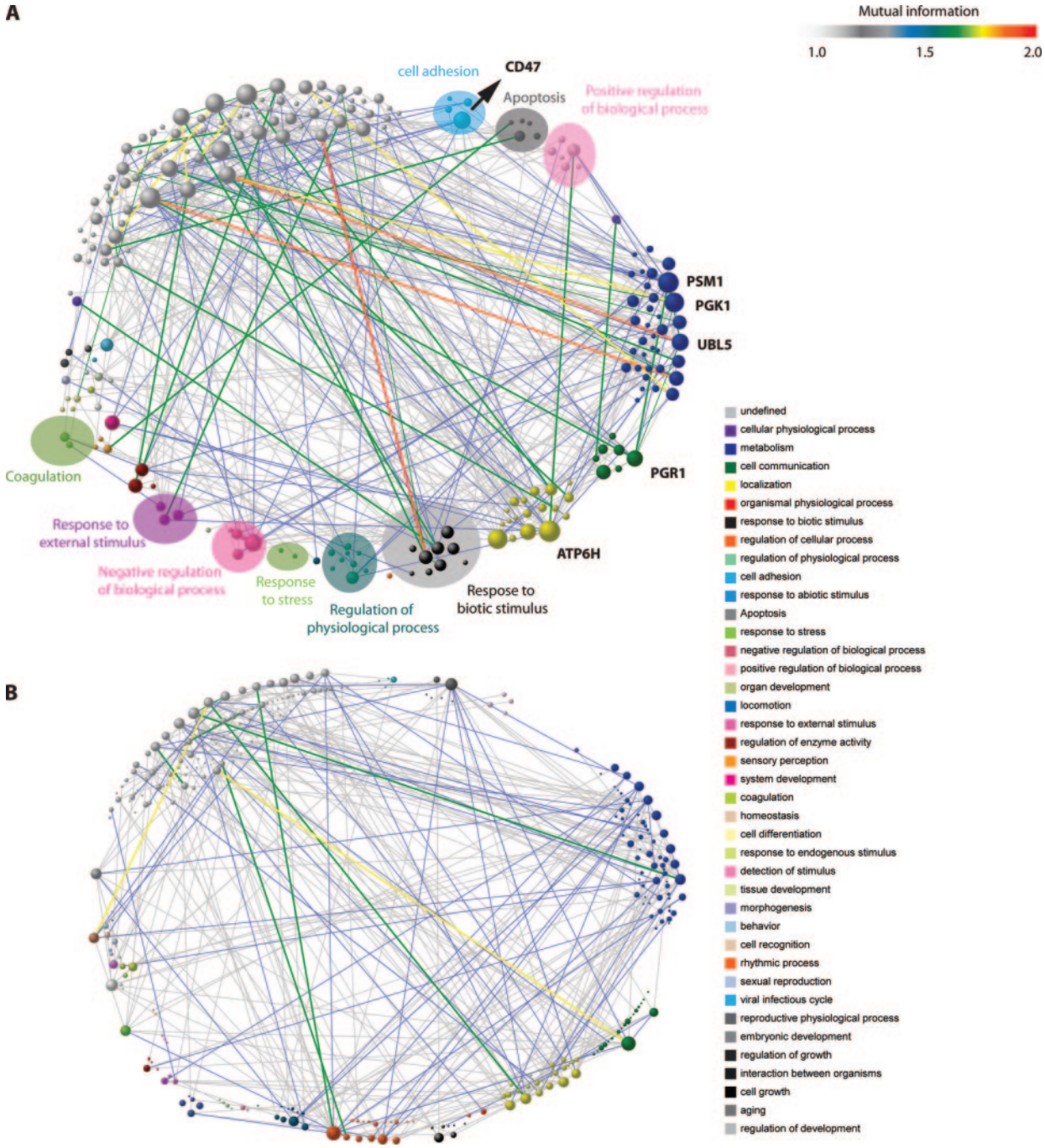

FIGURE 4. Relevant networks. Gene pairs in either the treated (*A*) or control (*B*) expression matrix showing a MI value above the TMI were linked by an edge, the color of which represents the magnitude of MI (from high to low: red, orange, yellow, green, and subsequently lighter shades of gray). Node size is proportional to the number of connections linking that gene with others. Node (gene) colors represent level 3 GO categories. Nodes from categories showing statistically significant representation across the two networks are highlighted in a circle. When present in both networks, matching GO categories are located in the same relative position to aid visualization.

