BBA - Gene Regulatory Mechanisms 1863 (2020) 194430

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Gene regulatory network inference resources: A practical overview☆

Daniele Mercatelli[a], Laura Scalambra[a], Luca Triboli[b], Forest Ray[c], Federico M. Giorgi[a,*]

[a] Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy
[b] Centre for Integrative Biology (CIBIO), University of Trento, Italy
[c] Department of Systems Biology, Columbia University Medical Center, New York, NY, United States

A B S T R A C T

Transcriptional regulation is a fundamental molecular mechanism involved in almost every aspect of life, from homeostasis to development, from metabolism to behavior, from reaction to stimuli to disease progression. In recent years, the concept of Gene Regulatory Networks (GRNs) has grown popular as an effective applied biology approach for describing the complex and highly dynamic set of transcriptional interactions, due to its easy-to-interpret features. Since cataloguing, predicting and understanding every GRN connection in all species and cellular contexts remains a great challenge for biology, researchers have developed numerous tools and methods to infer regulatory processes. In this review, we catalogue these methods in six major areas, based on the dominant underlying information leveraged to infer GRNs: Coexpression, Sequence Motifs, Chromatin Immunoprecipitation (ChIP), Orthology, Literature and Protein-Protein Interaction (PPI) specifically focused on transcriptional complexes. The methods described here cover a wide range of user-friendliness: from web tools that require no prior computational expertise to command line programs and algorithms for large scale GRN inferences. Each method for GRN inference described herein effectively illustrates a type of transcriptional relationship, with many methods being complementary to others. While a truly holistic approach for inferring and displaying GRNs remains one of the greatest challenges in the field of systems biology, we believe that the integration of multiple methods described herein provides an effective means with which experimental and computational biologists alike may obtain the most complete pictures of transcriptional relationships. This article is part of a Special Issue entitled: Transcriptional Profiles and Regulatory Gene Networks edited by Dr. Federico Manuel Giorgi and Dr. Shaun Mahony.

1. Introduction

The biology of living organisms can be interpreted in terms of information: information is stored in DNA, processed to RNA and is subsequently translated into proteins to perform several functions through a cyclic process involving the replication and propagation of information itself. This stream of information from DNA to protein through RNA transcription has been defined as the “central dogma” of molecular biology [1]. The growing knowledge of biological systems has revealed that transcription is structured in complex gene expression networks carrying out molecular processes in living cells [2]. The complexity of living organisms is therefore the result of a multilevel, dynamically controlled network of gene expression regulation [3] which shapes and maintains individual phenotypes and also modulates adaptive responses to environmental changes [4]. Transcriptional regulation is a fundamental biological process and much effort has been spent worldwide to dissect it and identify its key players, in order to understand how its molecular components collaborate to regulate gene expression levels [5]. Transcription of a gene is usually initiated when *cis*-regulatory regions on DNA are bound by a sequence-specific Transcription Factor (TF), and the transcriptional machinery is recruited to initiate the process [6]. Transcription represents a major control point in gene expression, but it is only a part of the complex mechanism used by cells to regulate their molecular repertoire and shape their phenotype. It has been estimated that the total collection of human TFs consists of around 1600 elements [7], accounting for almost 8% of all human protein-coding genes, further divided into > 30 families on the basis of their DNA-binding molecular motifs [8]. Generally, all kingdoms of life show similar proportions of TF-encoding genes over the total of genes, with 285 TF-coding genes in *Escherichia coli* (7%) [9], 301 in *Saccharomyces cerevisiae* (5%) [10] and 754 in *Drosophila melanogaster* (5%) [11].

Decoding the architecture of regulatory interactions has become one of the main tasks of modern biology [2]. TFs bind genomic DNA often in complexes to regulate the levels of transcription for each Target Gene (TG, i.e. any gene whose transcription is influenced by a regulator). The ensemble of these binding events forms a regulatory network,

Abbreviations: ChIP, Chromatin Immunoprecipitation; GRN, Gene Regulatory Network; MI, Mutual Information; NGS, Next Generation Sequencing; ODE, Ordinary Differential Equation; PPI, Protein-Protein Interaction; TEP, Transcript Expression Profile; TF, Transcription Factor; TG, Target Gene; TSS, Transcription Start Site

☆ This article is part of a Special Issue entitled: Transcriptional Profiles and Regulatory Gene Networks edited by Prof. Federico Manuel Giorgi and Dr. Shaun Mahony.

* Corresponding author.
E-mail address: federico.giorgi@unibo.it (F.M. Giorgi).

https://doi.org/10.1016/j.bbagrm.2019.194430
Received 31 May 2019; Received in revised form 6 September 2019; Accepted 9 September 2019
Available online 31 October 2019
1874-9399/

constituting a wired diagram of cellular transcriptional regulation [12]. Understanding how a limited cohort of TFs can finely orchestrate the broad diversity of gene expression patterns in several cellular types and conditions is a crucial task for life scientists, and the development of several molecular biology techniques have opened the way for dissecting the complexity of cellular regulatory networks [13]. Along with the advent of high-throughput measurement platforms, the need to develop new methods for interpreting large datasets emerged from the problem of handling the increasingly copious amounts of biological data acquired during experiments. A useful way to describe the complexity behind the regulatory mechanisms of biological systems is to construct mathematical models to interpret and reverse-engineer cellular functions [14]. This purpose can be fulfilled by creating an evidence-based description of all the relevant interactions occurring in a certain system, which can explain its architecture by providing a description of its function and a predictive model to infer its dysfunction [15].

Thus, Gene Regulatory Networks (GRNs) emerged from this effort as a powerful tool to describe and computationally reconstruct the complex scheme of interactions behind transcriptional regulation of gene expression [16]. GRNs are topological maps representing and predicting relationships between molecular entities. Often, these comprise the regulatory interactions between a TF and its TGs, but also between noncoding RNAs and TGs [17]. In general, the definition of “regulator element” (TF, miRNA, regulatory RNA, enhancer, etc.) and “regulated element” (a TG) is subjective to the scientist building and analyzing the specific GRN representation. While this review (and the bulk of scientific literature on GRNs) focuses on coding TF-TG GRNs, some methods and resources (e.g. coexpression-based tools, see later) allow the user to manually define the regulators to tailor their analysis.

