

# Bioinformatic Analysis Reveals cRel as a Regulator of a Subset of Interferon-Stimulated Genes

Lai Wei,<sup>1</sup> Meiyun Fan,<sup>1</sup> Lijing Xu,<sup>2,3</sup> Kevin Heinrich,<sup>4</sup> Michael W. Berry,<sup>4</sup>  
Ramin Homayouni,<sup>2,3</sup> and Lawrence M. Pfeffer<sup>1</sup>

Interferons (IFNs) are critical to the host innate immune response by inducing the expression of a family of early response genes, denoted as IFN-stimulated genes (ISGs). The role of tyrosine phosphorylation of STAT proteins in the transcription activation of ISGs is well-documented. Recent studies have indicated that other transcription factors (TFs) are likely to play a role in regulating ISG expression. Here, we describe a novel integrative approach that combines gene expression profiling, promoter sequence analysis, and literature mining to screen candidate regulatory factors in the IFN signal transduction pathway. Application of this method identified the nuclear factor  $\kappa$ B (NF $\kappa$ B) protein, cRel, as a candidate regulatory factor for a subset of ISGs in mouse embryo fibroblasts. Chromatin immunoprecipitation (ChIP) and real-time PCR assays confirmed that cRel directly binds to the promoters of several ISGs, including Cxcl10, Isg15, Gbp2, Ifit3, and Ifi203, and regulates their expression. Thus, our studies identify cRel as an important TF for ISGs, and validate the approach of using Latent Semantic Indexing (LSI)-based methods to identify regulatory factors from microarray data.

## Introduction

INTERFERONS (IFNs) WERE DISCOVERED by virtue of their antiviral activity; however, IFNs are multifunctional proteins that also affect cell proliferation, cell differentiation, apoptosis (programmed cell death), and the immune response. Type I IFNs, consisting of IFN- $\alpha$ , IFN- $\beta$ , IFN- $\omega$ , IFN- $\epsilon$ , IFN- $\kappa$ , IFN- $\delta$ , and IFN- $\tau$ , regulate their diverse cellular functions by modulating the expression of IFN-stimulated genes (ISGs) through the activation of a signal transduction pathway involving the JAK tyrosine kinases and STAT proteins (Friedman and Stark 1985; Larner and others 1986; Schindler and others 1992; Darnell and others 1994). Although the JAK/STAT signal transduction pathway is critical in mediating IFNs' antiviral and antiproliferative activities, IFNs also activate the nuclear factor  $\kappa$ B (NF $\kappa$ B) signaling pathway, which also plays an important role in the biological actions of IFN (Yang and others 2000; Yang and others 2001; Pfeffer and others 2004).

The NF $\kappa$ B transcription factor (TF) family regulates the expression of genes involved in cell survival and immune responses (Beg and others 1995; Beg and Baltimore 1996;

Van Antwerp and others 1996; Wang and others 1996). In mammals, the NF $\kappa$ B family of related proteins includes NF $\kappa$ B1 (p105 processed to p50), NF $\kappa$ B2 (p100 processed to p52), RelA (p65), RelB, and cRel. Both p50 and p52 lack a transcription activation domain, and as homodimers function as repressors. In contrast, p65, cRel, and RelB have a transcription activation domain, and thus when complexed with p50 or p52 are capable of activating transcription. Although p50:p65 and p52:RelB heterodimers are the NF $\kappa$ B complexes most often observed in cells, other Rel heterodimers also form. Recent studies identified that IFN induces NF $\kappa$ B activation through both a classical pathway that results in the formation of p50:NF $\kappa$ B dimers through I $\kappa$ B degradation, and an alternative pathway that results in the formation of p52:NF $\kappa$ B dimers through a NF $\kappa$ B-inducing kinase (NIK)/tumor necrosis factor (TNF) receptor-associated factors (TRAF)-dependent pathway (Yang and others 2001; Pfeffer and others 2004; Yang and others 2005a; Wei and others 2006). Moreover, a subset of ISGs are regulated by NF $\kappa$ B, and appear to play important roles in IFNs' biological actions.

<sup>1</sup>Department of Pathology and Laboratory Medicine, and the Center for Cancer Research, University of Tennessee Health Science Center, Memphis, Tennessee.

<sup>2</sup>Department of Biology and <sup>3</sup>Bioinformatics Program, University of Memphis, Memphis, Tennessee.

<sup>4</sup>Department of Computer Science, University of Tennessee, Knoxville, Tennessee.

The coordinated regulation of gene expression by extracellular signals requires the interplay between multiple TFs that selectively bind to gene promoters with spatial and temporal precision (Bluthgen and others 2005). Comprehensive analysis of motifs in gene promoters by phylogenetic footprinting or Bayesian clustering has been useful in constructing regulatory networks in mammalian cells (Qin and others 2003; Xie and others 2005). Combining gene expression profiling with promoter analysis may significantly improve identification of novel gene regulatory networks based on the assumption that co-regulated genes share common TF-binding sites and regulatory factors. However, due to the complexity of mammalian promoters, genes often contain potential binding sites for several TFs, making identification of regulatory programs very difficult. More sophisticated probabilistic graphical models have been explored to identify critical regulators, which are themselves transcriptionally regulated by differing stimuli (Segal and others 2003; Li and others 2005). Also, combination of promoter analysis and gene function information in human curated Gene Ontology database has had some success in identifying meaningful gene regulatory mechanisms (Bluthgen and others 2005). However, as with any human curated index, Gene Ontology may have limited usefulness because it is incomplete and contains broad functional index terms. Therefore, integration of functional information extracted from the primary literature would substantially improve these approaches.

Classical information retrieval and information extraction methods have recently been employed to mine the biomedical literature to elucidate gene/protein function and regulatory networks (Krallinger and others 2005; Rebholz-Schuhmann and others 2005). Information retrieval involves term matching and may include Boolean method or lexical matching methods common to natural language processing (Jenssen and others 2001; Yandell and Majoros 2002). However, these methods are limited in that they rely solely on known relationships. On the other hand, information extraction involves mathematical modeling of text, for example, by vector-space or statistical (Bayesian) approaches (Shatkay and Feldman 2003), which can deduce relationships from the literature even in the absence of a direct link. We previously developed a method using Latent Semantic Indexing (LSI) to identify gene relationships with high precision from titles and abstracts in MEDLINE citations (Homayouni and others 2005). In addition, we showed that this method identified both known (explicit) as well as unknown (implicit) gene relationships from the literature.

In the present report, we have applied a novel algorithm that combines expression profiling, promoter analysis and LSI to infer relationships between co-regulated genes and the TFs that are shared in gene promoters (outlined in Fig. 1). To test this method, we interrogated a dataset of ISGs from mouse embryonic fibroblasts (MEFs) (Pfeffer and others 2004). Our algorithm identified the cRel as a potential TF for a subset of ISGs. Chromatin immunoprecipitation (ChIP) assays demonstrated that cRel directly binds to the promoters of many of the ISGs identified, including Cxcl10, G1p2/Isq15, Gbp2, Ifit3, and Ifi203. Moreover, real-time PCR

analysis demonstrated that induction of these ISGs by IFN was diminished in cRel-deficient MEFs as compared to MEFs from their wild-type (WT) littermates.

