Systems biology

Using high-throughput multi-omics data to investigate structural balance in elementary gene regulatory network motifs

Alberto Zenere 1, Olof Rundquist 2, Mika Gustafsson 2, and Claudio Altafini 1,∗

1 Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-58183 Linköping, Sweden.
2 Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, SE-58183 Linköping, Sweden.
∗To whom correspondence should be addressed.

Abstract

Motivation: The simultaneous availability of ATAC-seq and RNA-seq experiments allows to obtain a more in-depth knowledge on the regulatory mechanisms occurring in gene regulatory networks (GRNs). In this paper, we highlight and analyze two novel aspects that leverage on the possibility of pairing RNA-seq and ATAC-seq data. Namely we investigate the causality of the relationships between transcription factors (TFs), chromatin and target genes and the internal consistency between the two omics, here measured in terms of structural balance in the sample correlations along elementary length-3 cycles.

Results: We propose a framework that uses the a priori knowledge on the data to infer elementary causal regulatory motifs (namely chains and forks) in the network. It is based on the notions of conditional independence and partial correlation, and can be applied to both longitudinal and non-longitudinal data. Our analysis highlights a strong connection between the causal regulatory motifs that are selected by the data and the structural balance of the underlying sample correlation graphs: strikingly, > 97% of the selected regulatory motifs belong to a balanced subgraph. This result shows that internal consistency, as measured by structural balance, is close to a necessary condition for 3-node regulatory motifs to satisfy causality rules.

Contact: claudio.altafini@liu.se

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

One of the trends in the field of GRN inference is to increase the inference power of the data by combining multiple omics techniques. For instance, in recent years the integration of RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) data has given promising results, see [4, 23, 1]. This integration can be carried out in different ways. Some studies use a two-step approach, where for instance ATAC-seq is used to obtain a large set of candidate interactions and then RNA-seq is used to prune this set and to identify a reliable subset of high-confidence TF-target gene interactions, see e.g. [20] and [14]. Alternatively, many studies analyze the correlation between chromatin peaks and target genes, see e.g. [12] and [26]. Unlike these studies, we propose to consider simultaneously three layers of transcription: TFs, chromatin peaks and target genes. The first and third level are quantified via RNA-seq and the second via ATAC-seq. In other words, we consider not only the correlation between peak and target gene, but also between peak-TF and TF-target gene. We show that this method can be used to identify new cross-layers elementary regulatory motifs involving TFs, chromatin peaks and target genes. A necessary condition for performing this analysis is to have paired ATAC-seq and RNA-seq data, as it is in our case.

More precisely, we decided to focus on two classes of three-node regulatory motifs, formed by transcription factors, chromatin peaks and target genes. The regulatory motifs we have chosen to work with are the chains and forks shown...
in Fig. 1, because they encode conditional independence relationships: two nodes in these regulatory motifs become independent once their values are conditioned on that of the third node [7]. Such conditional independence can be explored in a systematic way using partial correlation [2]. Partial correlation has been used extensively in the context of causal inference in GRNs to investigate gene-gene interactions, see [29, 22]. Here instead we use sample partial correlations as a tool to screen all possible three-node regulatory motifs, based on a computational map of all possible interactions among TFs, chromatin peaks and target genes we constructed. Only chains and forks that pass the partial correlation test are considered as “selected” regulatory motifs based on the data.

Another concept that has been used in biological networks is structural balance (hereafter simply denoted balance), see [9, 13, 19]. Notice that, in the context of signed networks, balance is synonym to coherence, although the latter assumes different meanings in other fields (e.g. [3]). Balance is associated to signed cycles, in particular a cycle is balanced if it has an even number of negative edges and unbalanced otherwise. In the studies [9, 13, 19] the focus was on counting balanced motifs in a given biological network, and the common result was that balanced motifs were enriched over unbalanced ones. Here balance is instead associated to the sample correlations of triplets of nodes that belong to different omics, which form our elementary regulatory motifs. Interestingly, in our analysis we also find a similar property: the triplets of correlations selected by the data for our chain and fork regulatory motifs tend to be enriched for balanced triangles, while the percentage of unbalanced triangles is significantly lower than in random data, suggesting that the notion of balance can be observed in experimental data, even when these span different omics.