Conceptually, GRN representations visualize the real underlying GRN. This real network controls every aspect of transcriptional regulation, generating phenomena through a “shadowplay” effect [18] that can be observed and measured (qualitatively or quantitatively) as biological data. The process of inferring the original system by creating a GRN representation based on data is commonly called “gene network reverse engineering” [14] (Fig. 1A).

GRN representations are formed by Nodes and Edges. Nodes usually represent genes, both coding and noncoding (miRNAs, lncRNAs, etc.), or regulatory elements like enhancers, cis-acting genomic regions, etc. Edges represent various regulatory connections (activation, repression, modulation, et cetera) and they can be weighted to describe the strength of the regulatory relationship [18] (Fig. 1B).

Unlike other biological networks (e.g. Protein-Protein Interaction networks), GRNs tend to contain *directed* edges. While undirected edges simply state there is a regulatory relationship between any two nodes (N), directed edges define a clear relationship between a Cause node (C) and an Effect node (E) (Fig. 1C). This derives from the nature of regulatory relationships, often characterized by an identified regulatory node (TF, miRNA, etc.) influencing a target node (TG). Causality can be defined when GRNs are derived from e.g. perturbation experiments (e.g. silencing a TF and observing the regulatory effects on TGs). Causal relationships are however not always evident when GRNs are derived from large quantitative datasets, such as in the case of a correlated pair of TFs, since correlation does not underlie causation, nor a specific causal directionality between the correlated genes [19]. For example, GRN reconstruction methods based on coexpression (see later) provide a network of relationships with no information on the directionality of the identified interactions. In other words, the correlation between two gene expression levels does not imply a specific causal relationship between the two, nor a direct one. However, causality can be inferred using several approaches based on quantitative measurements. One such method is Granger Causality [20], that requires time-series data to define directed relationships. Another example is Pearl Causality [21] which can derive causal relationships in acyclic graphs also given non-time-series data. Another simple heuristic principle for causality, implemented by several gene network inference methods, hypothesize a causal relationship between any two correlated genes wherever the first (cause) is a TF and the second (effect) is not [22]. Furthermore, a supervised analysis introducing prior knowledge could be applied to identify or refine directionality in the inferred interactions [23].

In other scenarios, GRNs may provide a distinction between direct edges (e.g. a TF directly binding the promoter of a TG) and indirect ones (e.g. a TF inducing an overexpression of a downstream TG, without a proven direct binding on the TG promoter) [24]. The distinction between these two edge categories is not always obvious, as it requires extensive experimental validation to prove a direct connection. Moreover, several GRN reconstruction methods (such as those based on simple correlation thresholds) cannot distinguish direct from indirect edges. In fact, in a real network one TF may control the expression of a TG by regulating a second TF, or a chain of them, resulting in all these gene expression levels being correlated (Fig. 1D). In this case, indirect interactions should be weaker than direct ones and mediated via intermediate nodes [25]. Conditioning methods such as partial correlation, or information theory methods like Data Processing Inequality (DPI) can be applied to prune network edges unlikely mediated by a direct (TF/promoter) interaction [26]. Conditional methods can provide at the same time causal inferences for some of the GRN edges [24].

Accurate inference of GRNs is crucial to obtain a systematic understanding of molecular mechanisms governing normal cells and disease states to further identify new potential therapeutic targets. Understanding how gene expression is regulated under different conditions has a tremendous impact in several life science subjects, with a particular focus in basic and applied biomedical research, medicine and drug discovery, as many dysregulated gene expression programs are at the basis of a wide range of diseases [27].

The inference of GRNs has proven to be a powerful tool in the study of essential processes occurring in cellular systems, such as development, differentiation, metabolism, adaptation and signaling. Many applications of GRNs are bringing innovation to research fields such as human health [28–30] and plant science [31,32].

For example, the reconstruction of causal maps of molecular interactions can be used to derive novel functional hypotheses about regulatory interactions that can be further investigated and validated experimentally. Networks can also serve as biomarkers in complex disorders where the dysregulation of several pathways can be better predicted from a network perspective and used as a prognostic marker [33]. To date, there is a huge amount of publicly available large datasets that can serve as input for GRN inference. Since different GRN reconstruction methods have their own advantages and challenges depending on the type and quality of input data and the purpose of inference [34], many efforts have been undertaken by researchers worldwide to develop tools that help scientists in characterizing complex relationships between biological entities. Nowadays, several computational tools are available to make GRN inference easier, reducing computational complexity and obtaining faster GRN reconstruction algorithms.

This Review outlines currently available approaches, resources and software for GRN reconstruction, giving some practical hints on how to choose appropriate tools to analyze available datasets and where to find them. It is intended for biologists and life sciences researchers who are new to computational methods for GRN inference, but it is also a useful reference for experienced users willing to get an updated overview on GRN tools, methods and software. To this purpose, all reviewed methods and resources are summarized in Supplementary Table 1. Our review won't focus deeply on the more algorithmic aspects of GRN reconstruction, considering that several comprehensive reviews exist on the subject, especially for coexpression methods [17,34–36].

We categorize GRN-inference resources in six groups, according to the approaches taken and the underlying data upon which they operate (Fig. 2). We further provide a paragraph for peculiar approaches not falling within any of the six groups, and finally a paragraph highlighting best practices for GRN reconstruction via an example.