## Materials and Methods

### *Biological reagents and cell culture*

Highly purified recombinant rat IFN- $\beta$  was obtained from Biogen-Idec, Inc. ( $2 \times 10^8$  IU/mg protein) (Arduini and others 2004). Polyclonal anti-cRel was obtained from Santa Cruz Biotechnology (Santa Cruz, CA, USA). 3T3 immortalized MEFs generated from E12.5 to 14.5 embryos from C57BL6/J mice and cRel-deficient littermates were generously provided by Dr. Alexander Hoffmann (University of California, San Diego) (Hoffmann and others 2003). MEFs were plated at  $3 \times 10^5$  cells/60-mm dish every 3 days in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS) (Hyclone Laboratories, Logan, UT, USA), and 100  $\mu$ g/mL of penicillin G and streptomycin.

### *Microarray analysis*

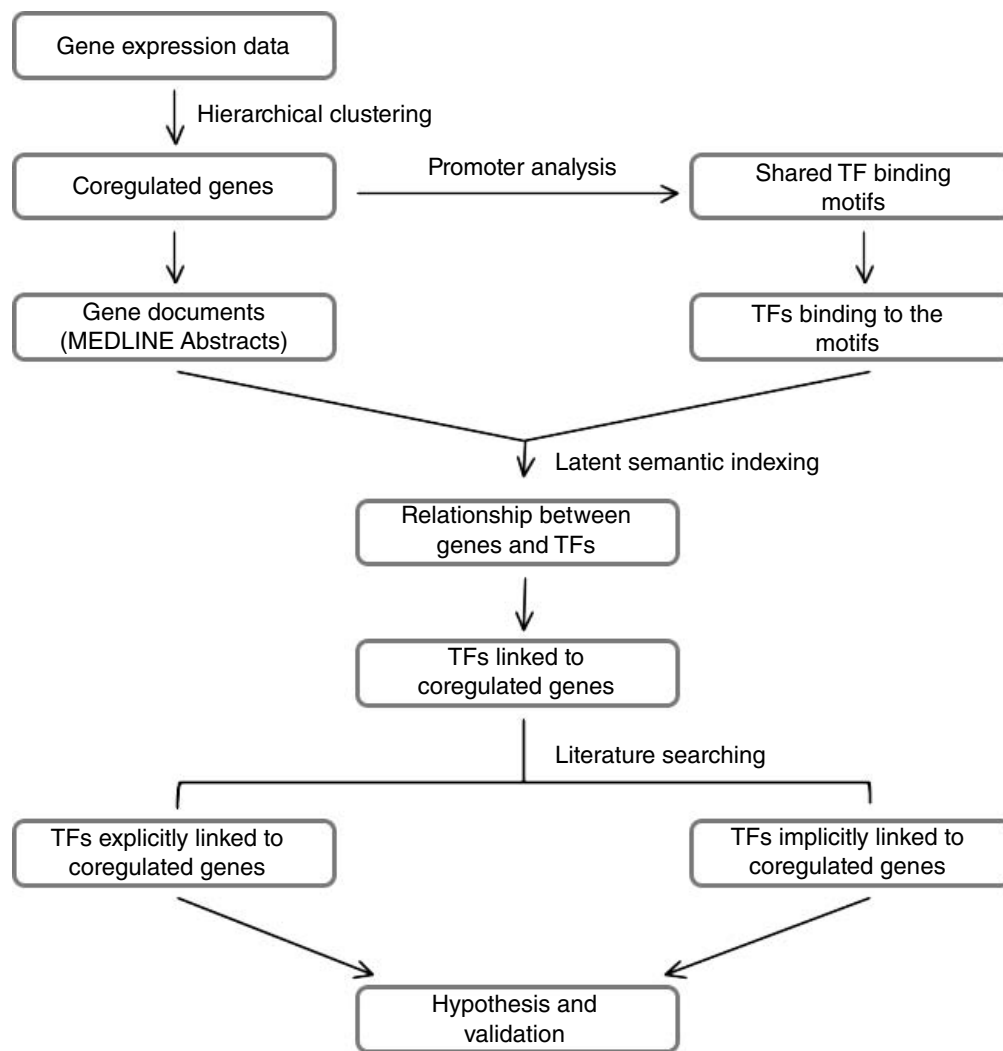
Microarray experiments and data analysis were performed as described previously (Pfeffer and others 2004). In brief, total cellular RNA from untreated and IFN- $\beta$ -treated (2500 units/mL for 5 h) MEFs was extracted with TRIzol reagent (Invitrogen) and submitted to Genome Explorations Inc. (Memphis, TN, USA) for labeling and hybridization to the murine U74Av2 GeneChip (Affymetrix Inc.). The data were deposited in Gene Expression Omnibus (GEO) at NCBI (Accession GSM76653-76658). Expression values were determined using Affymetrix Microarray Suite Version 5.0 software and analyzed using GeneSpring 7.2 (Silicon Genetics, Inc.). A subset of genes were selected whose expression was altered greater than 2-fold ( $p < 0.05$ , Welch's  $t$ -test,  $n = 3$  for each group) and then subjected to hierarchical clustering using standard correlation coefficients.

### *Promoter sequence analysis*

Genomic sequences corresponding to 3-kb upstream and 50-bp downstream of transcription start site (TSS) of each gene was obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/>). DNA-binding motifs and TF-binding sites were identified using MOTIF (<http://motif.genome.jp>) and the information in TRANSFAC 6.0 database (Matys and others 2003).

### *LSI-based literature analysis*

The text representation was performed as described previously (Homayouni and others 2005). In brief, the information for each gene was generated by concatenation of titles and abstracts in the MEDLINE citations cross-referenced in the mouse, rat and human Entrez Gene entries. To lower the false-positive rate, all references to sequencing or other high-throughput projects were removed from the database. Relationships between ISGs and TFs were determined by



**FIG. 1.** Overview of the method used to identify candidate regulatory transcription factors (TFs) from microarray data. Small groups of genes that are similarly affected by experimental treatments are selected for promoter sequence analysis to identify shared TF-binding motifs. A matrix consisting of pairwise literature relationships between each TF and co-regulated gene is constructed by Latent Semantic Indexing (LSI). The relationship *d*values for genes and TFs are then subjected to hierarchical clustering and the TFs that have high literature relationship to the co-regulated genes are selected for manual analysis of the literature to determine an explicit or implicit relationship.

calculating the cosine of vector angles between gene document vectors derived by a rank-300 approximation [single value decomposition (SVD)-based] to the original term-by-gene document matrix. The relationship of each gene to potential TFs was normalized by calculating the Z-score distribution of the similarity scores. The normalized similarity scores were imported into GeneSpring 7.2 and subjected to hierarchical clustering using standard correlation coefficients.

