HHS Public Access

Author manuscript

Curr Opin Genet Dev. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

Curr Opin Genet Dev. 2016 April ; 37: 101–108. doi:10.1016/j.gde.2016.02.002.

Understanding transcriptional regulatory networks using computational models

Bing He[1,*] and **Kai Tan**[1,2,*]

[1]Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA 52242, USA

[2]Department of Internal Medicine, University of Iowa, Iowa City, IA 52242, USA

Abstract

Transcriptional regulatory networks (TRNs) encode instructions for animal development and physiological responses. Recent advances in genomic technologies and computational modeling have revolutionized our ability to construct models of TRNs. Here, we survey current computational methods for inferring TRN models using genome-scale data. We discuss their advantages and limitations. We summarize representative TRNs constructed using genome-scale data in both normal and disease development. We discuss lessons learned about the structure/function relationship of TRNs, based on examining various large-scale TRN models. Finally, we outline some open questions regarding TRNs, including how to improve model accuracy by integrating complementary data types, how to infer condition-specific TRNs, and how to compare TRNs across conditions and species in order to understand their structure/function relationship.

Introduction

Gene expression can be regulated at multiple steps along the process, including transcriptional initiation and elongation, RNA stability, and accessibility and rate of translation. In this review, we focus on regulation of transcriptional initiation by the action of transcription factors (TFs) and *cis*-regulatory DNA elements. For the purpose of discussion, we define transcriptional regulatory networks (TRNs) as regulatory interactions among TFs and their target genes. Edges in a TRN thus represent direct interactions between a TF and its target genes. Models of TRNs provide systems-level explanation of developmental and physiological functions. Accurate knowledge about TRNs can benefit a range of basic and applied biomedical researches. It can help us better understand the molecular mechanisms of development and cellular reprogramming, which can lead to better strategies to generate various cell types for regenerative therapies. Mechanisms of diseases that are characterized by dysfunction of TRNs can also be elucidated. Knowledge about TRNs can also guide

Corresponding author: Tan, Kai (tank1@email.chop.edu).

*Current address: Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia; Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

selection of novel drug targets and development of efficient strategies for cellular engineering.

TRNs are highly complex as reflected by the number of regulatory components and complicated connectivity patterns among the components. They typically exhibit highly dynamic and often nonlinear behaviors in response to external or internal signals. For this reason, computational modeling is an essential component in TRN research.

In recent years, high-throughput technologies have greatly expanded our ability to collect complementary data types that can be used for computational modeling of TRNs. This article reviews recent developments and discusses their implications for future research. We start by reviewing computational methods for inferring TRN models using genome-scale data sets. TRNs inferred using large-scale data tend to have less precision but can be used to infer network components and wiring, which is particularly critical for largely uncharacterized TRNs. We discuss their relative advantages and limitations. Next, we summarize representative TRNs constructed using large-scale approaches in both normal and disease development. We then discuss insights into the structure/function relationship of TRNs, based on examining various large-scale TRN models. Finally we outline some open questions regarding TRNs, including how to integrate heterogeneous data types to improve model accuracy, how to infer condition-specific TRNs, and how to compare TRNs under different conditions and across species to understand their structure/function relationship.

Computational approaches for constructing genome-wide TRN models

Current approaches to constructing computational models of TRNs can be grouped into three classes based on the type of data used for inference (Figure 1). A list of methods discussed in this review along with their availability is provided in Table 1. The first class utilizes gene expression data as the only input[1–8]. Because they start with the regulatory output (e.g. expression level), collectively, this class of methods is known as the reverse engineering approach. A number of methods in this class have been developed using various computational frameworks, including linear regression[5,9–13], statistical correlation[2,3,14–16], and Bayesian network[17–21]. The basic assumption of regression-based approaches is that the expression levels of the TFs that directly regulate a target gene are the most informative, among all TFs, to predict the expression level of the target gene. When the expression level of a target gene is regressed on the expression levels of TFs, a non-zero regression coefficients indicate statistical dependency which in turn is interpreted as a regulatory interaction between the TF and the gene. Because there are many candidate TFs to consider in a regression, to identify the regulating TFs, a feature selection procedure is typically applied using regularized regression techniques.

