CHAPTER 54

OVERVIEW AND NEW PERSPECTIVES ON HUMAN TRANSCRIPTIONAL REGULATORY NETWORK

Aysegul CALISKAN, Kazim Yalcin ARGA

INTRODUCTION

Transcriptional regulation of gene expression is an essential cellular process that is arranged by transcription factors (TFs), non-coding RNAs (miRNAs and lncRNAs), and their target genes through a variety of mechanisms (Cao *et* al., 2015). Gene regulatory networks (GRNs) studies reveal complex life events in terms of gene interaction, an important area of research in system biology (Emmert-Streib *et* al., 2014). On the last years, with the development of high-throughput screening techniques and next generation sequencing (NGS) technologies, lots of data on gene expressions and their regulations were accumulated, and several methods have been developed to construct GRNs. These regulatory data are used in medicine and molecular biology applications such as identification of genetic diseases, identification of metabolic pathways, the discovery of new drugs, reducing side-effects of treatments, the study of expression patterns of genes with unknown function and gain ideas about their performance.

Here we aim to provide an extensive review of the transcriptional regulatory information in various databases, the methods used in network reconstruction, and present transcriptional regulatory networks of *Homo sapiens* in literature. In addition, we will demonstrate new insights on a comprehensive, genome-scale, human transcriptional regulatory network.

DATA AVAILABILITY

Pairwise data on three distinct regulatory relationships (i.e., miRNA-gene, miRNA-TF, and TF-gene interactions) at transcriptional and post-transcriptional regulation of gene expression in human are available in several publicly available databases (Table 1). The majority of the databases (i.e., Pazar, Trrust, Tfacts, HTRIdb, mir2disease, miRTarbase, TransMir, miRecords) present experimental data, whereas computational predictions are also stored in several databases, including iRegulon, miRWalk, Targetscan and miRDB. Furthermore, a few databases, such as Regnetwork and Tfacts, integrate regulatory data from different sources with computational and/or experimental origin.

Table 1. Lists of databases

Database Name (Web Link)	# of Interactions	**Regulatory relationships**	**Confidence identifier**	**References**
Regnetwork (www.regnetworkweb.org)	368573	TF/gene/ miRNA	experimentally validated / computational predictions	Liu *et al.* 2015
Pazar (www.pazar.info)	9472	TF/gene	experimentally validated	Portales-Casamar *et al.* 2008
Trrust (www.grnpedia.org/trrust/)	9398	TF/gene	experimentally validated	Han *et al.* 2017
Tfacts (www.tfacts.org)	4319	TF/gene	experimentally validated	Essaghir *et al.* 2010
HTRIdb (www.lbbc.ibb.unesp.br/htri.)	52467	TF/gene	experimentally validated	Bovolenta *et al.* 2012
iRegulon (Cytoscape plug-in)	928864	TF/gene	computational predictions	Janky *et al.* 2014
mir2disease (www.mir2disease.org)	809	miRNA/ gene	experimentally validated	Jiang *et al.* 2008
Targetscan (www.targetscan.org)	225210	miRNA/gene	computational predictions	Agarwal *et al.* 2015
miRTarBase (mirtarbase.mbc.nctu.edu.tw)	502652	miRNA/gene	experimentally validated	Chou *et al.* 2018
TransMir (www.cuilab.cn/transmir)	649	TF/miRNA	experimentally validated	Wang *et al.* 2009
miRecords (c1.accurascience.com/miRecords/)	2115	miRNA/gene	experimentally validated	Xiao *et al.* 2008
miRWalk (zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/)	825914	miRNA/gene	computational predictions	Dweep *et al.* 2011
miRDB (www.mirdb.org)	2100000	miRNA/gene	computational predictions	Wong & Wang, 2014