#### ChIP assays

ChIP experiments were performed using the ChIP-IT™ Chromatin Immunoprecipitation Kit (Active Motif, Carlsbad, CA, USA) according to the manufacturer's instructions. In brief, cells were fixed with 1% formaldehyde at

22°C for 15 min to generate protein–DNA cross-links, and chromatin DNA was fragmented to an average size of 600 bp. Immunoprecipitation with anti-cRel or anti-immunoglobulin (Ig) of sheared chromatin was performed on precleared cell lysates for 1 h at 4°C. PCR was performed for 38 cycles using 2 μL out of a 100 μL DNA extraction and the following forward and reverse primers corresponding to the upstream regions of ISGs flanking cRel-binding sites that are close to TSSs were used:

Cxcl10: 5'-CCTGTAAACCGAGGGCATTG-3', 5'-CACGCTT TGGAAAGTGAAAC-3';

Isg15: 5'-CCTTCTCTCTCCACTTTG-3', 5'-AGGTGAGAT GGGAGGTAGAG-3';

Gbp2: 5'-GTCTCAGTTTTGACAGTGGC-3', 5'-GTGGAGTT TCCAGTCATTTG-3';

Ifit3: 5'-CTGTCAGGCTGGAGGAAATG-3', 5'-TCAACCAG AAGAGGAAAGTG-3';  
 Ifit1: 5'-TGATGCAGAGAACACAGCCA-3', 5'-CTTCTTTCC TTTGGTCTTC-3';  
 Ifi203: 5'-CTTGGAACCCATGAAATTG-3', 5'-TTTTGGAA TGAAAGTAACCA-3'

### Quantitative real-time PCR

Total RNA was isolated from untreated and IFN- $\beta$ -treated (1000 IU/mL, 5 h) MEFs using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Quantitative RT-PCR was performed on an iCycler (BioRad) with 60 ng of total RNA in 15  $\mu$ L of reaction mixture using the AccessQuick<sup>TM</sup> RT-PCR system (Promega, Madison, WI, USA), and SYBR green I (Molecular Probes, Eugene, OR, USA) according to the manufacturer's instructions. The one-step RT-PCR cycling was as follows: reverse transcription at 48°C for 45 min, denaturation at 95°C for 2 min, amplification for 35 cycles at 94°C for 30 s, and 62°C for 30 s. The product size was initially monitored by agarose gel electrophoresis and melting curves were analyzed to control for the specificity of PCR reactions. Gene expression data was normalized to the expression of the housekeeping gene  $\beta$ -actin. The relative units were calculated from a standard curve, plotting three different concentrations against the PCR cycle number at

the cycle threshold (with a 10-fold increment equivalent to  $\sim$ 3.1 cycles).

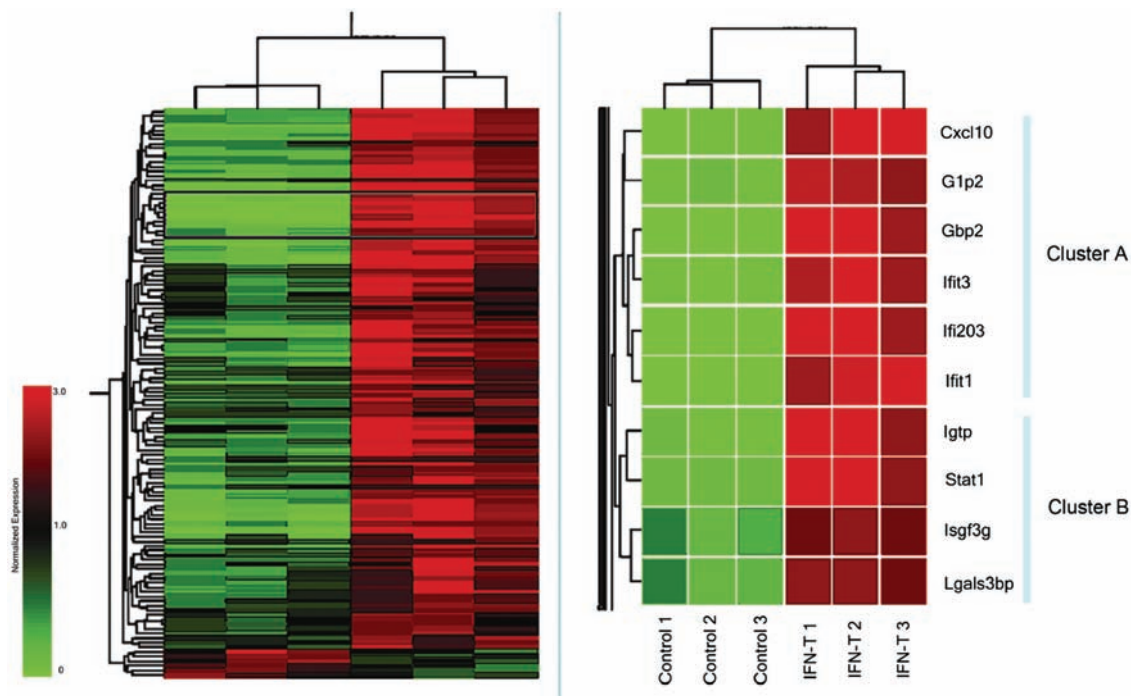
The following forward and reverse primers were used for each gene, respectively:

Cxcl10: 5'-CGTGTGAGATCATTGCCAC-3', 5'-TTAAGGA GCCCTTTTAGACC-3'  
 Gbp2: 5'-CCTCTTCCTCAAATGAGAC-3', 5'-GTGTTTCA ACAACATCTGCC-3'  
 Ifi203: 5'-AGTGGTGGTTTATGGACGAC-3', 5'-CTGTGCCT TACAGACCTCAG-3'  
 Ifit1: 5'-AAGAGAAGTCCTTTGCTTGG-3', 5'-TGCCCTTTC AGTTTGTAGAC-3'  
 Ifit3: 5'-GACGATTAACGATGGAGTTC-3', 5'-GGGCTCTC CTTACTGATGAC-3'  
 Isg15: 5'-ATGAGTCTTTCTGACGCAG-3', 5'-AGCAGCTC CTTGTCCTCCAT-3'  
 $\beta$ -actin: 5'-AAGGAGATTACTGCTCTGGC-3', 5'-ACATCT GCTGGAAGGTGGAC-3'.

## Results

### Identification of promoter elements in co-regulated genes

In a previous study, we identified 124 genes whose expression levels were significantly changed in MEFs after



**FIG. 2.** Gene expression clustering of 124 ISGs. ISGs (rows) were subjected to hierarchical clustering based on expression values obtained from three IFN-treated and three IFN-untreated samples (columns). Two related gene clusters (boxed in left panel) were identified as Cluster A and Cluster B gene clusters (right panel). Color bar indicates the normalized expression values below (green) and above (red) the median expression (set to 1.0). The brackets at the top of each panel reveal the similarities among samples, showing that control and IFN-treated triplicates clustered together, respectively.

IFN- $\beta$  treatment (Pfeffer and others 2004). Hierarchical clustering using standard correlation coefficients revealed several clusters of ISGs with highly similar expression profiles (Fig. 2). The basic hypothesis for our approach is that genes displaying closely related expression profiles are regulated by a common set of TFs. However, neighboring gene clusters are distinguished from one another by their utilization of different TFs. To test this approach, we selected two closely related clusters that were identified from this ISG dataset for promoter sequence analysis. Cluster A included the ISGs: Cxcl10, Isg15/G1p2, Gbp2, Ifit3, Ifi203, and Ifit1. Cluster B included the ISGs: Igtf, Stat1, Irf9, and Lgals3bp. The expression of the ISGs in both clusters were highly induced by IFN, ranging from 8-fold for Irf9 to 669-fold for Ifit1 (Pfeffer and others 2004). The expression values for these ISGs were then normalized across all of the arrays. The ISGs were clustered based on the expression pattern rather than the amplitude of the expression changes. As illustrated in the schematic in Figure 1, we next examined the 3-kb upstream region of each gene in Clusters A and B using the genomic sequence information at UCSC Genome Browser and the web-based search tool MOTIF. The gene promoters contained an average of 381 (ranging from 334 to 420) motifs, corresponding to 67 (ranging from 62 to 74) different TF-binding sites. Forty-eight TFs were common to at least 5 out of the 6 genes in cluster A, and 46 TFs were common to 3 out of 4 genes in the Cluster B. Interestingly, only 11 TFs were unique to Cluster A, and 9 TFs were unique to Cluster B.