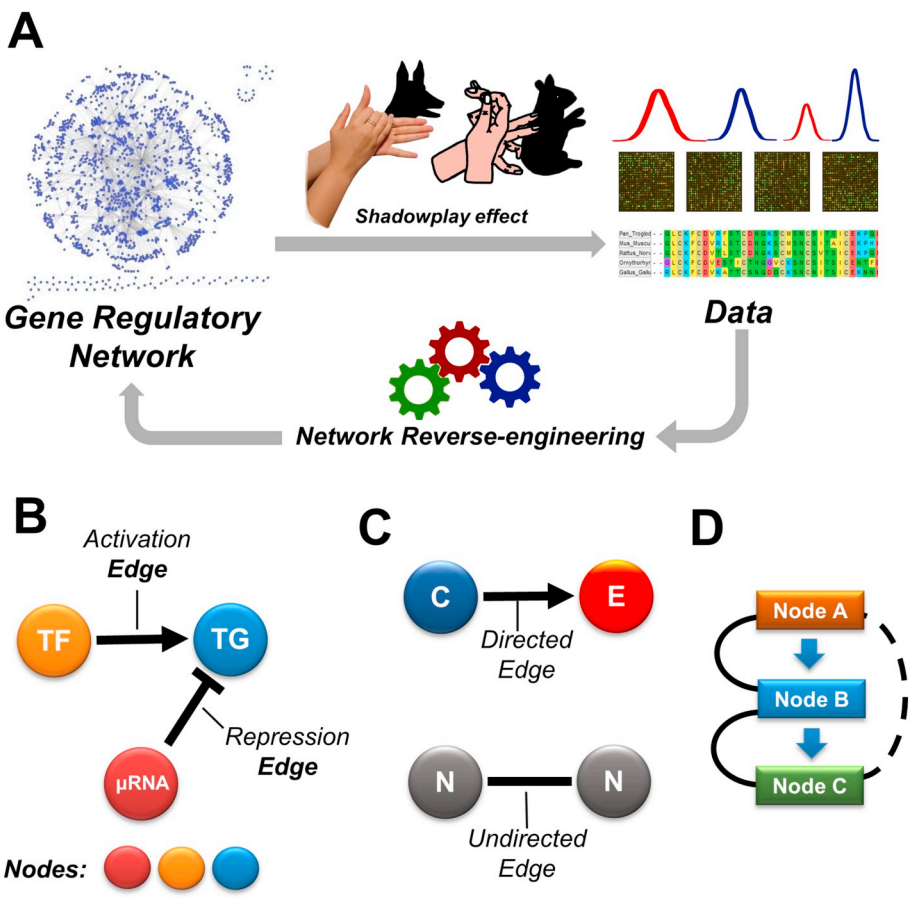

Fig. 1. Gene Regulatory Networks. (A) The principle of GRN reverse engineering. A real GRN (left) controls all aspects of transcriptional regulation in living cells. What is detectable from it is a metaphorical "shadowplay" that generates data from qualitative and quantitative empirical measurements, sometimes as the result of an experimental perturbation (right). Data takes many forms, for example ChIP-Seq-derived TF binding locations on the genome, Transcript Expression Profiles (TEPs) or sequence conservation information. (B) Basic nomenclature for GRNs: nodes represent molecular elements, such as TFs or TGs, while edges describe their relationships. (C) Causality in GRNs: edges can be directed, defining cause-effect relationships (top) or undirected (bottom). (D) A simple GRN with three nodes, where node A activates node B, which in turn activates node C (real functional relationship depicted by blue arrows). All three nodes may be correlated and therefore edges may be drawn between the three of them, but some methods are able to distinguish direct edges (solid lines) from indirect ones (dashed line).

1. Coexpression-based resources
2. Sequence motif-based resources
3. ChIP-based resources
4. Orthology-based resources
5. Literature-based resources
6. Transcriptional complexes protein-protein interaction (PPI) tools
7. Other Resources
8. Best Practices and Example

Other tools outside these six categories exist, as well as tools based on the integration of the aforementioned approaches. In fact, the combination of multiple GRN inference approaches has been proven to be successful in several cases [37,38]. However, we believe that the optimal selection of GRN inference tool(s) ultimately depends on the specific scientific context investigated and must be tailored on data availability on the specific species, TF, perturbation or cellular context.

2. Coexpression-based resources

Two genes are deemed coexpressed when a significant dependency between their Transcript Expression Profiles (TEPs) is determined [31]. Coexpression-based tools to infer gene regulatory relationships have been widely adopted after the introduction of transcriptome-scale quantification methods of transcript abundances [39]. These tools collect TEPs across multiple data samples generated by high-throughput platforms such as DNA microarrays, Serial Analysis of Gene Expression (SAGE) or RNA sequencing (RNA-seq) [40]. TEPs provide multi-sample quantitative information relative to several transcript species: commonly messenger RNAs (mRNA) but also noncoding RNAs (ncRNA) and microRNAs (miRNA) [41].

There are several available tools for the reverse engineering of GRNs based on coexpression. Most tools are available as Shell/R/Python data analysis pipelines, but some are also distributed as plugins or standalone software. There are also online interfaces to help non-expert users to take advantage of their services.

Some tools assess simple linear Pearson correlation between TEPs derived from transcriptomics data [42]. One of the simplest and most widely used implementation of Pearson correlation for application to biological TEPs is the R function *cor*(). However, since quantitatively non-linear interactions are often present in biological systems, several tools implement measurements that allow the detection of monotonic but non-linear dependencies, such as Spearman Correlation [31], or entirely non-linear relationships, such as pairwise Mutual Information [43]. A great advantage of coexpression tools is their simplicity and low computational complexity, and their capability of inferring genome-wide networks even when a relatively low number of samples ($n = 50$) is available [31].

2.1. Gene coexpression online databases

High-throughput gene expression data are commonly collected on publicly accessible primary databases, the most known of which are the European Bioinformatics Institute's ArrayExpress [44], the Gene Expression Omnibus (GEO) [45], supported by the National Center for Biotechnology Information, and the DNA Data Bank of Japan [46]. Using datasets retrieved from primary databases for GRNs inference requires a certain expertise in bioinformatics, but this task has now become easier thanks to the availability of several secondary or derivative tools, which contain curated information extracted from primary resources and provide a query-driven approach to analyze data [47]. Coexpression databases offer tools to derive gene coexpression scores across thousands of samples starting from a query gene (or a list of genes), guiding the user through a user-friendly interface also providing options to subsequently draw the network.