METHODS USED IN NETWORK RECONSTRUCTION

Several methods employing gene expression data have been developed to construct human GRNs so far. The pioneer studies in this field utilized microarray and promoter sequence data. For instance, Computational Ascertainment of Regulatory Relationships Inferred from Expression (CARRIE) is a web server that utilizes microarray and promoter sequence data for prediction of significant TF-target gene interactions by reconstructing GRNs in response to a specific stimulus (Haverty *et* al., 2004). Similarly, Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), which is an information theoretic algorithm for the reverse engineering of transcriptional networks from microarray data, was initially designed for reconstruction of human transcriptional network for B cell malignancies (Margolin *et* al., 2006). The ARACNE algorithm computes

pairwise mutual information between TFs and potential target genes and tests the significance of these interactions.

Besides gene expression data, chromatin immunoprecipitation coupled sequencing (ChIP-Seq) data was also employed to construct GRNs in the following years. For instance, the ChIP-Array was designed as a web server that integrates ChIP-Seq and gene expression data to identify direct and indirect target genes regulated by a TF of interest and to help functional characterization of a TF (Qin *et* al., 2011). The another reconstructing GRN technique that uses Chip-seq data is an integrative method that possesses three major types of regulatory interactions: TF-gene, TF-miRNA and miRNA-gene (Cheng *et* al., 2011). Target genes and target miRNAs are inferred by using RNA-Seq profiles. Cscan is another web server that uses a collection of different ChIP-Seq experiments performed on TFs, histone modifications, RNA polymerases and others to identify putative common regulators in a number of genes and to assess their transcriptional and epigenetic profiles (Zambelli *et* al., 2012). Similarly, to unravel hierarchical interactions at different regulatory levels (i.e., interrelationships among TFs, epigenetic modifications, and genes), Guan and coworkers (2014a) proposed the constructing multilevel gene regulatory networks (CMGRN) approach and presented as an integrative web server to construct hierarchical GRN structures. It enables biologists to analyze standard formatted data at ChIP-seq and gene expression levels without the much need for bioinformatics skills. More recently, disease-specific transcriptomic and epigenomic data are integrated to construct disease-relevant GRN for identifying non-coding risk variants (Gao *et* al., 2018). The constructed GRN consists of EP edges representing interactions between enhancers and target genes, and also FI edges representing the functional associations between target genes.

Guan and colleagues (2014b) reconstructed a post-translational hierarchical gene regulatory network (PTHGRN) model that aims to unravel relationships among post-translational modifications, TFs, epigenetic modifications and gene expression. Integrating protein–protein interactions (PPIs), ChIP-seq and gene expression data, it is possible to generate and score all possible interactions of protein–TF and TF/epigenetic modification-gene through the PTHGRN web server in order to capture essential regulation features. Another GRN construction approach included TF binding sites (TFBS) and PPIs among TFs to derive GRNs from mRNA expression profiles exposed to genetic perturbations by use of a Bayesian variable selection (BVS) algorithm (Santra, 2014). The main hypothesis behind this approach was that integrating PPIs among TFs with TFBS data increases the predictive power of the inference process, especially in a variable selection setting. This was demonstrated by inferring a liver-specific transcription regulatory network and the gene regulation program of human breast epithelium, and evaluating the accuracy of the inferred networks based on known interactions.

Considering the role of epigenetic regulations and post-transcriptional regulators such as non-coding RNAs, data at post-transcriptional level was also considered in reconstruction strategies in recent years. For instance, transcriptional, post-transcriptional and epigenetic data were integrated to identify specific miRNA-regulated transcription factors that explain the impact of miRNA perturbation on gene expression (Gosline *et* al., 2016). The IntegraMiR

aims to predict specific types of deregulated miRNA/TF-mediated regulatory mechanisms and networks that appear in a statistically over-represented manner in GRNs at the transcriptional, post-transcriptional and signaling levels (Afshar *et* al., 2014). For this purpose, it utilizes mRNA/miRNA expression data, sequence-based miRNA-target information, known information about mRNA and miRNA targets of TFs available in existing databases, certain three-node motifs in GRNs, and known molecular subtyping information available with gene expression data. MAGIA is another web server that uses miRNA and gene expression data to infer specific targets of miRNAs and TFs by reconstructing post-transcriptional regulatory networks (Bisognin *et* al., 2012). Very recently, Chiu and coworkers (2018) presented Cupid, which is an integrative method that uses sequence based data and RNA/miRNA expression analysis to infer potential miRNA binding sites on target gene and associated competitive endogenenous RNA interactions (Chiu *et* al., 2018).