Correlation-based approaches examine variation in gene expression across different conditions. Variation of gene expression in a large set of conditions provides a means to correlate statistically the expression of a specific TF with the set of expressed genes. The most commonly used correlation measures are Pearson and Spearman correlations. However, they cannot capture nonlinear relationship between two random variables. For this purpose, mutual information has been introduced[3,14]. Neither correlation measures nor

mutual information can distinguish indirect dependencies in a gene-triplet, i.e. dependency between two genes because the expression of both genes is dependent on a third gene. To remove indirectly dependencies, partial correlation coefficient and data processing inequality (DPI) measure have been introduced for correlation- and mutual information-based approaches.

Bayesian networks (BNs) are a class of probabilistic graphical models that represent statistical dependencies among a set of random variables. Formally, a BN is defined by a directed acyclic graph, a set of conditional probability distributions and their parameters, which collectively specify a joint distribution over the set of random variables. In the context of TRN modeling using BNs, nodes in a BN are genes and edges between nodes indicate the conditional dependencies between them. Learning BNs involves two steps: determining the network structure (i.e. structure learning) and estimating the conditional probability values associated with nodes (i.e. parameter learning). The computational challenge underlying BN-based methods is to exhaustively search the space of possible conditional dependencies (i.e. regulatory relationships). Given the numbers of genes and regulatory relationships among them in a typical metazoan species, a full implementation of a BN is intractable. Thus, heuristic approximation methods have been developed that used locally constraint search techniques, making the computational complexity manageable.

The major advantage of the first class of methods is their broad applicability because of the minimal requirement for input data. However, to achieve good performance, these methods typically require a large number of transcriptome profiles, usually at least as many as the number of TFs under study[22]. However, in most studies, sample size is much smaller than the number of TFs due to high experimental cost. Limited sample size makes the correlations between genes sensitive to noise, and thus highly correlated gene pair needs not imply a true regulatory relationship. A recent study assessed 35 reverse engineering methods using both experimental (*E. coli* and *S. cerevisiae*) and computationally simulated data[22]. The study revealed that no single inference method performs optimally across all data sets. In contrast, integration of predictions from multiple inference methods shows robust and high performance across diverse data sets.

Chromatin immunoprecipitation (ChIP) coupled with high-throughput techniques, such as sequencing or microarray (ChIP-Seq/Chip, hereafter refer to as ChIP-X) can provide genome-wide occupancy information for a given TF. Such data has become increasingly abundant in recent years. Although helpful, the utility of ChIP-X data alone for inferring regulatory interactions is limited because binding events detected by ChIP-X is only necessary but not sufficient for functional regulatory interactions. To address this shortcoming, the second class of methods combines gene expression profiling data with TF ChIP-X data to infer TRNs[23–30]. These methods fall into two categories in terms of their assumptions and approaches. In the first category, the methods identify subsets of ChIP-X binding sites for which the regulated genes have highly correlated expression profiles, and thus are co-regulated[23,28]. Methods in the second category use various regression techniques to fit ChIP-X binding data to the observed gene expression profiles in order to infer regulatory interactions[24–26,30]. Common to these approaches, a linear relationship between gene expression changes and TF binding affinities is assumed.

All class II methods except that by Maienschein-Cline *et al.* use the gene nearest to a ChIP-X binding site as the candidate regulated gene. This approach works well for compact genomes (e.g. bacteria and yeasts) and promoter-proximal TF binding sites. However, for metazoan species, functional TF binding sites may not reside next to their targets. A number of recent studies using chromosome confirmation capture (3C) based technologies have shown that the majority of enhancers do not target genes closest to them[31–33]. For this reason, accurate identification of enhancer-promoter interactions becomes a critical first step towards constructing accurate TRNs. Once the enhancer-promoter pairing is predicted, DNA motif analysis of the enhancer sequences is used to infer regulatory interactions between TFs that bind the enhancers and the paired gene promoters. The enormous amount of public data generated by projects such as ENCODE and Epigenomics Roadmap has opened up the door for integrative approaches to constructing TRNs for metazoan species. These integrative approaches represent the third class of methods for modeling TRNs. At the core of these methods is a strategy for linking transcriptional enhancers with target promoters. A general assumption of these methods is that the chromatin states of *bona fide* enhancers-promoter pairs tend to be correlated across cell/tissue types. Under this general assumption, different genome-wide chromatin state data have been used, including histone modifications[34] and chromatin openness (as measured by DNase I hypersensitivity)[35]. Further development along this line has correlated chromatin state of enhancers with expression profiles of promoters[36,37]. He *et al.*[38] took this approach further and identified three additional genomic features in addition to chromatin state correlation, including co-expression between promoter and genes encoding transcription factors that occupy the enhancer under consideration, sequence co-evolution between enhancer and promoter. These features when combined with chromatin state correlation were shown to significantly improve the inference accuracy of enhancer-promoter pairs.