We have gathered four publicly available datasets of paired RNA-seq and ATAC-seq experiments on human immune cells, see Tab. 1. Datasets A and B represent time-series of primary human naive CD4⁺ T cells during early T-helper type 1 differentiation [18]. The difference between A and B is that in the latter the activation was performed in the presence of progesterone. Datasets C and D are time-series experiments on human monocyte-derived dendritic cells under infection with HIV-1, where the latter serves as mock experiment [14]. For details, see the corresponding publications.

| Index | Cell type | Availability and Reference |
|-------|-----------|---------------------------|
| A     | human Th1 E-MTAB-7775, E-MTAB-10444 | [18] |
| B     | human Th1 E-MTAB-10418, E-MTAB-10444 | |
| C     | human DC GSE125817, [14] |
| D     | human DC GSE125918, [14] |

Table 1. List of paired RNA-seq and ATAC-seq datasets used in this study.

for more details. The result is two sets of interactions: between chromatin regions and target genes (A-G), and between TFs and chromatin regions (T-A). From there, we can retrieve a third set of interactions, between TFs and target genes (T-G), by connecting TFs and target genes that share at least one common chromatin region in the computational templates A-G and A-T. Altogether, the three combined interaction mappings form what we call a multi-omics TF-peak-target gene map.

Such mapping typically contains a significant amount of false positives, as highlighted in [28]. In this work, we address the issue by combining the notions of dynamical correlation, partial correlation and balance, which we now introduce.

2.2 Dynamical Correlation

Calculating correlation coefficients in longitudinal studies requires appropriate tools to take into account the dependency between (often irregularly spaced) time points as well as latent factors, see [31] and [11]. Failing to do so will introduce bias in the correlation coefficients and create false connections between the data. One of the approaches to render the data normally distributed is to use the notion of dynamical correlation. In particular we focus on the definition introduced by [21] and reviewed in the Materials and Methods in the Supplementary Information. From now on the adjective “dynamical” will be implicitly assumed when dealing with correlation or partial correlation.

2.3 Partial correlation

A partial correlation reflects the strength of a linear relationship between two variables after controlling for potential effects coming from other variables. The concept has received wide attention in different fields, such as GRN inference [32, 29, 22] and brain functional connectivity [24]. We denote the partial correlation coefficient between the variables $X$ and $Y$ given $Z$ with $R(X, Y | Z)$, which is expressed in formula by

$$ R(X, Y | Z) = \frac{R(X, Y) - R(X, Z)R(Y, Z)}{\sqrt{(1 - R(X, Z)^2)(1 - R(Y, Z)^2)}} $$

In particular, partial correlations can be used to test causal interactions in the data. To illustrate its usefulness, consider the simplest case of three variables: $X$, $Y$ and $Z$. Assume $X$, $Y$ and $Z$ are part of a regulatory chain, for instance $X$ regulates $Z$, which in turn regulates $Y$. $X \rightarrow Z \rightarrow Y$, see Fig. 1b, left. This common regulatory motif is characterized by the fact that the dependence between $X$ and $Y$ is mediated by $Z$ and that $X$ and $Y$ become independent once we “project away” the information due to $Z$ [2]. More formally, if we consider $X$, $Y$ and $Z$ as (Gaussian) random variables, the joint probability distribution of the regulatory motif $X \rightarrow Z \rightarrow Y$ factorizes as $p(X, Y, Z) = p(X|Z)p(Z)p(Y|Z)$ where $p(X)$ is the probability distribution of the variable $X$ and $p(Z|X)$ is the conditional
Investigating structural balance in multi-omics regulatory elementary regulatory motifs

