Loregic: Logic-circuits based computational approach to characterize cooperativity of transcription factors in gene regulatory hierarchical network

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. *Correspondence to: pi@gersteinlab.org

Abstract. Gene regulatory factors cooperatively control gene expression. We present a novel computational method, Loregic to identify and further characterize cooperativity of transcription factors (TFs) using logic circuits based on dynamic expression changes. We model the cooperativity of two that co-regulate a target gene as a two-input-one-output logic gate, and map it to 16 types of gates with scores. The high score means that two TFs highly likely regulate their target gene coordinated in a way of the corresponding logic gate. We first apply our method to yeast data, and find that the previously identified highly cooperative TFs have higher scores of cooperative logic gates than non-cooperative ones. Moreover, the yeast TFs of cooperative logic gates are found to have significantly larger expression fold-changes in TF deletion experiments. We further identify logic operations among human TFs by integrating human ChIP-seq and RNA-seq data from ENCODE for two cell lines K562 (erythroleukemic) and GM12878 (lymphoblastoid). Finally, we compare network positions of cooperative vs. non-cooperative TFs in gene regulatory hierarchical network, and find that the transcription factors between top and middle levels work more cooperatively than others. In summary, our method provides a valuable integrated framework to reveal complex gene regulator mechanisms, and can be also extended to analyze cooperativity among other regulatory elements such as enhancers and non-coding RNAs.

adfa, p. 1, 2011.

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 transcription level is controlled by transcription factors (TFs), histone modifications, activators, repressors, enhancers, and non-coding RNAs. Among them, transcription factors play key roles in transcriptional regulation, and have been found by various experimental and computational approaches to normally work together. Those approaches study collaborations among TFs 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 [1-3]. Also, TFs coordinate with other factors such as with microRNAs to co-regulate gene expression [4, 5]. Although RFs appear to cooperate, previous efforts have not further characterized how RFs co-regulate the target genes.

Regulatory factors control gene expression in a discrete way in many cases so that logic processing broadly exists in gene regulation [6-13]. 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 regulatory factors. The simple binary operations in the Boolean model only need a limit 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. Moreover, because the hierarchical structure in gene regulatory networks provide a system point of view of the cooperativity of RFs on genome wide [5, 14-17], it is necessary to better characterize gene regulatory cooperativity based on network structure by using proper computational models.

In this paper, we developed a novel computational method, referred to as Loregic, which integrate 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 to different target genes, and also across different levels in gene regulatory hierarchical network.

Materials and Methods

Overview

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: converting gene expression values to Boolean values based on their dynamic changes across conditions;

Step 2: as for every TF pairs that co-regulate a target gene (TF-TF-target triplet), mapping and scoring it using 16 two-input-one-output logic gates base on Boolean values;

Step 3: testing score significances to remove random effects;

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

Gene expression and transcription factors datasets

The gene expression datasets in yeast cell cycle have been well studied. Moreover, the systematic gene regulatory relationship in yeast was revealed. Thus, for yeast, we use the gene expression microarray data of the mitotic cell cycle with 17 time points at 0, 10, 20,…, 160 minutes in [18], and transcription factors (TFs) with their target genes identified by ChIP-chip from [19]. After yeast, we further apply Loregic to the human gene expression and regulatory data generated by next generation sequencing techniques. For human, we use the latest gene expression data (RPKM values in RNA-seq) for two cell lines, K562 (10 samples, erythroleukemic) and GM12878 (8 samples, lymphoblastoid), and their TFs with target genes found by ChIP-seq in ENCODE [5, 20, 21].

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 methods, 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. Actually, in design of logic circuits, output signals are normally triggered by dynamic changes of input signals such as Edge-triggered latches [22]. Thus, we converted gene expressions to Boolean values (1 or 0) as follows. Given a gene's n expression values $\{e_i, i=1,2,\ldots,n\}$, where e_i is its expression value at ith condition/time point, this gene's Boolean values are given by $\{b_i=1$ if $e_{i+1}>e_i$, and $b_i=1$ if $e_{i+1}\leq e_i$, i=1,2,…,n-1$\}$ for time-series gene expression data, and $\{b_{i,j}=1$ if $e_j>e_i$,

and $b_{i,j}$=1 if $e_j \leq e_i$, $(i,j) \in (1,2,..n)\}$ for non-time-series gene expression data. Our conversion method captures dynamic changes of gene expressions between two adjacent time points or two condition pairs if no temporal order implied.