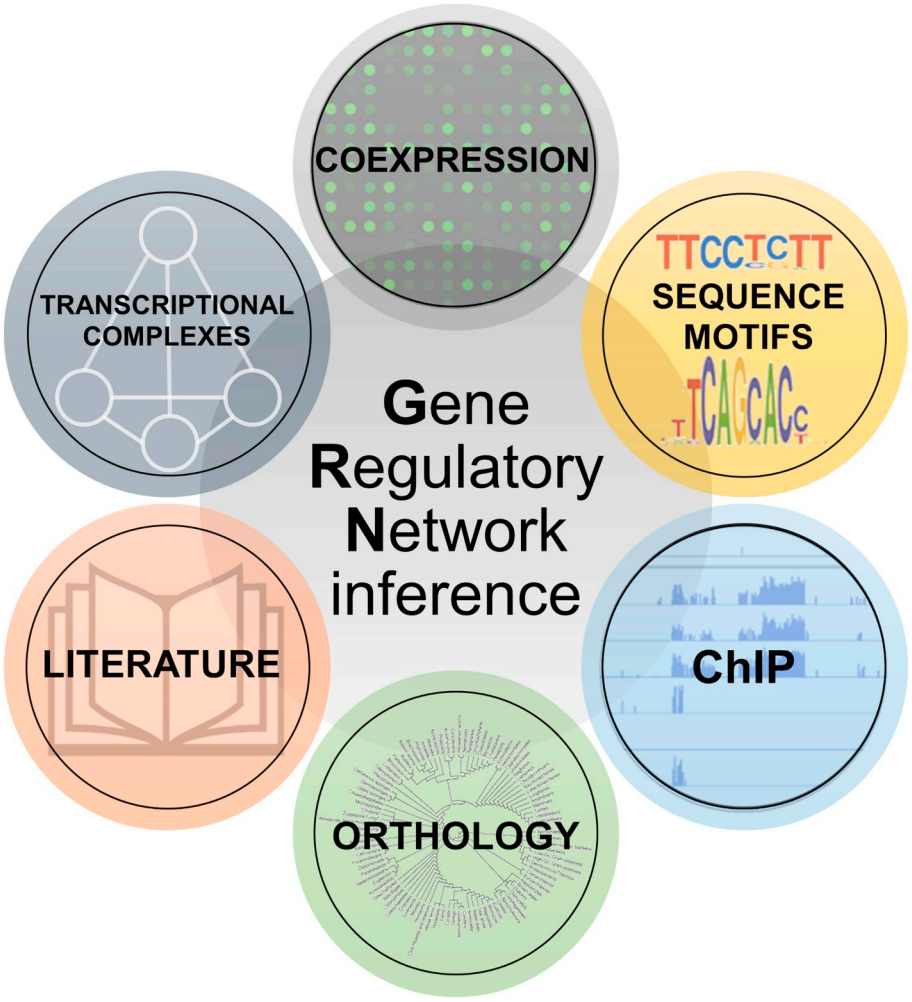

Fig. 2. The classification of GRN-inference tools followed by this review.

2.1.1. Xena Browser

The *Xena Browser* is a large functional online resource collecting genomics and transcriptomics data from 1617 human cancer datasets [48]. Broadly speaking, it allows the user to detect associations between variables measured in such datasets, i.e. somatic mutations, copy number alterations and TEPs at the level of individual genes or genomic regions. For the purpose of GRN inference, *Xena* can be quickly used to assess coexpression between any human TF-TG pair in cancer datasets. Furthermore, it allows to detect if somatic genomic alterations targeting a TF are associated to altered expression profiles of candidate TGs, similarly to KO experiments perturbing a TF, canonically used to infer candidate downstream TGs. Therefore, this integration of genomics and transcriptomics data allows *Xena* to extend itself beyond simple coexpression.

2.1.2. ALCOdb, ATTED-II and CORNET

Among the resources specializing in plant networks, *ALCOdb* is a coexpression data collection supporting information on two model algae, *Chlamydomonas reinhardtii* and *Cyanidioschyzon merolae* [49]. Exploration of coexpression networks of plant genomes is available at *ATTED-II* [50] and *CORNET* 3.0 [51], which provides coexpression platforms for nine plant species.

2.1.3. COXPRESdb, GeneFriends and COEXPEDIA

Moving to the animal kingdom, *COXPRESdb* is a web-based resource enabling the user to infer and draw gene coexpression relationships in model organisms (including human, mouse, rat and zebrafish) that returns also information about the functional enrichment among the coexpressed genes [52]. Similar functions are also offered by *GeneFriends* [53], while *COEXPEDIA* limits coexpression inference using individual studies rather than aggregating multiple datasets, and also associates coexpressions to indexed MeSH terms to help researchers in generating biomedical hypotheses [54].

2.1.4. SEEK

SEEK (Search-based Exploration of Expression Kompendia) is a query-based cross-platform search system containing thousands of human transcriptomic datasets. It is implemented in a user-friendly interactive web interface (see Supplementary Table 1), where users can also find instructions to install it on a local computer. By entering a query gene set, *SEEK* returns a graphical representation of coexpressed genes, which can be further refined in condition-specific contexts (e.g. for a given tissue, cell type or disease state) [55].

2.2. Mutual Information tools

Mutual Information (MI) is a measure of dependencies between two variables, developed in information theory. In the context of GRN reconstruction, MI can be used to infer relationships between all gene pairs (or all TF-TG pairs). The most important advantage of MI is its ability to infer non-linear relationships between TEPs. However, since MI was designed for categorical data, the continuous TEPs need to be reduced into discrete categories or bins, in a process called binning [56].

2.2.1. ARACNe

ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) is a MI-based algorithm for the reverse engineering of transcriptional networks from transcriptome-wide gene expression data [14,43]. The *ARACNe* approach relies on the identification of TF-TG gene pairs that exhibit similar transcriptional fluctuations by measuring the MI between their expression profiles. The TEPs in the original *ARACNe* implementation are binned using a Fixed-Bandwidth (FB) approach that divides the joint TF-TG distribution into evenly spaced categories. Initially, ARACNe is provided with a gene expression dataset and with a list of TFs (every other gene is considered a TG). All pairwise TF-TG MIs are then calculated, and the edges above a certain global significance threshold (defined by the distribution of MIs in multiple permutations of the dataset) are kept. To cope with possible associations deriving from indirect interactions (Fig. 1D), *ARACNe* applies an information-theory step known as Data Processing Inequality. Briefly, the algorithm evaluates all possible scenarios where there is support for a TG to be regulated by two or more different TFs (TF1 and TF2). Within each TF1-TF2-TG triplet, the TF-TG edge with the smallest MI is pruned, while the most likely TF-TG interaction is kept. This process is repeated for all possible triplets within the GRN. Then, bootstrapping of the initial expression matrix is performed, taking (by default 100) bootstraps (i.e. new data matrices with the same size as the initial one but with individual samples omitted or repeated once or multiple times by chance). The MI significance and DPI are recalculated in all bootstraps. Some edges can be rescued in these bootstrapping steps if they survive pruning in a significant number of bootstraps (the significance is assessed by a Poisson distribution), and in the final weighted GRN each TG can be connected to more than one TF. *ARACNe*'s inferred GRN prediction has been experimentally validated for the MYC TF by ChIP in human B cells [57] and it was shown that inferred networks recover known functionally characterized TF modules, while reliably predicting novel module components [58]. It is possible to use *ARACNe* from the command line using cross-platform code available at [43] and as an online graphical *ARACNe* included in the *geWorkbench* website [59]. There are several R implementations of the MI algorithm used by *ARACNe* within the R Bioconductor packages *minet* [60] and *RTN* [61].

