Loregic: Logic-circuits based method to characterize co-operativity of regulatory factors

Daifeng Wang[1,2], Koon-Kiu Yan[1,2], Chao Cheng[4], Joel Rozowsky[1,2], Mark Gerstein[1,2,3,*]

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA.
[2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA.
[3]Department of Computer Science, Yale University, New Haven, CT, USA.
[4]Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA.

ABSTRACT

Regulatory factors (RFs) cooperatively control gene expression. We present Loregic, a computational method, to identify and further characterize this cooperativity using logic-circuit models based on expression changes. We study the cooperativity of multiple RFs with common target genes, and specifically focus on two RFs co-regulating a target gene using a two-input-one-output logic gate model. We match binarized gene expression values to the inputs and outputs of all 16 (2^4) possible types of logic gates, assigning a score to closeness of the match. We first apply our method to yeast cell cycle data, and find that the transcription factors (TFs) are highly cooperative matching cooperative logic gates (e.g. AND-like gate) with better scores. As expected, in TF deletion experiments, the targets of TFs involved in cooperative logic gates have larger expression fold-changes. Next, we shift to the cooperativities between/among TFs and miRNAs in human leukemia cells using the ChIP-seq data from ENCODE K562, and the RNA/smRNA-seq expression data from Acute Myeloid Leukemia samples in TCGA. We find that c-Myc, pervasively amplifying target gene expressions, does not cooperate with other TFs and many miRNAs from its matched logics. However, we find that some miRNAs interact with c-Myc in cooperative logic gates, which is consistent with that miRNAs and c-Myc down-regulate each other in leukemia. In addition, we check the promoter motifs of TFs from cooperative logics to reveal their potential indirect bindings when their motifs are absent at target promoters. Finally, we compare the positions of cooperative vs. non-cooperative TFs in a hierarchy built from the regulatory network. We find that the TFs at middle levels tend to need more cooperativities. In summary, our method provides a valuable framework to reveal complex gene regulatory mechanisms, and can be extended to analyze cooperativity amongst other regulatory elements such as enhancers.

Contact: pi@gersteinlab.org

---

*To whom correspondence should be addressed.

---

1 INTRODUCTION

Regulatory factors (RFs) coordinately control gene expression through multiple dimensions including both space and time. For example, multiple transcription factors bind to the promoter region of their target gene in a spatial dimension. The gene regulatory network controls gene expression during embryo development in a temporal dimension. Moreover, gene expression is regulated in ways at the genomic levels from DNA, transcriptome to proteome. Multi-dimensional gene regulatory factors work cooperatively, rather than independently to determine correct gene expressions in various cell types. For example, gene expression at the transcriptional level can be controlled by various factors such as transcription factors (TFs), histone modifications, enhancers, and non-coding RNAs. Among them, transcription factors play key roles in transcriptional regulation, and have been found by various experimental and computational efforts to normally work together. Those approaches study TF-TF relationships from various aspects such as protein-protein interactions, sequence motifs in TF binding sites in *cis*-regulatory modules, co-associations of TFs in binding sites, and co-expressions of TF target genes (Banerjee and Zhang, 2003; Hardison and Taylor, 2012; Karczewski, et al., 2011). Also, TFs coordinate with other factors such as with miRNAs to co-regulate gene expression (Gerstein, et al., 2012; Poos, et al., 2013). However, previous efforts have focused on the identification of the wiring relationships between TFs like co-binding, co-association and co-expression, but the cooperative patterns of TFs driving actual biological functions behind the wiring diagrams have not yet been fully explored. In reality, for instance in electronic circuits, we can wire different electronic elements such as resistors, capacitors, and inductors together to generate different functions (Rabaey, et al., 2003). Similarly, different TFs identified to be wired may produce different biological functions. Thus, beyond finding wiring, it is necessary to further identify cooperative patterns of TFs, which has not been much researched via computational approaches yet to the best of our knowledge.