#### Literature-based prediction of candidate TFs

As outlined in Figure 1, the next task in our approach was to identify which TFs are already known to regulate the genes in Clusters A and B, and which TFs are likely to regulate them based on implied information in the literature. Previous studies have established a role for STAT and IFN regulatory factor (IRF) proteins in ISG regulation (Barnes and others 2002; Pfeffer and others 2004). However, the differences in the expression patterns among the ISGs suggest that other TFs are also involved in their regulation. To predict which TFs may play a role in ISG expression, we implemented LSI to identify both explicit and implicit (probable) relationships from MEDLINE abstracts for subsets of ISGs and the TFs that are shared in their promoters. First, an abstract document was constructed for each ISG and TF by concatenating titles and abstracts of the citations cross-referenced in their Entrez Gene entries. The number of abstracts used in our corpus for the ISGs ranged from 1 (Ifi203) to 162 (Stat1) and for the TFs ranged from 4 (Tcfap4) to 1025 (Trp53). The gene documents were parsed into a dictionary of terms (tokens) and weighted frequencies (mathematical values used to describe the correlation between terms and the corresponding MEDLINE documents), which were used to construct a term-by-gene document (sparse) matrix.

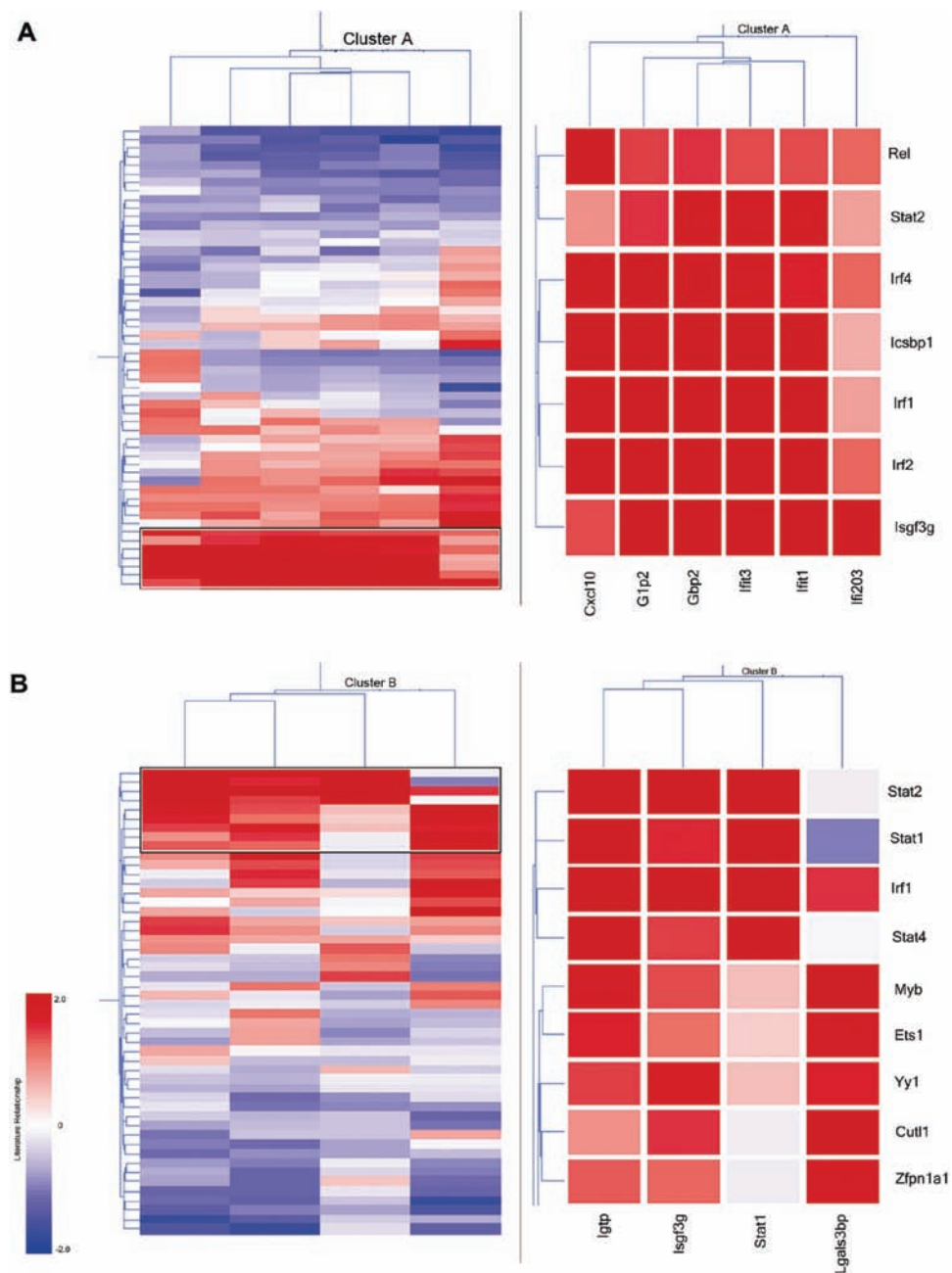
In the LSI model, term and document vectors are generated by truncating the singular value decomposition of the term-by-gene document matrix to a preselected number of factors. Fewer factors may be used for broad (more

conceptual) comparisons, whereas a larger number of factors may be used for specific (more literal) comparisons (Berry and others 1999). Thus, LSI produces a rank-reduced space in which two gene documents can be compared at different conceptual levels. Landauer and colleagues have previously demonstrated that maximal performance of LSI is achieved between 250 and 400 dimensions (factor space) (Landauer and others 2004). Therefore, in the present study, we calculated the relationships between ISGs and TFs using 300 factors. The resulting relationship values, which describe rank distribution, were then normalized and subjected to hierarchical clustering. We found that cRel, Stat2, Irf4, Icsbp1, Irf1, Irf2, and Irf9/Isgf3g were highly associated in the literature with the genes in Cluster A (Fig. 3A). On the other hand, STAT1, STAT4, IRF1, Myb, Ets1, Yy1, Cutl1, and Zfpn1a1 appeared to be highly associated with the genes in Cluster B (Fig. 3B). These associations appeared to be specific because several well-studied TFs such as Trp53, myc, and NF $\kappa$ B1, which have large abstract representations, were not found to be highly associated with either ISG cluster. The classical type I IFN signaling pathway leads to the formation of the ISGF3 complex that consists of STAT1, STAT2, and IRF9. ISGF3 binds to the highly conserved ISRE promoter element present in ISGs, which leads to their transcriptional activation. Therefore, the identification of STAT and IRF proteins with co-regulated genes was expected. Remarkably, there is no known regulatory relationship between the other candidate TFs and co-regulated genes in Clusters A and B, suggesting that there is only an implicit relationship between them as determined by LSI.