2.2.2. minet, CLR and MRNET

On top of providing ARACNe implementation, *minet* is an R package that collects several methods to perform mutual information calculation for inferring GRNs. In particular, *minet* provides two extra MI-based algorithms: *CLR* (Context Likelihood of Relatedness) and *MRNET* (Maximum Relevance/Minimum Redundancy NETwork inference method) [60]. Differently from *ARACNe*, which uses a global threshold for graph pruning, *CLR* allows to establish individual thresholds for each gene pair. *MRNET* incorporates the minimum redundancy maximum relevance feature selection method, which iteratively searches for direct interactions.

2.2.3. TimeDelay-ARACNe, hARACNe, ARACNe-AP, GPU-ARACNe and SJARACNe

Different extensions of the *ARACNe* algorithm have been released: *TimeDelay-ARACNe* can be used for GRN inference from time-course expression data [62]. The *hARACNe* algorithm uses higher-order Data Processing Inequality resulting in a significantly improved inference performance [63]. *ARACNe-AP* extends the fixed-bandwidth original *ARACNe* algorithm for binning expression profile with an adaptive partitioning strategy for MI estimation, achieving a dramatic improvement in computational performance [43]. *GPU-ARACNe* is a parallel implementation of the original algorithm using the NVIDIA CUDA framework to accelerate computing by taking advantage of multi-core and multi-level parallelism [64]. Recently, *SJARACNe*, a new scalable solution implemented in C++ and Python has been presented. *SJARACNe* further improves the computational performance over *ARACNe-AP*, consistently reducing memory usage to let researchers efficiently run large dataset analysis on common laptops [65].

2.2.4. RTN and RTNsurvival

RTN is an R package to infer gene regulatory networks using MI or other methods (such as standard Pearson Correlation) and to associate GRN expression with genomic features, offering several plotting tools to focus on specific TF clusters [61]. *RTNsurvival* is a recently released R/Bioconductor package able to integrate GRNs inferred by *RTN* with survival analysis [66]. Given the averaged expression of all TGs connected to a specific TF in a defined patient cohort, *RTNsurvival* performs Kaplan-Meier analyses and Cox Proportional Hazards regressions based on TF-centered GRNs rather than individual TEPs. This is particularly interesting, since genes are often not individually important for patient survival while appearing to be highly predictive in Kaplan-Meier analyses, due to the fact of being correlated to the causal determinant of clinical outcome [67].

2.2.5. C3NET

Finally, *C3NET* is a causal GRN inference algorithm available as an R package showing lower computational complexity and similar performance compared to *ARACNe* and *minet* package methods [68].

2.3. Bayesian network structure learning tools

Bayesian networks are probabilistic models combining Bayesian models, probability theory and graph theory. Considering a group of random variables, a Bayesian network aims at representing their conditional dependencies with a directed acyclic graph describing also the local joint probability distributions of all the interactions, i.e. the relationships between all transcripts that have been measured in the same samples.

A directed acyclic graph is a network with directed edges indicating causal dependencies between nodes and no cycles or loops [69]. Bayesian modelling requires two steps. 1) Parameter learning for each node, i.e. the formalization of TEPs fluctuations in the dataset by themselves (marginal distribution) and related to other TEPs (conditional distribution). 2) General GRN structure learning, i.e. the identification of the best topology that explains the observed data [70]. Bayesian network approaches offer a powerful tool to combine different data types obtaining information on the causation relationships among nodes, and are less sensitive to noisy data [37]. Some limitations need to be taken into consideration when inferring a network applying Bayesian methods: the number of possible network topologies shows a super-exponential increase with the number of genes, which reduces the possibility to compute the structure of all possible networks and requires significant computational power. Moreover, traditional Bayesian approaches cannot model cycles (i.e. biological regulatory loops), because they are generally designed for Directed Acyclic Graphs (DAGs) [71]. A way to overcome this limitation is to use dynamic Bayesian modelling. Paying the cost of an increased computational complexity and longer times required, dynamic Bayesian networks can model cyclic interactions by introducing temporal relationship between a sequence of acyclic events [72]. Several Bayesian network structure learning tools are available for GRN inference.

2.3.1. Banjo

Banjo (Bayesian Network Inference with Java Objects) is a command line program written in Java that uses Bayesian network frameworks to result in a directed inference network from discrete variables, but also implements simple discretization functions for continuous variables using either quantile or interval discretization methods. Multiple threading has been added in version 2.1, while the current version (2.2) introduced the possibility to perform parallel search on a large computer cluster [73].

2.3.2. CatNet, G1DBN and GRENITS

CatNet is an R package collecting functions for discrete Bayesian network structure learning and parameter estimation using likelihood-based criteria. This method focuses on reconstructing GRNs from incomplete data [74]. *G1DBN* (Graph 1 Dynamic Bayesian Network) is an R package whose inference pipeline is performed using first order partial dependencies to remove indirect relationships similarly to partial correlation. While powerful even on a small number of observations, *G1DBN* is limited to the reconstruction of directed acyclic networks [75]. *GRENITS* (Gene Regulatory Network Inference Using Time Series) is a Bioconductor package implementing several Bayesian methods to infer linear interactions, linear interactions in the presence of experimental noise, and nonlinear interactions. It focuses on inferring GRNs based on time course data, where the behavior of TFs in early data points can be related to candidate TGs early on [76].

2.3.3. GeneNet

Last but not least, *GeneNet* is a very efficient and easy-to-use Bayesian partial-correlation-based method to reconstruct GRNs. It can calculate full order partial correlation even in datasets with more genes (features) than samples, through a pseudoinverse operation, and it has been successfully tested on large-scale gene association networks [77]. *GeneNet* is available as an R package, which provides a useful *cor2pcor* function to convert correlation matrices in partial correlation matrices (a useful transformation to remove indirect edges). The current version of *GeneNet* allows also inferring directionality for some of the edges in the network.

2.4. Differential equations-based methods

Dynamic systems can be accurately modelled by applying ordinary differential equation (ODE) approaches. In GRN reconstruction, ODEs are learned from gene expression data, from multiple samples (both time-series and non-sequential data) and upon specific TF perturbations, with the purpose to fully discover all rates of transcript generation, half-life and degradation. ODE methods are naturally suited to model also non-linear relationships, as they model what are essentially RNA chemical reactions that can show a wide range of kinetic behaviors [78]. ODE methods can be extremely computationally intensive as they model multiple solutions to explain observed TEP fluctuations in the data. Introducing constraints, in the form of known kinetic parameters or prior knowledge of GRN structure can be extremely beneficial to ODE-based methods [17].

