This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

REPA: Applying Pathway Analysis to Genome-wide Transcription Factor Binding Data

Pranjal Patra, *Department of Computer Science, Memorial University of Newfoundland,*
Tatsuo Izawa, *Department of Biology, Memorial University of Newfoundland,*
and Lourdes Peña-Castillo, *Departments of Computer Science and Biology, Memorial University of Newfoundland*

Abstract—Pathway analysis has been extensively applied to aid in the interpretation of the results of genome-wide transcription profiling studies, and has been shown to successfully find associations between the biological phenomena under study and biological pathways. There are two widely used approaches of pathway analysis: over-representation analysis, and gene set analysis. Recently genome-wide transcription factor binding data has become widely available allowing for the application of pathway analysis to this type of data. In this work, we developed REPA (Regulatory Enrichment Pathway Analysis) to apply gene set analysis to genome-wide transcription factor binding data to infer associations between transcription factors and biological pathways. We used the transcription factor binding data generated by the ENCODE project, and gene sets from the Molecular Signatures and KEGG databases. Our results showed that 54% of the predictions examined have literature support and that REPA's recall is roughly 54%. This level of precision is promising as several of REPA's predictions are expected to be novel and can be used to guide new research avenues. In addition, the results of our case studies showed that REPA enhances the interpretation of genome-wide transcription profiling studies by suggesting putative regulators behind the observed transcriptional responses.

Index Terms—Transcription factor binding data, pathway analysis, gene set analysis.

1 Introduction

Pathway analysis identifies pathways or functions related to biological phenomena under study by associating gene functional annotations in at least one biological knowledge base, such as the KEGG [1] or the Gene Ontology (GO) [2], with gene expression patterns observed in a genome-wide transcriptional profiling study [3]. The result of pathway analysis is a list of biological pathways or functions (generally referred to as gene sets) which provides an insight into the biology underlying the phenomena under study. There are two widely used approaches for pathway analysis: Over-representation analysis (ORA) and functional class scoring (FCS) also called gene set analysis (GSA). In the ORA approach, a list of genes of interest (for example, those genes deemed differentially expressed in a study) is created, and a statistical test is applied to determine whether the fraction of genes in the list associated with a given gene set (e.g., pathway or function) is larger than expected by chance; that is, whether genes associated to a given gene set are over-represented among the genes in the input list. In the FCS or GSA approach, all genes in the experiment are first grouped in gene sets (a gene set consists of a group of genes sharing a functional annotation); then the expression measurements for all genes in the experiment are used to calculate a summary statistic per gene set; and the statistical significance of each gene set statistic is assessed. Pathway analysis (either the ORA or the GSA approach) has become an intrinsic part of most gene expression data analysis workflows [3] and there is a plethora of tools for pathway analysis of gene expression data [4].

Most of the focus of pathway analysis has been the analysis of gene expression data; however, genome-wide transcription factor binding data obtained from Chip-Seq or Chip-ChIP experiments has become increasingly available. Transcription factor binding data (TFBD) provides genome-wide measurements of the strength of binding of a transcription factor (TF) to genomic locations. These genomic locations are likely transcription factor binding sites that can then be associated to genes based on their proximity to coding regions. TFBD needs to be further analyzed to be used for further biological investigations. Applying pathway analysis to TFBD is a logical step to associate transcription factors to gene sets. Lachmann *et al.* [5] developed the system Chip Enrichment Analysis (ChEA) to apply ORA to TFBD. ChEA links a transcription factor to a user-generated list of genes by statistically assessing whether targets of the transcription factor are over-represented among the genes in the input list [5].

To the best of our knowledge, ChEA has been the only application of pathway analysis to TFBD. In this work, we investigated the use of GSA to associate transcription factors to gene sets based on genome-wide TFBD. To do this, we collected TFBD from the ENCODE project [6], developed REPA (Regulatory Enrichment Pathway Analysis) to apply GSA to this data, estimated REPA's precision and recall, and demonstrated the utility of REPA in two cases studies.

• *Corresponding author: L. Peña-Castillo, Departments of Computer Science and Biology, Memorial University of Newfoundland, St. John's, NL, Canada.*
E-mail: lourdes@mun.ca

Manuscript received XX; revised XX.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

2 Materials and Methods

2.1 Transcription factor binding data (TFBD)

Under the ENCODE project [6], numerous Chip-Seq and Chip-ChIP experiments were performed, and genome-wide TFBD for 160 transcription factors in 44 cell lines was produced for a total of approximately 1.4 million transcription factor - target-gene interactions [7]. This data was processed and made freely available for download on the ChEA2 web site by Kou *et al.* [7]. We downloaded the ENCODE TFBD in BED format from ChEA2. This TFBD contained the start and end genomic coordinates of the transcription factor binding sites, score and signal. The score and signal values represent the strength of the binding between the transcription factor and the DNA.

To link the transcription factor binding sites to the genes that the transcription factor likely regulates, we intersected the genomic coordinates in the TFBD with the promoter regions of all human genes coding for lincRNA, miRNA and proteins using the BEDtools' command intersectBed [8]. The promoter region of a gene was defined as 3000 bp upstream of the transcription start site (TSS). Genomic coordinates of the promoter regions of all human genes were downloaded from Ensembl's BioMart [9] (Ensembl release 71, human genome GRCh37.p10). At the end of this step we have a linked-TFBD containing data for 131 transcription factors.

2.2 Gene sets

We obtained gene sets from the KEGG database (downloaded in March 2014) [1], [10] and from the Molecular Signatures Database version 5 (MSigDB) [11]. MSigDB consists of eight collections: 50 hallmark gene sets, 326 positional gene sets (C1), 4725 curated gene sets (C2), 836 motif gene sets (C3), 858 computational gene sets (C4), 1454 GO gene sets (C5), 189 oncogenic signatures (C6) and 1910 immunologic signatures (C7). Collection C2 of the MSigDB contains 1330 canonical pathways derived from the BioCarta, KEGG, and Reactome [12] pathway databases. Since we noticed a discrepancy between MSigDB KEGG-derived gene sets and those gene sets obtained directly from KEGG, we used the KEGG database (instead of MSigDB) to define the KEGG gene sets. We excluded all gene sets with more than 350 genes as they are usually too general to be informative to guide further biological research. There were 484 such gene sets.

2.3 Description of REPA

We developed REPA (Regulatory Enrichment Pathway Analysis) to apply GSA to linked-TFBD with the aim of inferring associations between transcription factors and gene sets. These associations are based on how strongly a transcription factor (TF) binds to the promoter regions of the genes in a gene set. As measurement of the binding strength of a transcription factor with a specific genomic location, we used the signal value provided in the TFBD from the ENCODE project. To test whether a transcription factor binds preferentially to genes in a given gene set, a competitive test [13] was used. That is, the transcription factor binding signal values for the genes in a gene set were compared against that transcription factor binding signal values for genes randomly selected to form gene sets of the same size as the original gene set. The null hypothesis is that the transcription factor binds with the same strength to the promoter regions of the genes in a gene set as to those promoter regions of the genes in the random gene sets.

For each transcription factor, REPA first gets the transcription factor binding signal values for the genes in each gene set using the linked-TFBD. If the linked-TFBD contains more than one binding signal value for a given gene (i.e., more than one transcription binding site in the promoter region of that gene or signal values corresponding to different cell lines), REPA takes the maximum binding signal value available for that gene. For those gene sets with at least 5 genes with a binding signal value, a competitive permutation test is performed where the binding signal values of the genes in the gene set are compared against permuted gene sets of the same size as the real gene set. Permuted gene sets are formed by randomly selecting genes. The binding signal values of the genes in a permuted gene set are obtained in the same way as the signal values of the genes in the real gene set. Then the Mann-Whitney U test [14], a non-parametric test, is used to test whether the median signal value of the real gene set is statistically significantly higher than the median signal value of the permuted gene set, based on a given threshold of significance. One thousand permutations are performed for each gene set - transcription factor combination. At the end of the permutations, REPA reports the number of statistically significant tests nst. The permutation based p-value is then calculated as $max(1 - \frac{nst}{numberofpermutation}, \frac{1}{numberofpermutations})$. Signal values from different cell lines are combined to maximize the amount of binding data available per TF.

2.4 Pathway analysis of gene expression data

To illustrate the use of REPA to enhance the interpretation of transcriptional profiling studies, we combined REPA's predictions with the results of performing GSA on gene expression data. To perform GSA on expression data, we used the Bioconductor package GAGE (version 2.14.4) [15] and set the q-value for statistical significance to 0.0001. As for gene sets, we used canonical pathways and hallmark gene sets.

For the first case study, we used gene expression data from a study comparing gene expression of patient-matched histologically normal (HN) and ductal carcinoma in situ (DCIS) tissue from patients with estrogen receptor positive sporadic breast cancer (GEO accession number GSE16873) [16]. The normalized version of this data was obtained from the Bioconductor package gageData (version 2.2.0). To perform GSA on this data, the gage function was executed with default parameters.

For the second case study, we used gene expression data from a study profiling host transcriptional responses of well-differentiated, primary human bronchial epithelial cells during infection of influenza A viruses (GEO accession number GSE48466) [17]. The normalized version of this data was obtained from GEO using the Bioconductor package GEOquery (version 2.30) [18]. We identified gene sets differentially expressed between uninfected cells and cells infected with H1N1pdm isolates from a nonfatal case

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

(A/KY/136/09). For this case study, the parameter *compare* of the gage function was set to *unpaired*, and all other parameters were set to default values.

To reduce the amount of redundancy in GAGE results, we compared every pair of differentially expressed gene sets. We obtained the number of genes in common between each pair and removed those gene sets with a significant overlap with another gene set (p-value < 0.0001 using the hypergeometric distribution) and at least 50% of their genes lying on the intersection between both gene sets. This filtering reduced the number of gene sets by 70%; however, the number of associated TFs decreased by only 7%. This indicates that REPAs predictions are replicated in different annotation scheme.

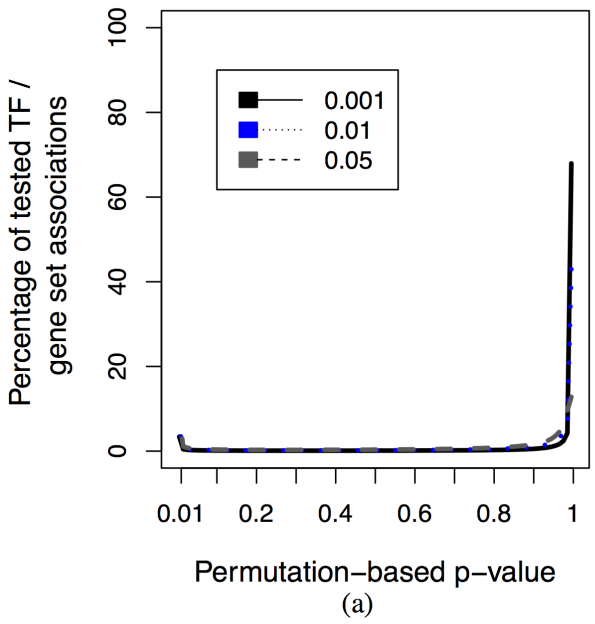
To obtain a list of statistically differentially expressed genes in each case study, we used the R/Bioconductor package limma [19] (version 3.20.8). For the breast cancer study we performed a moderated paired t-test, and for the influenza case study we performed a simple two sample comparison. Both of these approaches are described step by step in limma's user's guide.

3 Results and Discussion

3.1 Inferring associations between transcription factors and gene sets

REPA tested 803,711 transcription factor - gene set pairs and predicted 68,008 (or 8.5%) of these pairs to be true associations. REPA's predictions contained associations for 88 TFs and 8,277 gene sets. To explore the effect of the significance threshold used in the Mann-Whitney U test in the distribution of permutation-based p-values, we set the significance threshold at 0.001, 0.01 and 0.05. Figure 1 shows the distribution of permutation-based p-values corresponding to each of these significance thresholds. The distribution of permutation-based p-values has features that indicate a successful differentiation between non-associated TF - gene set pairs and TF - gene set associations with strong binding signal. The peaks on the left side (Figure 1b) contain TF - gene set associations with strong signal for which the null hypothesis was rejected. TF - gene set pairs with a p-value of less than 0.015 were predicted by REPA to be true associations. The TF -gene set pairs concentrated in the peaks on the right side or uniformly distributed in the interval $[0.015, 1)$ correspond to TF - gene set pairs for which the null hypothesis was accepted; i.e., these TF - gene set pairs did not have a strong binding signal.