Fig. 1. Workflow of this paper. (a) Schematic depiction of two key events that lead to gene expression: (1) the chromatin region around the promoter is loose and accessible to TFs binding, and (2) available TFs bind to specific DNA sequences in the promoter region of the gene. (b) Possible regulatory motifs. Which event precedes the other is still under investigation, thus several causal three-node regulatory motifs can be associated to represent the regulatory interactions between TFs, chromatin regions and target genes. (c) Balanced and unbalanced cycles corresponding to the undirected graph of $A \rightarrow T \rightarrow G$ (i.e. chromatin $\rightarrow$ TF $\rightarrow$ gene). Plus and minus signs denote positive and negative correlation values between the corresponding nodes.

### 2.4 Structural balance

Given three variables $X$, $Y$ and $Z$ let us compute their pairwise correlations $R(X,Y)$, $R(X,Z)$ and $R(Y,Z)$. These three correlations form an undirected cycle of length three (i.e. a triangle). We say that such a cycle is balanced if the product $R(X,Y) \cdot R(X,Z) \cdot R(Y,Z) > 0$. In the following section, balance will be used as a test of internal consistency among the variables involved in the chains and forks basic regulatory motifs.

### 3 Results

#### 3.1 Elementary gene regulatory motifs and their conditional independence

The approach we follow in this paper is to break down the complexity of GRNs by analyzing elementary causal regulatory motifs. In particular, we start our analysis by modeling the interplay between TF and chromatin accessibility, which leads to gene expression. We show that it can be represented as two regulatory motifs, $T \rightarrow A \rightarrow G$ and $A \rightarrow T \rightarrow G$.

Chromatin accessibility at the promoter region can enable (or amplify) the effect of TFs on gene expression. Consider the example of a gene with a unique transcriptional activator: it is...
Table 2. Overview of the datasets. (Upper) We report the total number of regulatory motifs (and the percentage of balanced ones) present in the TF-peaks target map. Since \( A \to T \to G \) and \( T \to A \to G \) correspond to the same undirected graph we use the more general term “Chains” to denote \((A, T, G)\) triplets. (Middle) Next, we test if each regulatory motif is characterized by a statistically large balance ratio (see Sec. 3.2 for details on how the statistical test was built); fold change indicates the ratio between the value observed in the data and the mean of the null distribution. (Lower) Lastly, we report the number of regulatory motifs that pass the conditional independence test described in the Materials and Methods of Supplementary Information and how many of them belong to an unbalanced cycle.

| Dataset | Number of selected regulatory motifs (of which unbalanced) |
|---------|----------------------------------------------------------|
| A       | \( A \to T \to G \) \( T \to A \to G \) \( A \to T \to G \) |}

| Dataset | Enrichment of balanced regulatory motifs: p-value, fold change |
|---------|-------------------------------------------------------------|
| A       | \(< 10^{-16}, 1.11\) \(< 10^{-16}, 1.11\) \(< 10^{-16}, 1.19\) |
| B       | \(3.60 \cdot 10^{-7}, 1.04\) \(1.16 \cdot 10^{-7}, 1.04\) \(< 10^{-16}, 1.11\) |
| C       | not significant \(2.50 \cdot 10^{-3}, 1.02\) \(< 10^{-16}, 1.07\) |
| D       | \(< 10^{-16}, 1.08\) \(< 10^{-16}, 1.16\) \(10^{-16}, 1.16\) |

Table 2. Overview of the datasets. (Upper) We report the total number of regulatory motifs (and the percentage of balanced ones) present in the TF-peaks target map. Since \( A \to T \to G \) and \( T \to A \to G \) correspond to the same undirected graph we use the more general term “Chains” to denote \((A, T, G)\) triplets. (Middle) Next, we test if each regulatory motif is characterized by a statistically large balance ratio (see Sec. 3.2 for details on how the statistical test was built); fold change indicates the ratio between the value observed in the data and the mean of the null distribution. (Lower) Lastly, we report the number of regulatory motifs that pass the conditional independence test described in the Materials and Methods of Supplementary Information and how many of them belong to an unbalanced cycle.
Investigating structural balance in multi-omics regulatory elementary regulatory motifs