process when only 133 would be expected by chance (expected) based on our sample size ($p \leq 3.86\ 10^{-5}$). A similar finding was seen for the macromolecular metabolism (O = 82, E = 57, $p = 8.6\ 10^{-5}$). These findings can be visually examined in Fig. 4, where nodes are colored according to the most statistically significant GO categories. As expected, genes from the patients network not present in any of the controls were strongly enriched in GO categories reflecting IFN-β activity such as apoptosis (O = 19, E = 9.6, $p = 0.003$) or immune response (O = 53, E = 14.9, $p = 2.11\ 10^{-16}$). The most enriched GO category from the list of 81 genes shared by patients and controls was response to stress (O = 10, E = 3.54, $p = 0.002$). Our results suggest that information theory principles can be applied to high frequency longitudinal transcriptional datasets to capture not only stimulus-induced changes but also normal physiological fluctuations.

Among the 10 most connected genes in the patients' networks were *CD47* and *GPR153* (*PGR1*), genes involved in cellular communication, as well as *PSMB1*, *PGK1*, *ATP6V0E* (*ATP6H*), and *UBL5*, involved in protein metabolism (Fig. 4). In contrast, seven of the 10 most connected nodes from the controls network belonged to the cellular physiological process category. This further suggests that the

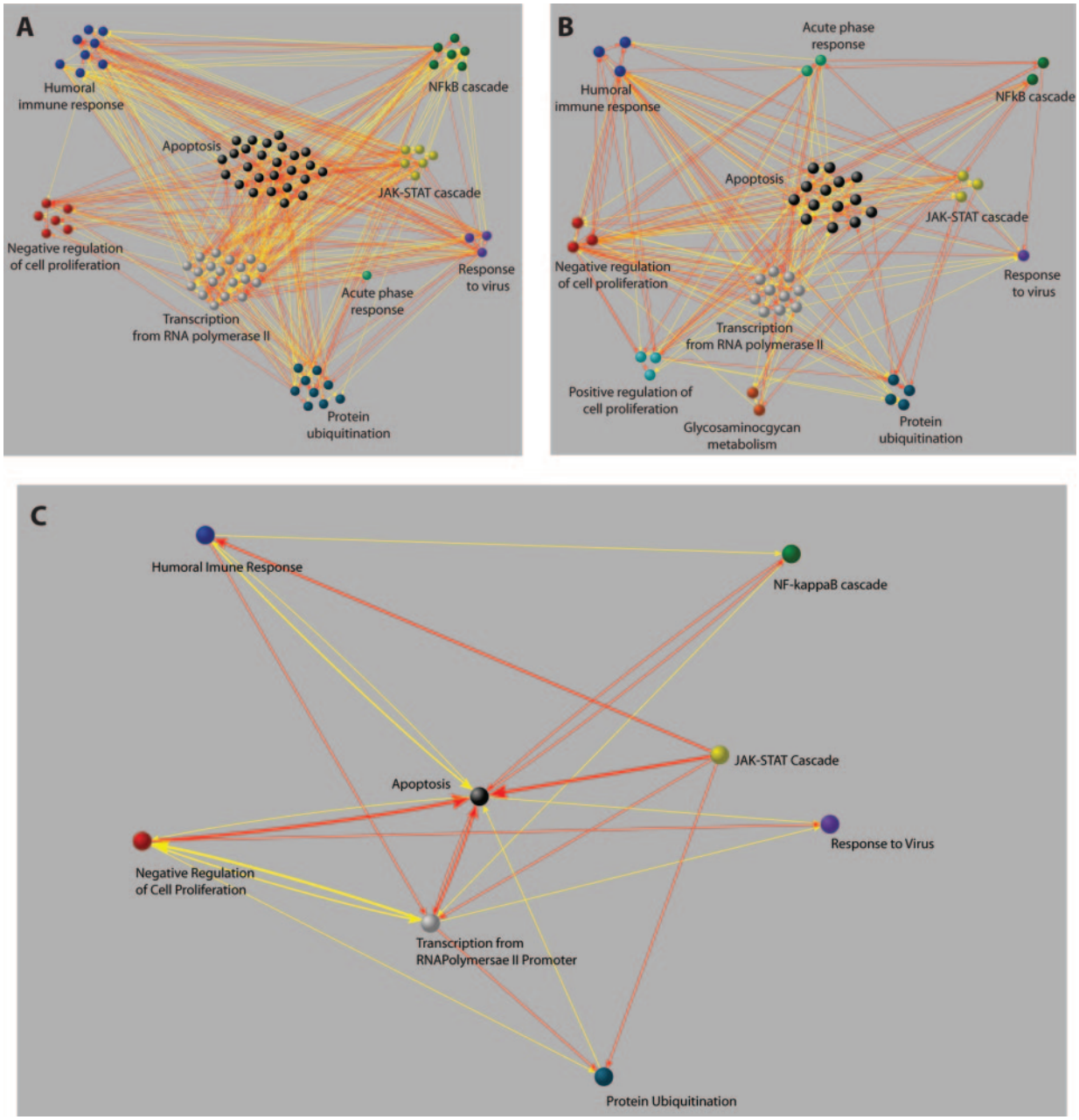

FIGURE 5. Reverse engineered networks of metagenes and gene ontologies reveal the IFN-β mechanism. *A* and *B*, Genes showing similar expression profiles over time were grouped into metagenes. Based on their component genes, each metagene was assigned a GO category known to be involved in IFN-β signaling. The longitudinal expression of each metagene was used to reverse engineer the most likely set of regulatory connections on each treated individual (T1 in *A* and T2 in *B*). Metagenes belonging to the same GO are represented as colored circles. *C*, An intersection of both networks is shown where the metagenes are replaced by their corresponding GO category. Yellow and red arrows denote interactions identified in at least 50 and 75% of the iterations, respectively.

coexpression network identified in the healthy individuals strongly reflects coordinated physiological gene activity, whereas in the IFN-β-treated individuals most of the coordinated transcription is in response to the drug.

Reverse engineering