Mapping and scoring a TF-TF-target triplet to 16 logic gates based on their Boolean values (Figure 2)

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 $2^{\wedge 4}$ different combinations of Z values, we can obtain 16 different logic gates in Figure 2. Given a TF-TF-target triplet with two transcription factors X and Y along with their target gene, Z, we model it using a logic gate, and obtain their Boolean values after conversion so that we have m (X, Y, Z) binary vectors ($m=n-1$ for time series data, $m=n(n-1)/2$ for non-time-series data). In this step, we need to map these m vectors to 16 logic gates, and find which logic gate(s) have most vectors mapped. For example in Fig. 2, suppose a TF-TF-target triplet, (X, Y, Z) has m=20 binary vectors. 17 out of 20 vectors highlighted by solids lines can be mapped to 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 be mapped to the OR gate, Z=X+Y; i.e., either X or Y presents to activate Z to express. Thus, we suspect that the cooperativity between X and Y in this example may be more likely to be the AND type than OR.

We propose a quantitative method to give scores for 16 logic gates for a triplet of (X, Y, Z), and thus, the logic gates with high scores imply the cooperative type of TFs X and Y. 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 values (1^{st} row: X=0, Y=0; 2^{nd} row: X=0, Y=1; 3^{rd} row: X=1, Y=0; 4^{th} row: X=1, Y=1), and the four elements at the same column share the same Z value (1^{st} column: Z=0; 2^{nd} column: Z=1). For the element at i^{th} row, j^{th} column, we count its appearances out of m (X, Y, Z) binary vectors, and denote as $c_{i,j}$. If it happens that both elements at the same row miss, i,e., $c_{i,1}=c_{i,2}=0$, we assume that both elements are possible so that we reinforce $c_{i,1}=c_{i,2}=1$. Any one of 16 logic gates corresponds to a unique pathway from 1^{st} row to 4^{th} row that have 4 elements from different rows. For example, the AND gate is marked as the solid pathway, and the OR gate is the dashed one in Fig. 3. We assign a weight, w to each logic gate, which is product of four elements' counts on the corresponding pathway, which represents the number of realizations of the logic gate in the triplet; e.g., w(AND)= $c_{1,1}$* $c_{2,1}$* $c_{3,1}$* $c_{4,2}$, and w(OR)= $c_{1,1}$* $c_{2,2}$* $c_{3,2}$* $c_{4,2}$ (See all logic gates in 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 k^{th} logic gate, G_k. The high scores suggest that the corresponding logic gates appear more frequently than

others in the triplet, so that the two TFs are more likely to cooperate in the way that the logic gates imply.

Testing score significances of triplets by replacing their target genes by random genes

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 k^{th} 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, $p(G_k)$=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.

Constructing gene regulatory hierarchical network

The gene regulatory network is structured in a hierarchical way [5, 15, 16]. 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 construct the gene regulatory hierarchical network using the feed-forward loops, a particular type of TF-TF-target triplets in which the first transcription factor X also regulates the second transcription factor Y. The transcription factors from FFLs are assigned to three levels (top, middle, bottom) in hierarchy using the simulated annealing method in [5]. We characterized the cooperativities for all FFLs in the network using Loregic, and analyzed the network positions of different cooperative types (e.g., AND vs. OR FFLs) among hierarchical levels. We summarized the numbers of FFLs and TFs at different hierarchical levels in Table 2.

Results

Previously known cooperative TFs are predicted with higher scores on cooperative logic gates

There were 31 TF pairs that were identified to significantly cooperate in gene regulation [1] in yeast's cell cycle. The cell-cycle expressions of the target genes regulated by both TFs of a cooperative pairs were found to correlate significantly higher than ones regulated by one of two TFs only. We predicted the logic relationships of those cooperative TF pairs using Loregic, and identified the logic gate with the highest score (s>0.1, p<0.1) for each TF pair. Out of those highest-score logic gates, we found that the cooperative logic gates are significantly more than non-cooperative ones (t-test p-$value$<0.05)

(Fig. 4). 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. But, we also found that the highest-score logic gates varied in different target genes even for a same cooperative TF pair. Therefore, our algorithm further characterized the target-gene-specific cooperativity of TFs, rather than the overall cooperativity measurement based on correlations among their target genes in [1].