T — A — G and T — G and checking if the triangle (T, A, G) has balanced correlations. When this does not happen, then our data shows internal inconsistency, i.e. the signs of the three correlations R(T, A), R(A, G) and R(T, G) are incompatible. A similar construction can be carried out for the other regulatory motifs mentioned above and we can then proceed to checking the balance (i.e. internal consistency) of the resulting triangles see Fig. 1(c).

Table 3. Contingency table between the number of selected T→A→G and A→T→G regulatory motifs in dataset A. See Results in Supplementary Information for the contingency tables of datasets B, C and D.

|   | A→T→G |   |
|---|---|---|
| Selected | 302 | 20840 |
| Non-selected | 18973 | 367973 |

It is interesting to observe that the data appears to be significantly consistent, as measured by the percentage of balanced cycles in the network. To retrieve the null distribution of the percentage of balanced cycles, we used a bootstrapping approach. Namely, we generated a population of 50,000 triplets of Gaussian random signals, having the same number of time-points as the data. Thereafter we extracted 10,000 sub populations, comprising 10,000 triplets each, and we calculated their balance ratio, thus leading to the null distribution. Balanced regulatory motifs appear to be significantly over-represented in the data; as can be seen in Tab. 2, both chain and fork regulatory motifs are enriched for balance in almost all the datasets.

Not only balanced triplets are over-represented in the data, they also consist of edges corresponding to the correlations in the network with the highest absolute values. To formalize this observation, we have associated each triplet to scalar measures that quantify the magnitude of the corresponding correlations. We have chosen three measures: geometric mean, minimum and maximum; although similar results can be obtained using other measures, such as mean and harmonic mean. A Kolmogorov-Smirnov test reveals that the distribution of each measure differs significantly (every p-value is <10−16) between balanced and unbalanced regulatory motifs, where the former show higher average values, as seen in Fig. 2.

3.3 Structural balance is a necessary condition for conditional independence

The categorization of triplets into balanced and unbalanced sheds light also on the conditional independence of the variables involved. As can be seen in Fig. 3, the distributions of partial correlation values in chains differ significantly between balanced and unbalanced cycles. In particular the latter distributions are characterized by a “drop” around zero, meaning that unbalanced cycles rarely lead to conditional independence. A similar observation holds for fork regulatory motifs as well, see Results in the Supplementary Information. What stands out from the analysis is that balance is “almost” a necessary condition for conditional independence. Strikingly, for all four datasets, > 97% of the selected chain and fork regulatory motifs belong to a balanced cycle. The enrichment of balance among selected regulatory motifs is statistically confirmed by an hypergeometric test that compares the balance ratio among the selected regulatory motifs and among all the regulatory motifs in the network (p-value<10−16).

3.4 Balanced and selected regulatory motifs are conserved under different cell stimuli

Datasets A and B come from the same cell type under partially similar stimuli. Both datasets have been generated from Th1 cells differentiated under Th1 polarizing conditions, with the difference that for dataset B the Th1 polarization was done in presence of progesterone. Accordingly, they are characterized by similar TF-peak-target gene mappings: ~ 50% of A → T → G and T → A → G, ~ 40% of T1 ∼ A → T2 and ~ 80% of G1 ← A → G2 regulatory motifs are shared by the two datasets. When we focus on this pool of common regulatory motifs we observe that a significant portion is balanced in both datasets. More precisely, there is a mild but significant overlap between the regulatory motifs that are balanced in A and those that are balanced in B, see Tab. 4. Interestingly, the relationship becomes stronger when we look at those regulatory motifs (except A → T → G) that are selected in dataset A and B. A similar comparison can also be carried out between datasets C and D, see Results in Supplementary Information, leading to similar conclusions.
4 Discussion

