Modulation of Gene Expression Regulated by the Transcription Factor NF-κB/RelA

Xueling Li¹,²,³,⁶, Yingxin Zhao¹,²,⁵, Bing Tian¹,²,⁴, Mohammad Jamaluddin¹,², Abhishek Mitra¹,², Jun Yang²,⁴, Maga Rowicka¹,²,³, Allan R. Brasier¹,²,⁴,⁵, Andrzej Kudlicki¹,²,³

¹Institute for Translational Sciences, ²Sealy Center for Molecular Medicine, ³Department of Biochemistry and Molecular Biology, ⁴Department of Internal Medicine, ⁵Center for Clinical Proteomics, University of Texas Medical Branch, Galveston, Texas, United States of America. ⁶Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, P.R. China.

*Running title: Modulation of NF-κB/RelA Dependent Gene Expression

To whom correspondence should be addressed: Andrzej Kudlicki, askudlic@utmb.edu

Key words: Transcription regulation; Modulatory network; NF-κB transcription factor; Transcription target genes; Action mode; Computational biology; Proteomics; Functional Genomics; Gene expression.

Background: Interacting proteins modulate the activity of NF-κB/RelA transcription factor and expression of its targets.

Results: By analyzing gene expression, protein binding and DNA binding, we inferred and characterized 8349 such modulations.

Conclusion: Different modulator groups affect separate pathways.

Significance: We provide new insight into the activity of NF-κB/RelA. Our inference model can be applied to other processes and pathways.

ABSTRACT

Modulators (M) are proteins that modify activity of transcription factors (TFs) and influence expression of their target genes (TGs). To discover modulators of NF-κB/RelA, we first identified 365 NF-κB/RelA binding proteins using Liquid Chromatography/Tandem Mass Spectrometry (LC-MS/MS) data, published RelA modulators and TGs, and compendium of gene expression profiles. Hierarchical clustering of the derived modulatory network revealed functional subnetworks and suggested new pathways modulating RelA transcriptional activity. The modulators with the highest number of TGs and most non-random distribution of action modes (measured by Shannon entropy) are consistent with published reports. Our results provide a repertoire of testable hypotheses for experimental validation. One of the NF-κB/RelA modulators we identified is STAT1. The inferred (STAT1, NF-κB/RelA, TG) triplets were validated by LC-SRM-MS and the results of STAT1 deletion in human fibrosarcoma cells. Overall, we have identified 562 NF-κB/RelA modulators, which are potential drug targets, and clarified mechanisms of achieving NF-κB/RelA multiple functions through modulators. Our approach can be readily applied to other TFs.

http://www.jbc.org/cgi/doi/10.1074/jbc.M113.539965 The latest version is at http://www.jbc.org/
NF-κB is a key human transcription factor for stress associated processes. It is involved in many cellular processes including innate immunity (1), inflammation (2), stress response (3), regulation of apoptosis (4), and regulation of proliferation and differentiation (5). Dysregulation of NF-κB activity is linked to inflammatory disorders (6), autoimmune (7), metabolic diseases (8) and cancers (9). Such multiple NF-κB functions are achieved through modulation (10,11) and cooperative binding with other transcription factors (12). RelA (p65) is one of five components of the NF-κB transcription factor complex in mammals (10). Due to the importance and complexity of the cellular processes regulated by NF-κB pathways, elucidating and characterizing modulators of NF-κB will provide important insights into the role of modulation in fine-tuning transcriptional regulations and conferring specificity, as well as identify potentially promising drug targets.

Gene expression profiles measured under various conditions provide important information on the regulatory relations between genes. Gene regulatory networks have been constructed using reverse engineering methods based on correlation of gene expression profiles (13). These methods include conditional information (13,14), correlation coefficient, machine learning (15) and statistical models (16). Recently, heterogeneous data integration of promoter sequence and gene expression profiles has also been used to infer gene regulation patterns (17,18).

Modulators (M) are proteins that modify the activity of a transcription factor (TF). Modulators often act by binding directly to the TFs, to affect the expression of a target gene. The outcome of modulation can be classified into the six possible action modes: inhibition attenuation, inhibition enhancement, inhibition inversion, activation inversion, activation enhancement, and activation attenuation (19). Including analysis of the modulators facilitates elucidation of a biologically realistic regulatory network. Therefore, instead of pairs (TF, TG) that lead to traditional regulatory networks, triplets (M, TF, TG) and the more complex regulatory and modulatory networks inferred based on significant triplets should be considered.

Several computational methods for inferring (M, TF, TG) triplets from gene expression profiles have been developed in recent years, including three-way mutual information (20), MINDy (21), GEM (19), MIMOSA (22), and MONSTER (23). MINDy uses the difference between conditional mutual information at low and high modulator expression, and has been applied to identify genome wide modulators of MYC. However, it has a major limitation in that it cannot identify modulators with more than one mode of activity (activator or repressor). MIMOSA applies a mixture model and maximum expectation method to predict modulators of STAT1. The method is computationally intensive and cannot differentiate between modulators and targets of the TF. Finally, MONSTER only considers kinases as potential modulator candidates. Neither MIMOSA, nor MONSTER, has been experimentally validated. GEM (19) aims to identify modulators of the androgen receptor, AR, by constructing a non-linear model of the three-way interaction between the activity levels of the two ‘inputs’, (TF and M), and the ‘output’ (TG).

In this work, we predicted modulators and the corresponding TGs of the NF-κB/RelA transcription factor, based on the approach in (19). As input for modulatory triplet inference, we used de novo identified RelA binding proteins and its target genes, identified respectively by affinity tandem mass spectrometry (LC–MS/MS) and Chromatin Immunoprecipitation followed by next-generation sequencing (ChIP-Seq). The new data were supplemented with published data on binding proteins and target genes of the NF-κB/RelA pathway.

**EXPERIMENTAL PROCEDURES**

**Probabilistic models for triplet prediction**

The probabilistic model is similar to Babur et al (19). The expected expression of the target gene }
Here, $F, M$ and $TG$ are the transcription factor, its modulator and the affected target gene, respectively. $H_c$ is the basal expression level of the TG, $H_f$ and $H_m$ represent the effect of $F$ and $M$ alone on $TG$, respectively. $G$ represents the effect of interaction between $M$ and $F$ on $TG$ and varies for different genes. Its magnitude may be interpreted as the modulation effect. Our task was to find genes with $G$ significantly different from zero. Accordingly, we approximated the probability that TG is highly expressed, $P(TG = 1)$, as:

$$P(TG = 1) = \alpha_c + \alpha_f F + \alpha_m M + \gamma M \times F$$  \hspace{1cm} (2)$$

where $\alpha_c$ represents the basal expression of $TG$, $\alpha_f$ is the effect of $F$ only (when $M$ low), $\alpha_m$ is the effect of $M$ only (at low $F$) and $\gamma$ is the interaction effect of $F$ and $M$ when both are highly expressed. $F, M$ and $TG$ are equal to 1 if the ranked expression of the corresponding gene is in the upper tertile. Conversely, if it is in the lower tertile, $F, M$ and $TG$ are set to zero. If $\gamma$ is not statistically significantly different from zero, the data do not have enough evidence to support the hypothesis that $M$ modulates the effect of $F$ on the target $TG$.

To estimate $\alpha_c, \alpha_f, \alpha_m$ and $\gamma$, we first calculated $P(TG = 1|M, F)$, the proportions of $TG = 1$ with different combination of states of $M$ and $F$. Specifically, $P(TG = 1|M = 1, F = 1)$, where $TG = 1$ with $M = 1$ and $F = 1$, serves as the observed case, while $P(TG = 1|M = 1, F = 0)$, $P(TG = 1|M = 0, F = 1)$ and $P(TG = 1|M = 0, F = 0)$ are the respective control cases with low $M$ or $F$ or both. $\alpha, \beta$ and $\gamma$ were then estimated by

$$\hat{\alpha}_c = P(TG = 1|M = 0, F = 0)$$  \hspace{1cm} (3)$$

Here, $\hat{\alpha}_c$ corresponds to the proportion of $TG = 1$ to $TG = 1$ plus $TG = 0$ when both $M$ and $F$ is low expressed.

**Classification of the triplet action**

$$\hat{\alpha}_f =$$

$$P(TG = 1|M = 0, F = 1) - P(TG = 1|M = 0, F = 0)$$  \hspace{1cm} (4)$$

$$\hat{\alpha}_m =$$

$$P(TG = 1|M = 1, F = 0) - P(TG = 1|M = 0, F = 0)$$  \hspace{1cm} (5)$$

$$\hat{\beta}_f =$$

$$P(TG = 1|M = 1, F = 1) - P(TG = 1|M = 1, F = 0)$$  \hspace{1cm} (6)$$

$$\hat{\beta}_m =$$

$$P(TG = 1|M = 1, F = 1) - P(TG = 1|M = 1, F = 0)$$  \hspace{1cm} (7)$$

$$\gamma = \hat{\beta}_m - \hat{\alpha}_m = \hat{\beta}_f - \hat{\alpha}_f = P(TG = 1|M = 1, F = 1)
- P(TG = 1|M = 0, F = 1)
+ P(TG = 1|M = 1, F = 0)
- P(TG = 1|M = 0, F = 0)$$  \hspace{1cm} (8)$$

Estimation of the above parameters can be found in (19).

Assuming that $M$ influences $TG$ via $M$, $M$ must act on $TG$ when $F$ is highly expressed, i.e., $\hat{\alpha}$ and $\hat{\beta}_m$ must be non-zero, while $\hat{\beta}_m$ is greater in absolute value or has a different sign than $\hat{\alpha}_m$.

$$H_1: \gamma = 0 \hspace{1cm} H_2: \hat{\beta}_m = 0 \hspace{1cm} H_3: \frac{\hat{\alpha}_m}{\hat{\beta}_m} \geq 1$$  \hspace{1cm} (9)$$

The estimation of the variance and p-value for rejecting the above three null hypotheses $H_1$, $H_2$ and $H_3$ can be found in (19). The predicted triplets are ordered according to the p-values of $\gamma$ and $\hat{\beta}_m$, the smaller the p-value, the higher the significance of the predicted triplet. The requirement for a significant $\gamma$ is crucial for detecting the interaction effect as opposed to a combination of independent actions of $M$ and $F$ separately.