#### The role of cRel in regulation of an ISG subset

We showed high literature association of six ISGs in Cluster A (Cxcl10, Isg15, Gbp2, Ifit1, Ifit3, and Ifi203) with seven TFs (cRel, Stat2, Irf1, Irf2, Irf4, Irf8/Icsbp1, and Irf9/Isgf3g). These TFs bind to three different DNA-binding motifs (Fig. 4). Since STAT and IRF proteins are well described in the regulation of ISGs in MEFs, we examined the potential role of a less well-described TF in ISG expression. Previous studies have determined that the family of NF $\kappa$ B proteins regulates ISG expression (Pfeffer and others 2004; Wei and others 2006; Yang and others 2007). However, the role of cRel in ISG expression is relatively unknown. We determined whether cRel interacted directly with the promoters of the six ISGs in Cluster A by CHIP assays. As shown in Figure 5, IFN induced the recruitment of cRel to promoters of Isg15, Gbp2, Ifit3, and Ifi203 within 30 min of addition. cRel remained bound to the promoters of these ISGs for different lengths of time, ranging from 1 h for Ifit3 and Isg15 to 4 h for Gbp2. In contrast, cRel was basally bound to the promoter of Cxcl10 and IFN induced cRel detachment from the Cxcl10 promoter by 1 h after addition. However, we were unable to detect cRel binding to the Ifit1 promoter either basally or upon IFN addition.

Since these data indicated a role for cRel in regulating ISG expression, quantitative real-time PCR assays were performed for Cxcl10, Isg15, Gbp2, Ifit1, Ifit3, and Ifi203 using RNA from WT and cRel-deficient MEFs treated with IFN- $\beta$ . As shown in Figure 6, the expression of all six ISGs in response



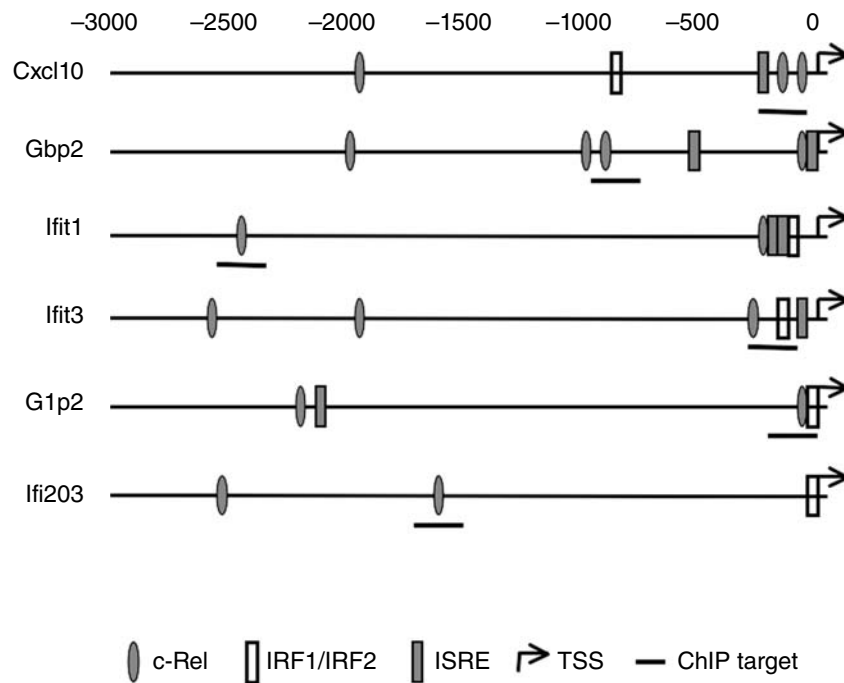
**FIG. 3.** Literature-based clustering of ISGs and TFs. (A) Hierarchical clustering of 6 Gbp2 co-regulated genes (columns) and 48 TFs common to their promoters (rows) using the normalized LSI literature similarity values. (B) Hierarchical clustering of 4 Stat1 co-regulated genes (columns) and 46 TFs common to their promoters (rows) using the normalized LSI literature similarity values. Right panel shows a magnified view of the boxed area in the left panel.

to IFN addition was reduced in cRel-deficient MEFs. For example, while Cxcl10 was induced over 3000-fold by IFN in WT MEFs, Cxcl10 was induced by IFN only 800-fold in cRel-deficient MEFs. Although we did not detect cRel binding to Ifi1 promoter as shown in Figure 5E, we found reduced induction of Ifi1 in cRel-deficient MEFs. Taken together these results suggest that cRel plays distinct roles in the regulation of the IFN-induced expression of this cluster of ISGs.

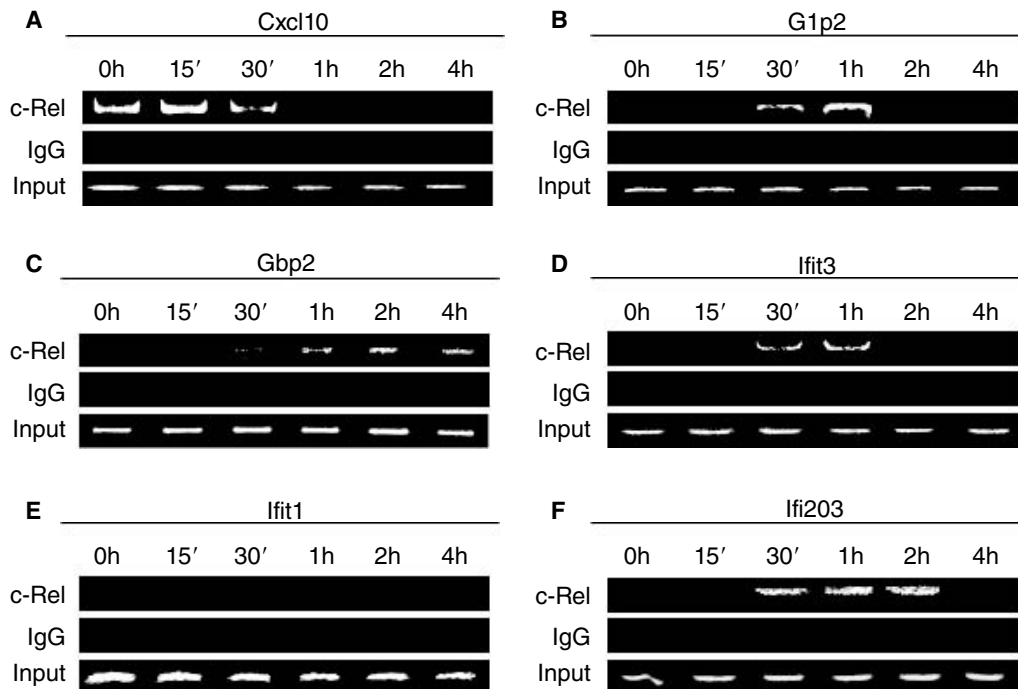
Moreover, since cRel was found to regulate the expression of all six ISGs in this cluster, our text-mining approach identified previously unknown regulatory relationships.