2.4.1. Inferelator

Inferelator is a method for deriving genome-wide transcriptional regulatory interactions, originally based on the inference of independent ODEs explaining the TEPs upon several experimental perturbations [79]. *Inferelator* simplifies ODE inference by introducing prior knowledge from manual gene annotations and by network structure inference through the generation of gene-centered linear or polynomial regressions with L1 constraints [80]. In other words, each TEP is explained by a combination of a limited subset of the other TEPs, a GRN where ODEs are defined to explain the observed data. A great advantage of this method is the capability of modelling not only TF/TG interactions, but also dynamic and kinetic parameters for transcription rates. However, while extremely successful in reconstructing unicellular GRNs, *Inferelator* becomes extremely computationally intensive to infer higher organisms' gene relationships.

2.4.2. TSNI

TSNI (Time Series Network Identification) is a differential equations-based GRNs inference method available as a MATLAB package. The purpose of this method is to infer the local network of gene–gene interactions surrounding a gene of interest by measuring TEPs at multiple time points focusing on the perturbation of only one (or a few) genes of interest [81].

2.5. Hierarchical clustering of gene co-expression networks

Hierarchical clustering is a common method applied to extract modules from GRNs to generate graph subunits for subsequent characterization. Clusters are identified by iteratively assigning nodes to clusters by weighting network vertices and progressively including neighboring vertices starting from high weights.

2.5.1. WGCNA

WGCNA (Weighted Correlation Network Analysis) generates gene coexpression networks based on pairwise TEP correlations. It became popular through its main R package implementation, which provides a collection of functions for network reconstruction and visualization with different thresholding and correlation methods [82]. *WGCNA* can be used to generate and explore clusters of highly correlated genes defining modules, to explore relationships between genes and modules or among modules (eigengene networks), to calculate network topological properties and to describe the correlation structure between gene expression and other high-dimensional data. Furthermore, *WGCNA* includes interfaces with commonly used bioinformatics tools for network visualization, such as Cytoscape [83].

2.5.2. HCCA

HCCA (Heuristic Cluster Chiseling Algorithm) is an algorithm written in Python which generates a gene-gene correlation matrix as a starting network, then for each node generates node vicinity networks by collecting all nodes within defined steps away from the seed node, and then iteratively removing nodes with higher connectivity to the outside. Filtering according to the desired gene group size is then applied to the resulting clusters, which are ranked by outside to inside connectivity ratios. Only non-overlapping clusters are retained in the resulting GRN [84].

While not specifically designed to predict direct TF-TG interactions, clustering methods such as WGCNA and HCCA can be directly applied for GRN inference by defining each cluster containing a TF as a coexpression-based GRN model: the TF will be the predicted regulator, and all the other genes will be the candidate TGs. When multiple TFs are present in a cluster, the direct regulatory structure can be calculated by local cluster full order partial correlation [77].

2.6. Single-cell GRN inference coexpression tools

Inference of GRNs using high-throughput single-cell RNA-Seq (scRNA-Seq) data can be achieved by using the tools and methods discussed above. However, due to the intrinsically noisy nature of such datasets, poor performances have been shown for most methods [88]. Drop-outs in single-cell expression datasets, i.e. genes that remains undetected due to low technical efficiency of mRNA capture, represent one of the major issues in inferring GRNs [85]. Thus, a compelling need for developing dedicated methods exists. To date, few tools specifically designed to handle single-cell data for GRN inference are available, which only allows for reconstruction of simple models [86].

2.6.1. ACTION

ACTION is a pipeline for single-cell RNA-seq analysis written C/C++, R and MATLAB. The *ACTION* framework identifies clusters of genes whose expression is predominant in defining the identity of each cell, and then infers which GRNs are responsible for the measured transcriptome. The collective GRN profiles from all single cells derived from tumor samples were used as network biomarkers to stratify melanoma patients according to clinical outcome [87].

2.6.2. BTR and SCNS

BTR and *SCNS* are Boolean network-based algorithms tailored for

single cell data. In brief, Boolean algorithms define interactions between predefined TFs and TGs using Boolean logic operators. A binary state is associated to each gene according to its expression profile: lower expression or silenced genes are assumed to be inactive (0), while up-regulation is associated with an active state (1) [88]. Every cell in a Boolean networks is classified into a state, defined on the TF expression, and cells having a limited number of differences are connected, obtaining a state-graph which infers key TFs involved in state changes [86]. Boolean models are quite robust in handling the effect of dropouts. *SCNS* (Single-Cell Network Synthesis) is a toolkit for the generation of Boolean GRNs from single-cell gene expression data. It was successfully used to model the GRN of blood development [89] and it is particularly suited to understand transition states and network rewiring upon perturbation of a cell population. *BTR* (BoolTraineR) is an R package available at CRAN. Differently from SCSN, which requires a connected state transition graph, *BTR* can infer both Boolean rules and network structure without needing information on trajectories through cell states [90].

2.6.3. MTL

MTL (Multi-Task Learning) is a recently developed method to infer GRNs from distinct datasets and then integrate them [91]. This provides an advantage over GRN reconstruction arising from a merged dataset, due to batch effects and different numbers of samples between conditions. The ideal data scenario for the application of *MTL* is that of scRNA-Seq, where thousands of cells carrying similar TF perturbations can be generated. This method was tested on a yeast (*Saccharomyces cerevisiae*) single-cell dataset, composed of TEPs from 38,000 cells with different genotypes and with multiple measurable TF knock-outs. The method was successfully benchmarked against the known yeast TF-TG interactions [92]. One promising aspect of *MTL* is that it can work on raw scRNA-Seq data, i.e. transcript counts without imputation (a common operation that tries to predict zero-count transcripts caused by dropout effects [93]).

2.6.4. nlnet

nlnet was designed to detect non-linear direct associations between TEPs (and specifically TFs vs TGs) in single-cell data, as well as classic linear relationships (such as those detected by Pearson correlation). Using the distance measure DCOL (Distance based on Conditional Ordered List) to efficiently compute transcriptome wide pairwise marginal non-linear associations, *nlnet* can infer a large quantity of candidate GRN edges [94]. This method is available as an R package at CRAN.