The number of REPA's predictions varied from 68,008 to 29,317 depending on the significance threshold used for the Mann-Whitney U test. Since the permutation-based p-value distribution remained stable at the different significance threshold for the Mann-Whitney U test, we set the significance threshold for the Mann-Whitney U test to 0.05 and predicted as true associations those 68,008 TF - gene set pairs with a permutation-based p-value of less than 0.015. Out of the 68,008 REPA's predictions, 8,948 (13.2%) involved either a canonical pathway, a GO gene set or a hallmark gene set. All REPA's predictions are available as supplementary material.



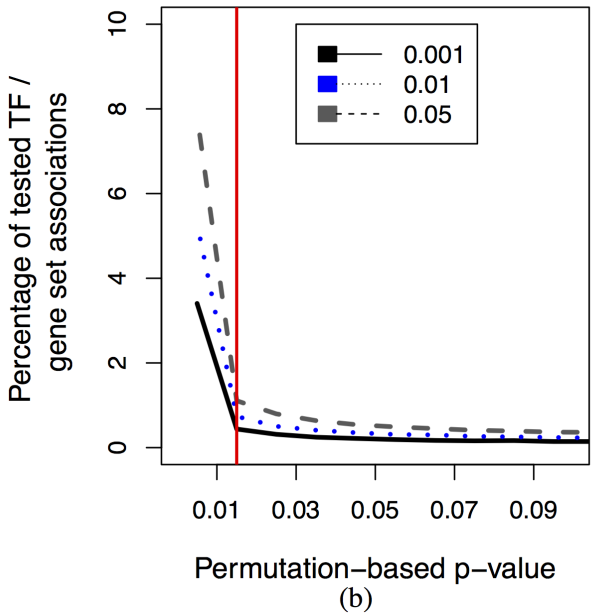


Fig. 1. Distribution of permutation-based p-values. (a) Percentage of TF - gene set pairs tested is shown as a function of REPA permutation-based p-value with different thresholds of significance used for the Mann-Whitney U test. The peak on the left side corresponds to REPA's predictions. That is, the TF - gene set pairs for which the null hypothesis was rejected (permutation-based p-value < 0.015). (b) Same as (a) but zooming in the peak on the left. REPA's predictions are those associations to the left of the vertical red line.

3.2 Number of REPA's predictions per transcription factor

We looked at the number of REPA's predictions per TF. For this analysis, we only considered predictions associating a TF with a canonical pathway, a GO gene set or a hallmark gene set, as these gene sets represent curated metabolic pathways or gene product functions. The number of REPA's predictions per TF is shown in Figure 2. There

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

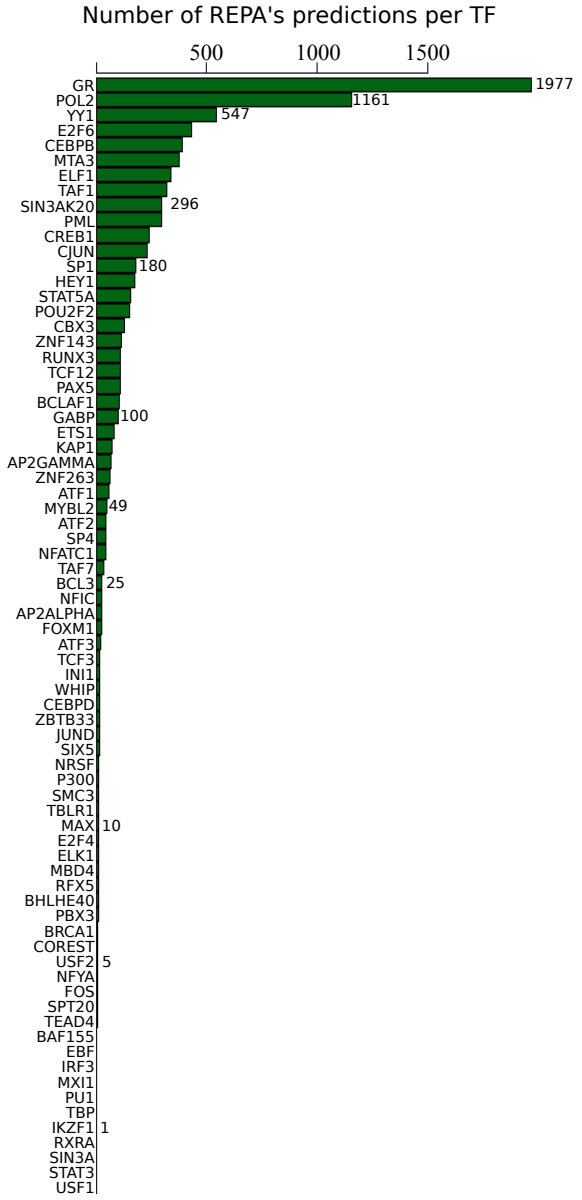


Fig. 2. Number of REPA's predictions per TF. Out of 88 TFs found among REPA's predictions, 74 (or 84%) are associated to at least one canonical pathway, GO gene set or hallmark gene set; 28 (or 37.8%) are associated to more than 50 canonical pathways, hallmark gene sets or GO gene sets; and three (3.4%), namely GR, POL2 and YY1, are associated to more than 500 of these gene sets.

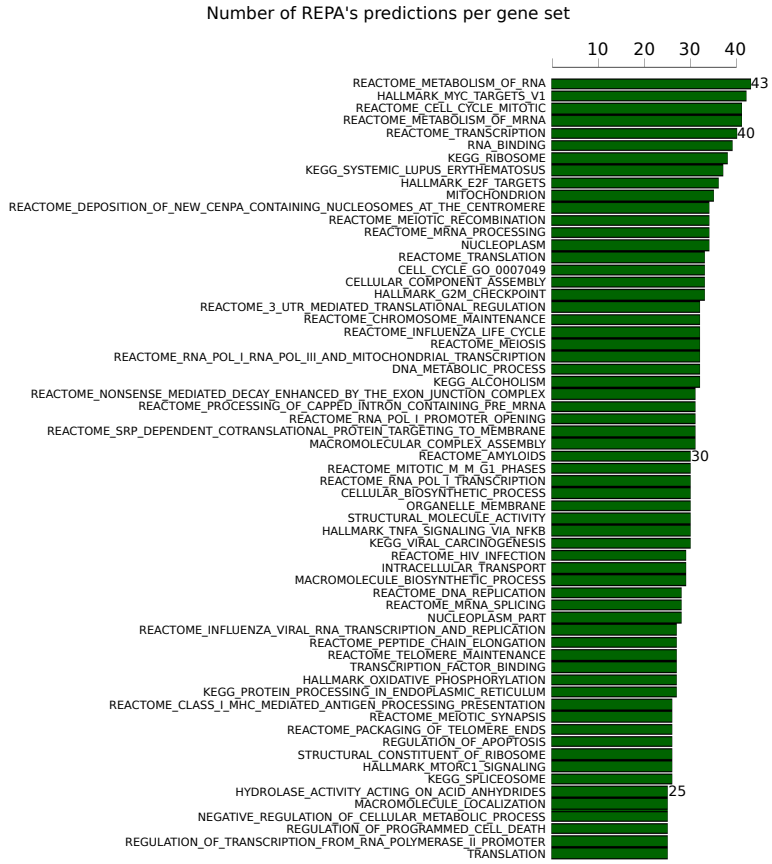


Fig. 3. Number of REPA's predictions per gene set (canonical pathways, GO gene sets and hallmark gene sets). The barplot shows the 63 genes sets predicted to be regulated by at least 25 TFs. 1,980 canonical pathways, GO gene sets and hallmark gene sets were predicted to be regulated by at least one TF, 30 (or 1.5%) were associated to more than 30 TFs and 240 (or 7.6%) were associated to more than 10 TFs.

are three TFs associated to more than 500 gene sets. These DNA-binding proteins are the glucocorticoid receptor (GR), the polymerase (RNA) II (DNA directed) polypeptide A (POL2), and the YY1 transcription factor. It is known that GR regulates diverse cellular functions (such as mitosis and apoptosis) and essential biological processes (such as growth, development, metabolism, and behaviour), and is expressed in most cell types [20], [21]. To explain how a single TF can regulate such a variety of processes, it has been proposed that different GR isoforms allow for regulation of genes in a cell type specific manner, and that each GR isoform regulates both a common and a unique group of genes in each cell type [20], [21]. The *POL2* gene encodes for the largest subunit of RNA polymerase II, the polymerase that synthesize mRNA in eukaryotes. POL2 is a general transcription factor that initiates transcription and is responsible for transcriptional regulation [22]. YY1 is an ubiquitously expressed transcription factor that regulates cell proliferation and differentiation, and is a multifunctional mediator of different signalling pathways that modulates an impressive and increasing list of genes [23]. Thus, one would expect that these three TFs regulate many different gene sets and REPA's results reflect this.

3.3 Number of REPA's predictions per gene set

We also looked at the number of REPA's predictions per gene set. There were 1,980 canonical pathways, GO gene sets and hallmark gene sets predicted to be regulated by at least one TF. Out of these 1,980, 30 (or 1.5%) were associated to more than 30 TFs. Figure 3 shows gene sets associated by REPA with at least 25 TFs. Many of the gene sets listed in Figure 3 are tightly regulated cellular processes. We examined the literature related to the regulation of the Reactome pathway and the KEGG pathway associated to the largest number of TFs.

The Reactome pathways metabolism of RNA and cell cycle mitotic were the Reactome pathways associated to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

largest number of TFs. The Reactome pathway metabolism of RNA has been deleted from Reactome since version 50 (current version is 52); thus, we investigated the cell cycle mitotic pathway which contains 325 genes. Cell cycle regulation is critical for growth and development, and its misregulation plays an important role in diseases such as cancer. There is a large variety of cell cycle programs within a single species that corresponds to specific cell types, developmental stages or physiological conditions [24]. In the budding yeast, cell cycle is controlled by a large and complex interacting network of regulatory proteins, and the general organization of this control system is conserved across the Eukaryota [25]. REPA associated the following 41 TFs with the Reactome cell cycle mitotic pathway (literature supporting the involvement of these TFs during the cell cycle is referred to after the corresponding TF): AP2ALPHA [26], AP2GAMMA, ATF1 [27], BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F4, E2F6 [25], ELF1, ELK1 [28], ETS1 [29], FOXM1 [30], GABP [31], GR [20], [21], HEY1, INI1 [32], KAP1 [33], MAX [34], MTA3, MXI1 [35], MYBL2 [36], PAX5, PBX3, PML, POL2, POU2F2 [26], RUNX3, SIN3AK20, SMC3, SP1, STAT3, STAT5A, TAF1, TBLR1, TCF12, WHIP, YY1, ZNF143, and ZNF263. Thus, we found literature support for 14 (or 34%) of these TFs. The central components of the cell-cycle control system, cyclin-dependent protein kinases (CDKs), are missing from this list because they were missing from the linked-TFBD used as REPA's input.

The KEGG ribosome pathway consists of 135 genes including the ribosomal proteins and ribosomal RNAs. The mechanisms regulating ribosome biogenesis are only partially understood, and they are the focus of current research. Recently, ribosome biogenesis has been linked to various diseases and aging, and studies have revealed an elaborate control of ribosome biogenesis that requires coordinate regulation of all three RNA polymerases and that includes feedback and feed-forward loops [37], [38]. A large number of TFs have been implicated in ribosome biogenesis; for instance, roughly 80 factors have been associated in the maturation of the 60S subunit [38]. REPA predicted the following 38 TFs to be associated with the KEGG ribosome pathway (supporting literature is referred to after the corresponding TF): ATF1, ATF2, ATF3, BCL3, BCLAF1, CBX3, CEBPB, CJUN, CREB1 [39], ELF1 [40], ETS1, GABP [41], GR [42], HEY1, IRF3, MTA3, MYBL2, NFATC1, NFIC, NRSF, PML [43], POL2 [37], POU2F2, RFX5, RUNX3, SIN3AK20, SIX5, SP1 [41], SP4, STAT5A, TAF1 [44], TAF7, TBLR1, TCF12, WHIP, YY1 [41], ZNF143, and ZNF263. Thus, we found literature support for 9 (or 24%) of these TFs.

We found an intriguing gene set (i.e., systemic lupus erythematosus (SLE)) among the top 10 genes sets shown in Figure 3, and thus decided to look at the TFs inferred by REPA to be associated with this disease. SLE is an autoimmune disease with more than 40 genes and loci identified as associated with this disease. However, these genes and loci only account for 10 to 20% of disease heritability. This indicates that there are many factors still to be identified [45]. REPA inferred 37 TFs associated with the KEGG systemic lupus erythematosus pathway which consists of 138 genes. Out of those 37 TFs predicted to regulate genes in the SLE pathway, we found literature support for 13 (or 35%). The TFs associated by REPA with SLE are the following (supporting literature is referred to after the corresponding TF): ATF2, ATF3 [46], BCLAF1, CBX3, CEBPB, CEBPD, ETS1 [47], FOS [45], FOXM1, GR [48], HEY1, INI1, JUND [49], MBD4 [50], MTA3, MYBL2, NFATC1 [49], NFIC, P300 [51], PAX5 [52], PML, POL2, POU2F2, RUNX3 [53], RXRA, SIN3AK20, SP1 [54], SP4, STAT5A, TAF1, TAF7, TBP [55], TCF12, TCF3, TEAD4, YY1 [56], and ZBTB33.

