Research Article

The Transcriptional Network Structure of a Myeloid Cell: A Computational Approach

Jesús Espinal-Enríquez,[1,2] **Daniel González-Terán,**[1] **and Enrique Hernández-Lemus**[1,2]

[1]*Computational Genomics Division, National Institute of Genomic Medicine, 14610 México City, Mexico*
[2]*Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 México City, Mexico*

Correspondence should be addressed to Enrique Hernández-Lemus; ehernandez@inmegen.gob.mx

Understanding the general principles underlying genetic regulation in eukaryotes is an incomplete and challenging endeavor. The lack of experimental information regarding the regulation of the whole set of transcription factors and their targets in different cell types is one of the main reasons to this incompleteness. So far, there is a small set of curated known interactions between transcription factors and their downstream genes. Here, we built a transcription factor network for human monocytic THP-1 myeloid cells based on the experimentally curated *FANTOM4* database where nodes are genes and the experimental interactions correspond to links. We present the topological parameters which define the network as well as some global structural features and introduce a relative inuence parameter to quantify the relevance of a transcription factor in the context of induction of a phenotype. Genes like ZHX2, ADNP, or SMAD6 seem to be highly regulated to avoid an avalanche transcription event. We compare these results with those of *RegulonDB*, a highly curated transcriptional network for the prokaryotic organism *E. coli*, finding similarities between general hallmarks on both transcriptional programs. We believe that an approach, such as the one shown here, could help to understand the one regulation of transcription in eukaryotic cells.

1. Introduction

Gene regulation is a key player in the development of living systems. Interactions amongst genes are critical to direct tissue-specific gene expression [1]. A dysfunctional process results in altered physiology, giving rise to malformations and diseases such as cancer. The gene regulatory networks that control gene expression are usually composed by several thousands of genes which are transcribed and translated to produce proteins that have a function in the cell. On the other hand, a reduced set of genes, called *transcription factors* (TFs), has the role of regulating the transcriptional program of the cells. These TFs enhance or repress the expression of other genes, which may be also TFs, or else *target genes*.

By looking at the number of genes that a particular TF could regulate, a hierarchical structure can be observed. In this sense, genes like MEF2C, mTOR, MYB, FOXM1, GATA3, FOXP3, BCL6, MNDA, POU2AF1, MEF2C, or SMAD3 have been reported to potentially regulate more than one thousand genes, which are transcriptional master regulators (TMRs) [2, 3]. The hierarchical character, observed in TF-driven regulatory networks, is likely due to the number of target genes that a TF may have. Another issue related to the relevance of such TFs is whether a given TF is regulated or not by another TF, in which case, a shadow effect could appear, indicating the primacy of a given TF above others [4, 5]. Deregulation of these TFs has been related to the development of cancer and other diseases [2].

Comprehensive studies regarding the extent of influence of TFs and the targets of the TFs have been made [6]. In such studies, phenomena like feedback, feed-forward loops, and other biological motifs [7] have been found. The said phenomena are indicative of a sophisticated machinery involved in the regulation of gene transcription: highly connected TFs are also regulated by others, which may (or may not) be highly connected. Hence, the importance of TFs in the whole regulatory program lies not only in the out-degree (the number of genes regulated by the TF) but also in the in-degree
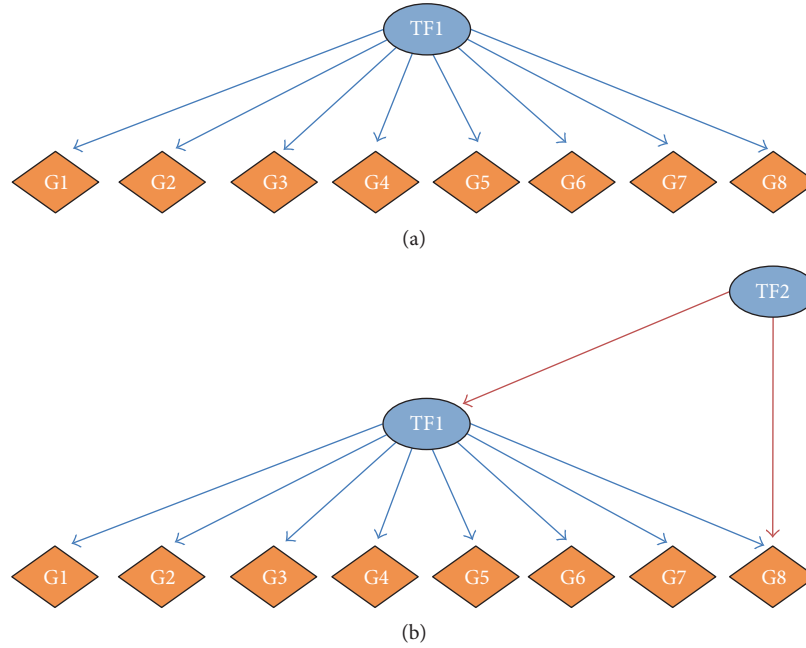
FIGURE 1: Representative schemes of TFs with high out-degree. (a) The TF1 is regulating its 8 target genes and it is not regulated by others. (b) The TF1 is regulating 8 targets, but at the same time, it is being regulated by TF2, which in turn is also regulating other target genes (G8).

(how many genes regulate the TF) (Figure 1). It is precisely the interplay of in- and out-degree distributions which shapes ultimately the delicate mechanisms of gene expression control, encoded in the topological structure of the transcriptional regulatory network.

The discovery of general patterns of cooperativity and coregulation for the transcriptional regulation program of eukaryotic cells will help to understand the control of genome-wide expression and how it influences the establishment of phenotypes. To this end, here, we develop a systematic strategy for the data mining curation of the whole set of interactions in *FANTOM4 Edge Express*: a comprehensive and authoritative catalog of transcriptional regulatory interactions in THP1 myeloid cells [1, 8].

We constructed a gene regulatory network for the whole set of genes and their interactions by developing a mining tool for the FANTOM4 database, adding each gene present in the said database as a node to a network, where the strength of connections between nodes corresponds to the intensity, or confidence, of the experimental interaction between any pair of genes. We obtained a network consisting in 9090 genes and 234,913 links. This transcriptional regulatory network contains a number of genes influencing many downstream genes, that is, a few genes are able to control the majority of the genes. As previously mentioned, those TFs are also regulated by other TFs. Taking this into account, we constructed a TF subnetwork, which contains 295 TF genes and 8483 links.

Acknowledging the relevance that in-degree and out-degree of TFs acquire in the context of gene regulation, we devise a *relative influence* parameter (see the corresponding subsection in Materials & Methods), related with the ratio of out-degree/in-degree of all TFs present in the network,

thus indicating the relative influence of the gene (and its regulators) in terms of how many targets the said gene is regulating and the number of genes which directly regulate it.

We finally present a network visualization highlighting the importance of those genes. Genes like NRF1, SPIB, GABPA, or TFAP2A present a high RI since they do not have regulators within the database. This could be indicative of a basal and transcription-independent level of regulation, being these genes a kind of default transcriptional regulators. On the other hand, genes like ZHX2, ADNP, or SMAD6 present a very high in-degree, while their out-degree is relatively low. Considering this, we argue that these genes are also relevant, acting as locks avoiding a *transcriptional avalanche*, that is, they can allow the transcription of genes related to rapid cell division or give place to differentiation of the cell. The lack of regulation of them could lead to diseases such as cancer or malformation during development.

As mentioned before, the network also contains experimentally obtained weighted values ($S_{i,j}$) of the interactions between any pair of genes which indicates whether the effect of a gene over any other connected to it is positive (activator) or negative (repressor), that is, there are transcription factors whose "preferential" activity is either of an activating or repressing nature. This fact may follow from the very physicochemical structure of the related protein, its interactions with other proteins and with certain regions in the genome that include, but are not restricted to, the promoter regions on their target genes. Of course, these functions are phenotype dependent and may be even context dependent.