Unmodulated TF activity was classified into three categories according to $\hat{\alpha}_f$: activation if
positive, inhibition if negative and inactive if zero. By comparing \( \hat{\alpha} \) and \( \hat{\beta} \), modulators are also classified into three categories of action: enhance, attenuate or invert the activity of the TF. Specifically, if \( \hat{\alpha} \) is non-negative and \( \hat{\beta} \) value is greater than \( \hat{\alpha} \), the action mode is activation enhancement. If \( \hat{\alpha} \) is positive and \( \hat{\beta} \) less than \( \hat{\alpha} \) and non-negative, it signifies activation attenuation, and if \( \hat{\alpha} \) is positive and \( \hat{\beta} \) is negative, activation inversion. If \( \hat{\alpha} \) is negative and \( \hat{\beta} \) value is greater than \( \hat{\alpha} \) and non-positive, the action mode is inhibition enhancement. If \( \hat{\alpha} \) is negative and \( \hat{\beta} \) greater than \( \hat{\alpha} \) and non-negative, the action mode is inhibition attenuation, and if \( \hat{\alpha} \) is negative and \( \hat{\beta} \) is positive, inhibition inversion (see Table I).

A pass-strategy based on Shannon entropy filter

We further filtered the predicted modulators by the Shannon entropy of the distribution across the six action modes by each modulator respectively predicted from binding data. Modulators with Shannon entropy significantly smaller than random ones were classified as specific. The null distribution of the Shannon entropy depends on the number of targets and is obtained by generating the six action modes given the number of TGs in a simulation repeated 10,000 times. The modulators were ranked according to the largest number of TGs. The calculation of the Shannon entropy of the action mode distribution for each modulator is as follows:

\[
H = - \sum_{i=1}^{6} p_i \log p_i, \quad p_i = \frac{pi}{\sum_{i=1}^{6} pi} \quad (10)
\]

where \( p_i \) is the proportion of the number of targets in each action mode \( i \), and \( H \) is the entropy of action mode distribution for each modulator.

Hierarchical clustering of M-TG networks

To identify the modules of the network composed by RelA modulators and their affected target genes, we applied hierarchical clustering (24) to matrices of modulator-target interactions. Hierarchical clustering has been chosen to reflect the coincidence between the modular structure of the modulation network and the multiple levels of functional annotations of the genes involved (see also Results). The matrix entries correspond to the value of \( \hat{\gamma} \), which represents the effect of an M when F is highly expressed, and \( \hat{\beta} \), the interaction effect of M and F when both are highly expressed. Here we used the Matlab clustergram function with default parameters and parsed the M-TG network into a hierarchy of groups according to the similarity of the value of \( \hat{\gamma} \) (the effect of M-F interaction) and the value of \( \hat{\beta} \) (the effect of M when F is high) that measures the common neighbors of M and T, respectively. The hierarchical clustering is agglomerative since we proceeded through the algorithm by adding links to the network. We applied the clustering procedure to triplets satisfying (\(-\log_{10} p\)-value of \( \hat{\gamma} \)) > 4 and (\(-\log_{10} p\)-value of \( \hat{\beta} \)) > 4 . To obtain the clustering presented in Fig. 4, we only used the modulators and TGs involved in significant predicted interactions and not involved in feedback loops. Specifically, we only used the 861 TGs for which there was at least one triplet satisfying the \( p \)-value thresholds. The modulators have been selected analogously, with at least one significant triplet, giving 493 candidate modulators of RelA.

Gene expression profiles

We used gene expression profiles of 2158 tumor samples published by expO (expression project for oncology) to characterize each gene. ExpO is based on the GeneChip Human Genome U133 Plus 2.0 array platform. The variety of the tumor samples guarantees enough perturbation for studying three-point (M-F-T) interaction. Since gene expression is noisy, as reported in (25), unsupervised discretization methods were usually used to discretize the gene expression profiles in gene regulatory network construction, which included Equal Width Discretization, Equal Frequency Discretization, Kmeans, Column Kmeans Discretization (Cokmeans), Bidirectional
Kmeans Discretization (Bikmeans). We discretized the expression values by rank-ordering across genes and dividing the ranked 2158 expression values of each gene across experiments into three bins, labeled as “1” (upper 719 experiments where the gene ranked highest), “0” (lower 719 experiments where the gene ranked lowest) and “middle” (720 remaining experiments). Based on the areas under ROC and Precision-recall curve in predicting RelA modulators (see Figure 1), we found that varying the bin size only weakly affected the prediction of RelA triplets. Thus, we used tertiles as bins. Unless specified otherwise, we predicted triplets from probeset profiles with the above discretization. For some of the genes, multiple probesets are available on the UG133Plus2 microarray. Since the individual probesets may correspond to different isoforms of the gene and thus be responsible for different functions, we treated each probeset of modulators and target genes separately in the prediction algorithm. Also, there are three probesets corresponding to RelA on the array platform: 201783_s_at, 209878_s_at and 230202_at. We considered the first two individually and ignored the third one, 230202_at, because of its low expression in the expO experiments.

Collection of binding proteins and target genes of RelA from literature

We have obtained the list of RelA-binding proteins from the PIANA database (http://sbi.imim.es/piana/). 287 of those 297 proteins (Supplementary Data File S1, sheet 1) have probesets available in the expO database. Here, we refer to these previously reported binding proteins as “existing”, as opposed to the newly identified proteins that bind to RelA. Similarly, the existing RelA targets (TGs) were collected from the Boston University website Cytoscape (27).

Identification of RelA binding protein by affinity tandem mass spectrometry

Material — All reagents were American Chemical Society (ACS) grade or higher. All solvents used, including water, methanol, acetonitrile (ACN) were LC/MS grade. Sequence grade modified trypsin was purchased from Promega (Madison, WI). Recombinant TNFα was from PeproTek (Rocky Hill, NJ). Antibodies used were: rabbit anti NF-κB/RelA (sc-372) Abs from Santa Cruz Biotech (Santa Cruz, CA).

Cell culture — Human A549 pulmonary epithelial cells (American Type Culture Collection [ATCC]) were grown in F12K medium (Gibco) with 10% fetal bovine serum, penicillin (100U/mL), and streptomycin (100 g/mL) at 37 °C in a 5% CO2 incubator as previously described (28).

Co-immunoprecipitation — 4-6 x10^6 A549 cells were transfected by electroporation in suspension with 10 mg/ml of polyIC (Sigma) according to the manufacturers recommendation (Amaxa). After 4 h, cells were washed twice with phosphate-buffered saline (PBS, 137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) and nuclei prepared by lysis in hypotonic buffer (29). Protein-chromatin crosslinking was performed using an optimized two-step protocol, with disuccinimidyl glutarate (Pierce) followed by formaldehyde (30). The nuclei were lysed in Lysis Buffer (1% SDS, 50 mM Tris, pH 8.0 and protease inhibitor cocktail).

Equal amounts of cross-linked protein were immunoprecipitated overnight at 4 °C with 4 μg rabbit anti RelA Ab diluted 1:4 in ChIP dilution buffer (50 mM Tris-HCl, pH 7.5, 1 mM EGTA, 1 mM EDTA, 1% Triton X-100, 0.27 M sucrose, 0.1% (v/v) β-mercapto-ethanol, 1 mM sodium orthovanadate, 50 mM sodium fluoride, 5 mM sodium pyrophosphate, complete proteinase inhibitor cocktail and 1 mM phenylmethylsulfonyl fluoride). Immunoprecipitates
were then collected with 40 μL protein-A magnetic beads (Dynal Inc) that were pre-blocked in 200 μL acetylated casein (0.2 mg/ml) at RT for 2 h [31]. Nonspecific IgG was used as a negative control. Immunoprecipitates were washed 5 times in PBS.

**On-beads Tryptic digestion** - The beads were resuspended in 30 μL of 50 mM ammonium hydrogen carbonate (pH 7.8). 20 μL of 0.1 μg/μL of trypsin was added. The samples were mixed and trypsinized by gentle shaking overnight at 37 °C. After digestion, the supernatant was collected. The beads were washed with 50 μL of 50% ACN three times and the supernatant was pooled, and dried [4].

**In-gel Tryptic digestion** — The protein mixtures obtained from immunoprecipitation of anti-RelA antibody and nonspecific IgG were separated on an SDS-PAGE gel. The proteins were visualized by Sypro® Ruby staining. The protein bands were cut from the gel and subjected to in-gel trypsin digestion as described previously [5]. Briefly, the gel slice was cut into small particles (∼1 mm³) using a scalpel. The resulting gel particles were destained in 1 mL of extraction solution (ACN/TFA/water, 50:5:45, v/v/v) for 60 min with a vortex and the extraction solution was carefully removed with a gel-loading pipet tip. Extraction was repeated once. The extracts were pooled and dried with a SpeedVac.

**LC-MS/MS** — HPLC/MS/MS analyses were performed in an LTQ velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) coupled on-line to a nano-HPLC system (Agilent 1100 Nano Pump, Agilent Technologies, San Jose, CA) and nanospray source. Three microliters of the peptide solution in buffer A (5% acetonitrile/94.9% water/0.1% acetic acid (v/v/v)) was manually injected and separated in a nano-HPLC C18 column (100 mm length × 75 μm inner diameter, 5-μm particle size), 100-A pore diameter) The peptides were eluted from the column with a linear gradient of 5–70% buffer B (99.9% acetonitrile/0.1% acetic acid (v/v/v)) in buffer A over 45 min. The eluted peptides were electrosprayed directly into the mass spectrometer. The MS/MS spectra were acquired in a data-dependent mode. The full MS scan was performed in Orbitrap with mass resolution of 60,000. The six strongest ions in each MS spectrum were automatically selected for CID and analyzed by LTQ velos.

**Protein identification and spectral counting** — The resulting spectra were used to identify protein candidates in the target-decoy Swissprot human protein database (downloaded on August 25th, 2011) with the Proteome Discoverer 1.1 search engine (Thermo Scientific, San Jose, CA). The cutoff of the false discovery rate (FDR) of peptide identifications was 1%. Normalized spectral abundance factor (NSAF) value for each protein was calculated as described [6],

\[
(\text{NSAF})_k = \frac{\sum_{i=1}^{N} \frac{\text{SpC}_i}{L}}{\sum_{i=1}^{N} \frac{\text{SpC}_i}{L}} \tag{11}
\]

in which the total number of tandem MS spectra matching peptides from protein k (SpC) was divided by the protein length (L), then divided by the sum of SpC/L for all uniquely identified proteins in the dataset.