2.6.5. SCODE

SCODE is an R implementation to infer GRNs from single-cell data through regulatory dynamics based on linear ODEs. It showed good inference performances in predicting GRNs across cell differentiation, but it requires continuous time expression data [95].

2.6.6. SINCERA

SINCERA is a computational pipeline implemented in R incorporating the *G1DBN* inference method optimized to handle single-cell RNA-seq data. It includes the possibility to integrate data, methods and external knowledge for GRN reconstruction [96].

2.6.7. VIPER

VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis) is a regulatory-network based approach for the accurate assessment of protein activity from both bulk [22] and single-cell gene expression data [30]. While not generating coexpression networks directly, *VIPER* can interrogate them (e.g. those generated by ARACNe) to infer which GRNs are responsible for a differential expression signature (e.g. cancer vs. normal tissue) or which GRNs underlie clustering structure in large transcriptomics datasets [97]. *VIPER* provides differential GRN-specific activity scores, which can be detected in a multi-sample fashion or in a sample-by-sample basis. The *VIPER* algorithm is available as an R-system package from Bioconductor containing two methods: the multi-sample *msVIPER* and the single sample *ssVIPER*. *ssVIPER* method allows transforming a gene expression matrix into a protein activity matrix representing the relative activity of each protein in each sample. Using these *VIPER*-inferred GRN aggregated profiles, rather than individual gene expression profiles, has been shown to be beneficial for noisy datasets such as low-coverage RNA-Seq or single cell RNA-Seq [22,30].

2.7. Other coexpression tools

2.7.1. BiRewire

BiRewire is an open-source Bioconductor package implementing functions for the randomization of bipartite graphs maintaining their node degrees and generating rewired versions of the same graphs [98]. The switching algorithm of *BiRewire* can generate randomized versions of a binary event matrix that for genomic data corresponds to a binary table in which the generic entry is equal to 1 if the gene in the corresponding sample is altered and is equal to 0 otherwise. *BiRewire* is specifically suited to model GRNs (with both indirect and directed edges) and GRN rewiring in cancer, where genomic alterations causally determine transcriptomic changes [30]. In these contexts, *BiRewire* is able to directly link and model genomic alterations to GRNs, with the optional implementation of prior knowledge of biological data. A recent implementation of BiRewire massively increases its speed in large transcriptomic datasets [99].

2.7.2. CCREPE

CCREPE (Compositionality Corrected by Permutation and Renormalization) is a correlation-based tool that can assess associations between features (e.g. gene TEPs) in compositional data (i.e. sparse datasets where not all the features are measured in all samples). *CCREPE* is available as a Bioconductor package [100] and it can be used to infer TF-TG associations in full and sparse expression datasets, such as those arising from a combination of different microarray platforms or single-cell RNA-Seq. *CCREPE* applies a specific pipeline that reduces spurious correlation arising from compositional datasets, where features can result as over-correlated simply because they are concurrently not measured. Finally, associations between genes are calculated using standard Pearson or Spearman correlation, or a specific correlation for sparse data, dubbed the N-dimensional Checkerboard (NC) score.

2.7.3. GENIE3

GENIE3 (GEne Network Inference with Ensemble of trees) is a decision tree-based method which has emerged as the best performer in the DREAM4 in silico multifactorial challenge. It is a GRN inference method based on variable selection with ensembles of regression trees. In contrast to linear regression models, tree-based approaches do not make any assumption about gene regulation nature, enabling the possibility to manage combinatorial and non-linear interactions. Random forest regression is one of the most prominent examples of tree-based approaches. It can produce directed graphs of regulatory interactions allowing for the presence of feedback loops in the network, thus obtaining realistic GRNs [101]. *GENIE3* is available as a MATLAB, Python and R package. Since its initial implementations, extensions have been proposed to handle time-series data (e.g., *dynGENIE3*) [102].

2.8. Validation of coexpression GRNs

Reconstructing realistic GRNs from expression data based on coexpression is a very difficult task. Validation and benchmarking of network inference methods is therefore a necessary step when trying to test the correctness and robustness of reconstructed GRNs. *NetBenchmark* is an open-source and freely available R/Bioconductor package offering

an easy and convenient benchmarking framework containing R implementations of several methods, like *ARACNe*, *C3NET*, *CLR*, *GeneNet*, *GENIE3*, and many others. While the *NetBenchmark* package code is mainly written in R, it contains some functions written in C++ that speed up time-consuming processes, optimizing resources usage [103]. Similarly, *GeneNetWeaver* is a tool to develop in silico GRNs and produce artificial gene expression data for performance profiling of inference methods [104]. Coexpression GRN validation can be manually performed by assessing specificity and sensitivity of the inferred edges with gold standards of validated species-specific interactions (some of which are described in the Literature-based chapter). In the end, the final validation of coexpression GRNs (and truly of all inferred GRNs), requires experimental validation in the appropriate conditions, via e.g. specific TF perturbations (knockout or overexpression) or alterations of TG's regulatory elements (promoters or enhancers) to define the molecular basis for each inferred regulatory edge.

3. Sequence motif-based resources

The identification of known and conserved DNA sequence motifs recognized by TFs in the regulatory region of genes is a widespread way to reconstruct GRNs [105]. A DNA-motif is a short conserved sequence located in the promoter region of a gene, acting as recognition site for transcriptional regulators. Each TF typically recognizes a collection of similar DNA sequences corresponding to its binding site motifs, commonly represented as a position-weight matrix (PWM) or as a Hidden Markov Model (HMM) describing the probability of a given nucleotide to hold a specific position in the DNA sequence [106]. The binding of co-factors and the sequence context, including flanking sequences and DNA shape, contribute to modulate specific TF-DNA recognition [8]. Collections of such sequence motifs are listed in databases such as JASPAR and TRANSFAC (see Section 7), and several computational methods for prediction and discovery of putative TF binding sites are available [107]. The characterization of DNA-motifs can help defining network modules, which are sets of genes co-regulated by the same *cis*-regulatory motif. Genes belonging to the same module are assumed to share the same properties. Regulatory motifs can be used to build GRNs where edges represent a predicted physical interaction between a TF and a TG (i.e. a physical regulatory network), but can be also used as prior knowledge to integrate network inference with expression data to understand TF-TF and TF-TG interactions supporting functional information [34,38]. Since the primary interest when reconstructing a transcriptional network is to focus on the interactions between regulators and targets, integrative approaches reduce the complexity of the inference problem by identifying regulators on the basis of common features, such as DNA-binding motifs, therefore reducing the number of parameters to be estimated. This approach has become increasingly common, as new computational methods and resources are available. Motif analysis is a powerful method to infer GRNs, since physical information is the most informative feature to infer a transcriptional network. However, the presence of a regulatory motif is only slightly predictive of gene expression, because the expression level of a single gene also relies on other factors, such as the occupancy state of multiple TF binding sites. Motif-analysis integrated with other omics data usually produce more robust models [34,38].