In terms of transcriptional regulation, a positive interaction represents that the source gene (TF) promotes the

transcription of the target gene (effector gene). Analogously, a negative interaction will represent that the TF inhibits the transcription of the target gene. The algebraic sum of all interactions of each gene allows us to observe that some TFs are mostly activators and others are mostly repressors.

In the first case, we have for instance NFYA, NFYB, and NFYC; whereas in the latter case, we have TFAP2A, TFAP2C, and MAZ. Arguably, regulation of those genes is fundamental to initiate or terminate a transcriptional cascade, likely related to events of differentiation or cell division. Experiments along this way are necessary. We argue that this framework may help to understand the importance of TFs in terms of the ratio between in- and out-degree, which could give new insights regarding the gene regulation.

2. Materials and Methods

Reproducibility of results and methodological clarity are fundamental in all scientific endeavors, but in the case of computational biology approaches, they gain even more relevance. In the present section, we will present detailed accounts of the methods used here, and in some cases (most notably, when we introduce novel concepts), we will even write down detailed calculations. Further details and custom-made computer code for this project are available at the following link: https://github.com/CSB-IG/fantomine.

2.1. Network Construction. We mined the FANTOM4 database [8] to construct a gene regulatory network. This database is based on genome-wide dynamics of transcription start site usage in the PMA-stimulated human monocytic cell line THP-1. To this end, we made a systematic approach to search all genes present in the database. We also tracked the genes which regulate the first one as well as the genes regulated by the searched gene. If a gene has no regulators, its in-degree is zero. Analogously, if the searched gene does not have target genes, its out-degree would be zero.

An important source of information about this network is the strength of the interaction between any pair of genes. This interaction is quantified by both experiments and sequence-based transcription factor binding site (TFBS) predictions. The range of intensities $S_{i,j}$ is a Z-score whose dynamic range goes from −10 to 12. With this experimentally obtained interaction value, we constructed a weighted and directed network of the FANTOM4-based gene regulatory network.

The network was built by using a specially devised crawler, following a breadth-first search strategy to find the connections between all genes (promoters and targets). The algorithm begins with a random gene and began to *walk* in the FANTOM EDGE-DB; for each promoter and target of that gene, it adds them to a queue of genes to explore and to a dictionary of the genes explored. If the current gene is new in the dictionary, it adds it to the database and the interaction (if it is a promoter or target); otherwise, it only verifies if the interaction is new or updates the weight (it only takes the biggest value). Since some of the predicted TFBS have different values for the same target

depending on the promoter region, we decided to use the highest value for all interactions (for further details regarding the interaction values, please see [8] and the FANTOM4 website: http://fantom.gsc.riken.jp/4/).

As an example of the last sentence, the SRF gene is regulated by SOX2, but this regulation have three different values in the FANTOM database, depending on the promoter region: 4.857, 0.834, and 0.845. We decided to establish a link with the highest value assuming that that interaction is the most plausible in an ideal context.

We also calculated different node centrality measures of the resulting network, such as their in-degree and out-degree distributions. The whole network depicted in Figure 2 is visualized by using Cytoscape v.3.2.2.

2.2. Relative Influence. To have a useful measure to retrieve information regarding the regulatory balance of each gene in the context of transcriptional process, we establish a parameter of *relative influence* RI_n for each gene n in the network. This parameter reflects the fact that there are some gene regulators that control the transcription of many targets but are in turn regulated by many genes, while other regulators may possess a smaller number of targets but are also regulated by fewer genes and thus may be of similar relative influence on the general transcriptional regulatory program. The RI_n is then obtained as follows:

$$RI_n = \frac{outDeg_n}{inDeg_n + outDeg_n} - \frac{\epsilon + inDeg_n}{\epsilon + outDeg_n}, \qquad (1)$$

where $outDeg_n$ and $inDeg_n$ are the out-degree and in-degree of the gene n, respectively. ϵ is a small variable (10^{-3}) to avoid division by zero. We define all negative RI_n values to be set equal to 0.

For the sake of clarity and result reproducibility, we will outline an explicit calculation as follows: For instance, the gene GABPB2 has 232 regulators (in-degree, inDeg) but 2829 targets (out-degree, outDeg) in the database. Hence, RI_{GABPB2} is calculated as follows:

$$RI_{GABPB2} = \frac{2829}{232 + 2829} - \frac{0.001 + 232}{0.001 + 2829} = 0.842196754.$$
$$(2)$$

Since most genes have zero out-degree, that is, they have no target genes, we constructed a subnetwork which contains only TFs; those are, in general terms, the genes which define the way in which the network is regulated.

2.3. The Algebraic Sum of Interactions. Arguably, the interaction strength ($S_{i,j}$) of the links in a transcriptional network is an abstract and important parameter to quantify the activity of a TFBS recognition motif on a given TF against all its potential targets, providing information about TF interactions at a gene-by-gene basis. This fact may mask the importance of certain TFs in the transcriptional regulatory programs of the cell.

In order to categorize the importance of different TFs in this network, this value ($S_{i,j}$) can be used. A TF may be a transcriptional activator or repressor or a combination of both
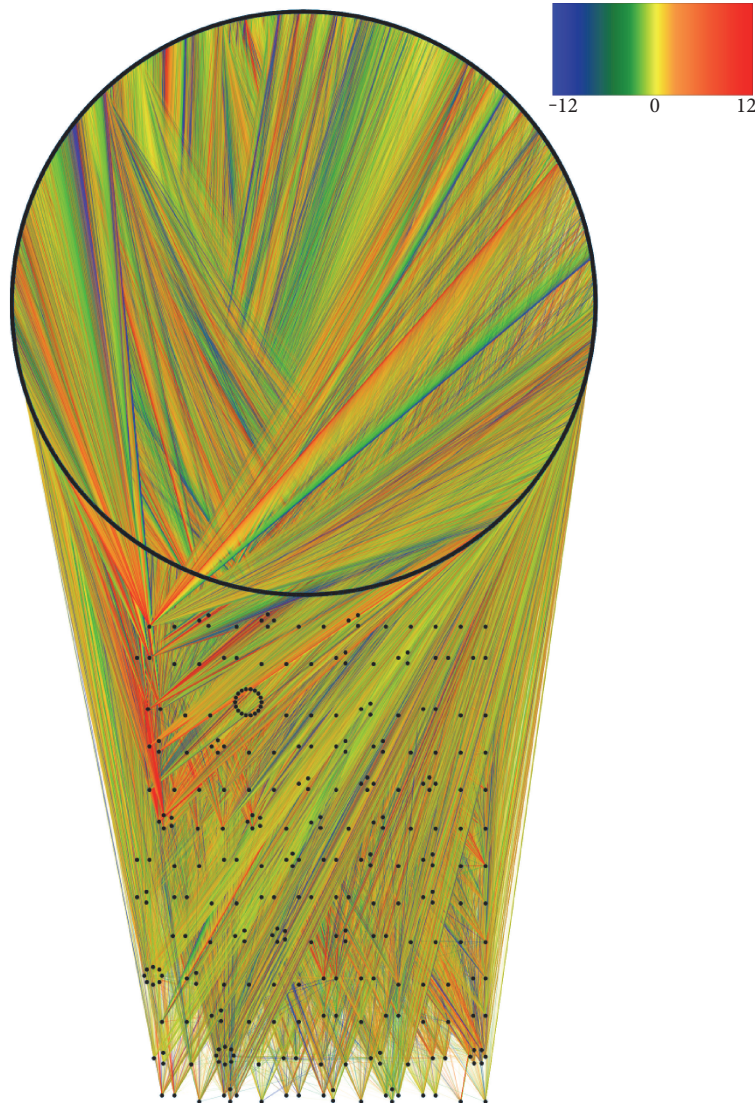
FIGURE 2: Transcription factor network constructed from the *FANTOM4 Edge Express* database. In this representation, nodes are represented by black dots. The color of the links depends on the value of the interaction: blue and green are repressions; yellow, orange, and red are for activations. The circle on top of the figure corresponds to target genes, that is, their out-degree = 0.