The Regulatory Network Enrichment Analysis (RNEA) tool is also based on a collection of regulatory interactions compiled from manually curated databases (Chouvardas *et* al., 2016). RNEA uses gene expression data to find differentially expressed genes, combines prior knowledge with standard statistical methods for the inference of active regulators, miRNAs and functional categories to construct a reference network of interactions, and then uses enrichment analysis coupled with a two-level hierarchical parsing of the network to infer the most relevant subnetwork (i.e., GRN topology) for a given experimental setting. Besides RNEA, several other methods considered topological features in GRN construction. For instance, Active Protein-Gene (APG) network model is designed to reveal transcriptional regulations among TFs and targ*et* genes through integrating both TF upstream-regulation and downstream-regulation high-throughput data (Wang *et* al., 2013). More recently, Grechkin and co-workers proposed a computational framework, Differential Sparse Regulatory Network (DISCERN), which allows to identify informative topological changes in GRNs inferred on the basis of mRNA expression datasets within distinct biological states. Two expression datasets are taken as input by DISCERN: an expression dataset of diseased tissues from patients with a disease of interest and another expression dataset from matching normal tissues. Comparing distinct regulator connectivity in the inferred GRNs for the disease and normal conditions, DISCERN estimates the extent to which each gene is perturbed (Grechkin *et* al., 2016).

Although gene regulation is a dynamic process, methods employing time course expression data to construct GRNs are quite limited in literature. The Bayesian Gene Regulation Model Inference (BGRMI) approach relies on the principles of Bayesian Model Averaging (BMA) for inferring GRNs from time course gene expression data (Martinez *et* al., 2016). BGRMI uses discretized ordinary differential equation based mathematical models to formulate the interactions between each gene and its regulators. It formulates the rate of change in a gene's expression as a function of the expressions of its regulators, takes basal expression and self-regulation into account.

FUTURE PERSPECTIVES

Despite several construction efforts in the last decade (Shalgi *et* al., 2007; Qiu *et* al., 2010; Chen *et* al., 2011), still a comprehensive, genome-scale, generic human transcriptional regulatory network is not available. Very recently, we set out to construct a comprehensive transcriptional and post-transcriptional GRN of *Homo sapiens* consisting of experimentally verified regulatory information on miRNAs, TFs, and their target genes (Gov & Arga, 2016). On the other hand, considering the rapid accumulation of data from state-of-the-art and frequent updates in databases, automatic or semi-automatic tools and web servers are needed for construction of a comprehensive transcriptional and post-transcriptional human GRN. In addition, these tools or servers should be able to (i) integrate data from genomics, epigenomics, transcriptomics, and proteomics levels, (ii) take into consideration regulatory information for distinct regulators (including TFs, miRNAs and lncRNAs) and epigenetic factors, and (iii) possess user friendly interface reducing the need for bioinformatics skills.

GRNs are needed in systems biomedicine applications such as identification of efficient systems biomarkers for diagnosis and prognosis of diseases (Gov & Arga, 2017; Turanli *et* al., 2017a; Sevimoglu *et* al., 2018), the discovery of new drugs or drug repositioning (Turanli & Arga, 2017; Turanli *et* al., 2017b), and identification of molecular mechanisms of complex diseases (Karagoz *et* al., 2016; Kori *et* al., 2016; Calimlioglu *et* al., 2015; Karagoz *et* al., 2015). However, this type of GRNs are available for limited number of diseases, such as schizophrenia (Guo *et* al., 2010), glioblastoma (Sun *et* al., 2012), and a few cancers (Yu *et* al., 2012; Sengupta & Bandyopadhyay, 2013; Gov & Arga, 2016). In addition, tissue or process-specific GRNs are not existent in literature within our knowledge. Further efforts are needed to develop methods in construction of disease, tissue, and process-specific GRNs.