3.4 Literature-based evaluation of REPA's precision

Since REPA is the first approach to systematically inferred relationships between TFs and gene sets, there is no benchmark available to assess REPA's precision. Therefore, to gain intuition into the quality of REPA's predictions, we performed a literature analysis on 35 randomly selected predictions. The number of predictions we could examine was limited by available resources as literature curation is a time consuming effort. For this analysis, we focused on predictions associating KEGG pathways to TFs. To avoid over-weighting particular TFs, we only allowed two predictions per TF. Related literature was searched using PubMed, Disgenet [80], and ChEA [5]. Table 1 contains the list of predicted TF - gene set associations investigated. The number of genes in the gene sets examined varied from 57 to 327, and the percentage of genes in the gene set with a binding signal value varied from 18% to 94% (see third column of Table 1).

We classified evidence found in the literature into four categories: direct, binding, indirect and refuting. Direct evidence indicates that current literature directly links a TF with a given gene set; for example, the association of the TF YY1 with oxidative phosphorylation is supported by promoter analysis for a complex I gene (NDUFS8) [77]. Binding evidence indicates that targets of a TF identified by a published Chip-Seq or Chip-ChIP study are over-represented in a given gene set. This type of evidence was found using ChEA. In the case of binding evidence, current literature does not directly discuss the link of that TF with the given gene set. Indirect evidence indicates that current literature suggests the involvement of a TF in the regulation of a given gene set; for example, Taqi *et al.* [66] suggested that a single nucleotide polymorphism (SNP) in the promoter region of PDYN, which is associated with alcohol-dependence, may impact PDYN transcription in the human brain and that this SNP is located within a regulatory region that may be targeted by the TF JUND. Finally, refuting evidence indicates that current literature contains experimental data against the prediction; for example, REPA associated STAT5A with herpes simplex infection; however, current literature [76] shows that STAT1, but not STAT5A, binds to the herpes simplex virus latency-associated transcript promoter.

Out of the 35 predictions investigated, 15 (42.9%) were supported by direct evidence and 4 (11.4%) by binding evidence (see Table 1). These 19 (54.3%) associations supported by direct or binding evidence were correct or likely to be correct associations. Out of the 35 associations examined, 15 had indirect evidence or could neither be confirmed nor refuted by current literature. One (2.9%) of the 35 associations investigated was considered to be incorrect or likely to be incorrect based on current literature. These results suggest that REPA's precision lies above 54%. This level of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 1
REPA's Predictions Evaluated based on the Literature. In the column type of evidence; B indicates binding, D direct, I indirect, NE "no evidence" and R refuting evidence.

TF	Gene Set (GS)	# of genes in GS with a binding signal value / # of genes in GS	Type of evidence	Ref.
ATF2	Ribosome	68/135	I	[57]
BCLAF1	Alcoholism	52/180	NE	
CEBPB	Spliceosome	69/131	B	[58]
CEBPB	Viral carcinogenesis	128/206	D	[59]
CJUN	FOXO signalling	36/133	D	[60]
CTCF	RAP1 signalling	107/213	NE	
E2F6	Pathways in cancer	145/327	D	[61]
ELF1	Huntington's disease	100/183	B	[62]
ELF1	Spliceosome	61/131	B	[62]
FOXM1	Alcoholism	49/180	NE	
GABP	Ribosome	64/135	D	[63]
GR	Acute myeloid leukemia	54/57	D	[64]
GR	Neurotrophin signalling	110/120	I	[65]
HEY1	RNA transport	75/164	NE	
JUND	Alcoholism	48/180	I	[66]
KAP1	miRNAs in cancer	69/296	D	[67]
MBD4	SLE	25/138	D	[50]
MTA3	Herpes simplex infection	106/188	NE	
MTA3	Ribosome biogenesis in eukaryotes	40/85	NE	
MYBL2	Alcoholism	52/180	NE	
NFATC1	SLE	45/138	D	[68]
NFIC	Viral carcinogenesis	107/206	I	[69]
PAX5	Epstein Barr virus infection	101/203	D	[70]
PML	HTLV I infection	129/263	D	[71]
POL2	Bacterial invasion of epithelial cells	65/76	D	[72]
POL2	Cytokine Cytokine receptor interaction	156/271	NE	
POU2F2	Transcriptional misregulation in cancer	47/179	NE	
RUNX3	Protein processing in endoplasmic reticulum	86/167	I	[73]
SP1	Alcoholism	63/180	D	[74]
SP1	Carbon metabolism	51/105	D	[75]
STAT5A	Herpes simplex infection	89/188	R	[76]
TAF1	Ribosome	65/135	D	[44]
YY1	Oxidative phosphorylation	61/133	D	[77]
YY1	Spliceosome	62/131	B	[78]
ZNF143	RNA Transport	70/164	I	[79]

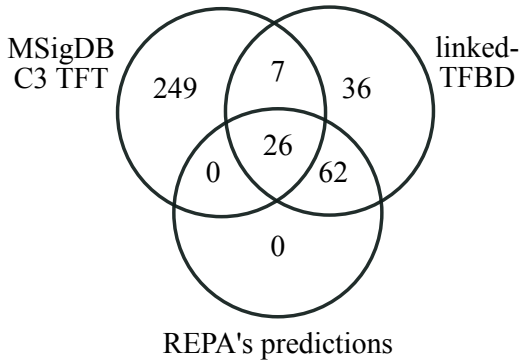


Fig. 4. Venn diagram indicating the number of TFs in common between MSigDB C3-TFT collection, ENCODE's TFBD and REPA's predictions.

precision is quite promising as several of REPA's predictions are expected to be novel and therefore lack literature support. Moreover, gene function predictions with similar precision levels have been successfully used in yeast [81] and mouse [82]. Based on this, we expect REPA's predictions to be an useful resource to guide further biological research.

3.5 Estimation of REPA's recall

Collection C3-TFT of the MSigDB consist of 615 gene sets formed by genes that share a DNA-binding motif defined in the TRANSFAC database [83]. Additionally, the TF known to bind the given motif is provided for 500 of these 615 gene sets. In total, 282 TFs are matched to a given DNA-binding motif in the C3-TFT collection. Out of these 282 TFs, 33 are also present in the linked-TFBD used as REPA's input, and REPA generated a prediction for 26 of them (see Figure 4). The 7 TFs for which REPA did not make a prediction had binding signal values for very few genes (four of them had binding signal values for less than 505 genes). To estimate REPA's recall, we counted the number of TFs associated by REPA with the corresponding C3-TFT gene set; for instance, since REPA associated SP1 to the V$SP1_01 gene set (which consists of genes whose promoter regions contain the motif GGGGCGGGGT which matches annotation for SP1) we counted SP1 as successfully retrieved by REPA. REPA associated 14 TFs to their corresponding C3-TFT gene set. Based on this, REPA's recall is around 53.8% (14/26).

3.6 Enhancing the interpretation of expression profiling studies

One of the goals of this work is to provide the research community with a resource to guide further biological research, and one logical application of REPA's predictions is to enhance the interpretation of expression profiling studies by suggesting putative regulators underlying the observed transcriptional responses. To illustrate this, we present two case studies where the results obtained from carrying out GSA on gene expression data are integrated with REPA's predictions. The purpose of these case studies is to demonstrate the utility of REPA's predictions and to evaluate the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 2
Combining REPA's Predictions with Pathway Analysis on Gene Expression Data. Statistically differentially expressed gene sets (FDR-corrected pvalue < 0.0001) between histologically normal and ductal carcinoma in situ samples from patients with estrogen receptor positive sporadic breast cancer were identified by GAGE and their putative regulators were obtained from REPA's predictions. The character between brackets before the gene set names indicates the database providing the gene set: K indicates KEGG, H MSigDB hallmark gene sets and R Reactome.

Differentially expressed gene sets	**REPA's predicted putative regulators**
(K) Protein processing in endoplasmic reticulum	ATF1, BCLAF1, BRCA1, CEBPB, CJUN, CREB1, E2F6, ELF1, ETS1, GR,HEY1, KAP1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SMC3, SP1, STAT5A, TAF1, TCF12, YY1, ZNF143, ZNF263
(K) Lysosome	BHLHE40, CEBPB, CREB1, E2F6, ELF1, ETS1, GR, MTA3, PAX5, POL2, SIN3AK20, USF1, YY1
(K) Protein export	GR, POL2
(K) Phagosome	CBX3, CEBPB, CREB1, E2F6, GR, MTA3, PML, POL2, POU2F2, SIN3AK20, STAT5A, YY1
(K) Antigen processing and presentation	BCLAF1, CEBPB, GR, MTA3, NFATC1, PML, POL2, POU2F2, STAT5A, YY1
(K) Focal adhesion	CJUN, E2F6, GR, MTA3, PML, POL2, SIN3AK20, YY1
(R) Class I MHC mediated antigen processing presentation	AP2GAMMA, BAF155, BCLAF1, CBX3, CEBPB, CJUN,CREB1, E2F6, ELF1, ETS1, GABP, GR, HEY1, MTA3, PAX5, PML, POL2,POU2F2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, TCF12, YY1, ZNF143
(R) Cell cycle mitotic	AP2ALPHA, AP2GAMMA, ATF1, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F4,E2F6, ELF1, ELK1,ETS1, FOXM1, GABP, GR, HEY1, INI1, KAP1, MAX, MTA3, MXI1,MYBL2, PAX5, PBX3, PML, POL2, POU2F2, RUNX3, SIN3AK20, SMC3, SP1, STAT3, STAT5A, TAF1, TBLR1, TCF12, WHIP, YY1, ZNF143, ZNF263
(R) Post translational protein modification	CEBPB, E2F6, ELF1, GR, HEY1, MTA3, PML, POL2, POU2F2, TAF1, YY1
(R) Membrane trafficking	CEBPB, CJUN, CREB1, ELF1, GABP, GR, PML, POL2, SIN3AK20, TAF1, YY1
(R) MHC class II antigen presentation	CREB1, ELF1, GR, POL2, SIN3AK20, YY1
(R) 3 UTR mediated translational regulation	ATF2, ATF3, BCL3, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F6, ELF1, ETS1, GABP, GR, HEY1, MTA3, NFATC1, NRSF, PML, POL2, POU2F2, RFX5, SIN3AK20, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, WHIP, YY1, ZNF143, ZNF263
(H) MTORC1 signalling	ATF1, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F6, ELF1, GR, HEY1, KAP1, MTA3, MYBL2, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, SP4, STAT5A, TAF1, TCF12, YY1, ZNF143
(H) E2F targets	AP2ALPHA, AP2GAMMA, ATF1, BCL3, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F4, E2F6, ELF1, ETS1, FOXM1, GR, HEY1, KAP1, MAX, MTA3, MYBL2, NFIC, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, YY1, ZNF143, ZNF263
(H) Oxidative phosphorylation	AP2GAMMA, CBX3, CEBPB, CJUN, COREST, CREB1, E2F6, ELF1, ELK1, ETS1, GR, HEY1, KAP1, MTA3, PML, POL2, RUNX3, SIN3AK20, SMC3, SP1, STAT5A, TAF1, TBLR1, TCF12, YY1, ZNF143, ZNF263
(H) G2M checkpoint	AP2ALPHA, AP2GAMMA, ATF1, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F4, E2F6, ELF1, ETS1, FOXM1, GABP, GR, HEY1, INI1, KAP1, MTA3, MYBL2, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, TCF12, YY1, ZNF143, ZNF263
(H) MYC targets V1	AP2ALPHA, AP2GAMMA, ATF2, ATF3, BCL3, BCLAF1, CBX3, CEBPB, CJUN, COREST, CREB1, E2F6, ELF1, ETS1, FOXM1, GABP, GR, HEY1, KAP1, MAX, MBD4, MTA3, MYBL2, NFATC1, NFIC, NRSF, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SMC3, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, YY1, ZNF143, ZNF263
(H) Estrogen response late	CEBPB, E2F6, GR, POL2
(H) Glycolysis	ATF1, CEBPB, CREB1, E2F6, ELF1, GR, HEY1, MTA3, PML, POL2, SIN3AK20, SP1, TAF1, YY1
(H) Protein secretion	CREB1, ELF1, GABP, GR, POL2, SIN3AK20, TAF1, YY1
(H) Interferon gamma response	CEBPB, CJUN, E2F6, EBF, ELF1, GR, HEY1, IKZF1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, SMC3, STAT5A, YY1
(H) Unfolded protein response	ATF2, BCLAF1, CEBPB, CJUN, CREB1, E2F6, ELF1, GR, HEY1, MTA3, PML, POL2, POU2F2, SIN3AK20, SP1, STAT5A, TAF1, TCF12, YY1, ZNF143
(H) Apical junction	CBX3, CEBPB, E2F6, GR, MTA3, PAX5, PML, POL2, SP1, STAT5A, YY1
(H) Myogenesis	E2F6, GR, POL2
(H) Epithelial mesenchymal transition	E2F6, GR, POL2

relevance of the associated TFs with the biological phenomena under study rather than to re-analyze the original gene expression data.