Large-scale mapping of TRNs during normal development

In the past several decades, through the effort of individual labs and large research consortia, large-scale TRN models have been constructed for various metazoan species. The TRN controlling sea urchin endomesoderm patterning is the largest and most extensively validated developmental TRN to date[39]. In *C. elegans*, TRN for the intestine has been constructed using yeast one-hybrid assay[40]. At a more global scale, ChIP-Seq was performed for 92 transcription factors spanning 11 developmental stages by the modENCODE consortium. Integration of ChIP-Seq and expression data produced a spatiotemporally resolved TRNs for this species[41]. In *D. melanogasters*, dorsal/ventral and anterior/posterior patterning have been studied extensively. TRNs for these two developmental processes have been mapped using ChIP-X assays and computational modeling[42–45]. Like *C. elegans*, the modENCODE project also generated TRN models based ChIP-Seq data for 38 fly TFs in different developmental stages and cell types[46]. In mammalian species, TRNs in various cell types have been constructed by individual labs, including dendritic cell[47], macrophage[48], embryonic stem cell[49–52], hematopoietic stem cell[53–55], B cells[1], Th17 cells[56,57] and T-cell fate specification[58]. By integrating TF ChIP-Seq data, gene expression profiles, and chromatin modification ChIP-Seq data, the

ENCODE and mouse ENCODE projects have also generated TRN models for various human and mouse cells/tissues[36,59–61]

TRNs in diseased cells

Given the key role of transcriptional regulation in development and cellular homeostasis, it is not surprising that perturbations to TRNs can lead to many diseases. Such perturbations include mutations in regulatory DNA sequences and in transcription factors, co-factors, and chromatin regulators. In order to understand the roles of mutations in pathogenesis, the underlying disease-specific TRNs in which the mutated factors operate need to be defined. Nowadays, the same high throughput technologies used to interrogate TRNs in normal cells are increasingly being applied to study TRNs underlying various human diseases especially cancer[62–69]. The developing insight is that the same TRNs in normal cells are either rewired or acquire altered activities in diseased cells to promote pathogenesis.

Reverse engineering approaches have been applied to construct TRNs in cancer cells. Carro *et al.* constructed a glioma-specific TRNs and identified two TFs (C/EBPβ and STAT3) as synergistic master regulators of oncogenic transformation[62]. Gatta *et al.*[67]decipher the oncogenic TRNs controlled by the two TFs, TLX1 and TLX3, in T cell acute lymphoblastic leukemia (T-ALL).

As ChIP-X technology became mature and more sensitive, they have been increasingly used to map TRNs in cancer cells. A recent study revealed that the oncogenic TF TAL1 forms an interconnected autoregulatory loop with two TF partners (RUNX1 and GATA3) in T-ALL. This circuitry contributes to the sustained activation of TAL1-regulated oncogenic program[66]. Many oncogenic TFs are generated by chromosomal translocation events. A well-known one is the RUNX1/ETO fusion TF, which is generated by the chromosomal translocation t(8;21)[70]. Using ChIP-Seq and expression profiling, Ptasinska *et al.*[68] showed that the transcriptional program underlying leukemic propagation is regulated by a dynamic equilibrium between the TRNs regulated by the RUNX1/ETO fusion TF and intact RUNX1 complexes. Using the same approach, TRNs in breast cancer[64], prostate cancer[69], and lung cancer[63] have also been studied.

Structure/function relationship of TRNs

Previous effort on mapping TRNs (both small-scale and large-scale) has yielded a large number of TRN models. Analysis of available TRN architectures has revealed the following organizational principles. First, TRNs have a global hierarchical topology[36,59,71]. TRNs for embryonic development of animal body parts, such as that for specification of endomesoderm in sea urchin[39], for specification of gut and mesoderm in *C. elegans*[72], and for specification of eye lens field in zebra fish[73,74], tend to have deep hierarchical organization. In comparison, TRNs for terminal fate choice from multipotent stem cells and physiological responses, such as that for specification of erythroid versus myeloid fates bifurcate[75,76], T helper versus killer cells diversification[77], innate immunity response[78], are relatively shallow. The structural difference between embryonic development TRNs and terminal fate TRNs reflects the difference in their functional requirement. Development of

the body plan requires a long sequence of progressive decisions that need to be made in different spatiotemporal domains. In contrast, terminal cell fate can be specified with fewer decisions.