Transcription factors control gene expression in a discrete way in many cases so that logic processing broadly exists in gene regulation (Albert and Othmer, 2003; Das, et al., 2009; Mangan and Alon, 2003; Peter and Davidson, 2011; Peter, et al., 2012; Shmulevich and Dougherty, 2007; Tu, et al., 2013; Xie, et al., 2011). The DNA sequence motifs follow the combinatorial logics (AND, OR and NOT) to match gene expression patterns (Beer and Tavazoie, 2004), but TFs can still interact without directly co-binding to regulatory sequence elements to control gene expressions (Farnham, 2009). Moreover, the combinatorial logics are much more than simple three logics (AND, OR and NOT); e.g., 16 (2^4) logic gates for two-input-one-output scenarios including all possible logic combinations between positive and

negative regulators. Thus, we need more sophisticated models that capture more possible combinatorial cooperativities between TFs and other RFs. As perturbation experiments such as TF knock-out provide Boolean regulatory data – the on and off of the TF, the Boolean model has been used to capture this logic processing, especially for logic combinatorial effects of different TFs working together (Somogyi and Sniegoski, 1996). The simple binary operations in the Boolean model only need a limited size of data, which is very computationally efficient. However, previous studies only focused on a small set of genes, which lack genome-wide identification and characterization of logic operations in gene regulation. TFs along with other regulatory factors interact with each other to form regulatory networks, which can be modeled as directed networks. We can further formulate those directed networks by hierarchical structures, and assign TFs to different hierarchical layers (Gerstein, et al., 2012). The hierarchical structure in gene regulatory networks provide a system point of view of the cooperativity of RFs on genome wide (Bhardwaj, et al., 2010; Bhardwaj, et al., 2010; Cheng, et al., 2011; Gerstein, et al., 2012; Jothi, et al., 2009), so it is necessary to better characterize gene regulatory cooperativity based on network structures by using efficient computational models.

In this paper, we developed a novel computational method called Loregic that integrates gene expression and regulatory data, to characterize the cooperativity of gene regulatory factors on genome wide using logic-circuit models (Fig. 1). We apply our method to transcription factors in yeast and human, and characterize the cooperativity of TFs and miRNAs to their target genes, and also across levels in gene regulatory hierarchical network.

2 MATERIALS AND METHODS

Loregic is a novel computational method based on logic-circuit models to characterize the cooperativity of regulatory factors using gene regulation and expression datasets. In this paper, we demonstrate Loregic by analyzing transcription factors and their target genes of yeast and human. Loregic consists of the following steps:

Step 1: convert gene expression values to Boolean values based on their dynamic changes across conditions;

Step 2: for every RF (TF/miRNA in this paper) pairs that co-regulate a target gene (RF-RF-target triplet), map and score it against all possible two-input-one-output logic gates (16) base on Boolean values;

Step 3: test score significances to remove random effects;

Step 4: characterize the cooperativity of two RFs in the triplet using the logic gate(s) with significantly high score(s);

Step 5: check TF sequence motifs at target regulatory regions such as promoters for all logics;

Step 6: compare the positions of cooperative vs. non-cooperative TFs in hierarchical regulatory networks.

2.1   Gene expression, transcription factor and miRNA datasets

For gene expression in yeast, we used three well-studied cell cycle datasets: 1) alpha-factor time course with 18 time points at 0, 7, 14, … , 119 minutes; 2) cdc15 time course with 24 time points at 10, 30, 50, 70, … , 290 minutes; 3) cdc28 time course with 17 time points at 0, 10, 20, … , 160 minutes in (Cho, et al., 1998; Spellman, et al., 1998). We combined all three datasets that include 5581 genes with 59 time points, and standardized gene expressions for each time point. For gene regulation in yeast, we used the transcription factors (TFs) with their target genes identified in (Harbison, et al., 2004; Jothi, et al., 2009), and found ~39k TF-TF-target triplets.

For gene expression in human leukemia, we downloaded RPKM expressions in RNA-seq for ~20k protein-coding genes (705 miRNAs) across 197 (188) samples with Acute Myeloid Leukemia (AML) from The Cancer Genome Atlas (TCGA) Data Portal (https://tcga-data.nci.nih.gov/tcga/). We standardized log(RPKM+1) across genes for each sample. For gene regulation in human leukemia, we identified 50865 TF-TF-target triplets using ChIP-seq data in ENCODE K562 cell line (Consortium, 2011; Djebali, et al., 2012; Gerstein, et al., 2012), and 56944 miRNA-TF-target triplets using confident miRNA-targets for human K562 cell line in (Chen, et al., 2014).