3.6.1 Case 1: Estrogen receptor positive sporadic breast cancer

We used gene expression data of patient-matched histologically normal (HN) and ductal carcinoma in situ (DCIS) sam-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

ples from 6 patients with estrogen receptor positive sporadic breast cancer [16]. In the original study, ORA was performed using DAVID [84]; while, we performed GSA using GAGE. As in the original study [16], focal adhesion was identified as one of the pathways differentially expressed between the HN and DCIS samples. Upon identification of the differentially expressed gene sets, we used REPA's predictions to list putative regulators of those differentially expressed gene sets (see Table 2). This list provides clues about likely regulatory mechanisms underlying the observed gene expression changes.

REPA's predictions associated 58 TFs with the gene sets identified as differentially expressed in the breast cancer expression profiling study. Out of these 58 TFs, 50 (or 86%) have previously been directly linked to breast cancer; namely, AP2ALPHA [85], AP2GAMMA [85], ATF1 [86], ATF2 [87], ATF3 [88], BAF155 [89], BCL3 [90], BCLAF1 [91], BHLHE40 [92], [93], BRCA1 [94], CBX3 [95], CEBPB [96], CJUN [97], COREST [98], CREB1 [99], E2F4 [100], E2F6 [61], EBF [101], ELF1 [102], ELK1 [103], ETS1 [104], FOXM1 [105], GABP [106], GR [107], HEY1 [108], IKZF1 [109], INI1 [110], KAP1 [111], MAX [112], MBD4 [113], MTA3 [114], MXI1 [115], MYBL2 [116], NFATC1 [117], NFIC [118], NRSF [119], PAX5 [120], PML [121], POL2 [122] RUNX3 [123], SMC3 [124], SP1 [125], SP4 [126], STAT3 [127], STAT5A [128], TAF1 [129], TBLR1 [130], TCF12 [131], USF1 [132], and YY1 [133], [134]. This result indicates that REPA's precision may be higher than the one suggested by our literature-based evaluation. Moreover, REPA's predictions suggested eight additional TFs that may play a role in breast cancer. These additional TFs are PBX3, POU2F2, RFX5, SIN3AK20, TAF7, WHIP, ZNF143, and ZNF263. Some of these eight additional TFs have already been implicated in other types of cancer.

We investigated whether the 58 TFs associated with the differentially expressed genes were themselves differentially expressed. None of the TFs was deemed statistically differentially expressed using a FDR-corrected p-value of 0.01 as significance threshold. The TF with the lowest FDR-corrected p-value was BCL3 (FDR-corrected p-value of 0.07). Nevertheless, the 58 TFs were highly expressed in the samples having half of them an average expression measurement higher than the expression measurement of 66% of genes in the experiment.

3.6.2 Case 2: Infection of influenza A viruses

We used gene expression data of uninfected cells and cells infected with H1N1pdm isolates from a nonfatal case (A/KY/136/09) from a study profiling host transcriptional responses of well-differentiated, primary human bronchial epithelial cells during infection of influenza A viruses [17]. The data set contains 3 biological replicates for each group. In the original study, pathway analyses were performed using Ingenuity Pathway Analysis (IPA, Ingenuity Systems) software. As in the original study, we identified gene sets related to cytokine signalling, interferon signalling, apoptosis, complement system, and antigen presentation. In addition, we identified several influenza-related gene sets (see Table 3).

Upon identification of the differentially expressed gene sets, we used REPA's predictions to list putative regulators of those differentially expressed gene sets (see Table 3). REPA's predictions associated 58 TFs with the gene sets identified as differentially expressed in the influenza infection transcriptional profiling study. These 58 TFs are the following (supporting literature is referred to after the corresponding TF): AP2ALPHA, AP2GAMMA, ATF1, ATF2 [135], ATF3 [136], BAF155, BCL3, BCLAF1, BRCA1, CBX3, CEBPB [137], CEBPD, CJUN [138], COREST, CREB1 [139], E2F4 [137], E2F6 [137], EBF, ELF1, ELK1 [140], ETS1, FOXM1, GABP, GR [141], HEY1, IKZF1, INI1, JUND, KAP1, MAX, MBD4, MTA3, MYBL2, NFATC1 [142], NFIC, NRSF, P300, PAX5 [143], PML [144], POL2, POU2F2 [145], PU1, RFX5, RUNX3, SIN3AK20, SMC3, SP1 [146], SP4, STAT5A, TAF1, TAF7, TBLR1, TCF12, TCF3, USF2, YY1, ZNF143, and ZNF263. We found literature linking with influenza infection 14 (or 24%) of these TFs. Additionally, ATF3 and SP1 were found to be expressed in cells infected with pandemic influenza A virus (H1N1pdm) but not in cells infected with seasonal influenza virus [17].

We looked into whether the 58 associated TFs were statistically differentially expressed in the transcriptional profiling study. We deemed genes with a FDR-corrected p-value of less than 0.01 to be differentially expressed. At this threshold of significance, 29 (or 50%) of the 58 listed TFs were found to be differentially expressed between the H1N1pdm infected and uninfected cells. Eight (ATF3, BAF155, BRCA1, CEBPB, HEY1, PAX5, PML, and PU1) of these 29 differentially expressed TFs had an absolute log2 fold change of at least 1.

3.7 Limitations and future work

One of the challenges we faced while doing this project is the lack of TFBD for all human transcription factors. There are over 1,391 known sequence-specific DNA-binding human transcription factors [147]; however, in this project we have TFBD of only 131 transcription factors, roughly a tenth of the total number of human TFs. As more TFBD is generated, REPA's accuracy and coverage will increase. Similarly, we have used a fixed number of base pairs upstream of coding regions as promoter regions; however, promoter regions are in fact of variable length. Using experimentally determined promoter sequences or promoter sequences obtained by promoter prediction programs may improve REPA's predictions. Additionally, REPA currently considers proximal promoter sequences; however, long-range regulatory interactions play an important role in mammalian genomes [148]. As our knowledge of functional long-range regulatory interactions and of the location of distal regulatory regions increases, this knowledge could also be incorporated into REPA.

4 Conclusion

We developed REPA to apply gene set analysis (GSA) to genome-wide transcription factor binding data (TFBD). To the best of our knowledge, this is the first application of gene set analysis to TFBD. The results show that currently available TFBD for human allow to infer associations for 88 TFs and 8277 gene sets. We estimated REPA's precision based on a literature assessment and REPA's recall using

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 3
Combining REPA's Predictions with Pathway Analysis on Gene Expression Data. Statistically differentially expressed gene sets (FDR-corrected pvalue < 0.0001) between uninfected and influenza infected cells were identified by GAGE and their putative regulators were obtained from REPA's predictions. The character between brackets before the gene set names indicates the database providing the gene set: K indicates KEGG, H MSigDB hallmark gene sets and R Reactome.

Differentially expressed gene sets	**REPA's predicted putative regulators**
(K) Viral carcinogenesis	AP2ALPHA, AP2GAMMA, ATF2, BCL3, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F6, ELF1, FOXM1, GR, HEY1, MTA3, NFATC1, NFIC, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, TBLR1, TCF12, TCF3, YY1
(K) Influenza A	BCLAF1, CEBPB, CJUN, E2F6, ELF1, GR, MTA3, NFATC1, PML, POL2, POU2F2, PU1, STAT5A, YY1
(K) NF KAPPA B signalling pathway	BCLAF1, CEBPB, E2F6, GR, MTA3, PML, POL2, POU2F2, STAT5A, YY1
(K) Antigen processing and presentation	BCLAF1, CEBPB, GR, MTA3, NFATC1, PML, POL2, POU2F2, STAT5A, YY1
(K) Viral myocarditis	GR, MTA3, NFATC1, POL2
(K) Hepatitis B	BCLAF1, CEBPB, CJUN, E2F6, ELF1, GR, HEY1, MTA3, PML, POL2, POU2F2, RUNX3, SIN3AK20, STAT5A, TAF1, TCF12, YY1
(K) Natural killer cell mediated cytotoxicity	CEBPB, GR, MTA3, POL2, STAT5A, YY1
(K) Cytokine cytokine receptor interaction	CEBPB, E2F6, GR, POL2
(K) Transcriptional misregulation in cancer	CEBPB, E2F6, GR, MTA3, NFIC, PML, POL2, POU2F2, SP1, TAF1, TCF12, YY1
(K) TNF signalling pathway	BCLAF1, CEBPB, E2F6, ELF1, GR, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, TCF12, YY1
(R) Cytokine signalling in immune system	BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F6, ELF1, GR, HEY1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, STAT5A, TAF1, YY1
(R) Innate immune system	AP2GAMMA, ATF1, CEBPB, CJUN, CREB1, E2F6, ELF1, ELK1, GR, HEY1, MTA3, PAX5, PML, POL2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, YY1, ZNF263
(R) Class I MHC mediated antigen processing presentation	AP2GAMMA, BAF155, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F6, ELF1, ETS1, GABP, GR, HEY1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, TCF12, YY1, ZNF143
(R) Influenza viral RNA transcription and replication	ATF2, BCL3, BCLAF1, CEBPB, CJUN, CREB1, ELF1, ETS1, GABP, GR, HEY1, MTA3, NFATC1, PML, POL2, POU2F2, RFX5, SIN3AK20, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, USF2, YY1, ZNF263
(R) Processing of capped intron containing pre mRNA	BCL3, BCLAF1, BRCA1, CBX3, CEBPB, CJUN, COREST, CREB1, E2F6, ELF1, ETS1, GABP, GR, HEY1, INI1, KAP1, MTA3, MYBL2, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, SP4, STAT5A, TAF1, TCF12, YY1, ZNF263
(R) Formation of tubulin folding intermediates by CCT TRIC	POL2
(R) Recruitment of mitotic centrosome proteins and complexes	CREB1, E2F6, ELF1, GR, MTA3, POL2, SIN3AK20, TAF1, YY1
(H) Interferon gamma response	CEBPB, CJUN, E2F6, EBF, ELF1, GR, HEY1, IKZF1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, SMC3, STAT5A, YY1
(H) TNFA signalling via NFKB	AP2GAMMA, ATF2, BAF155, BCL3, BCLAF1, CEBPB, CEBPD, CJUN, CREB1, E2F6, ELF1, GR, HEY1, JUND, MTA3, NFATC1, NFIC, P300, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, STAT5A, TAF1, TCF12, TCF3, YY1
(H) Inflammatory response	CEBPB, E2F6, GR, MTA3, POL2, POU2F2
(H) Complement	CEBPB, GR, HEY1, MTA3, PAX5, PML, POL2, POU2F2, RUNX3, YY1
(H) Apoptosis	CEBPB, E2F6, GR, MTA3, PML, POL2, POU2F2, SIN3AK20, TAF1, TCF12, YY1
(H) KRAS signalling up	E2F6, GR, MTA3, POL2, YY1
(H) Allograft rejection	ATF2, BCLAF1, CEBPB, CJUN, GR, MTA3, NFATC1, PML, POL2, POU2F2, RUNX3, STAT5A, TAF1, YY1
(H) MYC targets V1	AP2ALPHA, AP2GAMMA, ATF2, ATF3, BCL3, BCLAF1, CBX3, CEBPB, CJUN, COREST, CREB1, E2F6, ELF1, ETS1, FOXM1, GABP, GR, HEY1, KAP1, MAX, MBD4, MTA3, MYBL2, NFATC1, NFIC, NRSF, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SMC3, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, YY1, ZNF143, ZNF263
(H) E2F targets	AP2ALPHA, AP2GAMMA, ATF1, BCL3, BCLAF1, CBX3, CEBPB, CJUN, CREB1, E2F4, E2F6, ELF1, ETS1, FOXM1, GR, HEY1, KAP1, MAX, MTA3, MYBL2, NFIC, PAX5, PML, POL2, POU2F2, RUNX3, SIN3AK20, SP1, SP4, STAT5A, TAF1, TAF7, TCF12, YY1, ZNF143, ZNF263

gene sets consisting of genes sharing DNA-binding motifs. To demonstrate the utility of REPA we presented two case studies showing that by combining REPA's predictions with the results of GSA on expression data, it is possible to

enhance the interpretation of expression profiling studies by providing putative regulators underlying the observed transcriptional responses. Our results suggested that REPA's precision lies between 24% and 86%, and REPA's recall is around 54%. This level of precision is encouraging as some of REPA's predictions are expected to be novel. To facilitate the use of REPA's predicted regulatory associations, a web application to query, browse and download REPA's predictions is available at bengi.cs.mun.ca/repa. In this website, REPA's source code is also available.