Embedded in the global hierarchical topology of TRNs are many over-represented small connectivity patterns, the so-called network motifs[79,80]. Different types of network motifs are characterized by their different connectivity patterns and distinct functions associated with them. Some of the simplest and most prevalent motifs are feedback loops, either positive or negative. Positive feedback loops are often observed in systems that show switch-like behavior, memory, or bistability. Negative feedback loops are functionally associated with systems that show strong noise resistance to perturbations. A slightly more complex motif is feedforward loop (FFL). This motif consists of three genes: a TF, X, which regulates TF Y, and gene Z, which is regulated by both X and Y. FFL motifs fall into two classes depending on the net sign of the regulatory actions of the two arms of the motif. Coherent FFL motifs have two arms with the same net sign of actions whereas incoherent FFL motifs have two arms with different net signs of action. Coherent FFLs has been shown to filter out brief spurious pulses of signal. Thus, gene Z only respond in the presence of persistent signal that is over its threshold, instead of a brief signal[81,82]. In incoherent FFLs, the two arms of the FFL act in opposition. They have been shown to generate pulse-like response dynamics by Z after X is activated[83]. For detailed treatment of other types of network motifs and their structure/function relationship, readers are referred to a number of excellent reviews on this topic[80,84,85].

Summary and Outlook

A large number of studies have demonstrated the enormous value that the assembly of TRN models has in understanding normal and disease development. In this review, we only considered networks involving transcription factors and their target genes. However, these networks are part of a much larger and more complex cellular network composed of many other types of molecules and their interactions. For this reason, combining multiple, independently generated observations (such as gene expression, *in vivo* TF binding and chromatin modification states, protein abundance measure, physical and genetic interactions among genes) to infer network structure can strengthen the resulting models and provide novel insights. Although the majority of current computational methods use only gene expression and/or ChIP-X data as the input to infer TRNs, more integrative methods have been developed in the past few years. For instance, researchers have integrated chromatin modification ChIP-Seq data to construct TRNs (e.g. class III approach). Such integrative approaches will become increasingly powerful as more data becomes available.

Most computational methods and TRN models discussed in this review are global networks. That is, regulatory interactions in these networks are not specific (or not specific enough) to a particular phenotype under study. Such condition-specific-interactions are critical for better understanding the behavior of the network. Advanced methods (both computational and experimental) are needed to allow capturing more nuanced network models.

As models of TRNs start to accumulate rapidly, novel computational methods are needed to allow principled comparisons of TRNs to gain insights into the structure, function and evolution relationship of TRNs. For instance, comparing TRNs of normal and diseased cells will be particularly fruitful for understanding the molecular mechanisms of pathogenesis. Similarly, comparing developmental TRNs across species will provide valuable insights into their evolution and organization principles.

Acknowledgments

B.H. was supported by National Institutes of Health grants HG006130. K.T. was supported by National Institutes of Health grants HG006130, GM104369, GM108716.

References

Papers of particular interest, published within the period of review,

Have been highlighted as:

• **of special interest**

•• **of outstanding interest**

1. Basso K, et al. Reverse engineering of regulatory networks in human B cells. Nat Genet. 2005; 37:382–390. [PubMed: 15778709]
2. Faith JJ, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007; 5:e8. [PubMed: 17214507]
3. Margolin AA, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7(Suppl 1):S7. [PubMed: 16723010] This paper presents a popular reverse engineering method based on mutual information. The method has been used in a number of successful studies of deregulated TRNs in human cancers.
4. Belcastro V, et al. Transcriptional gene network inference from a massive dataset elucidates transcriptome organization and gene function. Nucleic acids research. 2011; 39:8677–8688. [PubMed: 21785136]
5. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. Science. 2003; 301:102–105. [PubMed: 12843395] This paper presents the first reverse engineering method to identify TRN governing the SOS response in *E. coli*.
6. Bonneau R, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome biology. 2006; 7:R36. [PubMed: 16686963]
7. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. Reconstructing dynamic regulatory maps. Molecular systems biology. 2007; 3:74. [PubMed: 17224918]
8. Segal E, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet. 2003; 34:166–176. [PubMed: 12740579]
9. Della Gatta G, et al. Direct targets of the TRP63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. Genome Res. 2008; 18:939–948. [PubMed: 18441228]
10. Yeung MK, Tegner J, Collins JJ. Reverse engineering gene networks using singular value decomposition and robust regression. Proc Natl Acad Sci U S A. 2002; 99:6163–6168. [PubMed: 11983907]
11. van Someren EP, et al. Least absolute regression network analysis of the murine osteoblast differentiation network. Bioinformatics. 2006; 22:477–484. [PubMed: 16332709]
12. Haury AC, Mordelet F, Vera-Licona P, Vert JP. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. BMC Syst Biol. 2012; 6:145. [PubMed: 23173819]