(for different targets and/or under different circumstances). To clarify this idea, we classify the TFs in terms of the total value of their interaction strength, that is, we sum all $S_{i,j}$ for each TF_i and observe if the value was positive or negative, under the assumption of an additive linear model. A positive value means the majority of their interactions are positive; thus, TF_i is an *overall activator*. On the other hand, a negative value indicates that the majority of their interactions are for repression; TF_i is acting then mostly as an *overall repressor*. As we have already stated, such terms are context and phenotype dependent, so that a TF whose main function is that of an activator in the context of myeloid cells, it well may be a repressor on a different cell type or cellular context.

Again, as an example, we will outline the explicit calculation for one instance of algebraic sum: TFAP2B has 687 negatively regulated targets and 965 positively regulated ones. The algebraic sum of all their targets is as follows:

$\sum_j S_{\text{TFAP2B},j} = 965 - 687 = 278$, which converts TFAP2B in a strong *overall activator*.

3. Results

3.1. Transcription Factor Network. The TF network is shown in Figure 2. 9090 genes with 234,970 interactions are depicted. From that network, we can observe several facts: The network is depicted ranking genes according to the *relative influence*. Upper part genes have a higher RI, meanwhile the lowest RIs are on the lower part. The black circle at the upper part of the network represents those genes with zero out-degree, that is, effector genes. Those genes have RI = 0 by definition. The interactions between genes are weighted according to their respective interaction strength and colored accordingly with a continuous scale, based on the *Z*-scores ($S_{i,j}$) for activity expression correlation: this way, the blue and green lines indicate inhibition from the source gene

TABLE 1: The top 10 highest RI values of TFs in the FANTOM network. The last column shows the average gene expression as given by the number of transcript counts (copy number) at $t = 0$, for each gene.

Gene	In-degree	Out-degree	RI	Avg. transcript count
NRF1	0	2175	0.99999954	1628.215
SPIB	0	1778	0.999999438	316.215
TFAP2A	0	1721	0.999999419	293.865
MYOD1	0	1713	0.999999416	0.02
TFAP2B	0	1652	0.999999395	0.03
ARNT2	0	1627	0.999999385	0.01
SNAI2	0	1543	0.999999352	7.485
MYOG	0	1435	0.999999303	203.985
MYF5	0	1435	0.999999303	0.985
MYF6	0	1435	0.999999303	0.265

TABLE 2: The bottom 10 lowest RI values of TFs in the FANTOM network. The last column shows the average gene expression as given by the number of transcript counts (copy number) at $t = 0$, for each gene.

Gene	In-degree	Out-degree	RI	Avg. transcript count
RUNX1	155	556	0.517749925	9382.91
ARID5B	114	392	0.49206167	12.885
TFAP4	190	632	0.483369442	642.65
POU6F1	87	268	0.442040072	229.68
ZIC2	135	407	0.438502347	10132.46
GTF2A1	100	253	0.357562642	1219.94
POU3F1	80	198	0.331108025	4.585
SMAD6	175	408	0.289214257	2936.75
ADNP	132	292	0.256880007	651.685

to the target, whereas yellow and red lines represent activation from the source gene to the target (see color scale inset in Figure 2).

We can see, for instance, how a somewhat small set of molecules is responsible for the concerted regulation of the whole transcriptional activity of the cells. The also important fact is that these molecules carry on their regulatory function by jointly regulating themselves. In this network, genes with the highest out-degree are SP1, MAZ, ELF4, ELF1, ELF2, SPI1, ELK1, ELK4, GABPA, and GABPB2. The subnetwork containing only the targets of those genes has 140,318 interactions with 7913 out of the total of genes.

Regarding the RI, the top10 out-degree genes (which are in fact TFs) have high values; however, they do not have the highest ones, since they are also regulated by other TFs. Table 1 shows those nodes with the highest RI, as well as their experimentally observed expression values at the starting of the experiments. The expression values for the whole time series are provided as a Supplementary material available online at https://doi.org/10.1155/2017/4858173.

As revealed by the RI parameter, TFs with the lowest values have a high out-degree, but at the same time, they present a high in-degree, such is the case of ZHX2, ADNP, SMAD6, POU3F1, GTF2A1, ZIC2, POU6F1, TFAP4, ARID5B, or RUNX1. In all cases, they have (at least) twice more targets than regulators. Table 2 shows the lowest RI genes.

The *relative influence* is a simple metric which provides relevant information about the regulatory activity of a transcription factor. Interestingly, even when a single instance of RI is not enough to give insight on coregulation and gene expression patterns, the full, genome-wide distribution of RI values does it so. A closer look at the top RI molecules as presented in Tables 1 and 2 unveiled interesting patterns. For instance, amongst the TF genes that are not under regulation (at this level of description, of course, see Table 1), we can find thousands of gene targets.

If we recall that the total number of genes in this study is about 9000, then having TFs regulating around 2000 of these is indeed a powerful indicator of strong coregulation. On the

other hand, it is also noticeable that the TF with the lowest RI (ZHX2) although being tightly regulated (with a total of 151 regulators) is able to participate in the regulation of 292 genes, evidencing the cooperative effect in TF regulation.

3.2. Transcription Factor Subnetwork. After eliminating all the effector genes (those genes with zero out-degree), the network is drastically reduced (295 genes and 8483 links between them). From this network, it can be observed that some genes are mainly activators meanwhile other TFs are inhibitors. This global activating/inhibiting nature of TFs will be discussed below. Regarding the structure of the subnetwork, it is interesting that several genes are highly regulated, even though they are transcription factors.

We can also see that by considering only the network formed by coregulated transcription factors, it is possible to unveil the presence of modules conformed by groups of TFs that not only regulate other target genes (absent in this network visualization) but also regulate each other. This phenomenon of regulatory modularization is an example of a phenomenon that cannot be seen from the FANTOM4 database consulted on a gene-by-gene basis but provides new biological insight on the whole genome regulatory patterns in this cell lineage.