**LC-SRM-MS** — The LC-SRM-MS analysis was performed with a TSQ Vantage triple quadrupole mass spectrometer equipped with nanospray source (Thermo Scientific, San Jose, CA). The online desalting and chromatography were performed using an Eksigent NanoLC-2D HPLC system (AB SCIEX, Dublin, CA). An aliquot of 10 μL of each trypsic digest was injected on a C18 peptide trap (Agilent, Santa Clara, CA), desalted with 0.1 % formic acid at a flow rate of 2 μL/min for 45 min. Peptides were eluted from the trap and separated on a reversed phase nano-HPLC column destained in 1 mL of water/methanol solution (50:50, v/v) containing 25 mM NH₄HCO₃ (pH 8.0) three times, with the solution changed every 10 min. The destained gel was washed in 1 mL of an acetic solution (acetic acid/methanol/water, 10:40:50, v/v/v) for 3 h, with the solution changed every 1 h. The resulting gel was soaked in 1 mL of water twice with the solvent changed every 20 min. The gel was then transferred to a 0.5-mL microcentrifuge tube and dehydrated by soaking the gel in 100% ACN until it became opaque white. The solution was removed and the gel was dried in a Speed Vac for 20–30 min. The dried gel was rehydrated with an adequate amount of trypsin digestion solution (10 ng of trypsin/μL in 50 mM NH₄HCO₃, pH 8.0). The digestion was carried out at 37 °C overnight. To extract tryptic digest, the gel was soaked in 40 μL of extraction solution (ACN/TFA/water, 50:5:45, v/v/v) for 60 min with a vortex and the extraction solution was carefully removed with a gel-loading pipet tip. Extraction was repeated once. The extracts were pooled and dried with a SpeedVac.
(PicoFrit™, 75 µm x 10 cm; tip ID 15 µm) packed in house using Zorbax SB-C18 (5-µm diameter particles, Agilent, Santa Clara, CA). Separations were performed using a flow rate of 500 nL/min with a 20-min linear gradient from 2-40% mobile phase B (0.1 % formic acid-90 % ACN) in mobile phase A (0.1 % formic acid), followed by 0.1-min gradient from 40-90% mobile phase B and 5-min 90% mobile phase B. The TSQ Vantage was operated in high-resolution SRM mode with Q1 and Q3 set to 0.2 and 0.7-Da Full Width Half Maximum. All acquisition methods used the following parameters: 1800 V ion spray voltage, a 275 °C ion transferring tube temperature, a collision-activated dissociation pressure at 1.5 mTorr, and the S-lens voltage used the values in S-lens table generated during MS calibration [4].

Identification of RelA target genes by ChIP-seq

We prepared the list of candidate target genes of RelA based on the result of our previous ChIP-seq experiment (32). To identify DNA sequences binding to RelA, covalent cross-linking, segmentation of DNA and anti-RelA antibody precipitation was performed in A549 cells. ChIP-seq data of RelA were obtained by Illumina sequencing, resulting in two fastq files, one for TNF treatment at 0 minutes (TNF-0) and one for the TNF treatment at 30 minutes (TNF-30). As reported in (32), our in-house Instant Sequencing software was used to trim the reads at the location where one base was called with confidence interval greater than 99% (corresponding to phred score $\geq 20$, p-value 0.01) (33) and only reads that were at least 35 bp long were preserved in the corresponding fasta file. The TNF-0 and TNF-30 samples contained 17,862,626 and 12,624,879 quality filtered reads, respectively. We used bowtie v0.12 software (34) to map the filtered reads to the human genome assembly GRCh37/hg19. We rejected reads that had mismatches or were non-uniquely mappable, leaving 77% of reads in the TNF-0 sample and 81% for the TNF-30 treated sample. ChIP-Seq binding peaks were called using MACS v0.14 (35), with parameters for calling peaks in human data, with TNF-0 sample as the control and TNF-30 sample as the treatment. In total, 20,733 peaks were called.

After parsing by Instant Sequencing, we selected the best 4195 peaks with a MACS significance score of at least 200. For gene assignment, Instant Sequencing was then used to find transcription start sites of known human genes proximal to ChIP-Seq peaks, using annotations from the GENCODE v12 database (36,37). Peaks whose summits were within a distance of 3 Kb upstream or downstream of a known transcription start site of a gene were identified. For each ChIP-Seq peak, a list of transcription start sites was ranked by the increasing order of their distance from the summit of a ChIP-Seq peak. Next, the gene with its transcription start site closest to the summit of a peak was found and reported, along with the MACS significance score of that peak. When multiple peaks were proximal to the same gene, the MACS significance scores for each peak were added together and the sum was reported as a single score for that gene. Finally a list of proximal genes ordered by their combined significance scores (Supplementary Data File S2, Sheet 1) was generated for finding enriched GO functions.

Enrichment analysis of predicted TGs composing the predicted triplets

Enrichment of the predicted TGs in newly identified and existing TGs was assessed by hypergeometric cumulative distribution function ($CDF$), using the $\text{hygecdf}$ function in Matlab. The function computes the hypergeometric cdf given the size of the population, the number of items with the desired characteristic and the number of samples. The final $p$-value is calculated by $1 - CDF$.

Experimental validation on STAT1-RelA-Target triplets

An interaction between STAT1 and RelA has been previously implicated by co-localization experiments based on confocal fluorescence microscopy (38). To verify the interaction between RelA and STAT1, we used immunoprecipitation and quantitative SID-SRM-MS to measure the enrichment of STAT1 in the sample obtained from the immunoprecipitation of anti-RelA antibody. The nuclear extract of A549 cells were incubated with anti-RelA antibody and nonspecific IgG
(negative control), respectively as depicted in the Methods. The proteins associated with the antibodies were pulled-down and subjected to LC-SRM-MS to quantify the level of STAT1 in each sample.

Targets of the STAT1-RelA interaction were validated by STAT1 knockout and poly(IC) stimulation. Specifically, the real action modes of the predicted triplets of STAT1-RelA-TGs were validated by comparing the expression level of a target gene candidate in four conditions, consisting of STAT1 wild type human 2f (STAT1+/+) cell line and knockout U3A (STAT1-/-) combined with the absence or presence of poly(IC) stimulation in human fibrosarcoma cells. The experimental action modes were then compared with the predicted action modes.

RESULTS
Implementation of the probabilistic model for (M, TF, TG) triplet prediction

We implemented an algorithm for predicting (M, TF, TG) triplets based on a compendium of 2,158 expression profiles. The underlying probabilistic model (19) of triplet action depends on four basic parameters $\alpha_e, \alpha_f, \alpha_m$ and $\gamma$, that correspond to the basal level of a TG, the dependence of TG expression on TF, the dependence on the M, and the interactive effect of the M and TF on the TG. For accurately predicting (M, TF, TG) triplets, two partial effects, and one total effect of the M and TF on the TG, were also estimated respectively by the parameters $\beta_m$, $\beta_f$, and $\alpha_f + \beta_m$ (Methods). The existence of a functional (M, TF, TG) triplet is inferred if both $\hat{\gamma}$ (the estimate of $\gamma$) and $\hat{\beta}_m$ (the estimate of $\beta_m$) are significantly different from zero. The corresponding $p$-values were estimated as described in the Methods section.

According to the estimated values of the parameters, $\gamma$, $\alpha_f, \beta_f, \beta_m$ for the given triplet, that correspond to effect of TF on TG and the interactive effect of M and TF, the inferred triplets were categorized into one of six action modes: inhibition attenuation, inhibition enhancement, inhibition inversion, activation inversion, activation enhancement, and activation attenuation. Approximately 2% of the triplets did not have an assigned category, because the estimated parameters did not pass the adopted significance thresholds. The unclassified triplets were not considered in further analysis. Among the modulators of NF-kB/RelA, 36% have prevalent action modes of their triplets, while other modulators act in many or all action modes.

We formalized the distinction between specific and general modulators using the entropy of the distribution of the action modes as a measure of specificity. The entropy is compared to the entropy expected of a non-specific modulator with the same number of affected targets (Methods).

In three of the six action modes (inhibition attenuation, inhibition inversion, and activation enhancement), the parameter $\hat{\gamma}$ is significantly positive, and the TG is predominantly upregulated by the modulator M; that is M is a dominant activator. Conversely, for inhibition enhancement, activation inversion and activation attenuation $\hat{\gamma}$ is significantly negative and the modulator is a dominant inhibitor.

Moreover, a modulator is classified as an agonist of NF-kB/RelA if it prevalently regulates the TGs in the same direction as NF-kB/RelA (up-up- or down-down-regulation), i.e., with inhibition enhancement and activation enhancement modes. Conversely, if M regulates its involving TGs prevalently in the opposite direction to NF-kB/RelA (up-down- or down-up regulation) it is considered an antagonist. Note that a dominant activator can be an agonist or antagonist of NF-kB/RelA, and likewise an inhibitor. For all four cases (activator, inhibitor, agonist, antagonist), we used the hypergeometric test to assess the significance of the dominant category.

Input data for inferring NF-kB/RelA modulatory network

Our goal was to characterize the dependencies between the modulators and corresponding target genes of NF-kB/RelA. We expected that the most significant (M, TF, TG) triplets would contain modulators and TGs with prior association with
NF-κB/RelA. Therefore, we focused on triplets in which the candidate modulator was independently indicated as an NF-κB/RelA interactor, and the candidate target was a target gene of NF-κB/RelA. Using such restricted sets of candidate Ms and TGs allowed minimizing the false discovery rate in inferred triplets. Before the triplet prediction and the following modulatory network inference, NF-κB/RelA modulators (NF-κB/RelA binding proteins) and the TG candidates were first collected as follows.

**Candidate modulators**

We postulated that the modulators of NF-κB/RelA would be enriched in proteins physically interacting with the NF-κB/RelA protein. To identify candidate NF-κB/RelA modulators, we performed two independent LC-MS/MS experiments, in which 365 NF-κB/RelA-associated proteins were identified (Supplementary Data File S1, Sheet 2-4). Additional RelA interactors were found by mining the PIANA database resource (http://sbi.imim.es/piana/, also see Methods), containing literature and other sources of data including the HPRD (39) and BioGrid databases (40), and Boston University website (http://www.bu.edu/nf-κb/). We thus collected 297 proteins that have been previously reported to interact with NF-κB/RelA (Supplementary Data File S1, sheet 1). We refer to these 297 proteins as “previously known binding proteins” of NF-κB/RelA. There are 31 proteins present in both our LC-MS/MS data and the PIANA dataset.

The functional categories enriched in the newly identified NF-κB/RelA associated proteins are shown in Table II. The mapped protein network of all newly identified NF-κB/RelA associated proteins is shown in Figure 2. Mapping the newly identified 365 NF-κB/RelA associated proteins into the global PPI network from IntAct database (41) by PPI spider (42), shows that the most enriched pathways are TNF receptor signaling pathway and TNF alpha/NF-κB (Supplementary Data File S3). It demonstrates the relevance of the experimentally identified proteins to the biological context of RelA-association. Functional analysis of the newly identified RelA-binding proteins (Supplementary Data File S3) confirms the established biological functions of NF-κB/RelA, such as enrichment in Gene Ontology (GO) categories: response to DNA damage stimulus; cell cycle progress; post-transcriptional regulation of gene expression; regulation of translation; regulation of protein ubiquitination, microtubule cytoskeleton and chromatin organization.