ACKNOWLEDGMENT

This research was funded by an NSERC discovery grant to LPC.

REFERENCES

[1] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Data, information, knowledge and principle: back to metabolism in KEGG," *Nucleic acids research*, vol. 42, no. Database issue, pp. D199–205, Jan 2014.

[2] G. O. Consortium, J. A. Blake, M. Dolan, H. Drabkin, D. P. Hill, N. Li, D. Sitnikov, S. Bridges, S. Burgess, T. Buza, F. McCarthy, D. Peddinti, L. Pillai, S. Carbon, H. Dietze, A. Ireland, S. E. Lewis, C. J. Mungall, P. Gaudet, R. L. Chrisholm, P. Fey, W. A. Kibbe, S. Basu, D. A. Siegele, B. K. McIntosh, D. P. Renfro, A. E. Zweifel, J. C. Hu, N. H. Brown, S. Tweedie, Y. Alam-Faruque, R. Apweiler, A. Auchinchloss, K. Axelsen, B. Bely, M. . Blatter, C. Bonilla, L. Bouguerleret, E. Boutet, L. Breuza, A. Bridge, W. M. Chan, G. Chavali, E. Coudert, E. Dimmer, A. Estreicher, L. Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, R. Hieta, C. Hinz, C. Hulo, R. Huntley, J. James, F. Jungo, G. Keller, K. Laiho, D. Legge, P. Lemercier, D. Lieberherr, M. Magrane, M. J. Martin, P. Masson, P. Mutowo-Muellenet, C. O'Donovan, I. Pedruzzi, K. Pichler, D. Poggioli, P. P. Millan, S. Poux, C. Rivoire, B. Roechert, T. Sawford, M. Schneider, A. Stutz, S. Sundaram, M. Tognolli, I. Xenarios, R. Foulgar, J. Lomax, P. Roncaglia, V. K. Khodiyar, R. C. Lovering, P. J. Talmud, M. Chibucos, M. G. Giglio, H. . Chang, S. Hunter, C. McAnulla, A. Mitchell, A. Sangrador, R. Stephan, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Bahler, A. Lock, P. J. Kersey, D. M. McDowall, D. M. Staines, M. Dwinell, M. Shimoyama, S. Laulederkind, T. Hayman, S. . Wang, V. Petri, T. Lowry, P. D'Eustachio, L. Matthews, R. Balakrishnan, G. Binkley, J. M. Cherry, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, B. C. Hitz, E. L. Hong, K. Karra, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, S. Weng, E. D. Wong, T. Z. Berardini, E. Huala, H. Mi, P. D. Thomas, J. Chan, R. Kishore, P. Sternberg, K. V. Auken, D. Howe, and M. Westerfield, "Gene ontology annotations and resources," *Nucleic acids research*, vol. 41, no. Database issue, pp. D530–5, Jan 2013.

[3] P. Khatri, M. Sirota, and A. J. Butte, "Ten years of pathway analysis: current approaches and outstanding challenges," *PLoS computational biology*, vol. 8, no. 2, p. e1002375, 2012.

[4] W. H. da, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic acids research*, vol. 37, no. 1, pp. 1–13, Jan 2009.

[5] A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, and A. Ma'ayan, "ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments," *Bioinformatics (Oxford, England)*, vol. 26, no. 19, pp. 2438–2444, Oct 1 2010.

[6] E. P. Consortium, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, Sep 6 2012.

[7] Y. Kou, E. Chen, N. Clark, Q. Duan, C. Tan, and A. Ma'ayan, "ChEA2: Gene-set libraries from ChIP-X experiments to decode the transcription regulome," in *Availability, Reliability, and Security in Information Systems and HCI*, ser. Lecture Notes in Computer Science, A. Cuzzocrea, C. Kittl, D. Simos, E. Weippl, and L. Xu, Eds. Springer Berlin Heidelberg, 2013, vol. 8127, pp. 416–430.

[8] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics (Oxford, England)*, vol. 26, no. 6, pp. 841–842, Mar 15 2010.

[9] R. J. Kinsella, A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek, "Ensembl biomarts: a hub for data retrieval across taxonomic space," *Database : the journal of biological databases and curation*, vol. 2011, p. bar030, Jul 23 2011.

[10] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic acids research*, vol. 27, no. 1, pp. 29–34, Jan 1 1999.

[11] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proc Natl Acad Sci U S A*, vol. 102, no. 43, pp. 15 545–50, Oct 2005.

[12] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio, "The reactome pathway knowledgebase," *Nucleic Acids Res*, vol. 42, no. Database issue, pp. D472–7, Jan 2014.

[13] F. Emmert-Streib and G. V. Glazko, "Pathway analysis of expression data: deciphering functional building blocks of complex diseases," *PLoS computational biology*, vol. 7, no. 5, p. e1002053, May 2011.

[14] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, 03 1947.

[15] W. Luo, M. S. Friedman, K. Shedden, K. D. Hankenson, and P. J. Woolf, "GAGE: generally applicable gene set enrichment for pathway analysis," *BMC bioinformatics*, vol. 10, pp. 161–2105–10–161, May 27 2009.

[16] L. A. Emery, A. Tripathi, C. King, M. Kavanah, J. Mendez, M. D. Stone, A. de las Morenas, P. Sebastiani, and C. L. Rosenberg, "Early dysregulation of cell adhesion and extracellular matrix pathways in breast cancer progression," *The American journal of pathology*, vol. 175, no. 3, pp. 1292–1302, Sep 2009.

[17] R. L. Gerlach, J. V. Camp, Y.-K. Chu, and C. B. Jonsson, "Early host responses of seasonal and pandemic influenza A viruses in primary well-differentiated human lung epithelial cells," *PLoS One*, vol. 8, no. 11, p. e78912, 2013.

[18] S. Davis and P. S. Meltzer, "GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor," *Bioinformatics*, vol. 23, no. 14, pp. 1846–7, Jul 2007.

[19] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Res*, vol. 43, no. 7, p. e47, Apr 2015.

[20] J. Zhou and J. A. Cidlowski, "The human glucocorticoid receptor: one gene, multiple proteins and diverse responses," *Steroids*, vol. 70, no. 5-7, pp. 407–417, May-Jun 2005.

[21] N. Z. Lu and J. A. Cidlowski, "Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes," *Molecular cell*, vol. 18, no. 3, pp. 331–342, Apr 29 2005.

[22] G. Orphanides, T. Lagrange, and D. Reinberg, "The general transcription factors of RNA polymerase II," *Genes and development*, vol. 10, no. 21, pp. 2657–2683, Nov 1 1996.

[23] Z. Deng, P. Cao, M. M. Wan, and G. Sui, "Yin Yang 1: a multifaceted protein beyond a transcription factor," *Transcription*, vol. 1, no. 2, pp. 81–4, 2010.

[24] H. Harashima, N. Dissmeyer, and A. Schnittger, "Cell cycle control across the eukaryotic kingdom," *Trends Cell Biol*, vol. 23, no. 7, pp. 345–56, Jul 2013.

[25] S. B. Haase and C. Wittenberg, "Topology and control of the cell-cycle-regulated transcriptional circuitry," *Genetics*, vol. 196, no. 1, pp. 65–90, Jan 2014.

[26] L. Prasov and T. Glaser, "Dynamic expression of ganglion cell markers in retinal progenitors during the terminal cell cycle," *Mol Cell Neurosci*, vol. 50, no. 2, pp. 160–8, Jun 2012.

[27] S. Bandyopadhyay, I. Dey, M. Suresh, and G. Sundaram, "The basic leucine zipper domain transcription factor Atf1 directly controls Cdc13 expression and regulates mitotic entry indepen-

dently of Wee1 and Cdc25 in Schizosaccharomyces pombe," *Eukaryot Cell*, vol. 13, no. 6, pp. 813–21, Jun 2014.

[28] O. Demir and I. A. Kurnaz, "Phospho-Ser383-Elk-1 is localized to the mitotic spindles during cell cycle and interacts with mitotic kinase Aurora-A," *Cell Biochem Funct*, vol. 31, no. 7, pp. 591–8, Oct 2013.

[29] T. Oikawa and T. Yamada, "Molecular biology of the Ets family of transcription factors," *Gene*, vol. 303, pp. 11–34, Jan 2003.

[30] I. Wierstra and J. Alves, "FOXM1, a typical proliferation-associated transcription factor," *Biol Chem*, vol. 388, no. 12, pp. 1257–74, Dec 2007.

[31] H. Imaki, K. Nakayama, S. Delehouzee, H. Handa, M. Kitagawa, T. Kamura, and K. I. Nakayama, "Cell cycle-dependent regulation of the Skp2 promoter by GA-binding protein," *Cancer Res*, vol. 63, no. 15, pp. 4607–13, Aug 2003.

[32] I. Versteege, S. Medjkane, D. Rouillard, and O. Delattre, "A key role of the hSNF5/INI1 tumour suppressor in the control of the G1-S transition of the cell cycle," *Oncogene*, vol. 21, no. 42, pp. 6403–12, Sep 2002.

[33] D. White, I. U. Rafalska-Metcalf, A. V. Ivanov, A. Corsinotti, H. Peng, S.-C. Lee, D. Trono, S. M. Janicki, and F. J. Rauscher, 3rd, "The ATM substrate KAP1 controls DNA repair in heterochromatin: regulation by HP1 proteins and serine 473/824 phosphorylation," *Mol Cancer Res*, vol. 10, no. 3, pp. 401–14, Mar 2012.

[34] B. Amati and H. Land, "Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death," *Curr Opin Genet Dev*, vol. 4, no. 1, pp. 102–8, Feb 1994.

[35] T. C. Lee and E. B. Ziff, "Mxi1 is a repressor of the c-Myc promoter and reverses activation by USF," *J Biol Chem*, vol. 274, no. 2, pp. 595–606, Jan 1999.

[36] M. Joaquin and R. J. Watson, "Cell cycle regulation by the B-Myb transcription factor," *Cell Mol Life Sci*, vol. 60, no. 11, pp. 2389–401, Nov 2003.

[37] H. Lempiainen and D. Shore, "Growth control and ribosome biogenesis," *Current opinion in cell biology*, vol. 21, no. 6, pp. 855–863, Dec 2009.

[38] E. Thomson, S. Ferreira-Cerca, and E. Hurt, "Eukaryotic ribosome biogenesis at a glance," *J Cell Sci*, vol. 126, no. Pt 21, pp. 4815–21, Nov 2013.

[39] N. Nosrati, N. R. Kapoor, and V. Kumar, "Combinatorial action of transcription factors orchestrates cell cycle-dependent expression of the ribosomal protein genes and ribosome biogenesis," *FEBS J*, vol. 281, no. 10, pp. 2339–52, May 2014.

[40] M. A. Algire, D. Maag, P. Savio, M. G. Acker, S. Z. Tarun, Jr, A. B. Sachs, K. Asano, K. H. Nielsen, D. S. Olsen, L. Phan, A. G. Hinnebusch, and J. R. Lorsch, "Development and characterization of a reconstituted yeast translation initiation system," *RNA*, vol. 8, no. 3, pp. 382–97, Mar 2002.

[41] R. P. Perry, "The architecture of mammalian ribosomal protein promoters," *BMC Evol Biol*, vol. 5, p. 15, 2005.

[42] O. J. Shah, J. A. Iniguez-Lluhi, A. Romanelli, S. R. Kimball, and L. S. Jefferson, "The activated glucocorticoid receptor modulates presumptive autoregulation of ribosomal protein S6 protein kinase, p70 S6K," *J Biol Chem*, vol. 277, no. 4, pp. 2525–33, Jan 2002.

[43] S. Vilotti, M. Biagioli, R. Foti, M. Dal Ferro, Z. S. Lavina, L. Collavin, G. Del Sal, S. Zucchelli, and S. Gustincich, "The PML nuclear bodies-associated protein TTRAP regulates ribosome biogenesis in nucleolar cavities upon proteasome inhibition," *Cell Death Differ*, vol. 19, no. 3, pp. 488–500, Mar 2012.

[44] C. Y. Lin, J. Tuan, P. Scalia, T. Bui, and L. Comai, "The cell cycle regulatory factor TAF1 stimulates ribosomal DNA transcription by binding to the activator UBF," *Current biology : CB*, vol. 12, no. 24, pp. 2142–2146, Dec 23 2002.

[45] E. A. Frangou, G. K. Bertsias, and D. T. Boumpas, "Gene expression and regulation in systemic lupus erythematosus," *European journal of clinical investigation*, vol. 43, no. 10, pp. 1084–1096, Oct 2013.