Although MI captures pairwise relationships, the resulting network does not provide information on their directionality, (i.e., whether gene A regulates the expression of gene B or vice versa). To explore this possibility we applied SA, a method used to infer causative relations in a complex set of interacting components (reverse engineering). To perform this analysis we started out from the same set of 1826 annotated genes that showed significant changes along time in any of the analyzed individuals. Because SA is an exhaustive method, a reduction in the number of inputs was mandatory. To accomplish this, we compiled transcripts with similar expression dynamics and biological function into metagenes (see *Materials and Methods*). Using the canonical IFN-β signaling pathway as a starting point, we identified 13 gene ontology categories (from the top-level category, biological function) associated with each gene in the pathway. Finally, we populated each preselected GO category with only those metagenes whose component transcripts completely belonged to that GO. This resulted in 63 and 44 metagenes for individual T1 and T2, respectively, corresponding to 13 GO categories characteristic of IFN-β response. A recurrent Elman neural network (see *Materials and Methods*) was

trained to predict the expression levels of a target metagene using the averaged expression levels of input metagenes (potential regulators of the target metagene). By systematically perturbing the expression profile of each input metagene and examining the resultant change in sum squared error (SSE) of the trained neural network, the significance of the jittered metagene in predicting the expression level of the target metagene was determined. After 20 iterations, one set of regulatory interactions supported in at least 10 (50%, yellow) and 15 (75%, red) independent runs was determined for each of the two IFN-treated individuals (Fig. 5, *A* and *B*). In both networks, interactions among metagenes did not seem randomly assigned. Indeed, interactions seemed more prominent among metagenes belonging to some GO categories and not others. For example several directed links were generated among transcription, regulation of cell proliferation (both positive and negative), and apoptosis. These relationships can be seen more clearly in Fig. 5*C*, where overlapping connections between the two networks are shown only at the GO level. Interestingly, the two categories with the most inputs are transcription and apoptosis, suggesting that these are two of the ultimate targets of IFN-β. Conversely, the GO category with the most outgoing connections was JAK/STAT cascade, known to be the earliest event after IFN-β binds to its receptor. These findings could well summarize the mechanism of action of IFN-β at the molecular level.