The combined set of LC-MS/MS data and exiting RelA associated proteins contains 632 unique proteins. Within this group, the mRNA expression profiles are available for 620 of the proteins, and these were selected for further analysis (Supplementary Data File S1, sheet 5). Several factors may be responsible for the relatively small overlap (10%) between the literature-derived set and our LC-MS/MS experiments. Our experimental identification of NF-κB/RelA interactors was restricted to nuclear extract, and the results may have been affected by different experimental techniques, interaction sampling variety and indirect binders of NF-κB/RelA identified by LC/MS spectrometry. However, as discussed above, the newly discovered set of RelA interactors exhibit properties characteristic of those previously known, which expands the set of experimentally verified NF-κB/RelA interactors by more than 2.5 times. We are confident that our work will be a very valuable contribution to elucidating NF-κB/RelA regulatory and modulatory networks.

**Candidate target genes**

The set of target genes used for triplet prediction has been compiled from two sources: our NF-κB/RelA ChIP-Seq experiments and literature mining. Our Chip-Seq experiments (Methods and (32)) have identified 830 TGs of NF-κB/RelA. These newly identified TGs, ranked by confidence scores, are provided in Supplementary Data File S2, Sheet 1. The Literature search (Methods), conducted to uncover possible additional TGs, yielded 424 targets of NF-κB/RelA (Supplementary Data File S2, Sheet 2). In total, 1194 unique NF-κB/RelA TGs were collected. 1182 of them had expression profiles available (Supplementary Data File S2, Sheet 3). GO analysis of the newly identified NF-κB/RelA TGs confirmed that they have functional properties and are involved in biological processes which are characteristic for NF-κB/RelA target genes, such as response to stress, regulation of apoptosis, cell
proliferation, differentiation and intracellular signaling pathways.

**Triplets predicted using the candidate NF-κB/RelA modulators and TGs of NF-κB/RelA**

Using the described above sets of candidate NF-κB/RelA modulators and TGs of NF-κB/RelA, we inferred the 8349 (M, NF-κB/RelA, TG) regulatory triplets of the six action modes (Supplementary Data File S4). At current threshold, for the triplets that satisfy the three hypotheses H1, H2 and H3, there are two percent (196) outliers among the predicted triplets that do not belong to any of the six action modes. In these cases, the $\hat{\alpha}_f$ and $\hat{\alpha}_f + \hat{\beta}_m$ are not significantly different from zero, while $\hat{\beta}_f$, $\gamma$ and $\hat{\beta}_m$ are significantly different from zero. The modulators have main and interactive effects at the present of the transcription factor, while the total effect is zero. Confounding factors may explain these cases. We excluded these type of modulators from all analyses and only considered the triplets with one of the six action modes. There is at least one predicted TG for 562 of the 620 binding proteins. The triplets and their parameters are listed in Supplementary Data File S4. The action modes among the 8349 triplets are 1179 inhibition attenuation, 839 inhibition enhancement, 1042 inhibition inversion, 724 activation inversion, 2580 activation enhancement, and 1985 activation attenuation. The modulators act as agonists of RelA in 3419 triplets (839 inhibition enhancement plus 2580 activation enhancement), and as antagonists in 4930 triplets (inhibition attenuation, inhibition inversion, activation inversion, and activation attenuation). Similarly, in 4801 triplets, modulators upregulate TGs, while in the remaining 3548 triplets, the modulators act as repressors to down regulate TGs.

The 20 modulators with the greatest number of TGs and their targets are shown in Figure 3. Among them, the top 10 specific modulators (filtered by the entropy of their action modes and not TGs of NF-κB/RelA themselves) are (1) Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH), (2) estrogen receptor alpha (ESR1), (3) phosphatidylinositol binding clathrin assembly protein (PICALM), (4) actin, gamma 1 (ACTG1), (5) radixin (RDX), (6) transcription factor 4 (TCF4), (7) heterogeneous nuclear ribonucleoproteins C1/C2 (HNRNPC), (8) ribosome protein L38 (RPL38), (9) microtubule-actin connecting factor 1 (MACF1), and (10) and ATP synthase subunit gamma (ATP5C1). The top 10 specific modulators and their distributions of action modes are visualized in Figure 4.

We closely examined these top 10 NF-κB/RelA modulators and found that there are literature reports in support of each of them as bona fide NF-κB/RelA modulators. Moreover, the literature generally supports the prevalent action modes identified here.

Specifically, GAPDH participates in NF-κB signaling (43,44). Recently, it has been reported that the mechanism is related to GAPDH’s role in TNF-induced NF-κB activation (45). This function may be disrupted during pathogen infection; as an important virulence strategy employed by attaching/effacing pathogens to inhibit the host NF-κB-dependent innate immune responses. Consistently, the predicted triplets demonstrate that as a modulator, GAPDH predominantly activates NF-κB TGs. Specifically, 155 triplets are of activation enhancement mode and 101 are inhibition inversion.

ESR1 and NF-κB have been found to repress each other in a cell-type specific manner in several contexts(46,47). Among the predicted 211 ESR1-RELA-TGs triplets, 91 are activation attenuation (P = 1.1e-10), consistent with the observed repression of NF-κB activity by ESR1.

In a yeast Alzheimer’s model induced by secretory amyloid-β (Aβ), a PICALM ortholog is involved in cellular toxicity by interacting with mitochondrion. Activation of the NF-κB pathway has been linked to Aβ neurotoxicity (48,49) in both neuronal cells and microglial cells. Consistently we predicted that the PICALM ortholog is a modulator of NF-κB with prevalent activation enhancement.

ACTG1 is one of the three isoforms of actin, and MACF1 is a microtubule and actin connecting factor involved in cytoskeleton dynamics homeostasis. Perturbation of cytoskeleton dynamics activates NF-κB (50,51), which can be mediated by IkappaB degradation, p38 MAP kinase activation (51) or NADPH-oxidase dependent pathways (51).

RDX, encodes Radixin, a component of the ezrin/radixin/moesin (ERM) complex that acts as a
linker between the actin cytoskeleton and plasma membrane proteins, and signal transducers involving cytoskeletal remodeling (52). In colon cancer cells, ERM is essential for cell-neural adhesion molecule (L1CALM) mediated metastasis through activation of NF-κB (53). In THP-1 monocytes treated with TNF-α, ERM rapidly translocates the RhoA small GTPase to the plasma membrane and induces actin polymerization through NF-κB (54). The predicted triplets of RDX demonstrate it is an agonist of NF-κB transcriptional activity, where 38 are inhibition enhancements and 31 activation enhancements (P=1.9e-6).

TCF4 encodes the E2-2 protein, a member of the E2 box protein family, an essential and specific regulator of plasmacytoid dendritic cell (pDC) development (55). One of the E2-2 associated proteins, Spi-B, regulates pDC survival through direct induction of the anti-apoptotic gene BCL2A1 (56), which is also an important target gene of NF-κB. Consistently, TCF4 predominantly activates NF-κB in 51 of 106 predicted triplets with activation enhancement (P=3.1e-8).

HNRNPC encodes the heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1/C2), which mainly mediates nuclear mRNA retention in 40S hnRNP particles by competing with exporting proteins as an mRNA length ruler (57). Direct modulation of hnRNPC1/C2 is not reported yet. There are however reports on modulation of NF-κB by alternative hnRNPs: hnRNPA1 enhances NF-κB activity through the mRNA stability of the modulator cIAP1 of NF-κB in HeLa cells (58). hnRNU stabilized the mRNA of the TGs regulated by NF-κB in the TLR signaling pathway in macrophages (59). hnRNU/hnRNPA1 is also involved in the stress response as components of a stress sensor complex B23/hnRNU/hnRNPA1 (60). hnRNPA1 attenuates NF-κB activity through demobilizing the mRNA level of cIAP1, while hnRNU, in our prediction, prevalently attenuates the NF-κB TGs, where 65 out of 76 triplets are of activation attenuation (P=0).

RPL38 was reported to be predominantly expressed in pancreatic ductal epithelium (61) and to perform transcript-specific translational control of HOX mRNA translation in mouse tissue patterning (62). No direct proof exists that RPL38 directly regulates NF-κB TGs. However, RPLS3 protein has been found to selectively regulate NF-κB TGs (63). Also, other ribosomal subunits have been previously implicated in selectively regulating NF-κB TGs (63).

Direct binding of ATP5C1 and RelA has not been reported. Nevertheless, there is indirect evidence of an interaction including that both ATP5C1 and RelA were reported to interact with MED15/ARC105, mediator complex subunit 15 (a RNA polymerase II transcription cofactor activity involved in transcription initiation). These interactions were detected by yeast two hybrid and pull down respectively in (64) and (65). In a recent study of the transcriptional response to relaxation training, ATP5C1 was one of the top upregulated critical molecules, and served as a hub in the interactive network of the relaxation response affected pathway. At the same time, NF-κB TGs were the top down-regulated hubs (66). Consistently, 48 of 65 ATP5C1-RelA-TG triplets showed ATP5C1 to be a repressor, which included 15 inhibition enhancement, 9 activation inversion, 24 activation enhancement (P=6.6e-8) (Figure 4).

Thus, for most of the predicted modulators, the functional association with NF-κB/RelA is supported by literature. These modulators may constitute key links in the cross-talk pathways modulating RelA/NF-κB activity in virus and bacteria infection, stress, immunity and cancer. The pathways include mitochondrion (ATP5C1), estrogen receptor pathway, actin cytoskeleton organization (ACTG1, MACF1 and RDX), clathrin mediated endocytosis (PICALM), and hnRNP C1/C2 mediated mRNA nuclear retention and genomic integrity.

The NF-κB/RelA modulatory network and its functional characterization

Many of the above triplets may be intertwined in the same pathway. To unravel the organization of transcription regulated by NF-κB/RelA and its modulators, we constructed a modulatory network, in which one type of nodes were modulators, and the other were TGs of RelA. This modulator-target interaction network can be represented as a matrix, whose rows are modulators, columns are TGs. The matrix elements correspond to the edges in the modulatory network, represented by the main effect of a modulator on a specific TG, $\beta_{m}$. In this
way, the association between any pair of modulators can be evaluated by their similar effect on all the TGs, and conversely, the association between two TGs can be assessed by their similar modulation by all the modulators. Subsequently, we used hierarchical biclustering to uncover the global organization of the modulatory network. The hierarchical approach to clustering was chosen because we expect that modulator-target interactions would reflect biological functions related to the NF-κB transcriptional activity, and such functions are known to form a nested structure. Such multi-level structure cannot be reflected by flat clustering methods, such as k-means, and most affinity propagation algorithms. Also, model-based clustering did not apply in this case, because no a-priori model of the complex and diverse interaction network is known.