REFERENCES

1. A. Bisognin, G. Sales,A. Coppe, S. Bortoluzzi,C. Romualdi, "MAGIA2: from miRNA and genes expression data integrative analysis to micro- RNA-transcription factor mixed regulatory circuits (2012 update)." Nucleic Acids Research, 40, W13–W21, 2012.
2. A. Essaghir, F. Toffalini, L. Knoops, A. Kallin, J. Helden, J. B. Demoulin, "Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data", Nucleic Acids Research, Vol. 38, No. 11, 2010.
3. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, A. Califano, "ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context, BMC Bioinformatics, 7(Suppl 1):S7, 2006.
4. A. S. Afshar, J. Xu, J. Goutsias, "Integrative identification of de- regulated miRNA/TF-mediated gene regulatory loops and networks in pros- tate cancer.", PLoS ONE, 9(6):1-24, 2014.
5. A. Y. Guo, J. Sun, P. Jia, "A novel microRNA and transcription factor mediated regulatory network in schizophrenia", BMC Systems Biology., 4, p.10, 2010.
6. B. Calimlioglu,K. Karagoz, T. Sevimoglu, E. Kilic, E. Gov, K. Y. Arga, "Tissue-Specific Molecular Biomarker Signatures of Type 2 Diabetes: An Integrative Analysis of Transcriptomics and Protein–Protein Interaction Data.", OMICS: A Journal of Integrative Biology, 19(9): 563-573, 2015.

7. B. Turanli, G. Gulfidan, K. Y. Arga, "Systems Biomedicine: From Genome Scale Molecular Networks to Efficacious Biomarkers and Therapeutic Targets." In Arapligirlioglu H.,Atik A., Elliott R.L., Turgeon E. (Eds.), Researches on Science and Art in 21st Century Turkey. Volume 2, 2380-2388 pp. Gece Kitapligi, Ankara, 2017a.

8. B. Turanli, G. Gulfidan, K. Y. Arga, "Transcriptomic-guided drug repositioning supported by a new bioinformatics search tool: GeneXpharma", OMICS: A Journal of Integrative Biology, 21(10):584-591, 2017b.

9. B. Turanli, K. Y. Arga, "Systems biomedicine acts as a driver for the evolution of pharmacology.", Ann Pharmacol Pharm, 2(16):1087, 2017.

10. C. Cheng, K. K. Yan, W. Hwang, J. Qian, N. Bhardwaj, "Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data.", PLOS Computational Biology, 7(11):1-15, 2011.

11. C. H. Chou, S. Shrestha, C. D. Yang, N. W. Chang, Y. L. Lin, K. W. Liao, W. C. Huang, T. H. Sun, S. J. Tu et al., "miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions", Nucleic Acids Research, Vol. 46, D296–D302, 2018.

12. C. Qiu, J. Wang, P. Yao, "microRNA evolution in a human transcription factor and microRNA regulatory network", BMC Systems Biology, 4, p.90, 2010.

13. C. Y. Chen, S.T. Chen, C.S. Fuh, "Coregulation of transcription factors and microRNAs in human transcriptional regulatory network", BMC Bioinformatics, 12, (Suppl 1), p. S41, 2011.

14. D. Guan, J. Shao, Y. Deng, P. Wang, Z. Zhao, Y. Liang, J. Wang, B. Yan, "CMGRN: a web server for constructing multilevel gene regulatory networks using ChIP-seq and gene expression data", Bioinformatics, Vol 30(8), p. 1190-1192, 2014a.