2.2   Converting gene expression changes over conditions to Boolean values

Previous Boolean models normally converted the gene expression to 1 or 0 based on whether its expression values are greater than a threshold (-> 1) or not (-> 0). This kind of conversion, however, is difficult to come up with a reasonable threshold, which may vary for genes or datasets. Moreover, the gene expression varies dynamically over conditions if their regulators express differently. Genes may have their specific thresholdings for highly or lowly expressions. Thus, we converted gene expressions to Boolean values (1 or 0) using BoolNet (Mussel, et al., 2010), which first clusters genes into co-expression modules using K-means, and discretizes gene expressions to binary values from co-expressed modular patterns across conditions.

2.3 Mapping and scoring a RF-RF-target triplet to 16 logic gates

A logic gate with two-input (X, Y) and one-output (Z) can be determined by a combination of four (X, Y, Z) binary vectors, (X=0, Y=0, Z), (X=0, Y=1, Z), (X=1, Y=0, Z), and (X=1, Y=1, Z) with specific Z values, also known as truth table. With 2^4 different combinations of Z values, we can obtain 16 different logic gates in Figure 2, where '~' denotes NOT (negative regulation), '*' denotes AND and '+' denotes OR logic operations.

For an example in Fig. 2, suppose a RF-RF-target triplet, (X, Y, Z) has *m*=20 binary vectors after conversion. 17 out of 20 vectors highlighted by solids lines can match the AND gate, Z=X*Y; i.e., both X and Y must present to activate Z to express. 10 out of 20 vectors highlighted by dash lines can match the OR gate, Z=X+Y; i.e., either X or Y presents to activate Z to express. Thus, the cooperativity between X and Y in this example is more likely to be the AND type rather than the OR type. Also, we can find match to other logic gates from those 20 vectors. We refer to the logic gates in which Z values are determined by AND operation between X/~X and Y/~Y as cooperative logic gates, and others as non-cooperative logic gates.

Here, for a triplet of (X, Y, Z), we develop a quantitative method to score the closeness of match to 16 logic gates. The logic gates with high scores imply the cooperative or non-cooperative behaviors of RFs X and Y to target Z. In Fig. 3, we build a matrix with 4 rows and 2 columns. The matrix elements cover all 8 different (X, Y, Z) binary vectors. The two elements at the same row share the same X and Y input values (1st row: X=0, Y=0; 2nd row: X=0, Y=1; 3rd row: X=1, Y=0; 4th row: X=1, Y=1), and the four elements at the same column share the same Z output value (1st column: Z=0; 2nd column: Z=1). We assign a binary *c*-value to each element to indicate if number of the element appearing in (X, Y, Z) vectors is greater than the other element at the same row with different output Z value. For the element at i^{th} row, 1st column, if its appearances out of *m* (X, Y, Z) binary vectors are more than the element at 2nd column (same X-Y inputs, different Z output), we let $c_{i,1}=1$ and $c_{i,2}=0$; i.e., with the same X and Y inputs, output Z is more likely to be zero. If less, we let $c_{i,1}=0$ and $c_{i,2}=1$; i.e., with the same X and Y inputs, output Z is more likely to be one. If equal or it happens that both elements at the same row miss, then we assume that both outputs are possible so that we let $c_{i,1}=c_{i,2}=1$; i.e., with the same X and Y inputs, output Z is equally likely to be one or zero. The truth table of any one of 16 logic gates corresponds to a unique pathway from 1st row to 4th row that have 4 elements from different rows. For example, the AND gate is the solid pathway, and the OR gate is the dashed one in Fig. 3. We assign a weight, *w* to each logic gate, as the product of *c*-values of four elements on its corresponding pathway. The weight,

also a binary number, indicates if four outputs (Z values) of its logic gate are no less than other logic gates in the triplet; e.g., $w(\text{AND})= c_{1,1}* c_{2,1}* c_{3,1}* c_{4,2}$, and $w(\text{OR})= c_{1,1}* c_{2,2}* c_{3,2}* c_{4,2}$ (Table 1). The score, s to a logic gate is then defined as its weight over the summation of weights of all 16 logic gates; i.e., $s(G_k)= w(G_k)/(w(G_1)+ w(G_2)+\ldots+ w(G_{16}))$ for the kth logic gate, G_k. The high scores suggest that the corresponding logic gates match the vectors of the triplet more likely than others, which indicates the cooperative (e.g., AND) or non-cooperative (e.g., OR) regulations of RFs X and Y to target Z.