Specifically, the rows and columns of this matrix were clustered using the agglomerative clustergram function from Matlab with default parameters. The M-TG network was parsed into a hierarchy of groups according to the similarity of the value of $\beta_m$. The hierarchical clustering dendrograms and the heatmaps for the $\beta_m$ matrix (Figure 5) show obvious large modules. As shown in Figure 5, modulators contain three dense clusters, A, B and C. Also the TG show three main clusters (I, II and III in Fig. 4). The lists of genes in these clusters are presented in Supplementary Data File S5. Clustering according to $\hat{\nu}$ did not produce modules as obvious as $\beta_m$ - and using both $\beta_m$ and $\hat{\nu}$ did not improve the results (data not shown). Therefore we based further analysis on clustering according to $\beta_m$.

The module composition and the significant triplets in each module are provided in Supplementary Data File S5. We found that modulators in the PCNMs tended to up-regulate the TGs in the same module, while the Modulators in the NCNMs down-regulated the TGs in the same module (Selected modules are presented in Figure 6). To identify the function associated with each module, we analyzed the enrichment of GO terms within clusters found in the network of all binding proteins and all RelA TGs. First we considered the network modules that show dominant positive correlations between the modulators from cluster A, B and C and TGs from cluster I, II and III (A-II, B-II, C-I and C-III in Figure 5), termed Positive Correlation Network Modules (PCNMs). For each PCNM, we find GO terms enriched in both the set of modulators and TGs, which demonstrates the functionality of the module (Table III). The functional annotations were characterized using BiNGO software (26) in Cytoscape (27).

As shown in Figure 5 and Supplementary Data File S6, the largest such module (C-I) consists of 296 annotated modulators and 261 positively modulated TGs (Figure 5). 34 of 50 enriched GO terms of the TGs overlap with 385 enriched GO terms of modulators. For the 296 modulators in module C-I, the most significantly enriched biological processes with enrichment fold ratio greater than 3.0 and $P<0.001$ and more than 20 modulators include: RNA processing and splicing, translation, interspecies interaction between organisms, protein complex biogenesis and assembly, response to DNA damage stimuli (29 genes, 3.5 fold), cellular macromolecular complex assembly, generation of precursor metabolites and energy, regulation of cell cycle (28 genes, 5.5 fold), and positive regulation of protein metabolic process. In many cases, the enrichments for modulators are consistent with the observed functional enrichments of TGs in this module. First, a number of significant GO terms of TGs are also enriched for modulators (see Supplementary Data File S6). These GO terms include: regulation of cell death (42 TGs, 2.7 fold and 37 Ms, 2.1 fold), regulation of transcription from RNA polymerase II promoter, positive regulation of apoptosis (22 TGs and 24 Ms), positive regulation of transcription from RNA polymerase II promoter, response to chemical stimulus (46 TGs and 44 Ms), regulation of transforming growth factor beta receptor signaling pathway (6 TGs and 5 Ms), response to stress (52 TGs and 55 Ms), cell death (26 TGs and 25 Ms), regulation of binding and aging (9 TGs and 9 Ms). Additional consistency exists between the functions of Ms and TGs in the same modules, even if not represented by the same GO term. For example, response to DNA damage stimuli (enriched in modulators) may result in cell death and affected cell proliferation, which are the most enriched GO
terms of the regulated TGs, mediated by genespecific transcription, RNA splicing and protein metabolism (Supplementary Data File S6) (67) in NF-κB pathways.

For module C-III, which is composed of the same cluster of the 296 Ms, the most enriched GO categories are cell cycle, mitotic cell cycle, cell cycle process, and negative regulation of macromolecule biosynthetic process.

Modules A-II and B-II are respectively composed of 44 and 32 annotated modulators as shown on the top of the map in Fig 4. These modulators have a prevalent effect opposite to cluster C; the TGs down regulated by cluster C (115 TGs in A-II and 165 TGs in B-II) are upregulated here. For module A-II, the cluster A is most enriched in translation, multi-organism process, ribosome biogenesis, cellular localization and RNA processing. The 115 TGs of A-II are most enriched in anti-apoptosis, cell activation and cell proliferation in immunity and inflammation processes. Consistently, the overlap between the categories enriched by the Ms and TGs composing module A-II includes positive regulation of transcription, regulation of primary metabolic process, protein metabolic process, and multi-organism process. For module B-II, cluster B (in green in Figure 5) is most enriched in anti-apoptosis, glucose homeostasis, fat cell differentiation, embryonic development, response to wounding, cell cycle arrest, response to stress and organ development. The 165 annotated TGs of B-II are mostly enriched in response to wounding, immune system process, defense response, cell migration, motility and adhesion. Consistently, the overlap between the categories enriched by the Ms and TGs composing module A-II include response to wounding, positive regulation of biosynthetic process, organ development, regulation of cell differentiation and anti-apoptosis.

The constructed modulation network also contains four main negatively correlated network modules (NCNMs): C-II, A-I, B-I and B-III, and were also analyzed, where the modulator clusters are respectively negatively correlated with the TG clusters (Table III and Supplementary Data File S6). Among the overlapping GO categories that are the top significant categories for the TGs for C-II are regulation of cell death (rank 52), cellular response to stimulus, regulation of cell cycles, multi-organism process, regulation of protein metabolic process, positive regulation of apoptosis. Besides, C-II is also enriched in other processes including immune system development, cellular response to chemical stimulus, hemopoiesis, regulation of protein metabolic process, regulation of cell cycle, multi-organism process, cell cycle arrest, placenta development, initiation of viral infection, maintenance of location, and phosphorylation.

As seen in Figure 5 and Table III, a finer hierarchy of submodules can be observed if deeper levels of clustering are considered for the modulation matrix. The functional annotations of these sub-modules do not differ significantly from their parent modules. Specifically, module C-II splits into C-II-1 and C-II-2 and further into C-II-1a and C-II-1b. The enriched GO categories for C-II-1 and C-II-2 are very similar and also consistent with the sub-sub-modules. For C-II-1a, the enriched GO categories are regulation of genespecific transcription, positive regulation of molecular function and protein modification process, and cellular response to stimulus. For C-II-1b, the enriched GO categories are regulation of apoptosis, multi-organism process, positive regulation of apoptosis, regulation of cell cycle, regulation of cellular protein metabolic process. Functional analysis shows that A-I is most enriched in transcription regulation, B-I in response to wounding and fat cell differentiation, while B-III in negative regulation of gene expression. Representative overlapping GO terms of A-I, B-I and B-III can be found in Supplementary Data File S6 and Figure 5. Our analysis suggests that the functional organization of the modulatory network may reflect the mechanisms of the cooperation between different processes.

Validation of the prediction using triplets predicted from the unconstrained TG candidates.

To validate the prediction method, we tested whether the predicted triplets were enriched in target genes obtained independently from the prediction. To this end, we applied the model allowing all genes as candidate targets of modulated regulation by NF-κB/RelA. Next, we selected all inferred target genes that had significant \( \hat{\gamma} \) and \( \hat{\beta} \), in at least one predicted
triplet. Using p-value thresholds between 1e-6 (for low false positive rate) and 1e-3 (for low false negative rate), we obtained between 4785 and 19115 candidate genes dependent on the modulated regulation by NF-κB/RelA. Over the entire range, the predicted TGs are significantly enriched in the 830 targets of RelA identified by Chip-seq (Table IV). Also the enrichments in the literature-derived and combined target lists are significant (see Table V–VI).

**Experimental validation of the STAT1-RelA-TGs triplets from the unconstrained TG candidates**

STAT1 (signal transducers and activators of transcription) is an important modulator of RelA as a pro-inflammatory effector. Acetylated STAT1 will repress the expression of anti-apoptosis NF-κB target genes via binding RelA and nuclear export (38) as revealed by in situ immunofluorescence analysis. To verify the interaction between RelA and STAT1 in the nucleus, we used immunoprecipitation and quantitative stable isotope dilution (SID)-selected monitoring (SRM)-mass spectrometry MS (68) in the sample obtained from the immunoprecipitation by anti-RelA antibody and nonspecific IgG (negative control) from the nuclear extract of A549 cells. The proteins associated with the antibodies were captured on protein-A beads and subjected to LC-SRM-MS to quantify the level of STAT1 in each sample. Compared to the negative control sample, STAT1 was enriched approximately 3.2 fold in the sample obtained from the immunoprecipitation by anti-RelA antibody, indicating that STAT1 was directly associated with RelA. Thus, the interaction between STAT1 and RelA is confirmed by our SRM experiment (Figure 7).

As a modulator of NF-κB/RelA-regulated expression, Stat1 is also predicted to modulate 2450 genes at a threshold of p-value of $\gamma$ and p-value of $\hat{\beta}$ smaller than 1e-2 in this study. We found that STAT1 modulates RelA-dependent gene transcription with multiple action modes depending on the TGs. To validate the predicted Stat1-RelA-TG triplets and their corresponding action modes, we selected 12 of the predicted TGs covering four action modes and p-values of $\gamma$ and $\hat{\beta}$ in the range from zero to 1e-2 (See Supplementary Data File S7). We predicted 31 triplets. The corresponding estimated false discovery rate is 0.19 at the thresholds. We conducted qPCR experiments measuring mRNA expression of the 12 TGs, in STAT1-deleted (U3A, STAT1-/-) and STAT-1 wild-type cells (2f) stimulated with poly(I:C) as described in Methods. We performed three independent quantitative PCR experiments for each triplet. 32% of the predicted triplets (10 out of 31 triplets as shown in Supplementary Data File S7) were validated by experiment (P= 0.016 with binomial distribution test). The validated 10 triplets involve six TGs: EGFR kinase substrate 8 (EPS8), Integrin, Alpha E (ITGA6), Short Stature Homeobox 2 (SHOX2), and CCCH-Type Containing 7B (ZC3H7B), Neurotrophin 4 (NTF4) and O-Sialoglycoprotein Endopeptidase-Like (OSGEPL1). Considering the estimated fold discovery ratio at the model thresholds is 0.19, our prediction approach is effective. Since each gene has several probesets, multiple triplets with different action modes may be predicted for each TG (Supplementary Data File S7). For each TG, only the most significant triplet was kept. In this way, four of the 12 TGs were validated by the experiment with the same action mode. As shown in Figure 8, four predicted TGs were fully confirmed along with the exact same action modes: EPS8, ITGA6, SHOX2, and ZC3H7B. Interestingly, the fully validated four TGs were among highest ranked according to their p-values of the probabilistic model parameter $\gamma$ with ranks 1, 2, 4, and 6 (the prevalence of highest ranking TGs is significant: P=0.049 with rank sum test). The experimental results here justify our lower p-value thresholds in triplet prediction from all the TGs and binding proteins of RelA.