15. D. Guan, J. Shao, Y. Deng, Z. Zhao, P. Wang, J. Qin, Y. Deng, K. R. Boheler, J. Wang, B. Yan, "PTHGRN: unraveling post-translational hierarchical gene regulatory networks using PPI, ChIP-seq and gene expression data", Nucleic Acids Research, Vol. 42, W130–W136, 2014b.

16. D. Sengupta, S. Bandyopadhyay, "Topological patterns in microRNA-gene regulatory network: studies in colorectal and breast cancer", Molecular BioSystems, 9,(6), pp.1360-1371, 2013.

17. E. Gov, K. Y. Arga, "Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer.", Scientific Reports, 7:4996, 2017.

18. E. Gov, K. Y. Arga, "Integrative cooperation and hierarchical operation of microRNA and transcription factor crosstalk in human transcriptional regulatory network.", IET Systems Biology, 10(6):219-228, 2016.

19. E. Portales-Casamar, D. Arenillas, J. Lim, M. I. Swanson, S. Jiang, A. McCallum, S. Kirov, W. W. Wasserman, "The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences", Nucleic Acids Research, Vol. 37, D54–D60, 2009.

20. F. Emmert-Streib, M. Dehmer, B. Haibe-Kains, "Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks", frontiers in Cell and Developmental Biology, Volume 2(38), 2014.

21. F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, T. Li, "miRecords: an integrated resource for microRNA–target interactions", Nucleic Acids Research, Vol. 37, D105–D110, 2009.

22. F. Zambelli, G. M. Prazolli, G. Pesole, G. Pavesi, "Cscan: finding common regulators of a set of genes by using a collection of genome-wide ChIP-seq datasets", Nucleic Acids Research, Vol. 40, W510–W515, 2012.

23. H. Dweep, C. Sticht, P. Pandey, N. Gretz, "miRWalk – Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes", Journal of Biomedical Informatics, 44: 839–847, 2011.

24. H. Han, J. W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C. Y. Kim, M. Lee, E. Kim, S. Lee, B. Kang, D. Jeong, Y. Kim, H. N. Jeon, H. Jung, S. Nam, M. Chung, J. H. Kim, I. Lee, "TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions", Nucleic Acids Research, Vol. 46, D380–D386, 2018.

25. H. S. Chiu, D. Llobet-Navas, X. Yang, W. J. Chung, A. Ambesi-Impiombato, A. Iyer, H.R. Kim, E.G. Seviour, Z. Luo, V. Sehgal, T. Moss, Y. Lu, P. Ram, J. Silva, G. B. Mills, A. Califano, P. Sumazin, "Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks.", Genome Research, 25, 257–267, 2018.

26. H. Yu, K. Tu, Y. J. Wang, "Combinatorial network of transcriptional regulation and microRNA regulation in human cancer", BMC Systems Biology, 6,(1), p.61, 2012.

27. J. Cao, Z. Luo, Q. Cheng, Q. Xu, Y. Zhang, F. Wang, Y. Wu, X. Song, "Three-dimensional regulation of transcription", Protein Cell, 6(4):241–253, 2015.

28. J. Quin, M. J. Li, P. Wang, M. Q. Zhang, J. Wang, "ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor ", Nucleic Acids Research, Vol. 39, W430–W436, 2011.

29. J. Sun, X. Gong, B. Purow, "Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma", PLoS Computational Biology, 8, 2012.

30. J. Wang, M. Lu, C. Qiu, Q. Cui, "TransmiR: a transcription factor–microRNA regulation database", Nucleic Acids Research, Vol. 38, D119–D122, 2010.

31. J. Wang, Y Sun, S. Zheng, X. S. Zhang, H. Zhou, L. Chen, "APG: an Active Protein-Gene Network Model to Quantify Regulatory Signals in Complex Biological Systems", Scientific Reports, 3 : 1097, 2013.