Cooperative relationships of TFs are consistent to target genes

The cooperative TFs identified by the expression correlations of their target genes in [1] were not able to give the cooperativity measurement for individual target genes. Since Loregic scores the cooperativity of individual TF-TF-target triplet to 16 logic gates, it can further classify the cooperativity of TFs according to their different target genes, and identify target-gene-specific cooperativity. For example in Fig. 5, we show that two cooperative TFs, SWI4 and SWI6, who forms a complex to control the transition from G1 to S phases in cell cycle [23], have high scores (darker colors) on AND gate to most of their target genes, though some targets genes have high scores on other types of logic gates.

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 [24, 25]. 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. As contrast, the target genes regulated by non-cooperative TFs such as OR relationship may not be affected by deleting one of TFs since either TF can initiate their expressions. We compared the expression fold changes of target genes of cooperative TFs with ones of non-cooperative TFs. We identified two groups of TFs from the AND and OR logic gates with significant scores (s>0.1, p<0.05). We found that deleting AND TFs gave rise to significantly larger fold changes of target gene expressions than deleting OR TFs (t-test *p-value*<0.05) in Fig. 6. Thus, this reveals that any one of cooperative TFs identified by Loregic is essential to maintain their target gene expressions, and a lack of one of them will potentially significantly change target gene expressions. For those non-cooperative TFs, their target gene expression may not be changed much if one of TFs absent.

Cooperativity of TFs in human K562 and GM12878 cell lines

We characterized cooperativities of TFs in human K562 and GM12878 cell lines, and identify the logic gates with highest scores (s>0.1) for TF-TF-target triplets (90256 in

K562, 14480 in GM12878). We found that four types of logic gates, AND, OR, Z=X, and Z=Y have most high scores, as Fig. 7 indicates that they take ~85% in K562 and ~78% in GM12878.

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 [5, 15, 16]. 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 using feed-forward loops (FFLs) [11], and assigned TFs to three hierarchical levels, top, middle and bottom [5]. For each logic gate, we identified the network edges associated with the FFLs that have significant scores (s>0.1, p<0.05) on it. For yeast (Fig. 8A), we found that 24.7% of AND edges were between top and middle levels, but none of OR edges were found. For two human cell lines (Figs. 8B and 8C) between top and middle levels, fractions of AND edges were also more than OR edges (K562: 35.5% vs. 22.5%, GM12878: 66.7% vs. 33.3%). The OR edges involving TFs at the bottom levels including both top-bottom and middle-bottom, however, took more fractions than the AND ones (yeast: 100% vs. 75.4%, human K562: 77.5% vs. 64.5%, human GM12878: 63.6% vs. 33.3%). Those observations suggest that the TFs between top and middle levels work more cooperatively than the ones between top/middle and bottom levels. Because the TFs at top and middle levels play core roles in gene regulatory system, they must work more coordinately to have genome run normally.

Discussion

Loregic is a computational approach to characterize the gene regulatory cooperativity using logic circuits model by integrating the gene expression and regulatory information. In this paper, we mainly focus on the cooperativity of transcription factors, and their network characteristics in gene regulatory network. We can also extend Loregic in future to study the regulatory coordination among other regulatory factors (RFs) such as enhancers, non-coding RNAs including miRNAs and pseudogenes since 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-TF-target triplets that includes two transcription factors and one target, but we should point out that Loregic could be also used to analyze the regulatory modules with multiple TF/RFs and multiple target genes as long as enough expression data support. For those regulatory modules with N_1 TFs and N_2 targets, we only

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 expression numerical values of every two 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. If we have enough size of data samples in future, in order to obtain more robust Boolean values, we can use Boolean values only coming from fold-changes between two samples greater than certain thresholds.

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.

To our knowledge, Loregic is the first approach to systematically characterize the regulatory cooperativity using logic-circuits model. 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.