Although STAT1 is both a binding protein and TG of RelA as newly identified by us, and thus involved in a feed-back pathway, our analysis validated the predicted triplet action modes of STAT1-RelA. Considering the high noise of the candidate targets, which actually are quantified by gene expression profiles from the genome-wide genes, our proposed computational model is effective in generating testable general triplet hypotheses.

**DISCUSSION**
We used a combination of experimental and computational approaches to characterize the modulation of transcription regulated by NF-κB/RelA. We selected candidate modulators of NF-κB/RelA using Liquid Chromatography/Tandem Mass Spectrometry (LC-MS/MS), identifying 365 binding proteins including 334 not previously reported. The identified modulators are functionally enriched in categories previously identified as related to NF-κB/RelA activity, as well as novel ones, such as ATP synthesis coupled electron transport, DNA helicase activity and conformation change, Ku70:Ku80 complex, RNA transport, oxidation reduction and cytoskeleton organization. Among the most significant enriched GO functions of the newly identified binding proteins and also of the existing modulators are: RNA processing (alternative splicing), cell cycle, mitochondrion and ribosome biogenesis, which may be candidate alternative processes interdependent with the NF-κB/RelA pathways.

We implemented a probabilistic model for modulatory network inference, which combines an earlier approach of GEM (Babur et al., 2010) with Shannon Entropy filtering. Our method is selective for modulators with well-defined prevalent action modes. Using this probabilistic model, we generated a network of 8349 M-TF-TG triplets. The results have been validated by enrichments of predicted TGs in NF-κB/RelA targets identified in a ChIP-Seq experiment and by functional analysis of the identified regulations.

We discovered modules by hierarchical clustering of the regulatory network built from all the binding proteins and TGs. Some modulators might affect the rate of transcription of genes of all or any of NF-κB/RelA complex proteins themselves. In this case, the corresponding modified modulators are target genes themselves. We removed the feedback loop to avoid this type of modulators in our network analysis. Additional validation of the method comes from literature-based analysis of the functions of the top predicted modulators.

One of the predicted RelA modulators is STAT1, a member of the Signal Transducers and Activators of Transcription (STAT) family of transcription factors, which modulate NF-κB mediated regulatory pathways in response to many cell stimuli and pathogens (38). Considering the important role of STAT1 in modulating RelA activity, we analyzed the predicted STAT1-RelA-TGs triplets and validated their action modes by STAT1 deletion and quantitative real-time PCR in human fibrosarcoma cells.

In conclusion, collecting all candidate binding proteins and the TGs of RelA (the main component of NF-κB complex), and predicting the M-TF-TG triplets resulted in an integrated functional interaction network, containing potential new mechanisms of NF-κB activity and specificity. Analysis of module organization of the modulation network may reveal mechanisms involved in the cooperation between different biological processes in immune responses, inflammation, cancer and others. Identification of new mechanisms of RelA-dependent gene expression, and related signaling pathways that cross talk with NF-κB, may provide guidance for the design of targeted therapeutics. The newly discovered 562 NF-κB/RelA modulators with distinct profiles of TGs may be potential drug targets. The modulatory network and its clusterings clarify mechanisms of achieving NF-κB/RelA specificity through modulators. Our computational and statistical approach can be readily applied to elucidate modulation of other transcription factors.

**AUTHOR CONTRIBUTIONS**

XL, AK and ARB conceived and designed the study. XL and AK performed the computational experiments and analyzed results. YZ performed Mass spectrometry based identification of new binding proteins of RelA. JY performed the ChIP-Seq experiments and library preparation. BT and MJ did the experimental validation of STAT1-RELA-TGs triplets, where BT designed the primers of TGs and MJ did the RT-PCR to validate the TGs in STAT1 modulated triplets. MR and AM analyzed ChIP-Seq data and identified novel TGs of RelA. XL, AK, ARB and MR wrote the paper.
CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Baeuerle, P. A., and Henkel, T. (1994) Function and activation of NF-kappa B in the immune system. *Annual review of immunology* **12**, 141-179

2. Barnes, P. J., and Karin, M. (1997) Nuclear factor-kappa: a pivotal transcription factor in chronic inflammatory diseases. *N Engl J Med* **336**, 1066-1071

3. Wang, T., Zhang, X., and Li, J. J. (2002) The role of NF-kappaB in the regulation of cell stress responses. *Int Immunopharmacol* **2**, 1509-1520

4. Jennewein, C., Karl, S., Baumann, B., Micheau, O., Debatin, K. M., and Fulda, S. (2012) Identification of a novel pro-apoptotic role of NF-kappaB in the regulation of TRAIL- and CD95-mediated apoptosis of glioblastoma cells. *Oncogene* **31**, 1468-1474

5. Parker, M. H., von Maltzahn, J., Bakkar, N., Al-Joubori, B., Ishibashi, J., Guttridge, D., and Rudnicki, M. A. (2012) MyoD-dependent regulation of NF-kappaB activity couples cell-cycle withdrawal to myogenic differentiation. *Skelet Muscle* **2**, 2044-5040

6. Tak, P. P., and Firestein, G. S. (2001) NF-kappaB: a key role in inflammatory diseases. *J Clin Invest* **107**, 7-11

7. Dale, E., Davis, M., and Faustman, D. L. (2006) A role for transcription factor NF-kappaB in autoimmunity: possible interactions of genes, sex, and the immune response. *Adv Physiol Educ* **30**, 152-158

8. Lumeng, C. N., and Saltiel, A. R. (2011) Inflammatory links between obesity and metabolic disease. *J Clin Invest* **121**, 2111-2117

9. Karin, M., Cao, Y., Greten, F. R., and Li, Z. W. (2002) NF-kappaB in cancer: from innocent bystander to major culprit. *Nat Rev Cancer* **2**, 301-310

10. Oeckinghaus, A., Hayden, M. S., and Ghosh, S. (2011) Crosstalk in NF-kappaB signaling pathways. *Nat Immunol* **12**, 695-708

11. Ghosh, S., and Hayden, M. S. (2008) New regulators of NF-kappaB in inflammation. *Nat Rev Immunol* **8**, 837-848

12. Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C. O., Forrest, A. R., Gough, J., Grimmond, S., Han, J. H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C. R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R. D., Tegner, J., Lenhard, B., Teichmann, S. A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D. A., Ideker, T., and Hayashizaki, Y. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-752

13. Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006) Reverse engineering cellular networks. *Nature Protocols* **1**, 663-672

14. Zhang, X. J., Zhao, X. M., He, K., Lu, L., Cao, Y. W., Liu, J. D., Hao, J. K., Liu, Z. P., and Chen, L. N. (2012) Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* **28**, 98-104

15. Bauer, T., Eils, R., and Konig, R. (2011) RIP: the regulatory interaction predictor--a machine learning-based approach for predicting target genes of transcription factors. *Bioinformatics* **27**, 2239-2247
16. Kuffner, R., Petri, T., Tavakkolkhah, P., Windhager, L., and Zimmer, R. (2012) Inferring gene regulatory networks by ANOVA. *Bioinformatics* 28, 1376-1382

17. Marbach, D., Roy, S., Ay, F., Meyer, P. E., Candeias, R., Kahveci, T., Bristow, C. A., and Kellis, M. (2012) Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. *Genome Res* 22, 1334-1349

18. Cheng, C., Yan, K.-K., Hwang, W., Qian, J., Bhardwaj, N., Rozowsky, J., Lu, Z. J., Niu, W., Alves, P., Kato, M., Snyder, M., and Gerstein, M. (2011) Construction and Analysis of an Integrated Regulatory Network Derived from High-Throughput Sequencing Data. *PLoS Comput Biol* 7, e1002190

19. Babur, O., Demir, E., Gonen, M., Sander, C., and Dogrusoz, U. (2010) Discovering modulators of gene expression. *Nucleic Acids Research* 38, 5648-5656

20. Luo, W. J., Hankenson, K. D., and Woolf, P. J. (2008) Learning transcriptional regulatory networks from high through put gene expression data using continuous three-way mutual information. *Bmc Bioinformatics* 9

21. Wang, K., Saito, M., Bisikirka, B. C., Alvarez, M. J., Lim, W. K., Rajbhandari, P., Shen, Q., Nemenman, I., Basso, K., Margolin, A. A., Klein, U., Dalla-Favera, R., and Califano, A. (2009) Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nature Biotechnology* 27, 829-U884

22. Hansen, M., Everett, L., Singh, L., and Hannenhalli, S. (2010) Mimosa: Mixture model of co-expression to detect modulators of regulatory interaction. *Algorithms for molecular biology* 5

23. Everett, L. J., Jensen, S. T., and Hannenhalli, S. (2011) Transcriptional regulation via TF-modifying enzymes: an integrative model-based analysis. *Nucleic Acids Research* 39

24. Girvan, M., and Newman, M. E. J. (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 7821-7826

25. Li, Y., Liu, L., Bai, X., Cai, H., Ji, W., Guo, D., and Zhu, Y. (2010) Comparative study of discretization methods of microarray data for inferring transcriptional regulatory networks. *Bmc Bioinformatics* 11, 520

26. Maere, S., Heymans, K., and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448-3449

27. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431-432

28. Brasier, A. R., Tian, B., Jamaluddin, M., and Lu, M. (2011) RelA Ser 276 Phosphorylation-Coupled Lys 310 Acetylation Controls Transcriptional Elongation Of Inflammatory Cytokines In RSV Infection. *Journal of Virology* 85, 11752-11769

29. Forbus, J., Spratt, H., Wiktorowicz, J., Wu, Z., Boldogh, I., Denner, L., Kurosky, A., Brasier, R. C., Luxon, B., and Brasier, A. R. (2006) Functional Analysis Of The Nuclear Proteome Of Human A549 Alveolar Epithelial Cells By HPLC- High Resolution 2D Gel Electrophoresis. *Proteomics* 6, 2656-2672

30. Tian, B., Yang, J., and Brasier, A. R. (2011) Two-step Crosslinking for Analysis of Protein-Chromatin Interactions. in *Transcriptional Regulation, Methods and Protocols*. pp