[46] X. Cai, Y. Xu, Y.-M. Kim, J. Loureiro, and Q. Huang, "Pikfyve, a class III lipid kinase, is required for TLR-induced type I IFN production via modulation of ATF3," *J Immunol*, vol. 192, no. 7, pp. 3383–9, Apr 2014.

[47] X. Lu, E. E. Zoller, M. T. Weirauch, Z. Wu, B. Namjou, A. H. Williams, J. T. Ziegler, M. E. Comeau, M. C. Marion, S. B. Glenn, A. Adler, N. Shen, S. K. Nath, A. M. Stevens, B. I. Freedman, B. P. Tsao, C. O. Jacob, D. L. Kamen, E. E. Brown, G. S. Gilkeson, G. S. Alarcón, J. D. Reveille, J.-M. Anaya, J. A. James, K. L. Sivils, L. A. Criswell, L. M. Vilá, M. E. Alarcón-Riquelme, M. Petri, R. H. Scofield, R. P. Kimberly, R. Ramsey-Goldman, Y. B. Joo, J. Choi, S.-C. Bae, S. A. Boackle, D. C. Graham, T. J. Vyse, J. M. Guthridge, P. M. Gaffney, C. D. Langefeld, J. A. Kelly, K. D. Greis, K. M. Kaufman, J. B. Harley, and L. C. Kottyan, "Lupus risk variant increases pSTAT1 binding and decreases ETS1 expression," *Am J Hum Genet*, Apr 2015.

[48] H. Chen, J. Fan, Q. Shou, L. Zhang, H. Ma, and Y. Fan, "Hypermethylation of glucocorticoid receptor gene promoter results in glucocorticoid receptor gene low expression in peripheral blood mononuclear cells of patients with systemic lupus erythematosus," *Rheumatol Int*, Apr 2015.

[49] K. Tenbrock, Y. T. Juang, V. C. Kyttaris, and G. C. Tsokos, "Altered signal transduction in SLE T cells," *Rheumatology (Oxford, England)*, vol. 46, no. 10, pp. 1525–1530, Oct 2007.

[50] E. Balada, J. Ordi-Ros, S. Serrano-Acedo, L. Martinez-Lostao, and M. Vilardell-Tarres, "Transcript overexpression of the MBD2 and MBD4 genes in CD4+ T cells from systemic lupus erythematosus patients," *Journal of leukocyte biology*, vol. 81, no. 6, pp. 1609–1616, Jun 2007.

[51] Y. T. Leung, L. Shi, K. Maurer, L. Song, Z. Zhang, M. Petri, and K. E. Sullivan, "Interferon regulatory factor 1 and histone H4 acetylation in systemic lupus erythematosus," *Epigenetics*, vol. 10, no. 3, pp. 191–9, Mar 2015.

[52] M. G. Dozmorov, J. D. Wren, and M. E. Alarcón-Riquelme, "Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes," *Epigenetics*, vol. 9, no. 2, pp. 276–85, Feb 2014.

[53] M. A. Jeffries, M. Dozmorov, Y. Tang, J. T. Merrill, J. D. Wren, and A. H. Sawalha, "Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus," *Epigenetics*, vol. 6, no. 5, pp. 593–601, May 2011.

[54] K. Hikami, A. Kawasaki, I. Ito, M. Koga, S. Ito, T. Hayashi, I. Matsumoto, A. Tsutsumi, M. Kusaoi, Y. Takasaki, H. Hashimoto, T. Arinami, T. Sumida, and N. Tsuchiya, "Association of a functional polymorphism in the 3'-untranslated region of SPI1 with systemic lupus erythematosus," *Arthritis and Rheumatism*, vol. 63, no. 3, pp. 755–763, Mar 2011.

[55] R. Chauhan, R. Handa, T. P. Das, and U. Pati, "Over-expression of TATA binding protein (TBP) and p53 and autoantibodies to these antigens are features of systemic sclerosis, systemic lupus erythematosus and overlap syndromes," *Clin Exp Immunol*, vol. 136, no. 3, pp. 574–84, Jun 2004.

[56] H. Zhao, W. Yang, R. Qiu, J. Li, Q. Xin, X. Wang, Y. Feng, S. Shan, Y. Liu, Y. Gong, and Q. Liu, "An intronic variant associated with systemic lupus erythematosus changes the binding affinity of Yinyang1 to downregulate WDFY4," *Genes Immun*, vol. 13, no. 7, pp. 536–42, Oct 2012.

[57] C. R. Johnson, T. Jiffar, U. M. Fischer, P. P. Ruvolo, and W. D. Jarvis, "Requirement for SAPK-JNK signaling in the induction of apoptosis by ribosomal stress in REH lymphoid leukemia cells," *Leukemia*, vol. 17, no. 11, pp. 2140–2148, Nov 2003.

[58] M. I. Lefterova, D. J. Steger, D. Zhuo, M. Qatanani, S. E. Mullican, G. Tuteja, E. Manduchi, G. R. Grant, and M. A. Lazar, "Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages," *Molecular and cellular biology*, vol. 30, no. 9, pp. 2078–2089, May 2010.

[59] N. Watanabe-Okochi, A. Yoshimi, T. Sato, T. Ikeda, K. Kumano, K. Taoka, Y. Satoh, A. Shinohara, T. Tsuruta, A. Masuda, H. Yokota, Y. Yatomi, K. Takahashi, J. Kitaura, T. Kitamura, and M. Kurokawa, "The shortest isoform of C/EBPbeta, liver inhibitory protein (LIP), collaborates with Evi1 to induce AML in a mouse BMT model," *Blood*, vol. 121, no. 20, pp. 4142–4155, May 16 2013.

[60] P. Xu, M. Das, J. Reilly, and R. J. Davis, "JNK regulates FoxO-dependent autophagy in neurons," *Genes Dev*, vol. 25, no. 4, pp. 310–22, Feb 2011.

[61] M. J. Oberley, D. R. Inman, and P. J. Farnham, "E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9," *The Journal of biological chemistry*, vol. 278, no. 43, pp. 42 466–42 476, Oct 24 2003.

[62] P. C. Hollenhorst, A. A. Shah, C. Hopkins, and B. J. Graves, "Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family," *Genes and development*, vol. 21, no. 15, pp. 1882–1894, Aug 1 2007.

[63] A. Donadini, F. Giacopelli, R. Ravazzolo, V. Gandin, P. C. Marchisio, and S. Biffo, "GABP complex regulates transcription of eIF6

(p27BBP), an essential trans-acting factor in ribosome biogenesis," *FEBS letters*, vol. 580, no. 8, pp. 1983–1987, Apr 3 2006.

[64] E. G. Haarman, G. J. Kaspers, R. Pieters, M. M. Rottier, M. L. D. Boer, G. E. Janka-Schaub, and A. J. Veerman, "In vitro glucocorticoid resistance in childhood leukemia correlates with receptor affinity determined at 37 degrees C, but not with affinity determined at room temperature," *Leukemia*, vol. 16, no. 9, pp. 1882–1884, Sep 2002.

[65] N. Adachi, T. Numakawa, M. Richards, S. Nakajima, and H. Kunugi, "New insight in expression, transport, and secretion of brain-derived neurotrophic factor: Implications in brain-related diseases," *World journal of biological chemistry*, vol. 5, no. 4, pp. 409–428, Nov 26 2014.

[66] M. M. Taqi, I. Bazov, H. Watanabe, F. Nyberg, T. Yakovleva, and G. Bakalkin, "Prodynorphin promoter SNP associated with alcohol dependence forms noncanonical AP-1 binding site that may influence gene expression in human brain," *Brain research*, vol. 1385, pp. 18–25, Apr 18 2011.

[67] D.-J. Min, T. Ezponda, M. K. Kim, C. M. Will, E. Martinez-Garcia, R. Popovic, V. Basrur, K. S. Elenitoba-Johnson, and J. D. Licht, "MMSET stimulates myeloma cell growth through microRNA-mediated modulation of c-MYC," *Leukemia*, vol. 27, no. 3, pp. 686–94, Mar 2013.

[68] M. C. Lu, N. S. Lai, H. C. Yu, S. C. Hsieh, C. H. Tung, and C. L. Yu, "Nifedipine suppresses Th1/Th2 cytokine production and increased apoptosis of anti-CD3 + anti-CD28-activated mononuclear cells from patients with systemic lupus erythematosus via calcineurin pathway," *Clinical immunology (Orlando, Fla.)*, vol. 129, no. 3, pp. 462–470, Dec 2008.

[69] E. R. Schuur, U. Kruse, J. S. Iacovoni, and P. K. Vogt, "Nuclear factor I interferes with transformation induced by nuclear oncogenes," *Cell growth and differentiation : the molecular biology journal of the American Association for Cancer Research*, vol. 6, no. 3, pp. 219–227, Mar 1995.

[70] R. Tierney, J. Nagra, I. Hutchings, C. Shannon-Lowe, M. Altmann, W. Hammerschmidt, A. Rickinson, and A. Bell, "Epstein-barr virus exploits BSAP/Pax5 to achieve the B-cell specificity of its growth-transforming program," *Journal of virology*, vol. 81, no. 18, pp. 10 092–10 100, Sep 2007.

[71] Y. Ariumi, T. Ego, A. Kaida, M. Matsumoto, P. P. Pandolfi, and K. Shimotohno, "Distinct nuclear body components, PML and SMRT, regulate the trans-acting function of HTLV-1 Tax oncoprotein," *Oncogene*, vol. 22, no. 11, pp. 1611–1619, Mar 20 2003.

[72] N. Lutay, I. Ambite, J. G. Hernandez, G. Rydstrom, B. Ragnarsdottir, M. Puthia, A. Nadeem, J. Zhang, P. Storm, U. Dobrindt, B. Wullt, and C. Svanborg, "Bacterial control of host gene expression through RNA polymerase II," *The Journal of clinical investigation*, vol. 123, no. 6, pp. 2366–2379, Jun 3 2013.

[73] D. M. Evans, C. C. A. Spencer, J. J. Pointon, Z. Su, D. Harvey, G. Kochan, U. Oppermann, U. Opperman, A. Dilthey, M. Pirinen, M. A. Stone, L. Appleton, L. Moutsianas, L. Moutsianis, S. Leslie, T. Wordsworth, T. J. Kenna, T. Karaderi, G. P. Thomas, M. M. Ward, M. H. Weisman, C. Farrar, L. A. Bradbury, P. Danoy, R. D. Inman, W. Maksymowych, D. Gladman, P. Rahman, Spondyloarthritis Research Consortium of Canada (SPARCC), A. Morgan, H. Marzo-Ortega, P. Bowness, K. Gaffney, J. S. H. Gaston, M. Smith, J. Bruges-Armas, A.-R. Couto, R. Sorrentino, F. Paladini, M. A. Ferreira, H. Xu, Y. Liu, L. Jiang, C. Lopez-Larrea, R. Díaz-Peña, A. López-Vázquez, T. Zayats, G. Band, C. Bellenguez, H. Blackburn, J. M. Blackwell, E. Bramon, S. J. Bumpstead, J. P. Casas, A. Corvin, N. Craddock, P. Deloukas, S. Dronov, A. Duncanson, S. Edkins, C. Freeman, M. Gillman, E. Gray, R. Gwilliam, N. Hammond, S. E. Hunt, J. Jankowski, A. Jayakumar, C. Langford, J. Liddle, H. S. Markus, C. G. Mathew, O. T. McCann, M. I. McCarthy, C. N. A. Palmer, L. Peltonen, R. Plomin, S. C. Potter, A. Rautanen, R. Ravindrarajah, M. Ricketts, N. Samani, S. J. Sawcer, A. Strange, R. C. Trembath, A. C. Viswanathan, M. Waller, P. Weston, P. Whittaker, S. Widaa, N. W. Wood, G. McVean, J. D. Reveille, B. P. Wordsworth, M. A. Brown, P. Donnelly, Australo-Anglo-American Spondyloarthritis Consortium (TASC), and Wellcome Trust Case Control Consortium 2 (WTCCC2), "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility," *Nat Genet*, vol. 43, no. 8, pp. 761–7, Aug 2011.

[74] S. Harada, T. Okubo, M. Tsutsumi, S. Takase, and T. Muramatsu, "A new genetic variant in the Sp1 binding cis-element of cholecystokinin gene promoter region and relationship to alcoholism," *Alcoholism, Clinical and Experimental Research*, vol. 22, no. 3 Suppl, pp. 93S–96S, May 1998.

[75] H.-H. Lin, S.-C. Lai, and L.-Y. Chau, "Heme oxygenase-1/carbon monoxide induces vascular endothelial growth factor expression via p38 kinase-dependent activation of Sp1," *J Biol Chem*, vol. 286, no. 5, pp. 3829–38, Feb 2011.