## Discussion

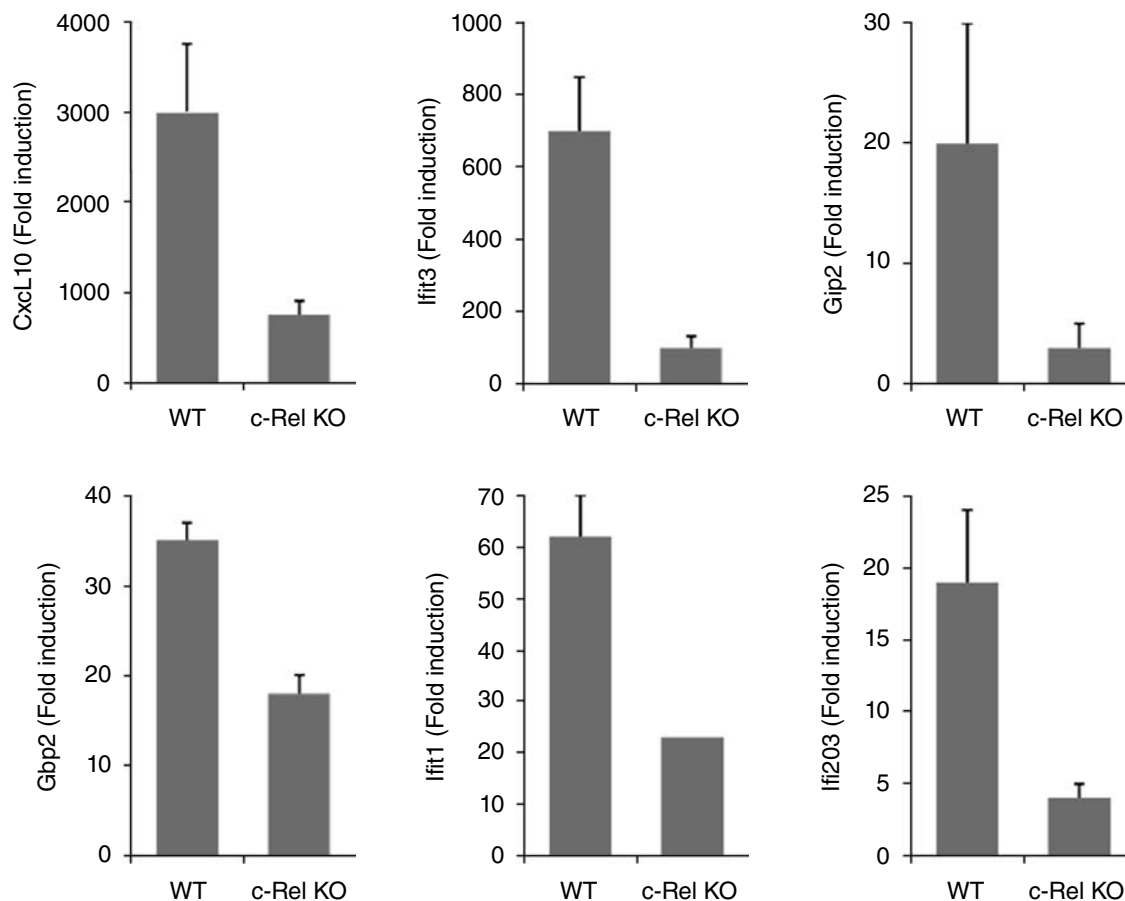
The cRel TF is a member of the NF $\kappa$ B family, which also includes NF $\kappa$ B1 (p105 processed to p50), NF $\kappa$ B2



**FIG. 4.** Putative TF-binding motifs in Cluster A ISGs. Genomic sequences corresponding to 3-kb upstream and 50-bp downstream of transcription start site (TSS) of each gene was obtained from the UCSC Genome Browser. TF-binding sites were identified using MOTIF program and the information in TRANSFAC 6.0 database.



**FIG. 5.** The binding of cRel to ISG promoters. ChIP assays were performed on extracts from MEFs treated with IFN for different durations. Specific antibodies were used to selectively precipitate cRel, or IgG (negative control). DNA was sheared to an average size of 600 bp fragments, and 38 cycles of amplification were used for PCR using primers that targeted upstream of Cxcl10 (A), G1p2/Isg15 (B), Gbp2 (C), Ifit3 (D), Ifit1 (E), and Ifi203 (F) transcription start sites. Bottom panel shows the amount of input DNA prior to ChIP analysis.



**FIG. 6.** Regulation by cRel of ISG induction. Real-time PCR was performed for Cxcl10, G1p2/Isg15, Gbp2, Ifit3, Ifit1, and Ifi203 using cDNAs prepared from WT and cRel-deficient MEFs treated with IFN- $\beta$  at 1000 U/mL for 5 h. Gene expression was normalized to actin expression in each sample. Data are shown as fold-induction relative to untreated fibroblasts, and are mean values  $\pm$  the SEM ( $n = 3$ ).

(p100 processed to p52), RelA (p65), RelB. The family of NF $\kappa$ B TFs regulates gene expression by binding to *cis*-acting  $\kappa$ B sites in its promoters (Beg and others 1995; Beg and Baltimore 1996; Van Antwerp and others 1996; Wang and others 1996). NF $\kappa$ B-regulated genes play important roles in immunity, inflammation, cell growth, and cell survival, which are all processes affected by IFN and this led us to examine a role for NF $\kappa$ B in IFN signal transduction. Both p50 and p52 lack a transcription activation domain, and as homodimers function as repressors. In contrast, p65, cRel, and RelB have a transcription activation domain, and thus when complexed with p50 or p52 are capable of activating transcription. Under most circumstances, NF $\kappa$ B is bound to I $\kappa$ B inhibitory proteins in the cytoplasm of unstimulated cells. Many cytokines including IFN promote the dissociation of the cytosolic inactive NF $\kappa$ B/I $\kappa$ B complexes via the serine phosphorylation and degradation of I $\kappa$ B, leading to NF $\kappa$ B translocation to the nucleus and DNA binding (Yang and others 2000), which is denoted as the classical NF $\kappa$ B pathway. Recent studies have identified an alternative NF $\kappa$ B signaling pathway, which does not

involve I $\kappa$ B degradation (Senftleben and others 2001; Xiao and others 2001; Claudio and others 2002; Coope and others 2002; Dejardin and others 2002; Pomerantz and Baltimore 2002; Muller and Siebenlist 2003; Luftig and others 2004; Xiao and others 2004). This pathway involves the linkage of TRAFs to the activation of the MAP3K-related kinase, NIK, which results in the ubiquitinylation and proteolytic processing of p100/NF $\kappa$ B2 protein and nuclear translocation of p52:RelB dimers to regulate specific NF $\kappa$ B target genes (Bonizzi and others 2004). We previously established that IFN induces NF $\kappa$ B activity in a variety of cells through both classical and alternative pathways, which promote cell survival (Yang and others 2000; Yang and others 2001; Pfeffer and others 2004; Yang and others 2005a; Yang and others 2005b; Wei and others 2006). In addition, we have identified a number of ISGs that are regulated through a NF $\kappa$ B-dependent pathway, but this is the first instance of identification of an ISG regulated by cRel (Pfeffer and others 2004; Wei and others 2006).