3.3. Some TFs Are Overall Repressors Meanwhile Others Are Activators. Taking into account the overall activating or inhibiting nature of TFs (in the context of the cell type and phenotype under consideration), we search in the network those genes with extreme positive and negative values of the sum of interaction strengths ($S_{i,j}$) for each TF: They are the most important activators as well as the most important repressors of the transcriptional program. In Figure 3, we show the top 3 activators and also the top 3 repressors along with their targets. Those genes are NFYB, NFYC, and NFYA for activators and MAZ, TFAP2C, and TFAP2A for repressors. The interplay between them may be of importance in shaping the phenotype. Regarding the MAZ gene, it has also a high RI (0.984): This means that it regulates thousands of genes but it is regulated by just a few. It mainly represses the genes that it regulates.
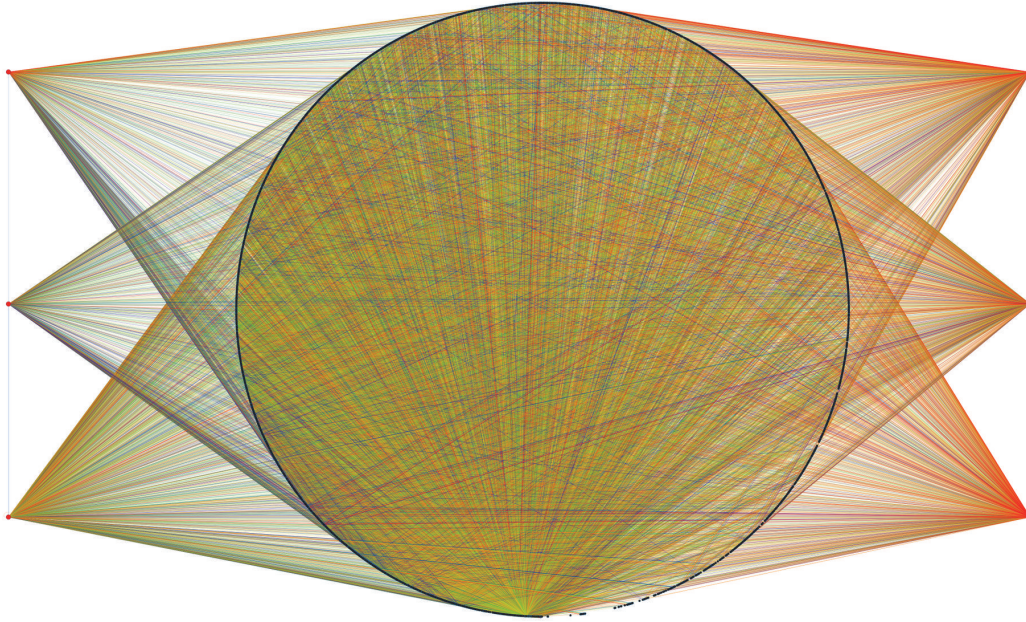
Figure 3: Top 3 overall activating TFs and top 3 overall repressing TFs. On the left side, the three genes which have more inhibiting links—TFAP2C, TFAP2A, and MAZ—are depicted. At the right side, the top 3 activating genes are shown: NFYA, NFYB, and NFYC. It can be observed that the predominance of the green and blue links at the left, meanwhile on the right side, the orange and red links, are more frequent. In the center of this network, we can see the first neighbors of those 6 genes. Links are colored as in Figures 2 and 5.
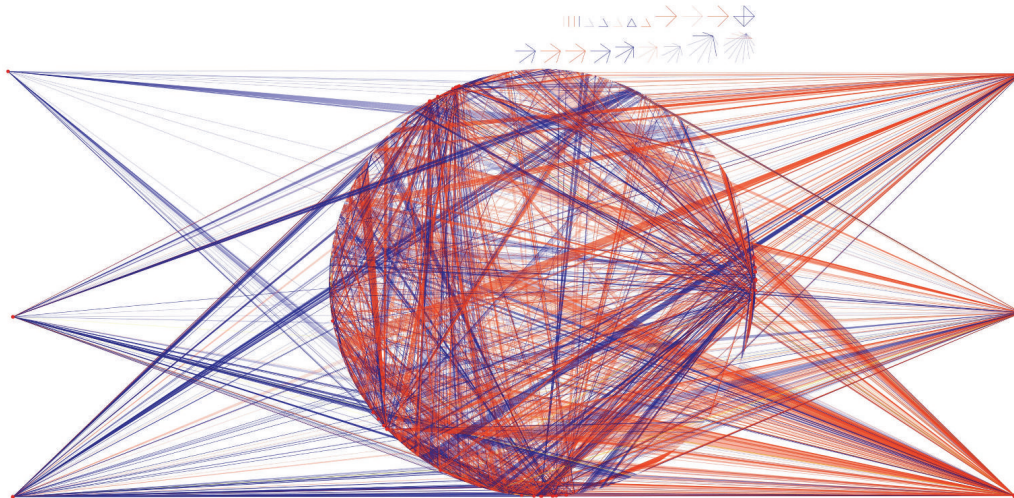

Figure 4: Top 3 overall activating TFs and top 3 overall repressing TFs in the *E. coli* network. On the left side, the three genes which have more inhibiting links—Crp, Fnr, and Fis—are depicted. At the right side, the top 3 activating genes are shown: Narl, Phob, and Lrp. It can be observed (analogous to the human transcriptional network of Figure 5) that the predominance of the blue links at the left, meanwhile on the right side the red links, are more frequent. In the center of this network, the first neighbors of those 6 genes are shown. It is also possible to observe some small disconnected components at the upper part of the figure.

3.4. The Transcriptional Network Structure of a Prokaryotic Cell: A Comparison. The topological and functional features, revealed in the transcriptional network of the myeloid cell with the approach performed here, were compared with a prokaryotic model, in order to provide a deeper understanding of the transcriptional regulatory program at the genome-wide level. The prokaryotic network architecture was obtained with the same methodology that we used for the

eukaryotic model, but we obtained information about the genetic interactions from RegulonDB [9], a rigorously curated database of a genome-wide transcriptional regulation in the bacteria *Escherichia coli*. This database contains information regarding the type of interaction between genes, positive, negative, or dual, as well as the direction of the said interactions. We decided to analyze the whole transcriptional network, observe the number of transcription factors,
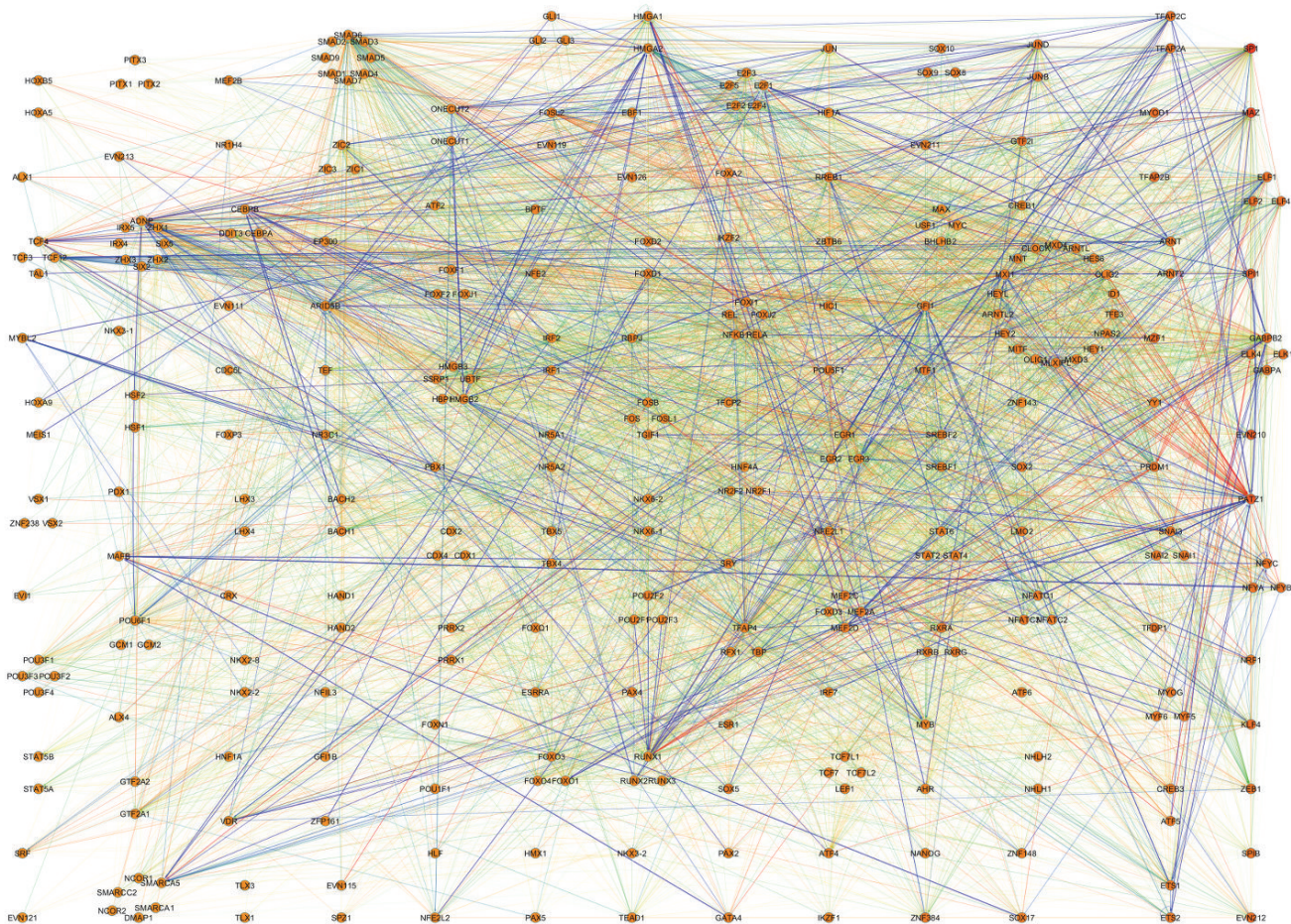
FIGURE 5: Transcription factor subnetwork. In this graph, the 298 TFs of the FANTOM4 network are depicted. Genes are sorted according to the gene out-degree (top left-to-bottom right); set of nodes describing a circle has the same out-degree. The color code of links is the same as in Figure 2.

calculate the RI for each gene, and obtain the overall activators and repressors.