Discussion

The global transcriptional response to IFN-β in humans provides an outstanding opportunity to analyze higher order properties of networks of coregulated genes of immunological interest. Although studies of gene expression profiling have been performed to evaluate the response to IFN-β, few were based on a longitudinal design, instead focusing on cross-sectional changes of large magnitude. In this study we generated in vivo longitudinal transcriptional profiles from two individuals immediately after IFN-β administration and from four untreated controls. This high-frequency longitudinal sampling strategy, focusing on the first week after IFN-β administration, allowed us to closely monitor the behavior of several known IFN-regulated genes as well as to identify novel targets. Even though a variety of genes were up-regulated in response to IFN-β, most genes experiencing detectable changes in expression were down-regulated. Among these, ribosomal genes, mitosis, and anti-apoptosis-related transcripts are of critical importance as they indicate that IFN-β slows down cellular activity and initiates commitment to programmed cell death. Although whole blood was used for these experiments, the clear reduction observed in CD3D suggests that most of the signals described above may originate from T cells.

Because of the limited sample size we were cautious not to draw general conclusions about specific genes or pathways and their involvement in response to IFN-β administration. However, we believe that our work provides a novel alternative to the study of large gene expression datasets in response to therapeutic drugs. One important aspect of our study is that we were able to assess the topological properties of in vivo gene expression networks from both IFN-β-treated and untreated individuals. Previous studies have demonstrated that some general properties of large networks are conserved across many different systems, from power grids to commercial flights to protein interactions. One such property is the connectivity distribution approximating a power law. In practical terms this means that while only a few nodes (genes) are highly connected, most will have very few connections. From a functional perspective these networks are more tolerant of the random failure of their components (robustness) and therefore compatible with evolutionary constraints. Because the IFN signaling pathway plays a critical role in the natural antiviral response in most mammals, it is reasonable to speculate that organisms evolved a robust mechanism for such a response. This suggests that the transcriptional signature engraved in blood cells (particularly T lymphocytes) is of such a magnitude that it was still detectable even in the presence of interfering patterns (noise) from other cell populations less affected by this drug.

Another important finding of this analysis is that the main regulatory effects of IFN-β were captured. The fact that transcriptional analysis was conducted from total blood makes this finding even more remarkable as gene regulation in response to an external stimulus likely diverges among cell types. The statistically significant gene associations that we were able to capture by this approach are also biologically relevant. This observation is supported by the fact that only the networks built from MI values near or above the threshold display scale-free properties. Furthermore, when smaller MI values were used the resulting connectivity profile displayed a Poisson distribution similar to what is found for random networks.

Although large networks can be also assembled using pairwise correlation, this metric performs poorly when the relationships involved are nonlinear. As growing evidence suggests, gene regulation is largely a nonlinear process. Hence MI, a method that does not assume linearity, is uniquely suitable for this type of analysis. An important limitation of the coexpression networks generated in this study resides in the fact that the directionality of the relationship cannot be specified. That is, a high MI value between two genes, A and B, indicates that these two genes are coregulated in time but it does not indicate whether one regulates the expression of the other or, if so, in which direction this happens. We point out, however, that this is not a limitation of the mutual information method per se but a result of our dataset having a limited number of time points. A directed network is obtainable if the expression of a gene at time t is correlated to the expression of all other genes at time $t + 1$. A high MI value between two such genes would imply a regulatory influence of the first gene over the second. Although potentially interesting, the number of time points needed for such an analysis would be in the order of 20–30, making this approach unfeasible, at least in vivo. We overcame this problem and explored the directionality of some of these interactions using SA. In this context, SA can be seen as systematic in in silico gene knockout experiments with the purpose of inferring genetic networks. SA has been successfully used in the past on a trained linear neural network to reverse engineer genetic networks using a single layer perceptron architecture (32). In this study we have expanded these methods by using the following: 1) a nonlinear transfer function that introduces needed nonlinearity while imposing an upper bound on the production rate of the mRNA; 2) a recurrent (Elman) neural network that adds explicit time-point expression memory to the system; and 3) a newly introduced novel systematic approach to dimension reduction in the gene space. This approach of combining dimensionality reduction with SA allowed us to capture biological relationships that would have gone unnoticed otherwise.

When the directed networks from the two treated individuals were superimposed, some interactions among biological functions were consistently observed (Fig. 5). For example, the JAK-STAT cascade influences the expression of genes from several other categories as evidenced by the large number of outgoing arrows. Similarly, apoptosis is a category receiving interactions from genes from most other categories, indicating that a variety of processes converge into programmed cell death after IFN-β administration.