Constitutive cRel expression has been previously found in cells of the mature monocytic and lymphocytic lineages



(Liou and others 1994). cRel is critical for the normal function of the host immune system as evidenced by the finding that cRel knockout mice develop normally but have severe defects in lymphocyte proliferation and humoral immunity (Kontgen and others 1995). cRel exists as homodimers or as heterodimers with p50 in a variety of cell types (Hoffmann and others 2003). cRel binds to a set of related 9–10 bp DNA sequences ( $\kappa$ B sites) and regulates the expression of genes involved in cell development, proliferation, and survival (Gilmore and others 2004). In this study, we implemented a novel integrative approach to identify cRel as a candidate regulatory factor for a subset of ISGs including *Isg15/G1p2*, *Gbp2*, *Ifit3*, and *Ifi203*. These genes were previously identified to be potential NF $\kappa$ B-regulated ISG targets, but the specific NF $\kappa$ B proteins responsible for the regulation of their expression was unknown (Ohmori and others 1994; Kumar and others 1997; Baker and others 2003; Pfeiffer and others 2004; Sizemore and others 2004; Wei and others 2006). Using ChIP assays we determined that IFN rapidly induced the recruitment of cRel to the promoters of *Isg15*, *Gbp2*, *Ifit3*, and *Ifi203*, and that cRel remained bound to the promoters of these ISGs for different lengths of time. Our preliminary studies suggest that the binding of cRel to the promoters of these genes correlates well with their transcriptional activation. In contrast, we determined that while cRel was basally bound to the *Cxcl10* promoter, IFN induced cRel detachment from the promoter so that by 1 h after IFN addition no cRel binding was observed. Since *Cxcl10* expression is induced within this time frame, our results are consistent with the hypothesis that cRel homodimers or heterodimers inhibit *Cxcl10* gene expression.

Our analysis revealed several novel TFs to be highly associated with ISG expression. It is important to note that not all of the relationships predicted by this method were confirmed, nor were all experimentally documented relationships in the literature predicted by our method. A possible source of false-positive prediction by LSI may be due to technical limitations of the follow-up ChIP assays, for instance in selection of appropriate primer pairs for the target gene promoters. On the other hand, false-negative predictions by LSI may be caused by the information biases in the scientific literature. For instance, some genes are studied in a very specific context whereas others may be studied in much broader context. Although the LSI model generally performs well in these circumstances, we have observed that broadly studied genes generally show low association scores with a large set of genes and lack high association scores with any specific set of genes.<sup>13</sup> Nonetheless, we have provided evidence that LSI-based algorithms can serve as a valuable exploratory tool for analysis of gene datasets.

Recent advances in genomic technologies allow investigators to identify novel associations between genes in a high-throughput manner. However, interpretation of genomic data is relatively low-throughput because of the requirement for expert knowledge and the ability of humans to identify implicit (hypothetical) connections between genes. Automated information extraction methods such as LSI will undoubtedly be useful in interpretation of high-throughput genomic data and may play a

significant role in taking discovery-based approaches to hypothesis-driven science.

## Acknowledgments

We thank Drs. E. Chesler and K. Manly for their helpful comments. We thank Dr. A. Hoffmann (University of California, San Diego) for providing MEFs, and Dr. D. Baker (Biogen-Idec) for providing rat IFN. This research was supported in part by UT Center for Genomics and Bioinformatics (R.H.), UT Center for Neurobiology of Brain Diseases (R.H.), NIH subcontract LM007292-03 (R.H.), CA73753 (L.M.P.), Muirhead Chair Endowment (L.M.P) and UT Center for Information Technology Research (K.H. and M.W.B.).

## References

- Arduini RM, Li Z, Rapoza A, Gronke R, Hess DM, Wen D, Miatkowski K, Coots C, Kaffashan A, Viseux N, Delaney J, Domon B, Young CN, Boynton R, Chen LL, Betzenhauser M, Miller S, Gill A, Pepinsky RB, Hochman PS, Baker DP. Expression, purification, and characterization of rat interferon-beta, and preparation of an N-terminally PEGylated form with improved pharmacokinetic parameters. *Protein Expr Purif* 34(2):229–242.
- Baker MS, Chen X, Rotramel AR, Nelson JJ, Kaufman DB. 2003. Interferon regulatory factor-1 down-regulates cytokine-induced IP-10 expression in pancreatic islets. *Surgery* 134(2):134–41.
- Barnes B, Lubyova B, Pitha PM. 2002. On the role of IRF in host defense. *J Interferon Cytokine Res* 22(1):59–71.
- Beg AA, Baltimore D. 1996. An essential role for NF- $\kappa$ B in preventing TNF- $\alpha$ -induced cell death. *Science* 274:782–784.
- Beg AA, Sha WC, Bronson RT, Ghosh S, Baltimore D. 1995. Embryonic lethality and liver degeneration in mice lacking the RelA component of NF- $\kappa$ B. *Nature* 376:167–170.
- Berry MW, Drmac Z, Jessup E. 1999. Matrices, vector spaces, and information retrieval. *SIAM Rev* 41:335–362.
- Bluthgen N, Kielbasa SM, Herzel H. 2005. Inferring combinatorial regulation of transcription in silico. *Nucleic Acids Res* 33(1):272–279.
- Bonizzi G, Bebién M, Otero DC, Johnson-Vroom KE, Cao Y, Vu D, Jegga AG, Aronow BJ, Ghosh G, Rickert RC and others. 2004. Activation of IKK $\alpha$  target genes depends on recognition of specific kappaB binding sites by RelB:p52 dimers. *EMBO J* 23(21):4202–4210.
- Claudio E, Brown K, Park S, Wang H, Siebenlist U. 2002. BAFF-induced NEMO-independent processing of NF- $\kappa$ B2 in maturing B cells. *Nat Immunol* 3(10):958–965.
- Coope HJ, Atkinson PG, Huhse B, Belich M, Janzen J, Holman MJ, Klaus GG, Johnston LH, Ley SC. 2002. CD40 regulates the processing of NF- $\kappa$ B2 p100 to p52. *Embo J* 21(20):5375–5385.
- Darnell JEJ, Kerr IM, Stark GR. 1994. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264:1415–1421.
- Dejardin E, Droin NM, Delhase M, Haas E, Cao Y, Makris C, Li ZW, Karin M, Ware CF, Green DR. 2002. The lymphotoxin-beta receptor induces different patterns of gene expression via two NF- $\kappa$ B pathways. *Immunity* 17(4):525–535.