13. Haynes BC, et al. Mapping functional transcription factor networks from gene expression data. Genome Res. 2013; 23:1319–1328. [PubMed: 23636944]
14. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000:418–429. [PubMed: 10902190]
15. Luo W, Hankenson KD, Woolf PJ. Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information. BMC Bioinformatics. 2008; 9:467. [PubMed: 18980677]
16. Soranzo N, Bianconi G, Altafini C. Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks: synthetic versus real data. Bioinformatics. 2007; 23:1640–1647. [PubMed: 17485431]
17. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol. 2000; 7:601–620. [PubMed: 11108481] This paper represents one of the first attemp of using Bayesian network to model TRNs using gene expression data in baker's yeast.
18. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. Pac Symp Biocomput. 2001:422–433. [PubMed: 11262961]
19. Pe'er D, Regev A, Elidan G, Friedman N. Inferring subnetworks from perturbed expression profiles. Bioinformatics. 2001; 17(Suppl 1):S215–S224. [PubMed: 11473012]
20. Husmeier D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. Bioinformatics. 2003; 19:2271–2282. [PubMed: 14630656]
21. Zou M, Conzen SD. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics. 2005; 21:71–79. [PubMed: 15308537]
22. Marbach D, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012; 9:796–804. [PubMed: 22796662] This paper presents a comprehensive evaluation of 35 reverse engineering methods for modeling TRNs. They demonstrated that community network prediction outperforms individual methods.
23. Bar-Joseph Z, et al. Computational discovery of gene modules and regulatory networks. Nat Biotechnol. 2003; 21:1337–1342. [PubMed: 14555958] This paper presents the GRAM algorithm, one of the first algorithm to infer TRNs by integrating gene expression profiles and TF ChIP-Chip data. The method was applied to 106 TF ChIP-Chip data and 500 expression profiles to construct a genome-wide TRN in baker's yeast.
24. Gao F, Foat BC, Bussemaker HJ. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. BMC Bioinformatics. 2004; 5:31. [PubMed: 15113405]
25. Boulesteix AL, Strimmer K. Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. Theor Biol Med Model. 2005; 2:23. [PubMed: 15978125]
26. Sanguinetti G, Rattray M, Lawrence ND. A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription. Bioinformatics. 2006; 22:1753–1759. [PubMed: 16632490]
27. Maienschein-Cline M, Zhou J, White KP, Sciammas R, Dinner AR. Discovering transcription factor regulatory targets using gene expression and binding data. Bioinformatics. 2012; 28:206–213. [PubMed: 22084256]
28. Wu G, Ji H. ChIPXpress: using publicly available gene expression data to improve ChIP-seq and ChIP-chip target gene ranking. BMC Bioinformatics. 2013; 14:188. [PubMed: 23758851]
29. Karlebach G, Shamir R. Constructing logical models of gene regulatory networks by integrating transcription factor-DNA interactions with expression data: an entropy-based approach. J Comput Biol. 2012; 19:30–41. [PubMed: 22216865]
30. Liao JC, et al. Network component analysis: reconstruction of regulatory signals in biological systems. Proc Natl Acad Sci U S A. 2003; 100:15522–15527. [PubMed: 14673099]