In this paper, we consider two alternative chain models to represent the interplay that exists between TFs and chromatin remodeling in regulating gene expression, differing for the causality direction between $A$ and $T$. Although the precise mechanisms are still unclear, several studies have showed that the regulation can happen in both directions: TFs affects chromatin accessibility and vice versa [15, 16, 25]. Hence we decided to consider both $A \rightarrow T \rightarrow G$ and $T \rightarrow A \rightarrow G$ as distinct plausible regulatory motifs. In our case, the two sets of $(A, T, G)$ triplets that fit the conditional independence hypothesis for these regulatory motifs are significantly disjoint. This is in accordance with the notion that in some physiological situations chromatin remodeling precedes TF binding whereas in other situations it is the TF binding that leads to chromatin remodeling [6].

![Diagram](image)

**Fig. 3.** (a) $T \rightarrow A \rightarrow G$ regulatory motif and corresponding distribution of $R(T, G|A)$, divided in balanced (blue) and unbalanced (red) cycles. The balanced and unbalanced distributions are normalized with respect to their total count independently. (b) Similar analysis for the $A \rightarrow T \rightarrow G$ regulatory motif and the corresponding distribution of $R(A, G|T)$.

In this work, we use balance as a consistency criterion. In the context of biological networks, multiple studies have already highlighted that GRNs are enriched for balanced patterns [9, 19] and altogether tend to be close to monotone systems [17]. However, the application of these ideas to sample correlations multi-omics data in particular has never been explored before, at least in the authors’ knowledge. Indeed, the observation that combined RNA-seq and ATAC-seq data is predominantly balanced provides evidence that it is for the most part internally consistent. It is interesting to couple this observation with the fact that $> 97\%$ of selected (i.e. conditionally independent) regulatory motifs were found to belong to a balanced cycle. Conditional independence is associated to low correlation values upon conditioning, thus it may be surprising that unbalanced cycles (characterized by lower correlation values) rarely lead to conditional independence.

We have also observed that the peaks that belong to chain regulatory motifs selected by the data are, on average, closer to the TSS of the corresponding target gene (see Sec. 2.4 in the Supplementary Information). From a biological perspective, this suggests that the regulation of gene transcription is primarily mediated by the remodeling of chromatin in near proximity of the TSS.

Another application of the ideas presented in this paper it to use conditional independence to identify relevant TF-target interactions from the data. A thorough analysis has been performed in Sec. 2.5 in the Supplementary Information, which shows that conditional independence highlights relevant interactions supported by the literature.

Lastly, it should be noted that the techniques presented in this paper can readily be applied to non-longitudinal data. In fact, chains and forks are also characterized by conditional independence in that case, and dynamical correlation reduces standard correlation in the case of steady-state data and multiple replicates (i.e. non-longitudinal data). Conceptually, the same remark can be made regarding single cell (sc) data, the only difference being that correlations must necessarily be computed across different cells. However, the limited depth of the currently available methods, especially for scATAC-seq [5], poses serious technical limitations.

### 5 Relationship between “balanced in A” and “selected in B”

| Relationship between “balanced in A” and “selected in B” |
|---------------------------------------------------------|
| Relationship between “balanced in B” and “selected in A” |
| $(A \rightarrow T \rightarrow G)$ | $(T \rightarrow A \rightarrow G)$ | $(T_1 \rightarrow A \rightarrow G_2)$ | $(G_2 \rightarrow A \rightarrow G_1)$ |
| $(p$-value, FC) | $(< 10^{-15}, 1.03)$ | $(< 10^{-15}, 1.03)$ | $(< 10^{-15}, 1.03)$ | $(< 10^{-15}, 1.03)$ |
| $(p$-value, FC) | $(< 10^{-15}, 1.36)$ | $(< 10^{-15}, 1.32)$ | $(< 10^{-15}, 1.03)$ | $(< 10^{-15}, 1.03)$ |

**Table 4.** To test if there exists a relationship between which regulatory motifs are balanced (resp. selected) in dataset A and B we performed an hypergeometric test that compares the ratio of balanced (resp. selected) regulatory motifs in dataset A with the same quantity but when we restrict only to regulatory motifs that are also balanced (resp. selected) in dataset B. FC indicates the fold change of the latter with respect to the former quantity.