- Friedman RL, Stark GR. 1985.  $\alpha$ -Interferon-induced transcription of HLA and metallothionein genes containing homologous upstream sequences. *Nature* 314:637–639.
- Gilmore TD, Kalaitzidis D, Liang MC, Starczynowski DT. 2004. The c-Rel transcription factor and B-cell proliferation: a deal with the devil. *Oncogene* 23(13):2275–2286.
- Hoffmann A, Leung TH, Baltimore D. 2003. Genetic analysis of NF- $\kappa$ B/Rel transcription factors defines functional specificities. *EMBO J* 22(20):5530–5539.
- Homayouni R, Heinrich K, Wei L, Berry MW. 2005. Gene clustering by latent semantic indexing of MEDLINE abstracts. *Bioinformatics* 21(1):104–115.
- Jenssen TK, Laegreid A, Komorowski J, Hovig E. 2001. A literature network of human genes for high-throughput analysis of gene expression. *Nat Genet* 28(1):21–28.
- Kontgen F, Grumont RJ, Strasser A, Metcalf D, Li R, Tarlinton D, Gerondakis S. 1995. Mice lacking the c-rel proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression. *Genes Dev* 9(16):1965–1977.
- Krallinger M, Erhardt RA, Valencia A. 2005. Text-mining approaches in molecular biology and biomedicine. *Drug Discov Today* 10(6):439–445.
- Kumar A, Yang YL, Flati V, Der S, Kadereit S, Deb A, Haque J, Reis L, Weissmann C, Williams BR. 1997. Deficient cytokine signaling in mouse embryo fibroblasts with a targeted deletion in the PKR gene: role of IRF-1 and NF- $\kappa$ B. *EMBO J* 16(2):406–416.
- Landauer TK, Laham D, Derr M. 2004. From paragraph to graph: latent semantic analysis for information visualization. *Proc Natl Acad Sci USA* 101 Suppl 1:5214–5219.
- Larner AC, Chaudhuri A, Darnell JEJ. 1986. Transcriptional induction by interferon: new protein(s) determine the length and extent of induction. *J Biol Chem* 261:453–459.
- Li H, Lu L, Manly KF, Chesler EJ, Bao L, Wang J, Zhou M, Williams RW, Cui Y. 2005. Inferring gene transcriptional modulatory relations: a genetical genomics approach. *Hum Mol Genet* 14(9):1119–1125.
- Liou HC, Sha WC, Scott ML, Baltimore D. 1994. Sequential induction of NF- $\kappa$ B/Rel family proteins during B-cell terminal differentiation. *Mol Cell Biol* 14(8):5349–5359.
- Luftig M, Yasui T, Soni V, Kang MS, Jacobson N, Cahir-McFarland E, Seed B, Kieff E. 2004. Epstein-Barr virus latent infection membrane protein 1 TRAF-binding site induces NIK/IKK  $\alpha$ -dependent noncanonical NF- $\kappa$ B activation. *Proc Natl Acad Sci USA* 101(1):141–146.
- Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV and others. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1):374–378.
- Muller JR, Siebenlist U. 2003. Lymphotoxin beta receptor induces sequential activation of distinct NF- $\kappa$ B factors via separate signaling pathways. *J Biol Chem* 278(14):12006–12012.
- Ohmori Y, Tebo J, Nedospasov S, Hamilton TA. 1994.  $\kappa$ B binding activity in a murine macrophage-like cell line. Sequence-specific differences in  $\kappa$ B binding and transcriptional activation functions. *J Biol Chem* 269(26):17684–17690.
- Pfeffer LM, Kim JG, Pfeffer SR, Carrigan DJ, Baker DP, Wei L, Homayouni R. 2004. The role of NF- $\kappa$ B in the antiviral action of interferon and interferon-regulated gene expression. *J Biol Chem* 279(30):31304–31311.
- Pomerantz JL, Baltimore D. 2002. Two pathways to NF- $\kappa$ B. *Mol Cell* 10(4):693–695.
- Qin ZS, McCue LA, Thompson W, Mayerhofer L, Lawrence CE, Liu JS. 2003. Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites. *Nat Biotechnol* 21(4):435–439.
- Rebholz-Schuhmann D, Kirsch H, Couto F. 2005. Facts from text—is text mining ready to deliver? *PLoS Biol* 3(2):e65.
- Schindler C, Fu X-Y, Improta T, Aebersold R, Darnell JEJ. 1992. Proteins of transcription factor ISGF-3: one gene encoded the 91- and 84-kDa ISGF-3 that are activated by interferon  $\alpha$ . *Proc Natl Acad Sci USA* 89:7836–7839.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34(2):166–176.
- Senftleben U, Cao Y, Xiao G, Greten FR, Krahn G, Bonizzi G, Chen Y, Hu Y, Fong A, Sun SC and others. 2001. Activation by IKK $\alpha$  of a second, evolutionary conserved, NF- $\kappa$ B signaling pathway. *Science* 293(5534):1495–1499.
- Shatkay H, Feldman R. 2003. Mining the biomedical literature in the genomic era: an overview. *J Comput Biol* 10(6):821–855.
- Sizemore N, Agarwal A, Das K, Lerner N, Sulak M, Rani S, Ransohoff R, Shultz D, Stark GR. 2004. Inhibitor of  $\kappa$ B kinase is required to activate a subset of interferon  $\gamma$ -stimulated genes. *Proc Natl Acad Sci USA* 101(21):7994–7998.
- Van Antwerp DJ, Martin SJ, Kafri T, Green DR, Verma IM. 1996. Suppression of TNF- $\alpha$ -induced apoptosis by NF- $\kappa$ B. *Science* 274:787–789.
- Wang C-Y, Mayo MW, Baldwin Jr. AS. 1996. TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF- $\kappa$ B. *Science* 274:784–787.
- Wei L, Sandbulte MR, Thomas PG, Webby RJ, Homayouni R, Pfeffer LM. 2006. NF $\kappa$ B negatively regulates interferon-induced gene expression and anti-influenza activity. *J Biol Chem* 281(17):11678–11684.
- Xiao G, Fong A, Sun SC. 2004. Induction of p100 processing by NF- $\kappa$ B-inducing kinase involves docking IkappaB kinase  $\alpha$  (IKK $\alpha$ ) to p100 and IKK $\alpha$ -mediated phosphorylation. *J Biol Chem* 279(29):30099–30105.
- Xiao G, Harhaj EW, Sun SC. 2001. NF- $\kappa$ B-inducing kinase regulates the processing of NF- $\kappa$ B2 p100. *Mol Cell* 7(2):401–409.
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434(7031):338–345.
- Yandell MD, Majoros WH. 2002. Genomics and natural language processing. *Nat Rev Genet* 3(8):601–610.
- Yang CH, Murti A, Basu L, Kim JG, Pfeffer LM. 2000. IFN $\alpha$ / $\beta$  promotes cell survival by activating NF- $\kappa$ B. *Proc Natl Acad Sci USA* 97:13631–13636.
- Yang CH, Murti A, Pfeffer SR, Kim JG, Donner DB, Pfeffer LM. 2001. Interferon  $\alpha$ / $\beta$  promotes cell survival by activating NF- $\kappa$ B through phosphatidylinositol-3 kinase and Akt. *J Biol Chem* 276:13756–13761.

- Yang CH, Murti A, Pfeffer LM. 2005a. IFN induces NIK/TRAF dependent NF- $\kappa$ B activation to promote cell survival. *J Biol Chem* 280(36):31530–31536.
- Yang CH, Murti A, Valentine WJ, Du Z, Pfeffer LM. 2005b. Interferon alpha activates NF- $\kappa$ B in JAK1-deficient cells through a TYK2-dependent pathway. *J Biol Chem* 280(27):25849–25853.
- Yang CH, Wei L, Pfeffer SR, Du Z, Murti A, Valentine WJ, Zheng Y, Pfeffer LM. 2007. Identification of CXCL11 as a STAT3-dependent gene induced by IFN. *J Immunol* 178(2):986–992.

Address reprint requests or correspondence to:  
*Lawrence M. Pfeffer*  
*Department of Pathology and Laboratory Medicine*  
*University of Tennessee Health Science Center*  
*930 Madison Avenue*  
*Room 530*  
*Memphis, TN 38163*  
*E-mail: lpfeffer@utm.edu*

Received 28 November 2007/Accepted 21 February 2008