This directed network contains 1988 genes and 4414 interactions. 202 of those genes correspond to transcription factors. In Figure 4, we show the transcriptional network of *E. coli*, depicting the links according to its type: blue for inhibition and red for activation. The top 3 overall activators and repressors are separated to illustrate that the general effect that these transcription factors exert on their targets is analogous to Figure 5. Crp, Fnr, and Fis genes are the top 3 repressors, meanwhile Narl, Phob, and Lrp correspond to the top 3 overall activators. Table 3 shows the top 5 and bottom 5 RI genes of the transcriptional network.

We also constructed the TF-only network for *E. coli*. This network is depicted in Figure 6. As it can be observed, coregulation of those TFs is evident.

4. Discussion

In this work, we developed a systematic search of the FANTOM4 database, which took into account several experimentally proven interactions between genes and transcription

TABLE 3: The top 5 highest and lowest values of TFs in the RegulonDB *E. coli* network.

Gene	In-degree	Out-degree	RI
Ihf	0	249	0.999999598
H-ns	0	186	0.999999462
NsrR	0	83	0.999998795
Flhdc	0	80	0.99999875
Narp	0	65	0.999998462
Uvry	1	2	0.166641668
Gadx	15	29	0.141847865
Dpia	6	11	0.101600146
Glcc	4	7	0.064928943
Gutm	4	7	0.064928943

factors. We constructed a network where the genes are nodes and their interactions correspond to links. The network is directed because it takes into account transcription factor binding sites to establish a relationship in which the TF is regulating to another gene (which could actually be another
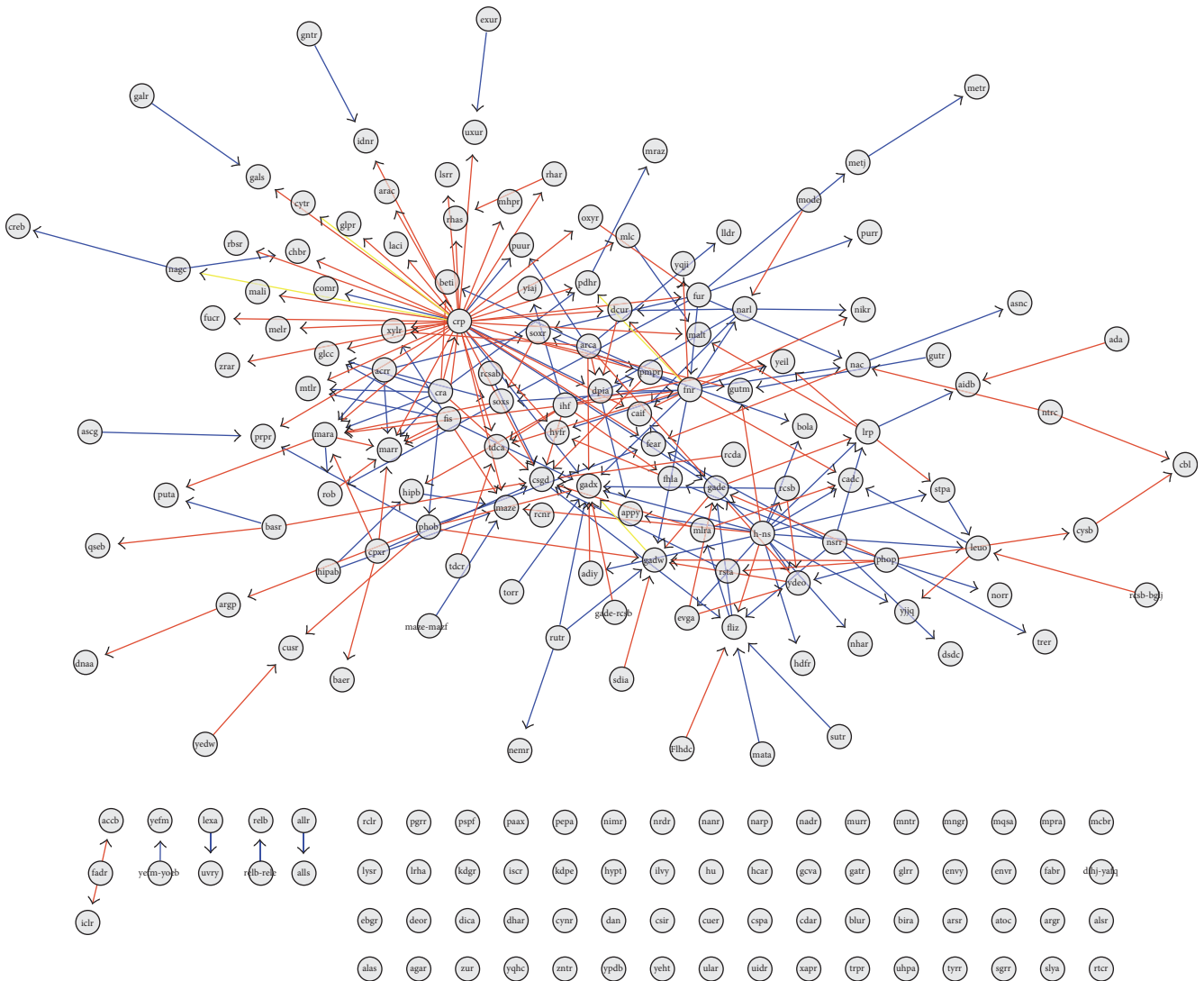
FIGURE 6: Transcription factor subnetwork. In this graph, the 203 TFs of the *E. coli* RegulonDB network are depicted. The color code of links is the same as that in Figure 4. As it was observed in the whole network, in this TF subnetwork, some disconnected components appear.

transcription factor). We found several nodes which do not regulate other genes, but they are highly regulated. On the other hand, there is a small group of TFs which are regulating several other genes, but they are not regulated by any gene. This could allow us to hypothesize that those genes are master regulators because they could be inducers of particular phenotypes, as is the case for NRF1, SPIB, or TFAP2A.

4.1. The Highest RI Genes May Have Crucial Roles in Cell Specificity. Nuclear respiratory factor 1 (NRF1) [10, 11] activates the expression of crucial metabolic genes related to responses to oxidative stress, endoplasmic reticulum stress, xenobiotic stress, and inflammation [12–14]. NRF1 has also been found misregulated in different carcinomas [15–18]. Despite NRF1 being a key player in the induction of several functions in the cell, there are no reports of transcriptional regulators of NRF1. It is regulated at the translational [19] and posttranslational [19–21] levels. This information reinforces our findings regarding the transcriptional

independence of NRF1 as well as the relevance that it acquires in the context of the regulatory network.

The second highest RI belongs to the SPiB gene. The spleen focus-forming virus integration site [22] gene is part of the ETS transcription factor family. It is involved in differentiation [23], immune response [22, 24], apoptosis [25], and activation of early viral expression [26], amongst others. Finally, the third highest RI gene is TFAP2A. This transcription factor is known as a tumor suppressor gene. Decreased expression of this gene has been related to many neoplasms [27–29], as well as other diseases, such as the brachio-oculo-facial syndrome [30, 31].

By observing the relevance of these three genes in the maintenance of correct cell behavior, it is interesting to observe that there are no regulators in our network for those genes. It could be due to the fact that the experimentally curated network is not complete or the available information regarding the regulators is still under construction. However, it is remarkable that the high number of targets contrasting

with the number of regulators of them is in instances such as this, where hierarchy acquires relevance. It is worth mentioning that this is a transcriptional network. We are not looking at posttranslational modifications, where these genes could be regulated. However, the transcriptional relevance of them is evidenced by this approach.