References

[1] Banerjee, N. and Zhang, M. Q. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic acids research*, 31, 23 (Dec 1 2003), 7024-7031.

[2] Karczewski, K. J., Tatonetti, N. P., Landt, S. G., Yang, X., Slifer, T., Altman, R. B. and Snyder, M. Cooperative transcription factor associations discovered using regulatory variation. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 32 (Aug 9 2011), 13353-13358.

[3] Hardison, R. C. and Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nature reviews. Genetics*, 13, 7 (Jul 2012), 469-483.

[4] Poos, K., Smida, J., Nathrath, M., Maugg, D., Baumhoer, D. and Korsching, E. How microRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation. *PLoS computational biology*, 9, 8 (Aug 2013), e1003210.

[5] Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A. P., Cayting, P., Charos, A., Chen, D. Z., Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'Geen, H., Ouyang, Z., Partridge, E. C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T. E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K. Y.,

Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P. J., Myers, R. M., Weissman, S. M. and Snyder, M. Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489, 7414 (Sep 6 2012), 91-100.
[6] Peter, I. S. and Davidson, E. H. Evolution of gene regulatory networks controlling body plan development. *Cell*, 144, 6 (Mar 18 2011), 970-985.
[7] Peter, I. S., Faure, E. and Davidson, E. H. Predictive computation of genomic logic processing functions in embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 41 (Oct 9 2012), 16434-16442.
[8] Das, D., Pellegrini, M. and Gray, J. W. A primer on regression methods for decoding cis-regulatory logic. *PLoS computational biology*, 5, 1 (Jan 2009), e1000269.
[9] Albert, R. and Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *Journal of theoretical biology*, 223, 1 (Jul 7 2003), 1-18.
[10] Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. and Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science*, 333, 6047 (Sep 2 2011), 1307-1311.
[11] Mangan, S. and Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 21 (Oct 14 2003), 11980-11985.
[12] Tu, S., Pederson, T. and Weng, Z. Networking development by Boolean logic. *Nucleus*, 4, 2 (Mar-Apr 2013), 89-91.
[13] Shmulevich, I. and Dougherty, E. R. *Genomic Signal Processing*. Princeton University Press, Princeton, 2007.
[14] Jothi, R., Balaji, S., Wuster, A., Grochow, J. A., Gsponer, J., Przytycka, T. M., Aravind, L. and Babu, M. M. Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Molecular systems biology*, 52009), 294.
[15] Bhardwaj, N., Kim, P. M. and Gerstein, M. B. Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. *Science signaling*, 3, 146 2010), ra79.
[16] Bhardwaj, N., Yan, K. K. and Gerstein, M. B. 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, 15 (Apr 13 2010), 6841-6846.
[17] Cheng, C., Yan, K. K., Hwang, W., Qian, J., Bhardwaj, N., Rozowsky, J., Lu, Z. J., Niu, W., Alves, P., Kato, M., Snyder, M. and Gerstein, M. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. *PLoS computational biology*, 7, 11 (Nov 2011), e1002190.
[18] Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J. and Davis, R. W. A

genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell*, 2, 1 (Jul 1998), 65-73.
[19] Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. and Young, R. A. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298, 5594 (Oct 25 2002), 799-804.
[20] Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigo, R. and Gingeras, T. R. Landscape of transcription in human cells. *Nature*, 489, 7414 (Sep 6 2012), 101-108.
[21] Consortium, E. P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology*, 9, 4 (Apr 2011), e1001046.
[22] Rabaey, J. M., Chandrakasan, A. P. and Nikoli*c, B. *Digital integrated circuits : a design perspective*. Pearson Education, Upper Saddle River, N.J., 2003.
[23] Koch, C., Moll, T., Neuberg, M., Ahorn, H. and Nasmyth, K. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science*, 261, 5128 (Sep 17 1993), 1551-1557.
[24] Hu, Z., Killion, P. J. and Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nature genetics*, 39, 5 (May 2007), 683-687.
[25] Reimand, J., Vaquerizas, J. M., Todd, A. E., Vilo, J. and Luscombe, N. M. Comprehensive reanalysis of transcription factor knockout expression data in Saccharomyces cerevisiae reveals many new targets. *Nucleic acids research*, 38, 14 (Aug 2010), 4768-4777.