[76] J. D. Kriesel, B. B. Jones, K. M. Dahms, and S. L. Spruance, "STAT1 binds to the herpes simplex virus type 1 latency-associated transcript promoter," *Journal of neurovirology*, vol. 10, no. 1, pp. 12–20, Feb 2004.

[77] P. Lescuyer, P. Martinez, and J. Lunardi, "YY1 and Sp1 activate transcription of the human NDUFS8 gene encoding the mitochondrial complex I TYKY subunit," *Biochimica et biophysica acta*, vol. 1574, no. 2, pp. 164–174, Mar 19 2002.

[78] E. M. Mendenhall, R. P. Koche, T. Truong, V. W. Zhou, B. Issac, A. S. Chi, M. Ku, and B. E. Bernstein, "GC-rich sequence elements recruit PRC2 in mammalian ES cells," *PLoS genetics*, vol. 6, no. 12, p. e1001244, Dec 9 2010.

[79] C.-C. Yuan, X. Zhao, L. Florens, S. K. Swanson, M. P. Washburn, and N. Hernandez, "CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription," *Mol Cell Biol*, vol. 27, no. 24, pp. 8729–38, Dec 2007.

[80] A. Bauer-Mehren, M. Bundschus, M. Rautschka, M. A. Mayer, F. Sanz, and L. I. Furlong, "Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases," *PloS one*, vol. 6, no. 6, p. e20284, 2011.

[81] W. T. Peng, M. D. Robinson, S. Mnaimneh, N. J. Krogan, G. Cagney, Q. Morris, A. P. Davierwala, J. Grigull, X. Yang, W. Zhang, N. Mitsakakis, O. W. Ryan, N. Datta, V. Jojic, C. Pal, V. Canadien, D. Richards, B. Beattie, L. F. Wu, S. J. Altschuler, S. Roweis, B. J. Frey, A. Emili, J. F. Greenblatt, and T. R. Hughes, "A panoramic view of yeast noncoding RNA processing," *Cell*, vol. 113, no. 7, pp. 919–33, Jun 2003.

[82] L. Peña-Castillo, M. Tasan, C. L. Myers, H. Lee, T. Joshi, C. Zhang, Y. Guan, M. Leone, A. Pagnani, W. K. Kim, C. Krumpelman, W. Tian, G. Obozinski, Y. Qi, S. Mostafavi, G. N. Lin, G. F. Berriz, F. D. Gibbons, G. Lanckriet, J. Qiu, C. Grant, Z. Barutcuoglu, D. P. Hill, D. Warde-Farley, C. Grouios, D. Ray, J. A. Blake, M. Deng, M. I. Jordan, W. S. Noble, Q. Morris, J. Klein-Seetharaman, Z. Bar-Joseph, T. Chen, F. Sun, O. G. Troyanskaya, E. M. Marcotte, D. Xu, T. R. Hughes, and F. P. Roth, "A critical assessment of Mus musculus gene function prediction using integrated genomic evidence," *Genome Biol*, vol. 9 Suppl 1, p. S2, 2008.

[83] E. Wingender, "The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation," *Brief Bioinform*, vol. 9, no. 4, pp. 326–32, Jul 2008.

[84] W. H. da, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature protocols*, vol. 4, no. 1, pp. 44–57, 2009.

[85] L. A. McPherson, G. W. Woodfield, and R. J. Weigel, "AP2 transcription factors regulate expression of CRABPII in hormone responsive breast carcinoma," *J Surg Res*, vol. 138, no. 1, pp. 71–8, Mar 2007.

[86] D. T. Jones, T. Lechertier, R. Mitter, J. M. J. Herbert, R. Bicknell, J. L. Jones, J.-L. Li, F. Buffa, A. L. Harris, and K. Hodivala-Dilke, "Gene expression analysis in human breast cancer associated blood vessels," *PLoS One*, vol. 7, no. 10, p. e44294, 2012.

[87] E. Lau and Z. A. Ronai, "ATF2 - at the crossroad of nuclear and cytosolic functions," *J Cell Sci*, vol. 125, no. Pt 12, pp. 2815–24, Jun 2012.

[88] X. Yin, C. C. Wolford, Y.-S. Chang, S. J. McConoughey, S. A. Ramsey, A. Aderem, and T. Hai, "ATF3, an adaptive-response gene, enhances tgfβsignaling and cancer-initiating cell features in breast cancer cells," *J Cell Sci*, vol. 123, no. Pt 20, pp. 3558–65, Oct 2010.

[89] L. Wang, Z. Zhao, M. B. Meyer, S. Saha, M. Yu, A. Guo, K. B. Wisinski, W. Huang, W. Cai, J. W. Pike, M. Yuan, P. Ahlquist, and W. Xu, "CARM1 methylates chromatin remodeling factor BAF155 to enhance tumor progression and metastasis," *Cancer Cell*, vol. 25, no. 1, pp. 21–36, Jan 2014.

[90] H. J. Choi, J. M. Lee, H. Kim, H. J. Nam, H.-J. R. Shin, D. Kim, E. Ko, D.-Y. Noh, K. I. Kim, J. H. Kim, and S. H. Baek, "Bcl3-dependent stabilization of CtBP1 is crucial for the inhibition of apoptosis and tumor progression in breast cancer," *Biochem Biophys Res Commun*, vol. 400, no. 3, pp. 396–402, Sep 2010.

[91] K. I. Savage, J. J. Gorski, E. M. Barros, G. W. Irwin, L. Manti, A. J. Powell, A. Pellagatti, N. Lukashchuk, D. J. McCance, W. G. McCluggage, G. Schettino, M. Salto-Tellez, J. Boultwood, D. J. Richard, S. S. McDade, and D. P. Harkin, "Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability," *Mol Cell*, vol. 54, no. 3, pp. 445–59, May 2014.

[92] Y. Wu, H. Sato, T. Suzuki, T. Yoshizawa, S. Morohashi, H. Seino, T. Kawamoto, K. Fujimoto, Y. Kato, and H. Kijima, "Involvement of c-Myc in the proliferation of MCF-7 human breast cancer cells induced by bHLH transcription factor DEC2," *International journal of molecular medicine*, Dec 17 2014.

[93] C. Cadenas, L. van de Sandt, K. Edlund, M. Lohr, B. Hellwig, R. Marchan, M. Schmidt, J. Rahnenführer, H. Oster, and J. G. Hengstler, "Loss of circadian clock gene expression is associated with tumor progression in breast cancer," *Cell Cycle*, vol. 13, no. 20, pp. 3282–91, Oct 2014.

[94] L. Wang and L.-J. Di, "BRCA1 and estrogen/estrogen receptor in breast cancer: where they interact?" *Int J Biol Sci*, vol. 10, no. 5, pp. 566–75, 2014.

[95] J. D. Choi, M. A. Park, and J.-S. Lee, "Suppression and recovery of BRCA1-mediated transcription by HP1γ via modulation of promoter occupancy," *Nucleic Acids Res*, vol. 40, no. 22, pp. 11 321–38, Dec 2012.

[96] M. M. Abreu and L. Sealy, "The C/EBPbeta isoform, liver-inhibitory protein (LIP), induces autophagy in breast cancer cell lines," *Exp Cell Res*, vol. 316, no. 19, pp. 3227–38, Nov 2010.

[97] Y. Xu, S.-T. Zou, R. Zhu, W. Li, C.-W. Gu, S.-H. Wei, J.-M. Xie, and H.-R. Wu, "Inhibition of proliferation of estrogen receptor-positive MCF7 human breast cancer cells by tamoxifen through cjun transcription factors," *Mol Med Rep*, vol. 7, no. 4, pp. 1283–7, Apr 2013.

[98] G. P. Vicent, A. S. Nacht, R. Zaurin, J. Font-Mateu, D. Soronellas, F. Le Dily, D. Reyes, and M. Beato, "Unliganded progesterone receptor-mediated targeting of an RNA-containing repressive complex silences a subset of hormone-inducible genes," *Genes Dev*, vol. 27, no. 10, pp. 1179–97, May 2013.

[99] N. T. T. Phuong, S. C. Lim, Y. M. Kim, and K. W. Kang, "Aromatase induction in tamoxifen-resistant breast cancer: Role of phosphoinositide 3-kinase-dependent CREB activation," *Cancer Lett*, vol. 351, no. 1, pp. 91–9, Aug 2014.

[100] S. S. Khaleel, E. H. Andrews, M. Ung, J. DiRenzo, and C. Cheng, "E2F4 regulatory program predicts patient survival prognosis in breast cancer," *Breast Cancer Res*, vol. 16, no. 6, p. 486, 2014.

[101] L.-Y. Geng, Z.-Z. Shi, Q. Dong, X.-H. Cai, Y.-M. Zhang, W. Cao, J.-P. Peng, Y.-M. Fang, L. Zheng, and S. Zheng, "Expression of SNC73, a transcript of the immunoglobulin alpha-1 gene, in human epithelial carcinomas," *World J Gastroenterol*, vol. 13, no. 16, pp. 2305–11, Apr 2007.

[102] A. Gerloff, A. Dittmer, I. Oerlecke, H.-J. Holzhausen, and J. Dittmer, "Protein expression of the Ets transcription factor Elf-1 in breast cancer cells is negatively correlated with histological grading, but not with clinical outcome," *Oncol Rep*, vol. 26, no. 5, pp. 1121–5, Nov 2011.

[103] A. Laliotis, T. Vrekoussis, M. Kafousi, E. Sanidas, J. Askoxilakis, J. Melissas, D. Mavroudis, E. Castanas, and E. N. Stathopoulos, "Immunohistochemical study of pElk-1 expression in human breast cancer: association with breast cancer biologic profile and clinicopathologic features," *Breast*, vol. 22, no. 1, pp. 89–95, Feb 2013.

[104] A. Furlan, C. Vercamer, F. Bouali, I. Damour, A. Chotteau-Lelievre, N. Wernert, X. Desbiens, and A. Pourtier, "Ets-1 controls breast cancer cell balance between invasion and growth," *Int J Cancer*, vol. 135, no. 10, pp. 2317–28, Nov 2014.

[105] C.-Y. Koo, K. W. Muir, and E. W.-F. Lam, "FOXM1: From cancer initiation to progression and treatment," *Biochim Biophys Acta*, vol. 1819, no. 1, pp. 28–37, Jan 2012.

[106] C. Thompson, G. MacDonald, and C. R. Mueller, "Decreased expression of BRCA1 in SK-BR-3 cells is the result of aberrant activation of the GABP beta promoter by an NRF-1-containing complex," *Mol Cancer*, vol. 10, p. 62, 2011.

[107] M. Vilasco, L. Communal, N. Mourra, A. Courtin, P. Forgez, and A. Gompel, "Glucocorticoid receptor and breast cancer," *Breast Cancer Res Treat*, vol. 130, no. 1, pp. 1–10, Nov 2011.

[108] V. Bolós, E. Mira, B. Martínez-Poveda, G. Luxán, M. Cañamero, C. Martínez-A, S. Mañes, and J. L. de la Pompa, "Notch activation stimulates migration of breast cancer cells and promotes tumor growth," *Breast Cancer Res*, vol. 15, no. 4, p. R54, 2013.

[109] L. Yang, Y. Luo, and J. Wei, "Integrative genomic analyses on ikaros and its expression related to solid cancer prognosis," *Oncol Rep*, vol. 24, no. 2, pp. 571–7, Aug 2010.

[110] J. L. Hornick, P. Dal Cin, and C. D. M. Fletcher, "Loss of INI1 expression is characteristic of both conventional and proximal-type epithelioid sarcoma," *Am J Surg Pathol*, vol. 33, no. 4, pp. 542–50, Apr 2009.

[111] J. B. Addison, C. Koontz, J. H. Fugett, C. J. Creighton, D. Chen, M. K. Farrugia, R. R. Padon, M. A. Voronkova, S. L. McLaughlin, R. H. Livengood, C.-C. Lin, J. M. Ruppert, E. N. Pugacheva, and A. V. Ivanov, "KAP1 promotes proliferation and metastatic progression of breast cancer cells," *Cancer Res*, vol. 75, no. 2, pp. 344–55, Jan 2015.

[112] D. Iliopoulos, A. Rotem, and K. Struhl, "Inhibition of miR-193a expression by Max and RXRalpha activates K-Ras and PLAU to mediate distinct aspects of cellular transformation," *Cancer research*, vol. 71, no. 15, pp. 5144–5153, Aug 1 2011.

[113] S. Cunha, Y.-C. Lin, E. A. Goossen, C. I. DeVette, M. R. Albertella, S. Thomson, M. J. Mulvihill, and A. L. Welm, "The RON receptor tyrosine kinase promotes metastasis by triggering MBD4-dependent DNA methylation reprogramming," *Cell Rep*, vol. 6, no. 1, pp. 141–54, Jan 2014.