**Funding**

This work was supported by the Swedish Foundation for Strategic Research [SB16-0011].
References

[1] A. M. Ackermann, Z. Wang, J. Schug, A. Naji, and K. H. Kaestner. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Molecular Metabolism*, 5(3):233–244, 2016.

[2] K. Baba, R. Shihata, and M. Sibuya. Partial Correlation and Conditional Independence As Measures of Conditional Independence. *Australian and New Zealand Journal of Statistics*, 46(4):657–664, 2004.

[3] J. A. Cadzow and O. M. Solomon. Linear modeling and the coherence function. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(1):19–28, 1987.

[4] D. Calderon, M. L. Nguyen, A. Mezger, A. Kathiria, D. A. Knowles, Z. Gao, F. Blanchie, A. V. Parent, T. D. Burt, M. S. Anderson, L. A. Criswell, W. J. Greenleaf, A. Marson, and J. K. Pritchard. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics*, 51(10):1494–1505, 2019.

[5] H. Chen, C. Lareau, T. Andreani, M. E. Vinyard, S. P. Garcia, K. Clement, M. A. Andrade-Navarro, J. D. Buenrostro, and L. Paciello. Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biology*, 20(1):1–25, 2019.

[6] A. Chomukrahall and P. Matthias. The interplay between chromatin and transcription factor networks during B cell development: Who pulls the trigger first? *Frontiers in Immunology*, 5(APR):1–11, 2014.

[7] M. B. Christopher. *Pattern recognition and machine learning*. Springer, 2006.

[8] M. R. Coree, J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti, and H. Y. Chang. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics*, 48(10):1201–1206, 2016.

[9] G. Facchetti, G. Iacono, G. De Palo, and C. Altafini. A rate-distortion theory for gene regulatory networks and its application to logic gate consistency. *Bioinformatics (Oxford, England)*, 29(9):1166–1173, 2013.

[10] J. P. Fullard, M. E. Haasberg, J. Bendil, G. Egereri, M. D. Cirmaru, S. M. Reach, J. Motl, M. E. Ehrlisch, Y. L. Hurd, and P. Rousson. An atlas of chromatin accessibility in the adult human brain. *Genome Research*, 28(8):1243–1252, 2018.

[11] C. W. Grainger. Spurious REGRESSIONS in ECONOMETRICS: A Companion to Theoretical ECONOMETRICS, 2:557–561, 2007.

[12] D. G. Hendrickson, I. Soifer, B. J. Wranik, D. Botstein, and R. Scott McIsaac. Simultaneous profiling of DNA accessibility and gene expression dynamics with ATAC-seq and RNA-seq. *Methods in Molecular Biology*, 1819:317–334, 2018.

[13] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini. Determining the distance to monotonicity of a biological network: A graph-theoretical approach. *IET Systems Biology*, 4(3):223–235, 2010.

[14] J. S. Johnson, N. De Veaux, A. W. Rives, X. Lahaye, S. Y. Lucas, B. P. Perot, M. Luka, V. Garcia-Paredes, L. M. Amorn, A. Watters, G. Abdessalem, A. Aderem, N. Manel, D. R. Littman, R. Bonneau, and M. M. Méner. A Comprehensive Map of the Monocyte-Derived Dendritic Cell Transcriptional Network Engaged upon Innate Sensing of HIV. *Cell Reports*, 30(3):914–931.e9, 2020.

[15] B. Li, M. Carrey, and J. L. Workman. The Role of Chromatin during Transcription. *Cell*, 128(4):707–719, 2007.