31. Li G, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. doi:S0092-8674(11)01517-0 [pii]10.1016/j.cell.2011.12.014. [PubMed: 22265404]
32. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012; 489:109–113. [PubMed: 22955621] This study applied chromosome conformation capture carbon copy to interrogate interactions between gene promoters and distal elements in 1% of the human genome. The authors found only ~27% of looping interactions are with the nearest gene, indicating that genomic proximity is a poor predictor for enhancer-promoter interactions.
33. Jin F, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013; 503:290–294. [PubMed: 24141950]
34. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. doi:nature09906 [pii]10.1038/nature09906. [PubMed: 21441907]
35. Thurman RE, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]
36. Gerstein MB, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012; 489:91–100. [PubMed: 22955619]
37. Corradin O, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 2014; 24:1–13. [PubMed: 24196873]
38. He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. Proc Natl Acad Sci U S A. 2014; 111:E2191–E2199. [PubMed: 24821768] This paper presents a state-of-the-art algorithm to predict enhancer-promoter interactions using multiple genomic features. Systematic analysis of enhancer-promoter interactome in 12 human cell types revealed global features of enhancer-promoter communication.
39. Peter IS, Davidson EH. Modularity and design principles in the sea urchin embryo gene regulatory network. FEBS Lett. 2009; 583:3948–3958. [PubMed: 19932099]
40. Deplancke B, et al. A gene-centered C. elegans protein-DNA interaction network. Cell. 2006; 125:1193–1205. doi:S0092-8674(06)00624-6 [pii] 10.1016/j.cell.2006.04.038. [PubMed: 16777607]
41. Araya CL, et al. Regulatory analysis of the C. elegans genome with spatiotemporal resolution. Nature. 2014; 512:400–405. [PubMed: 25164749]
42. Zeitlinger J, et al. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev. 2007; 21:385–390. [PubMed: 17322397]
43. Sandmann T, et al. A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev. 2007; 21:436–449. [PubMed: 17322403]
44. MacArthur S, et al. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol. 2009; 10:R80. doi:gb-2009-10-7-r80 [pii]10.1186/gb-2009-10-7-r80. [PubMed: 19627575]
45. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature. 2009; 462:65–70. [PubMed: 19890324]
46. Negre N, et al. A cis-regulatory map of the Drosophila genome. Nature. 2011; 471:527–531. doi:nature09990 [pii]10.1038/nature09990. [PubMed: 21430782]
47. Miller JC, et al. Deciphering the transcriptional network of the dendritic cell lineage. Nat Immunol. 2012; 13:888–899. [PubMed: 22797772]
48. Gilchrist M, et al. Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. Nature. 2006; 441:173–178. [PubMed: 16688168]
49. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. doi:S0092-8674(05)00825-1 [pii]10.1016/j.cell.2005.08.020. [PubMed: 16153702]
50. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008; 133:1106–1117. doi:S0092-8674(08)00617-X [pii]10.1016/j.cell.2008.04.043. [PubMed: 18555785]

51. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–1061. [PubMed: 18358816]
52. Wang J, et al. A protein interaction network for pluripotency of embryonic stem cells. Nature. 2006; 444:364–368. [PubMed: 17093407]
53. Wilson NK, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell. 2010; 7:532–544. doi:S1934-5909(10)00440-6 [pii]10.1016/j.stem.2010.07.016. [PubMed: 20887958]
54. Beck D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood. 2013; 122:e12–e22. [PubMed: 23974199]
55. Moignard V, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol. 2015; 33:269–276. [PubMed: 25664528]
56. Ciofani M, et al. A validated regulatory network for Th17 cell specification. Cell. 2012; 151:289–303. [PubMed: 23021777]
57. Yosef N, et al. Dynamic regulatory network controlling TH17 cell differentiation. Nature. 2013; 496:461–468. [PubMed: 23467089]
58. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell. 2012; 149:467–482. [PubMed: 22500808]
59. Neph S, et al. Circuitry and dynamics of human transcription factor regulatory networks. Cell. 2012; 150:1274–1286. [PubMed: 22959076]
60. Stergachis AB, et al. Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature. 2014; 515:365–370. [PubMed: 25409825]
61. Yue F, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014; 515:355–364. [PubMed: 25409824]
62. Carro MS, et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature. 2010; 463:318–325. [PubMed: 20032975]
63. Kumar MS, et al. The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. Cell. 2012; 149:642–655. [PubMed: 22541434]
64. Eeckhoute J, Carroll JS, Geistlinger TR, Torres-Arzayus MI, Brown M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. Genes Dev. 2006; 20:2513–2526. [PubMed: 16980581]
65. Palomero T, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci U S A. 2006; 103:18261–18266. [PubMed: 17114293]
66. Sanda T, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell. 2012; 22:209–221. [PubMed: 22897851] This study is an excellent example of combining gene expression profile and ChIP-Seq to map deregulated TRN in cancer cell. The authors identified an autoregulatory loop involving TAL-1 and two other TFs. This subnetwork contributes to the sustained activation of TAL1-regulated oncogenic program.
67. Della Gatta G, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. Nat Med. 2012; 18:436–440. [PubMed: 22366949] This study identified RUNX1 as a tumor suppressor gene in T-ALL. It demonstrates the power of reverse-engineering methods to identify key deregulated regulatory circuits in human cancer.
68. Ptasinska A, et al. Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. Cell Rep. 2014; 8:1974–1988. [PubMed: 25242324]
69. Tan PY, et al. Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. Mol Cell Biol. 2012; 32:399–414. [PubMed: 22083957]
70. Peterson LF, Zhang DE. The 8;21 translocation in leukemogenesis. Oncogene. 2004; 23:4255–4262. [PubMed: 15156181]

71. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010; 468:911–920. [PubMed: 21164479] This is an excellent review on structure/function relationship of TRNs for animal development.
72. Maduro MF. Structure and evolution of the C. elegans embryonic endomesoderm network. Biochim Biophys Acta. 2009; 1789:250–260. [PubMed: 18778800]
73. Chan TM, et al. Developmental gene regulatory networks in the zebrafish embryo. Biochim Biophys Acta. 2009; 1789:279–298. [PubMed: 18992377]
74. Morley RH, et al. A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation. Proc Natl Acad Sci U S A. 2009; 106:3829–3834. [PubMed: 19225104]
75. Zhang P, et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci U S A. 1999; 96:8705–8710. [PubMed: 10411939]
76. Stopka T, Amanatullah DF, Papetti M, Skoultchi AI. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. EMBO J. 2005; 24:3712–3723. [PubMed: 16222338]
77. Rothenberg EV. Decision by committee: new light on the CD4/CD8-lineage choice. Immunol Cell Biol. 2009; 87:109–112. [PubMed: 19079359]
78. Amit I, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science. 2009; 326:257–263. [PubMed: 19729616]
79. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002; 31:64–68. [PubMed: 11967538]
80. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8:450–461. doi:nrg2102 [pii]10.1038/nrg2102. [PubMed: 17510665] This is an excellent review on network motifs and their functions, with an emphasis on experimental studies.
81. Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J Mol Biol. 2003; 334:197–204. [PubMed: 14607112]
82. Kalir S, Mangan S, Alon U. A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. Mol Syst Biol. 2005; 1 2005 0006.
83. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R. Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A. 2004; 101:6355–6360. [PubMed: 15096621]
84. Sneppen K, Krishna S, Semsey S. Simplified models of biological networks. Annu Rev Biophys. 2010; 39:43–59. [PubMed: 20192769]
85. Lim WA, Lee CM, Tang C. Design principles of regulatory networks: searching for the molecular algorithms of the cell. Mol Cell. 2013; 49:202–212. [PubMed: 23352241]