2.4 Testing score significances of triplets by randomizing their targets

Due to the sample size limit in gene expression data, the random effects may bias the predicted scores. In order to overcome the random effects, given a triplet of (X, Y, Z), we calculate its significances of 16 logic gates' scores as follows. For the kth logic gate, G_k, we replace the target gene, Z by a randomly selected gene, obtain a score for G_k, and after repeating N times (e.g., $N=1000$), define its significance level, $p(G_k)=\text{Prob}(G_k>N$ random scores). We obtain the score significances for all 16 logic gates. Therefore, the random effects may drive the insignificant logic gates. In this paper, a score is considered to be significant if its significance level, p is less than 0.1.

2.5 Constructing gene regulatory hierarchical network

The gene regulatory network is structured in a hierarchical way (Bhardwaj, et al., 2010; Bhardwaj, et al., 2010; Gerstein, et al., 2012). The transcription factors from different hierarchical levels may play different roles in gene regulation. We are interested to look at the characteristic cooperativity among different hierarchical levels. We constructed the gene regulatory hierarchical network, assigned TFs to three levels (top, middle, bottom) in hierarchy using the simulated annealing method in (Gerstein, et al., 2012). We characterized the cooperativities for TFs in the network using Loregic, and analyzed the network positions of different cooperative types (e.g., cooperative vs. non-cooperative) among hierarchical levels.

3 RESULTS

3.1 Yeast TFs are cooperative during cell cycle

We identified ~39k TF-TF-target triplets with 176 TFs using TF-target predictions in (Harbison, et al., 2004; Jothi, et al., 2009). We characterized the cooperativity for each TF-TF pair across their common targets during cell cycle using 59 time points (See Methods) by Loregic. Generally, we found that coop-

erative logic gates, especially the gates of AND (i.e., Z=X*Y), Z=~X*Y and Z=X*~Y, have significantly higher scores than others (Fig. 4). In Fig. 5, we included 4126 TF-TF-target triplets with significant highest scores (*s*=1) for one logic gate. Out of those highest-score logic gates, we found more cooperative logic gates than non-cooperative ones. Moreover, among cooperative logic gates, we found that the AND gates were the most abundant, in which that both TFs have to be present to activate their target gene to express.

3.2 Deleting TFs with cooperative logic gates gives rise to significantly higher fold changes of target gene expression

The TF knockout experiments gave us the fold changes of gene expression after deleting single TF in yeast (Hu, et al., 2007; Reimand, et al., 2010). If a target gene is regulated by two cooperative TFs such as AND relationship, deletion of either TF may corrupt the cooperativity so that eventually it impacts gene expression. For example, for the triplets with significant highest scores at AND gate, we found that deleting either of their TFs gave rise to significantly down-regulated target genes, i.e., negative expression fold changes (*t-test p-value* =0.068). For non-cooperative TFs such as Z=X or Z=Y gates, i.e., one of TFs (dominate TF) fully determines target gene expression, we found that target genes are significantly down-regulated by deleting dominate TFs than by deleting the other TFs (*t-test p-value* < 0.05 for Z=X, <0.005 for Z=Y).

3.3 Cooperativity between TF-TF and miRNA-TF across targets in Acute Myeloid Leukemia

We further applied Loregic to the human leukemia. We first identified 50865 TF-TF-target triplets from ChIP-seq experiments for 70 TFs in ENCODE K562 cell line (Consortium, 2011; Djebali, et al., 2012; Gerstein, et al., 2012). Moreover, because miRNAs and TFs have been found to co-regulate common target genes (Cheng, et al., 2011; Gerstein, et al., 2012), we are interested to study their cooperativities. We obtained 222 miRNAs that have highly confident interactions with their targets in K562 cell line (Chen, et al., 2014). Thus, by integrating miRNA-target pairs and TF-target pairs in K562, we identified 56944 miRNA-TF-target triplets. The gene/miRNA expression datasets we used included RNA/smRNA-seq RPKM values for ~20k protein-coding genes across 197 samples and 705 miRNAs across 188 samples in TCGA Acute Myeloid Leukemia (AML). Thus, we characterized TF-TF and miRNA-TF cooperativities in the AML cancer by integrating ENCODE and TCGA datasets. For TF-TF-target triplets, we found that XXX triplets with significant highest scores (s=1) for a logic gate, out of

TRKCHG

which XXX gate is the most. For miRNA-TF-target triplets, we found that XXX triplets with significant highest scores (s=1) for a logic gate, out of which XXX gate is the most.