32. K. Karagoz, H. Lehman, D. Stairs, R. Sinha, K. Y. Arga, "Proteomic and metabolic signatures of esophageal squamous cell carcinoma.", Current Cancer Drug Targets, 16: 721-736, 2016.

33. K. Karagoz, R. Sinha, K. Y. Arga, "Triple Negative Breast Cancer: A Multi-omics Network Discovery Strategy for Candidate Targets and Driving Pathways.", OMICS: A Journal of Integrative Biology, 19(2):1-14, 2015.

34. L. A. Bovolenta, M. L. Acencio, N. Lemke, "HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions", BMC Genomics, 13:405, 2012.

35. L. F. I. Martinez, W. Kolch, T. Santra, "BGRMI: A method for inferring gene regulatory networks from time-course gene expression data and its application in breast cancer research", Scientific RepoRts, 6:37140, 2016.

36. L. Gao, Y. Uzun, P. Gao, B. He, X. Ma, J. Wang, S. Han, K. Tan, "Identifying noncoding risk variants using disease- relevant gene regulatory networks", Nature Communications, 9:702, 2018.

37. M. Grechkin, B. A. Logsdon, A. J. Gentles, S. I. Lee, "Identifying Network Perturbation in Cancer", PLOS Computational Biology, DOI:10.137, 2016.

38. M. Kori, E. Gov, K. Y. Arga, "Molecular Signatures of Ovarian Diseases: Insights from Network Medicine Perspective.", Systems Biology in Reproductive Medicine, 62(4):266-282, 2016.

39. N. Wong, X. Wang, "miRDB: an online resource for microRNA target prediction and functional annotations", Nucleic Acids Research, Vol. 43, D146–D152, 2015.

40. P. Chouvardas, G. Kollias, C. Nikolaou, "Inferring active regulatory networks from gene expression data using a combination of prior knowledge and enrichment analysis", BMC Bioinformatics, 17(Suppl 5):181, 2016.

41. P. M. Haverty, M. C. Frith, Z. Weng, "CARRIE web service: automated transcriptional regulatory network inference and interactive analysis." Nucleic Acids Research, 32, W213–W216, 2004.

42. Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang, Y. Liu, "miR2Disease: a manually curated database for microRNA deregulation in human disease", Nucleic Acids Research, Vol. 37, 98–D104, 2009.

43. R. Janky, A. Verfaillie, H. Imrichova, B. V. Sande, L. Standaert, V. Christiaens, G. Hulselmans, K. Herten, M. N. Sanchez, D. Potier, D. Svetlichnyy, Z. K. Atak, M. Fiers, J. C. Marine, S. Aerts, "iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections", PLOS Computational Biology, Volume 10(7), 2014.

44. R. Shalgi, D. Lieber, M. Oren, "Global and local architecture of the mammalian microRNA-transcription factor regulatory network", PLOS Computational Biology, 3, (7), p. e131, 2007.

45. S. J. C. Gosline, A. M. Gurtan, C. K. JnBaptiste, Y. S. Yap, P. A. Sharp, E. Fraenkel, "Elucidating MicroRNA Regulatory Networks Using Transcriptional, Post-transcriptional, and Histone Modification Measurements", Cell Reports 14, 310–319, 2016.

46. T. Santra, "A Bayesian framework that integrates heterogeneous data for inferring gene regulatory networks", frontiers in Bioengineering and Biotechnology, vol 2(13), 2014.

47. T. Sevimoglu, B. Turanli, C. Bereketoglu, K. Y. Arga, A.S. Karadag,"Systems biomarkers in psoriasis: Integrative evaluation of computational and experimental data at transcript and protein levels.", Gene, 647:157-163, 2018.

48. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, "Predicting effective microRNA target sites in mammalian mRNAs", Computational and Systems Biology, 2015.

49. Z. P. Liu, C. Wu, H. Miao, H. Wu, "RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse", The Journal of Biological Databases and Curation, Vol. 201, 2015.