[16] P. Li and W. J. Leonard. Chromatin accessibility and interactions in the transcriptional regulation of T cells. *Frontiers in Immunology*, 9(NOV):1–8, 2018.

[17] A. Ma'ayan, A. Lippestat, R. Iyengar, and E. Santag. Proximity of intracellular regulatory networks to monotone systems. *IET Systems Biology*, 2(3):103, 2008.

[18] R. Magnusson, O. Rundquist, M. J. Kim, S. Hellberg, C. H. Na, M. Benzon, D. Gomez-Cabero, I. Kockum, J. Tegnér, F. Pichl, M. Jagodic, J. Mebergård, C. Altafini, J. Ernérudh, M. Jenmalm, C. Neutor, M.-S. Kim, and M. Gustafsson. A validated strategy to infer protein biomarkers from RNA-Seq by combining multiple mRNA splice variants and time-delay. *bioRxiv*, 599573, 2019.

[19] S. Mangan and U. Alon. 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):11980–11985, 2003.

[20] E. R. Miraldi, M. Pokrovskii, A. Watters, D. M. Castro, N. De Vauxs, J. A. Hall, J. Y. Lee, M. Cifaini, A. Madar, N. Carriero, D. R. Littman, and R. Bonneau. Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Research*, 29(3):449–463, 2019.

[21] R. Opgen-Rhein and K. Strimmer. Inferring gene dependency networks from genomic longitudinal data: a functional data approach. *Revistat*, 4(1):53–65, 2006.

[22] R. Opgen-Rhein and K. Strimmer. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, 1, 2007.

[23] R. N. Ramirez, N. C. El-Alhi, M. A. Mager, D. Wyman, S. Mangan, and U. Alon. Structure and function of the feed-forward loop network motif. *Genome Research*, 29(3):449–463, 2019.

[24] A. T. Reid, D. B. Readley, R. D. Mill, R. Sanchez-Romero, L. Q. Uddin, D. Marinazzo, D. J. Lurie, P. A. Valdés-Sosa, S. J. Hanon, B. B. Biswal, V. Calhoun, R. A. Poldrack, and M. W. Cole. Advancing functional connectivity research from association to causation. *Nature Neuroscience*, 22(11):1751–1760, 2019.

[25] R. Sadhukhesh, E. Vidal, P. Serra, B. De Stefano, F. Le Dily, J. Quilez, A. Gomez, S. Collonhuet, C. Berenguer, Y. Cuartero, J. Hecht, J. G. Filion, M. Beato, M. A. Martí-Remon, and T. Graf. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nature Genetics*, 50(2):238–249, 2018.

[26] R. R. Starks, A. Biswas, A. Jain, and G. Tuteja. Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. *Epigenetics and Chromatin*, 12(1):1–16, 2019.

[27] J. Wu, J. Xu, B. Liu, G. Yao, P. Wang, Z. Lin, B. Huang, X. Wang, T. Li, S. Shi, N. Zhang, F. Duan, J. Ming, X. Zhang, W. Niu, W. Song, H. Jin, Y. Guo, S. Dai, L. Hu, L. Fang, Q. Wang, Y. Li, W. Li, J. Na, W. Xie, and Y. Sun. Chromatin analysis in human early development reveals epigenic transition during ZGA. *Nature*, pages 1807–1813, 2020.
[28] F. Yan, D. R. Powell, D. J. Curtis, and N. C. Wong. From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. *Genome biology*, 21(1):22, 2020.

[29] Z. Yiming, Y. Guoqiang, T. Mahlet G., and R. Habtom W. Biological network inference using low order partial correlation. *Methods*, 69(3):266–273, 2014.

[30] G. Yu, L. G. Wang, and Q. Y. He. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14):2382–2383, 2015.

[31] G. U. Yule. Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society*, 89(1):1, 1926.

[32] M. Zampieri, G. Legname, D. Segrè, and C. Altafini. A system-level approach for deciphering the transcriptional response to prion infection. *Bioinformatics*, 27(24):3407–3414, 2011.