3.4 c-Myc does not cooperate with other TFs

The transcription factor, MYC has been found to universally amplify target gene expressions in lympho-cytes (Nie, et al., 2012), which implies that it does not need the cooperation from other TFs. We found that there were 2153 MYC-TF-target triplets with 67 other TFs, and 905 out of 2153 triplets can be assigned the significant highest scores (s=1) for one logic gate. Two logic gates with the most triplets among 905 ones are Z=MYC (133 triplets) and OR (Z=MYC+Y) (211 triplets). Z=MYC means that the target gene expressions are fully determined by MYC, and Z=MYC+Y means that either expressed MYC or the second TF, Y can make targets express. Both scenarios suggest that MYC is able to determine target expressions by itself without any requirements of other TFs presenting, which supports the recent finding that MYC plays a universal amplifier role in gene expression.

3.5 miRNAs and c-Myc double down-regulate to each other

MYC and miRNAs have been found to down-regulate to each other by forming double down-regulatory feed-forward loops in leukemia (Tao, et al., 2014). We identifed 1805 miRNA-MYC-target triplets with 117 miRNAs, and 1143 out of 1805 triplets can be assigned the significant highest scores (s=1) for one logic gate. Out of 1143 triplets, 446 ones match Z=MYC, and 201 ones match Z=~miRNA+MYC, two most logic gates, both of which imply that MYC successfully down-regulates miRNAs so that target gene expressions can be turned on without down-regulations from miRNA. We also found that there were 56 triplets matching Z=~miRNA*MYC, and 16 triplets matching Z=~miRNA, which suggests that those targets cannot be expressed when miRNAs down-regulate MYC. In short, those matched logic gates support that the miRNAs and MYC indeed form a double-negative regulatory loop in leukemia (Fig. 6).

3.6 Promoter motif absences of cooperative TFs imply potential indirect binding activities

We checked promoter motifs of TFs at their target promoter regions (upstream 1000 bps (yeast) and 5000 bps (human) from TSS) (DebRoy, 2013; Lawrence, 2014; Li, 2014; Pages, 2014). It is interesting that there are many TFs do not have motifs (<90% PWM similarity) in target promoters even though they are predicted to match cooperative logic gates. For example, out of 1043 yeast TF-TF-target triplets with significant highest scores on AND gate, 362 ones have one TF whose motifs do not present at target promoters (545 out of 1616 for Z=X*~Y, 565 out of 1660 for Z=~X*Y). For human leukemia, we

also found that out of XXX TF-TF-target triplets with significant highest scores on AND gate, XXX ones have one TF whose motifs do not present at target promoters (XXX out of XXX for Z=X*~Y, XXX out of XXX for Z=~X*Y). Because an amount of TFs regulate target genes without interacting their DNA sequences (indirect binding), and ChIP experiments are not able to detect indirect binding signals, those TFs with motifs absent as above potentially regulate targets through indirect bindings by cooperating with direct binding TFs (Biddie, et al., 2011; Farnham, 2009; Gordan, et al., 2009; Neph, et al., 2012; Zhao, et al., 2012).