4.2. Misregulation of the Lowest RI Genes Is Associated with Several Malformations.
On the other hand, the three lowest RI genes, ZHX2, ADNP, and SMAD6 are highly regulated; regulation of them is probably more critical in order to avoid transcriptional avalanches. Let us recall that these genes, even though they are highly regulated (higher in-degree counts), are still active transcription factors (see Table 2). In the same sense, mutations of these genes could be involved in the development of several and important diseases. For instance, ZHX2 is considered as a transcriptional repressor [32], which binds the promoter regions, thus regulating transcription of their target genes. This TF also suppresses glypican 3 (GPC3) transcription [33]. Furthermore, ZHX2 has been found as a tumor suppressor [34]. This gene is transcriptionally suppressed by MSX1 and XBP1 [35]. This downregulation is crucial for progression of Hodgkin lymphoma. Concomitantly, there are at least 40 transcription factor binding sites downstream ZHX2 gene [35] which regulate its expression.

The second lowest RI belongs to the ADNP gene, which is essential for brain formation and correct neural development [36]. Mutations in this gene have been related to neuronal disorders such as autism [37, 38], Alzheimer's disease [39], or schizophrenia [40]. Moreover, ADNP interacts with HP1 regulating chromatin remodeling during embryogenesis [41]. Total absence of ADNP is lethal, thus indicating its crucial role in the regulation of transcriptional programs.

Finally, SMAD6 is a signal transducer that modulates multiple signaling pathways, such as the BMP and TGF-beta/activin signaling [42], erythropoiesis [43], or cell cycle [44]. Incorrect regulation of this gene is related to lung adenocarcinoma [42, 45], oral squamous cell carcinoma [46], ovarian cancer [47], or cardiovascular malformation [48].

The RI, introduced here to study the network, remarks the relevance of TFs with the highest values in terms of its master role as transcription factors (since they have no upstream transcriptional regulators); furthermore, this value unveils the importance of the lowest scored genes: those genes need to be highly regulated in order to maintain its correct transcriptional behavior. As it has been observed experimentally, misregulation of the lowest RI genes may result in lethal phenotypes or cancer.

The approach developed here may help to understand genomic regulation. The RI parameter introduced provides insight regarding the influence and importance of the genes in the context of maintenance of a particular phenotype.

4.3. Overall Activators and Repressors May Define Transcriptional Programs.
Finally, the activation or inhibitory nature of TFs is an important field of investigation. With an approach such as the one presented here, research can be guided to unveil the overall nature of certain TFs, by observing whether its overall effect is activation or inhibition. A TF, which positively regulates hundreds of downstream targets, could induce a particular phenotype by activating the said targets. On the contrary, if the majority of the targets is inhibited by the TF, this gene could prevent or stop a particular transcriptional program.

The following TFs are the top 3 overall activators. Interestingly, this set corresponds precisely with the so-called NF-Y complex: NFYA, NFYB, and NFYC, whose role is to bind a sequence in DNA to start the transcription process. They are involved in several basal activities, such as expression of human proteasome genes [49], transcriptional cascade via CDCA8 gene [50], and remodeling of chromatin [51]. Furthermore, the NFY complex has been associated with the coexpression of other TFs to start transcriptional cascades [52]. However, the separated subunit NFYB also can be an inhibitor of DNA topoisomerase II-α [53], indicating that the role of this complex is not restricted to be an activator, revealing thus another aspect of the complexity of the transcriptional program in eukaryotic cells.

As mentioned above, the lowest RI gene in the network is ZHX2. This TF interacts with the A subunit of nuclear factor-Y (NFYA). For instance, ZHX2 represses activation of MDR1 transcription mediated by NFYA [54]. Interestingly enough is the fact that during liver carcinoma, the normal transcriptional program of ZHX2 is highly altered. ZHX2 represses NFYA during liver carcinoma [54]. Taking into account the fact that NFYA is the most important overall activator in the network, we can argue that under a repression of NFYA via ZHX2, the consequence will be a general inhibition of the transcriptional program, which may result for instance, in the progression of liver carcinoma. This example highlights how crucial is the correct control of interconnections in this transcriptional network.

The following are the top 3 overall inhibitor genes and might be involved in the control of transcription by avoiding anomalous events: MAZ, TFAP2A, and TFAP2B. The Myc-associated zinc-finger protein (MAZ) was identified as participating in breast cancer cells by interacting with SAF-1 and inducing transcription of Ras [55]. MAZ also has a role in prostate cancer [56] by interacting with the androgen receptor. On the other hand, TFAP2A, B genes, known tumor suppressor genes, participate in the reduction of glioma progression, by downregulation of Bcl-xl, Bcl-2, c-IAP2, and survivin [27]. These genes have been encountered decreased in several types of cancer, glioma [57], prostate cancer [56], breast cancer [58], or testicular cancer [59]. This is a clear example that the absence of its transcriptional inhibition generates dramatic changes in phenotype.

4.4. Similarities with the Prokaryotic Cell Transcriptional Program.
Regarding the RI in the E. coli network, the higher one, Ihf (integration host factor), plays a crucial role in the survival of the cell, induction of acid resistance, and expression of several other factors [60–62]. H-ns acts on DNA binding of RNA polymerase [63]. The NSrR gene is a major transcriptional repressor in response to iron and also a negative regulator of motility [64–66]. On the other hand, the lowest RI gene is gutM, a crucial transcription factor involved

in the phosphotransferase system [67]. Mutations on those genes have a direct impact on the resulting phenotype.

To highlight the importance of the implications that the concepts of RI and the overall activator/repressor may have in the regulatory program, we provide a functional comparison of these measures in the prokaryotic transcriptional network; for this purpose, we can observe (for instance) the Fis gene in *E. coli*. This gene is the second most important overall repressor. Fis acts repressing the Crp gene, which is the most connected gene in the *E. coli* genome. Fis in turn is regulated positively by Ihf, the gene with the highest RI of *E. coli*. With this example, we highlight that the transcriptional regulation of the most influential genes could determine the phenotype of a cell via the transcriptional cascades generated by activation or repression of those influential genes.

4.5. Final Considerations. With this approach, we present a hierarchical transcription network built from a highly curated database, containing the values of the interactions between TFs and their targets. We observed that the large majority of genes is controlled by just a few TFs. Those few TFs are strongly coregulated between them, which is translated into a fine tuning in the transcription process. A way to quantify this is by the relative influence parameter (RI), in which the lesser regulated genes and highly regulated ones are relevant for global transcriptional control.

Finally, the extent to which genes are regulated is also important, whether regulation is positive or negative. The negative interaction means that the TF is a repressor of the target, meanwhile a positive interaction represents an activation given by the TF. A global value of each TF in terms of their regulatory values is presented. We observe that NFY subunits are overall activators, meanwhile MAZ and TFAP2A and B are overall repressors. Those genes must be important in the context of transcription. Arguably, regulation of those genes presented in this discussion is fundamental to initiate or terminate a transcriptional cascade, likely related to events of differentiation or cell division. Experiments along this way are necessary. We argue that this framework may help to understand the importance of TFs in terms of the ratio between in- and out-degree as well as their overall effect on targets, which could in turn give new insights regarding the gene regulation.

Additional Points

Network Data Availability. Cytoscape .cys files for the networks presented here are available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] T. Ravasi, H. Suzuki, C. V. Cannistraci et al., "An atlas of combinatorial transcriptional regulation in mouse and man," *Cell*, vol. 140, no. 5, pp. 744–752, 2010.

[2] K. Baca-López, M. Mayorga, A. Hidalgo-Miranda, N. Gutiérrez-Nájera, and E. Hernández-Lemus, "The role of master regulators in the metabolic/transcriptional coupling in breast carcinomas," *PLoS One*, vol. 7, no. 8, article e42678, 2012.