31. Zhao, Y., Widen, S. G., Jamaluddin, M., Tian, B., Wood, T. G., Edelh, B., and Brasier, A. R. (2011) Quantification of Activated NF-kB/RelA Complexes Using ssDNA Aptamer Affinity - Stable Isotope Dilution (SID)-Selective Reaction monitoring (SRM)- Mass Spectrometry. *Molecular and Cellular Proteomics* 10, 1-16

32. Yang, J., Mitra, A., Dojer, N., Fu, S., Rowicka, M., and Brasier, A. R. (2013) A probabilistic approach to learn chromatin architecture and accurate inference of the NF-kB/RelA regulatory network using ChIP-Seq. *Nucleic Acid Research*, In press
33. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**, 175-185
34. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25
35. Zhang, Y., Liu, T., Meyer, C., Eeckhoute, J., Johnson, D., Bernstein, B., Nusbaum, C., Myers, R., Brown, M., Li, W., and Liu, X. S. (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137
36. Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., FitzGerald, S., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Overduin, B., Pritchard, B., Riat, H. S., Rios, D., Ritchie, G. R., Ruffier, M., Schuster, M., Sobral, D., Spudich, G., Tang, Y. A., Trevanion, S., Vandoorne, J., Vilella, A. J., White, S., Wilder, S. P., Zadissa, A., Zamora, J., Aken, B. L., Birney, E., Cunningham, F., Dunham, I., Dunham, R., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Vogel, J., and Searle, S. M. (2011) Ensembl 2011. *Nucleic Acids Res* **39**, D800-806
37. Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C. K., Chrast, J., Lagarde, J., Gilbert, J. G., Storey, R., Swarbreck, D., Rossier, C., Ucla, C., Hubbard, T., Antonarakis, S. E., and Guigo, R. (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol* **7** Suppl 1, S4 1-9
38. Kramer, O. H., Baus, D., Knauer, S. K., Stein, S., Jager, E., Stauber, R. H., Grez, M., Pfitzner, E., and Heinzel, T. (2006) Acetylation of Stat1 modulates NF-kappaB activity. *Genes Dev* **20**, 473-485
39. Keshava Prasad, T. S., Goel, R., Kanakasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D. S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C. J., Kanth, S., Ahmed, M., Kashyap, M. K., Mohmood, R., Ramachandra, Y. L., Krishna, V., Rahman, B. A., Mohan, S., Ranganathan, P., Ramabadrn, S., Chaerkady, R., and Pandey, A. (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res* **37**, D767-772
40. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539
41. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuerman, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., and Hermjakob, H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res* **40**, D841-846
42. Antonov, A. V., Dietmann, S., Rodchenkov, I., and Mewes, H. W. (2009) PPI spider: A tool for the interpretation of proteomics data in the context of protein-protein interaction networks. *Proteomics* **9**, 2740-2749
43. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidirese, S., Hopf, C., Huhse, B., Mangano, R., Michon, A. M., Schirle, M., Schlegl, J., Schwab, M., Stein, M. A., Bauer, A., Casari, G., Drewes, G., Gavin, A. C., Jackson, D. B., Joberty, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004) A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* **6**, 97-105
44. Mookherjee, N., Lippert, D. N., Hamill, P., Falsafi, R., Nijnik, A., Kindrachuk, J., Pisticio, J., Gardy, J., Miro, P., Naseer, M., Foster, L. J., and Hancock, R. E. (2009) Intracellular receptor for human host defense peptide LL-37 in monocytes. *J Immunol* **183**, 2688-2696
45. Gao, X., Wang, X., Pham, T. H., Feuerbacher, L. A., Lubos, M. L., Huang, M., Olsen, R., Mushegian, A., Slawson, C., and Hardwidge, P. R. (2013) NleB, a bacterial effector with glycosyltransferase activity, targets GAPDH function to inhibit NF-kappaB activation. *Cell Host Microbe* **13**, 87-99
46. Stein, B., and Yang, M. X. (1995) Repression of the interleukin-6 promoter by estrogen receptor is mediated by NF-kappa B and C/EBP beta. Mol Cell Biol 15, 4971-4979

47. Gionet, N., Jansson, D., Mader, S., and Pratt, M. A. (2009) NF-kappaB and estrogen receptor alpha interactions: Differential function in estrogen receptor-negative and -positive hormone-independent breast cancer cells. J Cell Biochem 107, 448-459

48. Casal, C., Serratosa, J., and Tusell, J. M. (2004) Effects of beta-AP peptides on activation of the transcription factor NF-kappaB and in cell proliferation in glial cell cultures. Neurosci Res 48, 315-323

49. Longpre, F., Garneau, P., Christen, Y., and Ramassamy, C. (2006) Protection by Egb 761 against beta-amyloid-induced neurotoxicity: involvement of NF-kappaB, SIRT1, and MAPKs pathways and inhibition of amyloid fibril formation. Free Radic Biol Med 41, 1781-1794

50. Nemeth, Z. H., Deitch, E. A., Davidson, M. T., Szabo, C., Vizi, E. S., and Hasko, G. (2004) Disruption of the actin cytoskeleton results in nuclear factor-kappaB activation and inflammatory mediator production in cultured human intestinal epithelial cells. J Cell Physiol 200, 71-81

51. Kustermans, G., El Benna, J., Piette, J., and Legrand-Poels, S. (2005) Perturbation of actin dynamics induces NF-kappaB activation in myelomonocytic cells through an NADPH oxidasedependent pathway. Biochem J 387, 531-540

52. Ivetic, A., and Ridley, A. J. (2004) Ezrin/radixin/moesin proteins and Rho GTPase signalling in leucocytes. Immunology 112, 165-176

53. Gavert, N., Ben-Shmuel, A., Lemmon, V., Brabletz, T., and Ben-Ze’ev, A. (2013) Nuclear factor-kappaB signaling and ezrin are essential for L1-mediated metastasis of colon cancer cells. J Cell Sci 123, 2135-2143

54. Lim, S., Ryu, J., Shin, J. A., Shin, M. J., Ahn, Y. K., Kim, J. J., and Han, K. H. (2009) Tumor necrosis factor-alpha potentiates RhoA-mediated monocyte transmigratory activity in vivo at a picomolar level. Arterioscler Thromb Vasc Biol 29, 2138-2145

55. Cisse, B., Caton, M. L., Lehner, M., Maeda, T., Scheu, S., Locksley, R., Holmberg, D., Zweier, C., den Hollander, N. S., Kant, S. G., Holter, W., Rauch, A., Zhuang, Y., and Reizis, B. (2008) Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell 135, 37-48

56. Karrich, J. J., Balzarolo, M., Schmidlin, H., Libouban, M., Nagasawa, M., Gentek, R., Kamihira, S., Maeda, T., Amsen, D., Wolkers, M. C., and Blom, B. (2012) The transcription factor Spi-B regulates human plasmacytoid dendritic cell survival through direct induction of the antiapoptotic gene BCL2-A1. Blood 119, 5191-5200

57. McCloskey, A., Taniguchi, I., Shinmyozu, K., and Ohno, M. (2012) hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. Science 335, 1643-1646

58. Zhao, T. T., Graber, T. E., Jordan, L. E., Cloutier, M., Lewis, S. M., Goulet, I., Cote, J., and Holcik, M. (2009) hnRNP A1 regulates UV-induced NF-kappaB signalling through destabilization of cIAP1 mRNA. Cell Death Differ 16, 244-252

59. Zhao, W., Wang, L., Zhang, M., Wang, P., Qi, J., Zhang, L., and Gao, C. (2012) Nuclear to cytoplasmic translocation of heterogeneous nuclear ribonucleoprotein U enhances TLR-induced proinflammatory cytokine production by stabilizing mRNAs in macrophages. J Immunol 188, 3179-3187

60. Yao, Z., Duan, S., Hou, D., Wang, W., Wang, G., Liu, Y., Wen, L., and Wu, M. (2010) B23 acts as a nucleolar stress sensor and promotes cell survival through its dynamic interaction with hnRNPU and hnRNPA1. Oncogene 29, 1821-1834
61. Sahin, F., Qiu, W., Wilentz, R. E., Iacobuzio-Donahue, C. A., Grosmark, A., and Su, G. H. (2005) RPL38, FOSL1, and UPP1 are predominantly expressed in the pancreatic ductal epithelium. *Pancreas* **30**, 158-167

62. Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Xue, S., Ishijima, J., Shiroishi, T., and Barna, M. (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383-397

63. Wan, F., Anderson, D. E., Barnitz, R. A., Snow, A., Bidere, N., Zheng, L., Hegde, V., Lam, L. T., Staudt, L. M., Levens, D., Deutsch, W. A., and Lenardo, M. J. (2007) Ribosomal protein S3: a KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. *Cell* **131**, 927-939

64. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957-968

65. Naar, A. M., Beaurang, P. A., Zhou, S., Abraham, S., Solomon, W., and Tjian, R. (1999) Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**, 828-832

66. Bhasin, M. K., Dusek, J. A., Chang, B. H., Joseph, M. G., Denninger, J. W., Fricchione, G. L., Benson, H., and Libermann, T. A. (2013) Relaxation response induces temporal transcriptome changes in energy metabolism, insulin secretion and inflammatory pathways. *PLoS One* **8**, e62817

67. Schwerk, C., and Schulze-Osthoff, K. (2005) Regulation of apoptosis by alternative pre-mRNA splicing. *Molecular Cell* **19**, 1-13

68. Zhao, Y., Widen, S. G., Jamaluddin, M., Tian, B., Wood, T. G., Edeh, C. B., and Brasier, A. R. (April 20, 2011) Quantification of activated NF-kappaB/RelA complexes using ssDNA aptamer affinity-stable isotope dilution-selected reaction monitoring-mass spectrometry. *Mol Cell Proteomics* **10**, M111 008771

**Acknowledgements**—This study was conducted with the support of the Institute for Translational Sciences at the University of Texas Medical Branch. This work was partially supported by Clinical and Translational Science Award (UL1TR000071) from the National Center for Advancing Translational Sciences, National Institutes of Health, National Institutes of Health/NHLBI grants N01-HV-00245 and NIAID AI062885 (ARB), National Natural Science Foundation of China (grant No. 31271412 and 31371340) and Anhui Provincial Natural Science Foundation (grant no. 1208085MF96) (XL). The funders had no role in the preparation of the manuscript. We thank Thomas Wood for performing the sequencing experiments. We thank Heather Lander for comments and editing of the manuscript.
Figure legends

**Figure 1** ROC and Precision-recall curve in predicting RelA modulators.
A) Area under receiver operator curve of modulator prediction by varying the thresholds of bins (see Methods) with four types of proteins as benchmark positives: new binding proteins (newly identified binding proteins of RelA in this study), existing binding proteins (binding proteins collected from PIANA database and other sources), overlap of two new sets (overlapping binding proteins of RelA from our two LC-MS/MS experiments), All binding proteins (combined binding proteins of RelA of newly identified binding proteins in this study and those collected from database).
B) Area under precision-recall curve of modulator prediction by varying the thresholds of bins (see Methods) with all binding proteins.