3.7 Network positions of cooperative vs. non-cooperative TFs on gene regulatory hierarchical network

Hierarchical structure has been found in gene regulatory network. The transcription factors tend to regulate the genome at different hierarchical levels, and the ones at middle levels are found to play key regulatory roles (Bhardwaj, et al., 2010; Bhardwaj, et al., 2010; Gerstein, et al., 2012). We are interested to characterize the cooperativities among TFs at different levels, and identify network positional preferences for cooperative and non-cooperative TFs. Here, we constructed the hierarchical network, and assigned TFs to three hierarchical levels, top, middle and bottom (Gerstein, et al., 2012). For each logic gate, we identified the network edges associated with the triplets that have significant highest scores (s=1) on it. For yeast (Fig. 8A), we found that 58 cooperative and 24 non-cooperative edges were between top and bottom levels, 38 cooperative and 31 non-cooperative edges were between top and middle levels, and 65 cooperative and 26 non-cooperative edges were between middle and bottom levels. For human (Figs. 8B), we found that XXX cooperative and XXX non-cooperative edges were between top and bottom levels, XXX cooperative and XXX non-cooperative edges were between top and middle levels, and XXX cooperative and XXX non-cooperative edges were between middle and bottom levels. Those observations suggest that the TFs at top and middle levels need to work cooperatively when they regulate ones at bottom levels, which implies that the TFs at top and middle levels play core roles in gene regulatory system, so they must work more coordinately to make genome function correctly.

4 DISCUSSION

Loregic is a computational method to characterize the gene regulatory cooperativity using logic-circuit models by integrating the gene expression and regulatory information. In this paper, we mainly focus on the cooperativity of transcription factors and miRNAs, and their network characteristics in gene regulatory network. We can also extend Loregic in future to study the regulatory coordination among other

genomic elements relating to gene regulation such as enhancers because next generation sequencing technologies provide us more fruitful and accurate expression (e.g., RNA-seq, small RNA-seq), and regulation (e.g., Chip-seq, DNase-seq) datasets for them.

We demonstrated Loregic using TF/miRNA-TF-target triplets that includes two transcription factors and one target. We should also point out that Loregic could be also used to analyze the regulatory modules with multiple RFs and multiple target genes as long as enough expression data support (2^N samples, N is number of RFs in module). For those regulatory modules with N_1 TFs and N_2 targets, we need to expand the scoring matrix to the one with 2^{N1} rows and 2^{N2} columns, fill the matrix elements with all ($N_1 + N_2$)-dimension binary vectors, and calculate the scores of pathways associated with corresponding logic gates with N_1-input and N_2-output.

We converted the gene expression to Boolean values by comparing co-expression patterns across samples. For noisy expression data like microarrays in yeast, we may still obtain Boolean values even for noisy values, thus our step of testing significance is also designed to remove this effect. Loregic is also compatible to other discretization methods, and take their binarized gene expressions as direct inputs.

We also found that some triplets didn't have significant high scores for any logic gates, which may be caused by that the regulatory cooperativity for their TFs and targets might be random processes or driven by other stochastic biological activities, rather than deterministic ones, which couldn't be described as logic operations.

To our knowledge, Loregic is the first computational method to systematically characterize the regulatory cooperativity using logic-circuit models. It will have a widely variety of applications to study the regulatory mechanisms of increasing genomic elements such as ones annotated in ENCODE, and help to build the gene regulatory panoramagram.

ACKNOWLEDGEMENTS

REFERENCES

Albert, R. and Othmer, H.G. (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster, *Journal of theoretical biology*, **223**, 1-18.

Banerjee, N. and Zhang, M.Q. (2003) Identifying cooperativity among transcription factors controlling the cell cycle in yeast, *Nucleic acids research*, **31**, 7024-7031.

Beer, M.A. and Tavazoie, S. (2004) Predicting gene expression from sequence, *Cell*, **117**, 185-198.

Bhardwaj, N., Kim, P.M. and Gerstein, M.B. (2010) Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators, *Science signaling*, **3**, ra79.

Bhardwaj, N., Yan, K.K. and Gerstein, M.B. (2010) Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels, *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 6841-6846.

Biddie, S.C*., et al.* (2011) Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding, *Mol Cell*, **43**, 145-155.

Chen, D*., et al.* (2014) Dissecting the chromatin interactome of microRNA genes, *Nucleic Acids Res*, **42**, 3028-3043.

Cheng, C*., et al.* (2011) Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data, *PLoS Comput Biol*, **7**, e1002190.

Cho, R.J*., et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle, *Molecular cell*, **2**, 65-73.

Consortium, E.P. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE), *PLoS biology*, **9**, e1001046.

Das, D., Pellegrini, M. and Gray, J.W. (2009) A primer on regression methods for decoding cis-regulatory logic, *PLoS computational biology*, **5**, e1000269.