[3] J. H. Woo, Y. Shimoni, W. S. Yang et al., "Elucidating compound mechanism of action by network perturbation analysis," *Cell*, vol. 162, no. 2, pp. 441–451, 2015.

[4] H. Tovar, R. García-Herrera, J. Espinal-Enríquez, and E. Hernández-Lemus, "Transcriptional master regulator analysis in breast cancer genetic networks," *Computational Biology and Chemistry*, vol. 59, Part B, pp. 67–77, 2015.

[5] A. A. Margolin, I. Nemenman, K. Basso et al., "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics*, vol. 7, Supplement 1, p. S7, 2006.

[6] 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, 2009.

[7] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450–461, 2007.

[8] FANTOM Consortium, H. Suzuki, A. R. Forrest et al., "The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line," *Nature Genetics*, vol. 41, no. 5, pp. 553–562, 2009.

[9] S. Gama-Castro, H. Salgado, A. Santos-Zavaleta et al., "RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond," *Nucleic Acids Research*, vol. 44, no. D1, pp. D133–D143, 2015.

[10] J. Y. Chan, X. L. Han, and Y. W. Kan, "Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast," *Proceedings of the National Academy of Sciences*, vol. 90, no. 23, pp. 11371–11375, 1993.

[11] N. C. Andrews, H. Erdjument-Bromage, M. B. Davidson, P. Tempst, and S. H. Orkin, "Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein," *Nature*, vol. 362, no. 6422, pp. 722–728, 1993.

[12] M. Bugno, M. Daniel, N. L. Chepelev, and W. G. Willmore, "Changing gears in Nrf1 research, from mechanisms of regulation to its role in disease and prevention," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1849, no. 10, pp. 1260–1276, 2015.

[13] M. Mohrin, J. Shin, Y. Liu et al., "A mitochondrial UPR-mediated metabolic checkpoint regulates hematopoietic stem cell aging," *Science*, vol. 347, no. 6228, pp. 1374–1377, 2015.

[14] D. V. Ho and J. Y. Chan, "Induction of herpud1 expression by ER stress is regulated by Nrf1," *FEBS Letters*, vol. 589, no. 5, pp. 615–620, 2015.

[15] V. O. Okoh, N. A. Garba, R. B. Penney et al., "Redox signalling to nuclear regulatory proteins by reactive oxygen species

contributes to oestrogen-induced growth of breast cancer cells," *British Journal of Cancer*, vol. 112, no. 10, pp. 1687–1702, 2015.

[16] M. Biswas, D. Phan, M. Watanabe, and J. Y. Chan, "The Fbw7 tumor suppressor regulates nuclear factor E2-related factor 1 transcription factor turnover through proteasome-mediated proteolysis," *Journal of Biological Chemistry*, vol. 286, no. 45, pp. 39282–39289, 2011.

[17] J. Zhang, X. Zheng, and Q. Zhang, "Egln2 positively regulates mitochondrial function in breast cancer," *Molecular & Cellular Oncology*, vol. 3, no. 2, article e1120845, 2015.

[18] L. Zhao, M. Tang, Z. Hu et al., "mir-504 mediated down-regulation of nuclear respiratory factor 1 leads to radio-resistance in nasopharyngeal carcinoma," *Oncotarget*, vol. 6, no. 18, pp. 15995–16018, 2015.

[19] Y. Zhang and J. D. Hayes, "The membrane-topogenic vectorial behaviour of Nrf1 controls its post-translational modification and transactivation activity," *Scientific Reports*, vol. 3, no. 1, article 2006, 2013.

[20] Y. Zhang, J. M. Lucocq, and J. D. Hayes, "The Nrf1 CNC/bZIP protein is a nuclear envelope-bound transcription factor that is activated by t-butyl hydroquinone but not by endoplasmic reticulum stressors," *Biochemical Journal*, vol. 418, no. 2, pp. 293–310, 2009.

[21] Y. Zhang, Y. Ren, S. Li, and J. D. Hayes, "Transcription factor Nrf1 is topologically repartitioned across membranes to enable target gene transactivation through its acidic glucose-responsive domains," *PLoS One*, vol. 9, no. 4, article e93458, 2014.

[22] S. Gallant and G. Gilkeson, "ETS transcription factors and regulation of immunity," *Archivum Immunologiae et Therapiae Experimentalis*, vol. 54, no. 3, pp. 149–163, 2006.

[23] Y. Zhou, X. Liu, L. Xu et al., "Transcriptional repression of plasma cell differentiation is orchestrated by aberrant overexpression of the ETS factor SPIB in Waldenström macroglobulinaemia," *British Journal of Haematology*, vol. 166, no. 5, pp. 677–689, 2014.

[24] Y. Yang, A. L. Shaffer 3rd, N. C. Emre et al., "Exploiting synthetic lethality for the therapy of ABC diffuse large B cell lymphoma," *Cancer Cell*, vol. 21, no. 6, pp. 723–737, 2012.

[25] J. J. Karrich, M. Balzarolo, H. Schmidlin et al., "The transcription factor Spi-B regulates human plasmacytoid dendritic cell survival through direct induction of the antiapoptotic gene BCL2-A1," *Blood*, vol. 119, no. 22, pp. 5191–5200, 2012.

[26] L. J. Marshall, L. D. Moore, M. M. Mirsky, and E. O. Major, "JC virus promoter/enhancers contain TATA box-associated Spi-B-binding sites that support early viral gene expression in primary astrocytes," *Journal of General Virology*, vol. 93, Part 3, pp. 651–661, 2012.

[27] W. Su, J. Xia, X. Chen et al., "Ectopic expression of AP-2α transcription factor suppresses glioma progression," *International Journal of Clinical and Experimental Pathology*, vol. 7, no. 12, pp. 8666–8674, 2014.

[28] A. R. Hallberg, S. U. Vorrink, D. R. Hudachek et al., "Aberrant CpG methylation of the TFAP2A gene constitutes a mechanism for loss of TFAP2A expression in human metastatic melanoma," *Epigenetics*, vol. 9, no. 12, pp. 1641–1647, 2014.

[29] D. Shi, F. Xie, Y. Zhang et al., "TFAP2A regulates nasopharyngeal carcinoma growth and survival by targeting HIF-1α signaling pathway," *Cancer Prevention Research*, vol. 7, no. 2, pp. 266–277, 2014.

[30] H. Li, R. Sheridan, and T. Williams, "Analysis of TFAP2A mutations in branchio-oculo-facial syndrome indicates functional complexity within the AP-2α DNA-binding domain," *Human Molecular Genetics*, vol. 22, no. 16, pp. 3195–3206, 2013.

[31] T. I. Meshcheryakova, R. A. Zinchenko, T. A. Vasilyeva et al., "A clinical and molecular analysis of branchio-oculo-facial syndrome patients in Russia revealed new mutations in TFAP2A," *Annals of Human Genetics*, vol. 79, no. 2, pp. 148–152, 2015.

[32] H. Kawata, K. Yamada, Z. Shou et al., "Zinc-fingers and homeoboxes (ZHX) 2, a novel member of the ZHX family, functions as a transcriptional repressor," *Biochemical Journal*, vol. 373, Part 3, pp. 747–757, 2003.

[33] F. Luan, P. Liu, H. Ma et al., "Reduced nucleic ZHX2 involves in oncogenic activation of glypican 3 in human hepatocellular carcinoma," *The International Journal of Biochemistry & Cell Biology*, vol. 55, pp. 129–135, 2014.

[34] S. Nagel, B. Schneider, A. Rosenwald et al., "t(4;8)(q27;q24) in Hodgkin lymphoma cells targets phosphodiesterase PDE5A and homeobox gene ZHX2," *Genes, Chromosomes and Cancer*, vol. 50, no. 12, pp. 996–1009, 2011.