Our approach also allowed us to detect networks of coordinated gene activity in healthy individuals. Remarkably, in the absence of a strong external perturbation (an immunomodulatory drug), genes

involved in normal cell physiology and metabolism were identified. Not surprisingly, the MI values connecting genes in the control networks were smaller than those in the IFN-β-treated individuals (Fig. 4). This reflects a wider breadth of gene expression in controls, possibly due to circadian rhythms or interindividual metabolic variability. Despite this, when overlapping directed networks from untreated individuals were analyzed a set of interactions were also identified but differed markedly from the ones in Fig. 5. For example, no genes involved in the JAK-STAT cascade were consistently influencing the expression of other genes, and apoptosis was a much less visited node (data not shown).

If functional genomic variants are responsible for an individual's ability to respond satisfactorily to a therapeutic drug, this may well be reflected in that individual's expression profile. Although studies with larger cohorts are warranted, we hypothesize that by analyzing networks of coregulated genes in patients before and immediately after the initiation of therapy a correlation between the general properties of these networks and response to therapy, measured prospectively, can be drawn. This would set the stage for a global, transcription-based, personalized assay.

Acknowledgments

We are grateful to the individuals who participated in this study.

Disclosures

A.R.P. and K.N. have received research grants or honoraria from Berlex, Biogen Idec, and Serono.