[114] N. Fujita, D. L. Jaye, M. Kajita, C. Geigerman, C. S. Moreno, and P. A. Wade, "MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer," *Cell*, vol. 113, no. 2, pp. 207–19, Apr 2003.

[115] L.-P. Xu, Y. Sun, W. Li, L. Mai, Y.-J. Guo, and Q.-X. Fan, "MYC and MXI1 protein expression: potential prognostic significance in women with breast cancer in china," *Oncol Res Treat*, vol. 37, no. 3, pp. 118–23, 2014.

[116] H. Shi, M. Bevier, R. Johansson, K. Enquist-Olsson, R. Henriksson, K. Hemminki, P. Lenner, and A. Försti, "Prognostic impact of polymorphisms in the MYBL2 interacting genes in breast cancer," *Breast Cancer Res Treat*, vol. 131, no. 3, pp. 1039–47, Feb 2012.

[117] G. K. Yiu, A. Kaunisto, Y. R. Chin, and A. Toker, "NFAT promotes carcinoma invasive migration through glypican-6," *Biochem J*, vol. 440, no. 1, pp. 157–66, Nov 2011.

[118] J. Eeckhoute, J. S. Carroll, T. R. Geistlinger, M. I. Torres-Arzayus, and M. Brown, "A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer," *Genes Dev*, vol. 20, no. 18, pp. 2513–26, Sep 2006.

[119] M. W. Bronson, S. Hillenmeyer, R. W. Park, and A. S. Brodsky, "Estrogen coordinates translation and transcription, revealing a role for NRSF in human breast cancer cells," *Mol Endocrinol*, vol. 24, no. 6, pp. 1120–35, Jun 2010.

[120] C. B. Moelans, A. H. J. Verschuur-Maes, and P. J. van Diest, "Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer," *J Pathol*, vol. 225, no. 2, pp. 222–31, Oct 2011.

[121] A. Carracedo, D. Weiss, A. K. Leliaert, M. Bhasin, V. C. J. de Boer, G. Laurent, A. C. Adams, M. Sundvall, S. J. Song, K. Ito, L. S. Finley, A. Egia, T. Libermann, Z. Gerhart-Hines, P. Puigserver, M. C. Haigis, E. Maratos-Flier, A. L. Richardson, Z. T. Schafer, and P. P. Pandolfi, "A metabolic prosurvival role for PML in breast cancer," *J Clin Invest*, vol. 122, no. 9, pp. 3088–100, Sep 2012.

[122] Z. Han, L. Tian, J. Zhang, T. Huang, R. Mahiraju, and K. Huang, "Estrogen induced RNA polymerase II stalling in breast cancer cell line MCF7," in *Bioinformatics Research and Applications: 10th International Symposium, ISBRA 2014, Zhangjiajie, China, June 28-30, 2014, Proceedings*, ser. Lecture Notes in Computer Science, M. Basu, Y. Pan, and J. Wang, Eds. Springer International Publishing, 2014.

[123] J. Bai, H.-M. Yong, F.-F. Chen, W.-B. Song, C. Li, H. Liu, and J.-N. Zheng, "RUNX3 is a prognostic marker and potential therapeutic target in human breast cancer," *J Cancer Res Clin Oncol*, vol. 139, no. 11, pp. 1813–23, Nov 2013.

[124] M. Wernicke, P. Roitman, D. Manfre, and R. Stern, "Breast cancer and the stromal factor. the "prometastatic healing process" hypothesis," *Medicina (B Aires)*, vol. 71, no. 1, pp. 15–21, 2011.

[125] L.-M. Kong, C.-G. Liao, Y. Zhang, J. Xu, Y. Li, W. Huang, Y. Zhang, H. Bian, and Z.-N. Chen, "A regulatory loop involving miR-22, Sp1, and c-Myc modulates CD147 expression in breast cancer invasion and metastasis," *Cancer Res*, vol. 74, no. 14, pp. 3764–78, Jul 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2015.2453948, IEEE/ACM Transactions on Computational Biology and Bioinformatics

[126] F. Wu, I. Ivanov, R. Xu, and S. Safe, "Role of SP transcription factors in hormone-dependent modulation of genes in MCF-7 breast cancer cells: microarray and RNA interference studies," *J Mol Endocrinol*, vol. 42, no. 1, pp. 19–33, Jan 2009.

[127] J. Yu, Y. Wang, F. Yan, P. Zhang, H. Li, H. Zhao, C. Yan, F. Yan, and X. Ren, "Noncanonical NF-κβactivation mediates STAT3-stimulated IDO upregulation in myeloid-derived suppressor cells in breast cancer," *J Immunol*, vol. 193, no. 5, pp. 2574–86, Sep 2014.

[128] Y. Zeng, L. Min, Y. Han, L. Meng, C. Liu, Y. Xie, B. Dong, L. Wang, B. Jiang, H. Xu, Q. Zhuang, C. Zhao, L. Qu, and C. Shou, "Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis," *Carcinogenesis*, vol. 35, no. 10, pp. 2244–53, Oct 2014.

[129] D. P. McDonnell, S. L. Dana, P. A. Hoener, B. A. Lieberman, M. O. Imhof, and R. B. Stein, "Cellular mechanisms which distinguish between hormone- and antihormone-activated estrogen receptor," *Ann N Y Acad Sci*, vol. 761, pp. 121–37, Jun 1995.

[130] X. Li, W. Liang, J. Liu, C. Lin, S. Wu, L. Song, and Z. Yuan, "Transducin (β)-like 1 X-linked receptor 1 promotes proliferation and tumorigenicity in human breast cancer via activation of beta-catenin signaling," *Breast Cancer Res*, vol. 16, no. 5, p. 465, 2014.

[131] C.-C. Lee, W.-S. Chen, C.-C. Chen, L.-L. Chen, Y.-S. Lin, C.-S. Fan, and T.-S. Huang, "TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer," *J Biol Chem*, vol. 287, no. 4, pp. 2798–809, Jan 2012.

[132] L. A. deGraffenried, T. A. Hopp, A. J. Valente, R. A. Clark, and S. A. W. Fuqua, "Regulation of the estrogen receptor alpha minimal promoter by Sp1, USF-1 and ERalpha," *Breast Cancer Res Treat*, vol. 85, no. 2, pp. 111–20, May 2004.

[133] J. G. Lieberthal, M. Kaminsky, C. N. Parkhurst, and N. Tanese, "The role of YY1 in reduced HP1alpha gene expression in invasive human breast cancer cells," *Breast Cancer Res*, vol. 11, no. 3, p. R42, 2009.

[134] M. Wan, W. Huang, T. E. Kute, L. D. Miller, Q. Zhang, H. Hatcher, J. Wang, D. B. Stovall, G. B. Russell, P. D. Cao, Z. Deng, W. Wang, Q. Zhang, M. Lei, S. V. Torti, S. A. Akman, and G. Sui, "Yin Yang 1 plays an essential role in breast cancer and negatively regulates p27," *Am J Pathol*, vol. 180, no. 5, pp. 2120–33, May 2012.

[135] E. R. Hrincius, V. Wixler, T. Wolff, R. Wagner, S. Ludwig, and C. Ehrhardt, "CRK adaptor protein expression is required for efficient replication of avian influenza A viruses and controls JNK-mediated apoptotic responses," *Cell Microbiol*, vol. 12, no. 6, pp. 831–43, Jun 2010.

[136] M. M. Whitmore, A. Iparraguirre, L. Kubelka, W. Weninger, T. Hai, and B. R. G. Williams, "Negative regulation of TLR-signaling pathways by activating transcription factor-3," *J Immunol*, vol. 179, no. 6, pp. 3622–30, Sep 2007.

[137] W. Zhu, B. W. Higgs, C. Morehouse, K. Streicher, C. S. Ambrose, J. Woo, G. W. Kemble, B. Jallal, and Y. Yao, "A whole genome transcriptional analysis of the early immune response induced by live attenuated and inactivated influenza vaccines in young children," *Vaccine*, vol. 28, no. 16, pp. 2865–76, Apr 2010.

[138] G. Cannon, M. A. Callahan, J. Q. Gronemus, and R. J. Lowy, "Early activation of MAP kinases by influenza A virus X-31 in murine macrophage cell lines," *PLoS One*, vol. 9, no. 8, p. e105385, 2014.

[139] L. Liu, Z. Cao, J. Chen, R. Li, Y. Cao, C. Zhu, K. Wu, J. Wu, F. Liu, and Y. Zhu, "Influenza A virus induces interleukin-27 through cyclooxygenase-2 and protein kinase A signaling," *J Biol Chem*, vol. 287, no. 15, pp. 11 899–910, Apr 2012.

[140] N. Harii, C. J. Lewis, V. Vasko, K. McCall, U. Benavides-Peralta, X. Sun, M. D. Ringel, M. Saji, C. Giuliani, G. Napolitano, D. J. Goetz, and L. D. Kohn, "Thyrocytes express a functional toll-like receptor 3: overexpression can be induced by viral infection and reversed by phenylmethimazole and is associated with Hashimoto's autoimmune thyroiditis," *Mol Endocrinol*, vol. 19, no. 5, pp. 1231–50, May 2005.

[141] X. Ge, M.-A. Rameix-Welti, E. Gault, G. Chase, E. dos Santos Afonso, D. Picard, M. Schwemmle, and N. Naffakh, "Influenza virus infection induces the nuclear relocalization of the Hsp90 co-chaperone p23 and inhibits the glucocorticoid receptor response," *PLoS One*, vol. 6, no. 8, p. e23368, 2011.

[142] Y. Zhang, J. M. Reynolds, S. H. Chang, N. Martin-Orozco, Y. Chung, R. I. Nurieva, and C. Dong, "MKP-1 is necessary for T cell activation and function," *J Biol Chem*, vol. 284, no. 45, pp. 30 815–24, Nov 2009.

[143] D. Savitsky and K. Calame, "B-1 B lymphocytes require Blimp-1 for immunoglobulin secretion," *J Exp Med*, vol. 203, no. 10, pp. 2305–14, Oct 2006.

[144] W. Li, G. Wang, H. Zhang, D. Zhang, J. Zeng, X. Chen, Y. Xu, and K. Li, "Differential suppressive effect of promyelocytic leukemia protein on the replication of different subtypes/strains of influenza A virus," *Biochem Biophys Res Commun*, vol. 389, no. 1, pp. 84–9, Nov 2009.

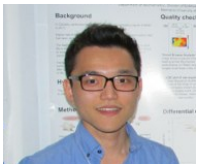
[145] D. Bussfeld, M. Bacher, A. Moritz, D. Gemsa, and H. Sprenger, "Expression of transcription factor genes after influenza A virus infection," *Immunobiology*, vol. 198, no. 1-3, pp. 291–8, Dec 1997.

[146] D. Barbier, I. Garcia-Verdugo, J. Pothlichet, R. Khazen, D. Descamps, K. Rousseau, D. Thornton, M. Si-Tahar, L. Touqui, M. Chignard, and J.-M. Sallenave, "Influenza A induces the major secreted airway mucin MUC5AC in a protease-EGFR-extracellular regulated kinase-Sp1-dependent pathway," *Am J Respir Cell Mol Biol*, vol. 47, no. 2, pp. 149–57, Aug 2012.

[147] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, "A census of human transcription factors: function, expression and evolution," *Nature reviews.Genetics*, vol. 10, no. 4, pp. 252–263, Apr 2009.

[148] A. Smallwood and B. Ren, "Genome organization and long-range regulation of gene expression by enhancers," *Current opinion in cell biology*, vol. 25, no. 3, pp. 387–394, Jun 2013.

Pranjal Patra is currently a graduate student in the Department of Computer Science at Memorial University of Newfoundland. He completed his Bachelor's in Computer Engineering from Bharati Vidyapeeth University in July 2007. During his graduate studies under the supervision of Dr. Lourdes Peña Castillo in the field of Bioinformatics, his main research interests were focused on pathway regulation and enrichment analysis.



Tatsuo Izawa is currently studying biology and physics at Memorial University in St. John's, NL, Canada. His research interests lie in the interdisciplinary fields of biology, physics, and computer science, all of which deal with "big data" that leads to scientific discovery.

Lourdes Peña Castillo obtained her bachelor's degree in Information Systems Engineering at the Technological Institute of Monterrey (ITESM) in Mexico, her Master's degree in Computer Science at the University of Alberta in Canada, and her PhD at the Otto-von-Guericke University Magdeburg in Germany. She continued with a postdoc in bioinformatics in the Banting and Best Department of Medical Research at the University of Toronto. She joined Memorial University in 2010 as an assistant professor in bioinformatics jointly appointed to the Departments of Computer Science and Biology with a cross-appointment to the Discipline of Genetics.