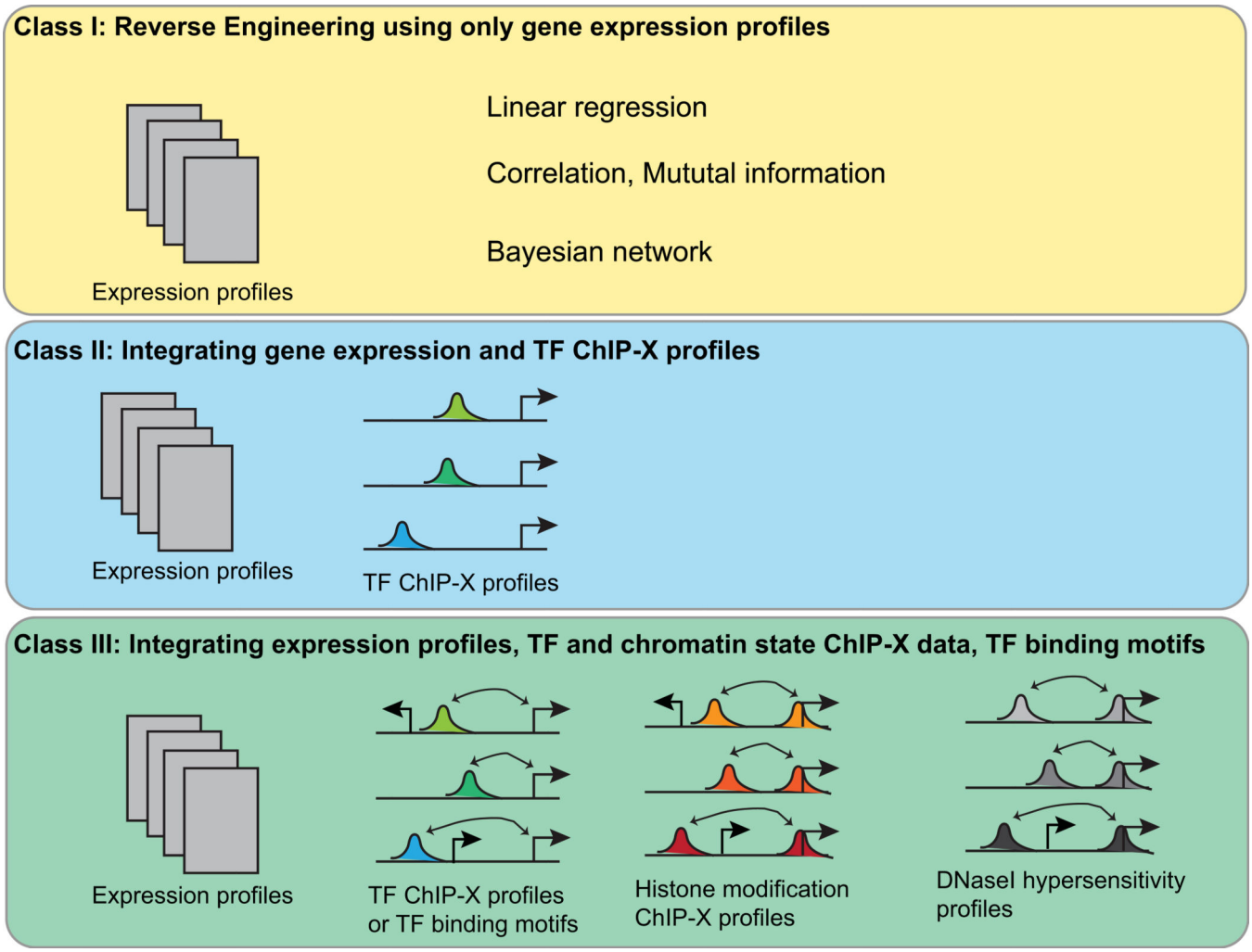

Figure 1. Three classes of computational methods for inferring transcriptional regulatory networks

ChIP-X represents ChIP-ChIP or ChIP-Seq protocol. In the panel representing class III methods, arcs over genome loci and transcription start sites indicate enhancer-promoter links.

Table 1
List of computational methods for modeling transcriptional regulatory networks

Class I, methods that only use gene expression profiles as their input data; Class II, methods that integrate both gene expression and transcription factor ChIP- X data. Class III, methods that integrate gene expression, transcription factor, and chromatin interaction ChIP-X data. These methods explicitly address a critical subproblem of TRN modeling, i.e. identifying enhancer-promoter interactions.

Method	Download link	Reference
	Class I	
ARACNe	http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/ARACNE	1
Inferelator	https://sites.google.com/a/nyu.edu/inferelator/home	6
DREM	http://www.sb.cs.cmu.edu/drem	7
GeneXPress	http://robotics.stanford.edu/~erans/module_nets	8
TSNI	http://dibernardo.tigem.it/softwares/time-series-network-identification-tsni	9
GENLAB	http://genlab.tudelft.nl/larna.html	11
TIGRESS	http://cbio.ensmp.fr/~ahaury/svn/dream5/html/index.htm	12
NetProphet	http://mblab.wustl.edu/software.html	13
MI3	http://sysbio.engin.umich.edu/~luow/downloads.php	15
NetRec	http://www.sissa.it/fa/altafini/papers/SoBiAl07	16
DBmcmc	http://www.bioss.ac.uk/people/dirk/Supplements	20
	Class II	
GRAM	http://groups.csail.mit.edu/cgs/onePageGraml	23
MA-Networker	http://bussemaker.bio.columbia.edu/papers/MA-Networker	24
plsgenomics	https://cran.r-project.org/web/packages/plsgenomics/index.html	25
PUMA	http://umber.sbs.man.ac.uk/resources/puma	26
EMBER	http://dinnergroup.uchicago.edu/downloads.html	27
ChIPXpress	http://www.biostat.jhsph.edu/~gewu/ChIPXpress	28
ModEnt	http://acgt.cs.tau.ac.il/modent	29
NCA	http://www.seas.ucla.edu/~liaoj/downloads.html	30
	Class III	
PreSTIGE	http://prestige.case.edu	37
IM-PET	http://www.healthcare.uiowa.edu/labs/tan/IM-PET.html	38