**Figure 2** Mapped protein network of the new identified RelA associated proteins. The proteins are color-coded according to their corresponding functions or pathways.

**Figure 3** The triplets and the action modes of the top 20 specific modulators. The color-coded edges denote the action mode. The pink ovals indicate top 20 modulators, which include all the listed top 10 specific modulators in the main text.

**Figure 4** The mode distribution of the analyzed 10 modulators. The six numbers in each bin chart represent the numbers of triplets of the corresponding six color-coded action modes.

**Figure 5** Hierarchical clustering of the M-TG network. The dendrograms of both modulators (left) and target genes (top) are generated according to the values of $\hat{\beta}_m$(heatmap). Outlined rectangles correspond to functional modules. Top enriched GO terms for clusters (A, B, C, I, II, III), and modules (A-I, A-II, etc) are listed in inserts.

**Figure 6** Action modes of the significant triplets in network modules A-I, A-II, B-I and B-III.
Rows represent modulators and columns are TGs. Small squares are color-coded according the action mode of the triplet composed by the corresponding modulator and the TG (color scheme is the same as in Fig. 2 and Fig. 3 of the main text: dark green – inhibition attenuation, light green – inhibition enhancement, yellow – inhibition inversion, light orange – activation inversion, dark orange – activation enhancement, red – activation attenuation). The numeric value in each square is the p-value of $\hat{\beta}_m$. Only statistically significant triplets are provided.

**Figure 7** Validation of the association of STAT1 with RelA by SRM experiment

**Figure 8** Experimental validation of the Stat1 modulated RelA-dependent genes and the corresponding action modes.
### Table I. Interpretation of the categories of modulation and the constraints that the categories satisfy

| Modulation Classification | Explanation | $\hat{\gamma}$ | $\hat{\alpha_f}$ | $\hat{\beta_f}$ | $\hat{\beta_m}$ | $\hat{\alpha_f} + \hat{\beta_m}$ |
|---------------------------|-------------|-----------------|------------------|-----------------|-----------------|-------------------|
| Inhibition attenuation    | TF, alone, inhibits TG; M attenuates TF activity | +               | 0/-              | -/+             | 0/-             |
| Inhibition enhancement    | modulated TF inhibits TG                       | -               | 0/-              | -/              | -/              |
| Inhibition inversion      | TF, alone, inhibits TG; M inverts TF activity   | +               | -/              | +/              | -/              |
| Activation inversion      | TF, alone, activates TG; M inverts TF activity  | -               | -/              | -/              | -/              |
| Activation enhancement    | modulated TF activates TG                       | +               | 0/+             | +/              | 0/+             |

The "+/" and "/-" signs indicate significantly positive and negative values, respectively. "0/-" denotes not significantly different from zero or significantly negative values, "0/+", not significantly different from zero or significantly positive values, "+/" and "/-", non-zero values. Moreover, the classification is for triplets where the null hypotheses in Eq. (10) were also rejected.

### Table II. Representative enriched functional categories of the new identified RelA associated proteins

| Functional categories                      | Protein name                                                                 | P-value   | Fold ratio |
|-------------------------------------------|------------------------------------------------------------------------------|-----------|------------|
| DNA binding factors and enzymes           |                                                                              |           |            |
| DNA metabolic process                     | XRCC5|SSRP1|AIFM1|XRCC6|SMC3|MCM5|NONO|RF C5|TFAM|SET|MCM7|PPIA|SFPQ|TOP2A | 0.013 | 2.6 |
| DNA recombination                         | NONO|XRCC5|PPIA|SFPQ|XRCC6 | 0.044 | 4.6 |
| DNA integration                           | XRCC5|PPIA|XRCC6 | 0.009 | 16.9 |
| Response to DNA damage stimulus           | RFC5|NONO|XRCC5|SSRP1|MCM7|MYO6|AIFM1|SFPQ|XRCC6|TOP2A|SMC3 | 0.032 | 2.7 |
| DNA ligation                              | XRCC6|TOP2A | 0.036 | 21.2 |
| DNA-dependent ATPase activity             | RFC5|XRCC5|G3BP1|XRCC6|TOP2A | 0.002 | 10.0 |
| DNA helicase activity                     | XRCC5|MCM7|G3BP1|XRCC6 | 0.009 | 9.8 |
| RNA binding factors and enzymes           |                                                                              |           |            |
| RNA binding                               | RALY|FUS|SRSF1|HNRPF|PNPT1|SRP68|HNRPR|RBPF5|RLAV1|NEB1|HNRPM|HNRPL|FUBP2|NONO|FUBP2|FUBP3|DDX17|DDX28|DDX3X|PCBP1|PCBP2|DDX21|HNRPC|G3BP1|PTBP1|DDX1|NPM|TIM50|LF3|NXF1|DDX5|RBMX|FXR1|SRFS2|PA2G4|PRTF|ILF2|SRFS6|SFPQ|PABP1|RBM39|MTR3|PUTF60|THOC1 | 3.6e-19 | 5.8 |
| RNA metabolic process                     | RALY|FUS|SRSF1|HCD2|HNRPF|PNPT1|HNRPR|SM2D1|ELAV1|NEB1|HNRPM|HNRPL|SF3B2|FUBP1|NONO|T FAM|FUBP2|SRRT|SF3B1|DDX17|PCBP1|PCBP2|DHX15|DDX21|CCAR1|HNRPC|PTBP1|DDX1|DDX5|RBMX|SF3A3|SRFS2|PA2G4|PRTF|SRFS6|SFPQ|PABP1|CPSF6|RBM39|PUTF60|THOC1 | 6.6e-13 | 4.1 |
| RNA helicase activity                     | DDX17|DDX3X|G3BP1|DHX15|DDX1|DDX21|DDX5|FUBP2|NUP93|NPM|NXF1|THOC1 | 2.6e-07 | 25.7 |
| Establishment of RNA localization         |                                                                               | 3.9e-02 | 4.7 |
| Nuclear matrix and cytoskeleton components|                                                                              |           |            |
| Cytoskeleton                              | TARA|TLN1|LIMA1|ENAH|LMNB1|LMNB2|CBX3|TP | 4.5e-07 | 2.7 |
| Modules  | Representatives of enriched GO categories                                                                 |
|---------|------------------------------------------------------------------------------------------------------------|
| C-I     | Regulation of cell-death, positive regulation of apoptosis, response to chemical stimulus, regulation of TGFβ receptor pathway, aging. |
| C-I-1   | Regulation of apoptosis, aging, cellular component organization and positive regulation of apoptosis.      |
| C-I-2   | Regulation of apoptosis, response to chemical stimulus.                                                   |
| C-II    | Regulation of cell death, cellular response to stimulus, regulation of cell cycle, multi-organism process, regulation of protein metabolic process, positive regulation of apoptosis,. |
| C-II-1  | Regulation of apoptosis, positive regulation of gene expression, protein modification process, and response to organic substance. |
| C-II-1a | Regulation of gene-specific transcription, positive regulation of molecular function, positive regulation of protein modification process, and cellular response to stimulus |
Table IV. Enrichment of the predicted TGs in the newly identified TGs from ChIP-Seq at different triplet detection thresholds.

| -log10 (the p-value of the model parameters $\hat{\gamma}$ and $\hat{\beta}$) | P-value of the enrichment | #Overlap | #Predicted TGs | Enrichment fold ratio |
|---|---|---|---|---|
| 0.001 | 0 | 811 | 19115 | 1.07 |
| 0.00032 | 0 | 780 | 16941 | 1.17 |
| 0.001 | 0 | 709 | 14180 | 1.27 |
| 3.2e-05 | 0 | 608 | 11263 | 1.37 |
| 1.0e-05 | 0 | 472 | 8570 | 1.40 |
| 3.2e-06 | 4e-13 | 348 | 6413 | 1.38 |
| 1.0e-06 | 1.3e-07 | 251 | 4785 | 1.33 |

Table V. Enrichment of the predicted TGs in the existing TGs collected from database and literature at different triplet detection thresholds.

| -log10 (the p-value of the model parameters $\hat{\gamma}$ and $\hat{\beta}$) | P-value of the enrichment | #Overlap | #Predicted TGs | Enrichment fold ratio |
|---|---|---|---|---|
| 0.001 | 1.2e-05 | 407 | 19115 | 1.06 |
| 0.00032 | 0.0014 | 364 | 16941 | 1.07 |
| 0.0001 | 0.0057 | 309 | 14180 | 1.08 |
| 3.2e-05 | 0.0043 | 253 | 11263 | 1.12 |
| 1.0e-05 | 0.024 | 192 | 8570 | 1.11 |
| 3.2e-06 | 0.042 | 145 | 6413 | 1.12 |
| 1.0e-06 | 0.031 | 112 | 4785 | 1.16 |
Table VI. Enrichment of the predicted TGs in the combined TGs at different triplet detection thresholds.

| -log10 (the p-value of the model parameters $\hat{\gamma}$ and $\hat{\beta}$) | P-value of the enrichment | #Overlap | #Predicted TGs | Enrichment fold ratio |
|---|---|---|---|---|
| 0.001 | 0 | 1148 | 19115 | 1.07 |
| 0.00032 | 0 | 1080 | 16941 | 1.14 |
| 0.0001 | 0 | 961 | 14180 | 1.21 |
| 3.2e-05 | 0 | 811 | 11263 | 1.28 |
| 1.0e-05 | 0 | 630 | 8570 | 1.31 |
| 3.2e-06 | 6.3e-12 | 466 | 6413 | 1.29 |
| 1.0e-06 | 4.6e-08 | 345 | 4785 | 1.28 |
Figures

Figure 1

A

B
Figure 5
Figure 7

![Bar chart showing Native-AChE pesticide ratio for IgG and anti-Ro52 antibody]
Figure 8
Modulation of Gene Expression Regulated by the Transcription Factor NF-κB/RelA
Xueling Li, Yingxin Zhao, Bing Tian, Mohammad Jamaluddin, Abhishek Mitra, Jun Yang, Maga Rowicka, Allan R. Brasier and Andrzej Kudlicki

J. Biol. Chem. published online February 12, 2014

Access the most updated version of this article at doi: 10.1074/jbc.M113.539965

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2014/02/12/M113.539965.DC1