DebRoy, H.P.a.P.A.a.R.G.a.S. (2013) Biostrings: String objects representing biological sequences, and matching algorithms, *R package version 2.28.0.*

Djebali, S*., et al.* (2012) Landscape of transcription in human cells, *Nature*, **489**, 101-108.

Farnham, P.J. (2009) Insights from genomic profiling of transcription factors, *Nat Rev Genet*, **10**, 605-616.

Gerstein, M.B*., et al.* (2012) Architecture of the human regulatory network derived from ENCODE data, *Nature*, **489**, 91-100.

Gordan, R., Hartemink, A.J. and Bulyk, M.L. (2009) Distinguishing direct versus indirect transcription factor-DNA interactions, *Genome Res*, **19**, 2090-2100.

Harbison, C.T*., et al.* (2004) Transcriptional regulatory code of a eukaryotic genome, *Nature*, **431**, 99-104.

Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory modules in animals, *Nat Rev Genet*, **13**, 469-483.

Hu, Z., Killion, P.J. and Iyer, V.R. (2007) Genetic reconstruction of a functional transcriptional regulatory network, *Nature genetics*, **39**, 683-687.

Jothi, R*., et al.* (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture, *Mol Syst Biol*, **5**, 294.

Karczewski, K.J*., et al.* (2011) Cooperative transcription factor associations discovered using regulatory variation, *Proc Natl Acad Sci U S A*, **108**, 13353-13358.

Lawrence, M.C.a.H.P.a.P.A.a.S.F.a.M.M.a.D.S.a.M. (2014) GenomicFeatures: Tools for making and manipulating transcript centric annotations, *R package version 1.12.4.*

Li, H.P.a.M.C.a.S.F.a.N. (2014) AnnotationDbi: Annotation Database Interface, *R package version 1.22.6.*

Mangan, S. and Alon, U. (2003) Structure and function of the feed-forward loop network motif, *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 11980-11985.

Mussel, C., Hopfensitz, M. and Kestler, H.A. (2010) BoolNet--an R package for generation, reconstruction and analysis of Boolean networks, *Bioinformatics*, **26**, 1378-1380.

Neph, S*., et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints, *Nature*, **489**, 83-90.

Nie, Z*., et al.* (2012) c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells, *Cell*, **151**, 68-79.

Pages, H. (2014) BSgenome: Infrastructure for Biostrings-based genome data packages, *R package version 1.28.0.*

Peter, I.S. and Davidson, E.H. (2011) Evolution of gene regulatory networks controlling body plan development, *Cell*, **144**, 970-985.

Peter, I.S., Faure, E. and Davidson, E.H. (2012) Predictive computation of genomic logic processing functions in embryonic development, *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16434-16442.

Poos, K*., et al.* (2013) How microRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation, *PLoS computational biology*, **9**, e1003210.

Rabaey, J.M., Chandrakasan, A.P. and Nikoli*c, B. (2003) *Digital integrated circuits : a design perspective*. Prentice Hall electronics and VLSI series. Pearson Education, Upper Saddle River, N.J.

Reimand, J*., et al.* (2010) Comprehensive reanalysis of transcription factor knockout expression data in Saccharomyces cerevisiae reveals many new targets, *Nucleic acids research*, **38**, 4768-4777.

Shmulevich, I. and Dougherty, E.R. (2007) *Genomic Signal Processing*. Princeton series in applied mathematics. Princeton University Press, Princeton.

Somogyi, R. and Sniegoski, C.A. (1996) Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation, *Complexity*, **1**, 45-63.

Spellman, P.T*., et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization, *Mol Biol Cell*, **9**, 3273-3297.

Tao, J., Zhao, X. and Tao, J. (2014) c-MYC-miRNA circuitry: a central regulator of aggressive B-cell malignancies, *Cell Cycle*, **13**, 191-198.

Tu, S., Pederson, T. and Weng, Z. (2013) Networking development by Boolean logic, *Nucleus*, **4**, 89-91.

Xie, Z*., et al.* (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells, *Science*, **333**, 1307-1311.

Zhao, Y*., et al.* (2012) Improved models for transcription factor binding site identification using nonindependent interactions, *Genetics*, **191**, 781-790.