[35] S. Nagel, B. Schneider, C. Meyer, M. Kaufmann, H. G. Drexler, and R. A. Macleod, "Transcriptional deregulation of homeobox gene ZHX2 in Hodgkin lymphoma," *Leukemia research*, vol. 36, no. 5, pp. 646–655, 2012.

[36] I. Gozes, "Activity-dependent neuroprotective protein: from gene to drug candidate," *Pharmacology & Therapeutics*, vol. 114, no. 2, pp. 146–154, 2007.

[37] G. Vandeweyer, C. Helsmoortel, A. Van Dijck et al., "The transcriptional regulator ADNP links the BAF (SWI/SNF) complexes with autism," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 166, no. 3, pp. 315–326, 2014.

[38] C. Helsmoortel, A. T. Vulto-van Silfhout, B. P. Coe et al., "A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP," *Nature Genetics*, vol. 46, no. 4, pp. 380–384, 2014.

[39] M. H. Yang, Y. H. Yang, C. Y. Lu et al., "Activity-dependent neuroprotector homeobox protein: a candidate protein identified in serum as diagnostic biomarker for Alzheimer's disease," *Journal of Proteomics*, vol. 75, no. 12, pp. 3617–3629, 2012.

[40] A. Merenlender-Wagner, A. Malishkevich, Z. Shemer et al., "Autophagy has a key role in the pathophysiology of schizophrenia," *Molecular Psychiatry*, vol. 20, no. 1, pp. 126–132, 2015.

[41] S. Mandel, G. Rechavi, and I. Gozes, "Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis," *Developmental Biology*, vol. 303, no. 2, pp. 814–824, 2007.

[42] H. S. Jeon, T. Dracheva, S. H. Yang et al., "SMAD6 contributes to patient survival in non-small cell lung cancer and its knockdown reestablishes TGF-β homeostasis in lung cancer cells," *Cancer Research*, vol. 68, no. 23, pp. 9686–9692, 2008.

[43] Y. J. Kang, J. W. Shin, J. H. Yoon et al., "Inhibition of erythropoiesis by Smad6 in human cord blood hematopoietic stem cells," *Biochemical and Biophysical Research Communications*, vol. 423, no. 4, pp. 750–756, 2012.

[44] K. Pardali, M. Kowanetz, C. H. Heldin, and A. Moustakas, "Smad pathway-specific transcriptional regulation of the cell

cycle inhibitor p21WAF1/Cip1," *Journal of Cellular Physiology*, vol. 204, no. 1, pp. 260–272, 2005.

[45] E. Frullanti, F. Colombo, F. S. Falvella et al., "Association of lung adenocarcinoma clinical stage with gene expression pattern in noninvolved lung tissue," *International Journal of Cancer*, vol. 131, no. 5, pp. E643–E648, 2012.

[46] F. R. Mangone, F. Walder, S. Maistro et al., "Smad2 and Smad6 as predictors of overall survival in oral squamous cell carcinoma patients," *Molecular Cancer*, vol. 9, no. 1, p. 1, 2010.

[47] J. Yin, K. Lu, J. Lin et al., "Genetic variants in TGF-β pathway are associated with ovarian cancer risk," *PLoS One*, vol. 6, no. 9, article e25559, 2011.

[48] H. L. Tan, E. Glen, A. Töpf et al., "Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation," *Human Mutation*, vol. 33, no. 4, pp. 720–727, 2012.

[49] H. Xu, J. Fu, S. W. Ha et al., "The CCAAT box-binding transcription factor NF-Y regulates basal expression of human proteasome genes," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1823, no. 4, pp. 818–825, 2012.

[50] C. Dai, C. X. Miao, X. M. Xu et al., "Transcriptional activation of human CDCA8 gene regulated by transcription factor NF-Y in embryonic stem cells and cancer cells," *Journal of Biological Chemistry*, vol. 290, no. 37, pp. 22423–22434, 2015.

[51] C. Romier, F. Cocchiarella, R. Mantovani, and D. Moras, "The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y," *Journal of Biological Chemistry*, vol. 278, no. 2, pp. 1336–1345, 2003.

[52] J. D. Fleming, G. Pavesi, P. Benatti, C. Imbriano, R. Mantovani, and K. Struhl, "NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors," *Genome Research*, vol. 23, no. 8, pp. 1195–1209, 2013.

[53] C. F. Chen, X. He, A. D. Arslan et al., "Novel regulation of nuclear factor-YB by miR-485-3p affects the expression of DNA topoisomerase IIα and drug responsiveness," *Molecular Pharmacology*, vol. 79, no. 4, pp. 735–741, 2011.

[54] H. Ma, X. Yue, L. Gao et al., "ZHX2 enhances the cytotoxicity of chemotherapeutic drugs in liver tumor cells by repressing MDR1 via interfering with NF-YA," *Oncotarget*, vol. 6, no. 2, p. 1049, 2015.

[55] A. Ray and B. K. Ray, "Induction of Ras by SAF-1/MAZ through a feed-forward loop promotes angiogenesis in breast cancer," *Cancer Medicine*, vol. 4, no. 2, pp. 224–234, 2015.

[56] L. Jiao, Y. Li, D. Shen et al., "The prostate cancer-up-regulated myc-associated zinc-finger protein (MAZ) modulates proliferation and metastasis through reciprocal regulation of androgen receptor," *Medical Oncology*, vol. 30, no. 2, pp. 1–8, 2013.

[57] R. Britto, S. Umesh, A. S. Hegde et al., "Shift in AP-2α localization characterizes astrocytoma progression," *Cancer Biology & Therapy*, vol. 6, no. 3, pp. 413–418, 2007.

[58] B. C. Turner, J. Zhang, A. A. Gumbs et al., "Expression of AP-2 transcription factors in human breast cancer correlates with the regulation of multiple growth factor signalling pathways," *Cancer Research*, vol. 58, no. 23, pp. 5466–5472, 1998.

[59] C. E. Hoei-Hansen, J. E. Nielsen, K. Almstrup et al., "Transcription factor AP-2 is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors," *Clinical Cancer Research*, vol. 10, no. 24, pp. 8521–8530, 2004.

[60] T. Nystorm, "Glucose starvation stimulon of *Escherichia coli*: role of integration host factor in starvation survival and growth phase-dependent protein synthesis," *Journal of Bacteriology*, vol. 177, no. 19, pp. 5707–5710, 1995.

[61] H. Bi and C. Zhang, "Integration host factor is required for the induction of acid resistance in *Escherichia coli*," *Current Microbiology*, vol. 69, no. 2, pp. 218–224, 2014.

[62] D. Charlier, M. Roovers, D. Gigot, N. Huysveld, A. Piérard, and N. Glansdorff, "Integration host factor (IHF) modulates the expression of the pyrimidine-specific promoter of the carAB operons of Escherichia coli K_{12} and Salmonella typhimurium LT_2," *Molecular and General Genetics MGG*, vol. 237, no. 1-2, pp. 273–286, 1993.

[63] S. S. Singh and D. C. Grainger, "H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions," *PLoS Genetics*, vol. 9, no. 6, article e1003589, 2013.

[64] N. P. Tucker, M. G. Hicks, T. A. Clarke et al., "The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster," *PLoS One*, vol. 3, no. 11, article e3623, 2008.

[65] D. F. Browning, D. J. Lee, S. Spiro, and S. J. Busby, "Down-regulation of the *Escherichia coli* K-12 nrf promoter by binding of the NsrR nitric oxide-sensing transcription repressor to an upstream site," *Journal of Bacteriology*, vol. 192, no. 14, pp. 3824–3828, 2010.

[66] J. D. Partridge, D. M. Bodenmiller, M. S. Humphrys, and S. Spiro, "NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility," *Molecular Microbiology*, vol. 73, no. 4, pp. 680–694, 2009.

[67] M. Yamada and M. H. Saier Jr., "Physical and genetic characterization of the glucitol operon in *Escherichia coli*," *Journal of Bacteriology*, vol. 169, no. 7, pp. 2990–2994, 1987.