References

1. Kappos, L., and H. P. Hartung. 2005. 10 years of interferon β-1b (Betaferon) therapy. *J. Neurol.* 252(Suppl. 3): iii1–iii2.
2. David, M. 2002. Signal transduction by type I interferons. *BioTechniques* Suppl.: 58–65.
3. Billiau, A., B. C. Kieseier, and H. P. Hartung. 2004. Biologic role of interferon β in multiple sclerosis. *J. Neurol.* 251(*Suppl. 2*): II10–II14.
4. Rep, M. H., H. M. Schrijver, T. van Lopik, R. Q. Hintzen, M. T. Roos, H. J. Ader, C. H. Polman, and R. A. van Lier. 1999. Interferon (IFN)-β treatment enhances CD95 and interleukin 10 expression but reduces interferon-γ producing T cells in MS patients. *J. Neuroimmunol.* 96: 92–100.
5. Minagar, A., A. Long, T. Ma, T. H. Jackson, R. E. Kelley, D. V. Ostanin, M. Sasaki, A. C. Warren, A. Jawahar, B. Cappell, and J. S. Alexander. 2003. Interferon (IFN)-β 1a and IFN-β 1b block IFN-γ-induced disintegration of endothelial junction integrity and barrier. *Endothelium* 10: 299–307.
6. Leppert, D., E. Waubant, M. R. Burk, J. R. Oksenberg, and S. L. Hauser. 1996. Interferon β-1b inhibits gelatinase secretion and in vitro migration of human T cells: a possible mechanism for treatment efficacy in multiple sclerosis. *Ann. Neurol.* 40: 846–852.
7. Kraus, J., R. Bauer, N. Chatzimanolis, B. Engelhardt, J. Tofighi, T. Bregenzer, B. S. Kuehne, E. Stolz, F. Blaes, K. Morgen, et al. 2004. Interferon-β1b leads to a short-term increase of soluble but long-term stabilisation of cell surface bound adhesion molecules in multiple sclerosis. *J. Neurol.* 251: 464–472.
8. Floris, S., S. R. Ruuls, A. Wierinckx, S. M. van der Pol, E. Dopp, P. H. van der Meide, C. D. Dijkstra, and H. E. De Vries. 2002. Interferon-β directly influences monocyte infiltration into the central nervous system. *J. Neuroimmunol.* 127: 69–79.
9. Wandinger, K. P., J. D. Lunemann, O. Wengert, J. Bellmann-Strobl, O. Aktas, A. Weber, E. Grundstrom, S. Ehrlich, K. D. Wernecke, H. D. Volk, and F. Zipp. 2003. TNF-related apoptosis inducing ligand (TRAIL) as a potential response marker for interferon-β treatment in multiple sclerosis. *Lancet* 361: 2036–2043.
10. Sharief, M. K., and Y. K. Semra. 2002. Down-regulation of survivin expression in T lymphocytes after interferon β-1a treatment in patients with multiple sclerosis. *Arch. Neurol.* 59: 1115–1121.
11. Sharief, M. K., Y. K. Semra, O. A. Seidi, and Y. Zoukos. 2001. Interferon-β therapy downregulates the anti-apoptosis protein FLIP in T cells from patients with multiple sclerosis. *J. Neuroimmunol.* 120: 199–207.
12. Arbour, N., E. Rastikerdar, E. McCrea, Y. Lapierre, J. Dorr, A. Bar-Or, and J. P. Antel. 2005. Upregulation of TRAIL expression on human T lymphocytes by interferon β and glatiramer acetate. *Mult. Scler.* 11: 652–657.
13. Van Weyenbergh, J., J. Wietzerbin, D. Rouillard, M. Barral-Netto, and R. Liblau. 2001. Treatment of multiple sclerosis patients with interferon-β primes monocyte-derived macrophages for apoptotic cell death. *J. Leukocyte Biol.* 70: 745–748.
14. Wandinger, K. P., C. S. Sturzebecher, B. Bielekova, G. Detore, A. Rosenwald, L. M. Staudt, H. F. McFarland, and R. Martin. 2001. Complex immunomodulatory effects of interferon-β in multiple sclerosis include the upregulation of T helper 1-associated marker genes. *Ann. Neurol.* 50: 349–357.
15. Bartholome, E. J., F. Willems, A. Crusiaux, K. Thielemans, L. Schandene, and M. Goldman. 1999. Interferon-β inhibits Th1 responses at the dendritic cell level. Relevance to multiple sclerosis. *Acta Neurol. Belg.* 99: 44–52.
16. Der, S. D., A. Zhou, B. R. Williams, and R. H. Silverman. 1998. Identification of genes differentially regulated by interferon α, β, or γ using oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* 95: 15623–15628.
17. Koike, F., J. Satoh, S. Miyake, T. Yamamoto, M. Kawai, S. Kikuchi, K. Nomura, K. Yokoyama, K. Ota, T. Kanda, T. Fukazawa, and T. Yamamura. 2003. Microarray analysis identifies interferon β-regulated genes in multiple sclerosis. *J. Neuroimmunol.* 139: 109–118.
18. Iglesias, A. H., S. Camelo, D. Hwang, R. Villanueva, G. Stephanopoulos, and F. Dangond. 2004. Microarray detection of E2F pathway activation and other targets in multiple sclerosis peripheral blood mononuclear cells. *J. Neuroimmunol.* 150: 163–177.
19. Weinstock-Guttman, B., D. Badgett, K. Patrick, L. Hartrich, R. Santos, D. Hall, M. Baier, J. Feichter, and M. Ramanathan. 2003. Genomic effects of IFN-β in multiple sclerosis patients. *J. Immunol.* 171: 2694–2702.
20. Geiss, G. K., V. S. Carter, Y. He, B. K. Kwieciszewski, T. Holzman, M. J. Korth, C. A. Lazaro, N. Fausto, R. E. Bumgarner, and M. G. Katze. 2003. Gene expression profiling of the cellular transcriptional network regulated by α/β interferon and its partial attenuation by the hepatitis C virus nonstructural 5A protein. *J. Virol.* 77: 6367–6375.
21. Bar-Joseph, Z. 2004. Analyzing time series gene expression data. *Bioinformatics* 20: 2493–2503.
22. Hood, L., J. R. Heath, M. E. Phelps, and B. Lin. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science* 306: 640–643.
23. Huang, S. 1999. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *J. Mol. Med.* 77: 469–480.
24. Albert, R., H. Jeong, and A. L. Barabasi. 2000. Error and attack tolerance of complex networks. *Nature* 406: 378–382.
25. Albert, R., and A. L. Barabasi. 2000. Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* 85: 5234–5237.
26. Barabasi, A. L., and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286: 509–512.
27. Bornholdt, S., and H. G. Schuster. 2003. *Handbook of Graphs and Networks: From the Genome to the Internet.* Wiley-VCH, Berlin.
28. Strogatz, S. H. 2001. Exploring complex networks. *Nature* 410: 268–276.
29. Butte, A. J., and I. S. Kohane. 2001. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* 418–429.
30. Eberwine, J. 1996. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *BioTechniques* 20: 584–591.
31. Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5: 101–113.
32. Krishna, A., A. Narayanan, and E. Keedwell. 2005. Reverse engineering gene networks with artificial neural networks. In *International Conference on Adaptive and Natural Computing Algorithms*, Coimbra, Portugal.