Motif-based inference of GRNs has not to be confused with transcriptional network motifs inference. A general topological property of biological networks is that they contain small recurring circuits of interaction patterns, or recurring subnetworks, known as “network motifs” [108], whose tractation is beyond the scope of this review.

3.1. Motif databases

3.1.1. MEME suite

The *MEME* suite is the most complete collection of tools for discovery and analysis of sequence motifs. It offers a set of different programs organized in a web server interface to let users perform five types of motif-based analysis. 1) De novo motif discovery on a set of sequences. 2) Enrichment analysis of known motifs and user defined motifs in the promoters of a gene set. 3) Finding matches to user provided motifs in a set of sequences. 4) Querying novel motifs sets for similarity with known motifs, even in other species. 5) Prediction of regulatory links between genomic loci and gene expression [109]. These tools can be focused to find motif-based candidate TGs for a specific TF, or to find a TF best explaining the common behavior of a gene set (e.g. genes upregulated upon drug perturbation). *MEME* suite toolkit and databases are available for download to be run on a local computer.

3.1.2. mirBASE

Since 2002, *mirBASE* is the primary public resource for miRNA sequences and annotation, and therefore a tool for reconstructing GRNs for miRNAs and their targets. *mirBASE* succeeded in establishing a uniform system to assign names to newly discovered miRNAs and to provide a comprehensive and freely searchable miRNA public registry. In its latest release (v22), it contains 38,589 entries from 271 organisms [110]. Its web-interface allows users to search by specific miRNAs, miRNA families or keywords, returning information on the predicted hairpin of the primary transcript and the mature miRNA, along with sequence homology, possible targets and links to literature references. All sequences and data contained in *mirBASE* are available for download.

3.1.3. MotifMap

The *MotifMap* tool allows to search model organism genomes (human, mouse, fruit fly, yeast and *Caenorhabditis*) for known TF binding motifs. It can be used as a GRN reconstruction tool in a TF-centric manner, through its “Motif Search” functionality, allowing to search for a specific TF or a specific motif over the genome, focusing on promoter regions of TGs. It can also be used as a TG-centered tool via the “Gene Search” mode, which identifies all TF putative binding sites in regions close to a user-defined Transcription Start Site (TSS) [111].

3.1.4. PlantPAN

PlantPAN (Plant Promoter Analysis Navigator) is a vast online repository describing GRNs for 17,230 TFs from 78 plant species [112]. For each TF family, it provides position-weight matrices to define the sequence binding motifs, which can be used to find putative novel TGs as new plant genomes become available. It can be queried with a specific TG, TF, or binding motif, and it can yield both a GRN model or likely TF cofactors via enrichment of binding motif co-localization.

3.1.5. TargetScan

TargetScan is a web-based tool that catalogues miRNA-target known interactions, while allowing predicting new targets, by querying for a gene or miRNA (family) of interest [113]. Extending the miRNA-TG GRNs contained in *TargetScan* is performed through predictive binding of miRNA seed sequences to 3′UTR regions of candidate TGs, a common target region for miRNA-mediated silencing. The tool provides also a companion package, *TargetScanTools*, that allows predicting expression changes in response to a miRNA perturbation and training models to predict miRNA targets.

3.1.6. CIS-BP

CIS-BP is an online repository of binding motifs for TFs, spanning 734 species (with specific focus on model organisms), 160,862 motifs (2.8% experimentally validated and the rest inferred by electronic annotation and orthology analysis) and a total of 384,465 distinct TFs [114]. The interface is extremely user-friendly, as it allows to quickly search for a TF identifier (using the full gene symbol or wildcards) specifying the organism, the level of evidence (direct, inferred or both) and the type of experiment from which the motif information was

derived. The database can be fully queried online or downloaded for a genome-wide analysis. However, while CIS-BP is one of the most complete DNA binding motif databases currently available, the conversion of the motif information into TF-TG relationships is not trivial, and it requires the user to perform the extra step of identifying the motif occurrence in the promoter/enhancer regions associated to TGs.

3.2. Motif-based tools

3.2.1. DNAShapeR

While most tools described here define "motif" as a specific (even if fuzzy) pattern of nucleotides, *DNAShapeR* defines sequence-based DNA motifs as DNA structural shape parameters [115]. This recently published method that predicts several topological features using known crystallized DNA structures and a machine learning algorithm that slides over long stretches of DNA. Its current implementation (available as an R package) is able to predict Minor Groove Width (MGW), Electrostatic Potential (EP), Propeller Twist (ProT), Helix Twist (HelT) and Roll (Supplementary Fig. S1). The notion that TFs can recognize specific DNA shapes in the promoters of genes has been leveraged by the *DNAShapeR* team to predict TGs of specific TFs and reconstruct GRNs in an implementation called TFBShape [116].

3.2.2. HOMER

HOMER (Hypergeometric Optimization of Motif EnRichment) is a collection of command line tools for de novo motif discovery supporting several model organisms. *HOMER* lets the user to analyze list of genes or genomic positions for enriched motifs. It can identify the most enriched sequences (de novo discovery), or the most enriched known motif in a target set of interest, and also perform gene ontology analysis [117]. The *HOMER* suite contains tools to analyze and annotate Next Generation Sequencing (NGS) data (ChIP/DNase/ATAC-seq) to find enriched peaks, regions and transcripts.

3.2.3. iRegulon and i-cis Target

iRegulon and *i-cis Target* [118] allow searching for overrepresented TF binding sites in a given input using a wide collection of position-weight matrices and ChIP-seq tracks from different species. While *iRegulon* detects potential regulators and their TGs from a user-loaded input list of co-expressed genes, *i-cis Target* allows the user to input either gene names or genomic regions, enabling the use of different datasets like ChIP-seq, DHS-seq, FAIRE-seq or ATAC-seq experiments. Both tools predict the most significant TFs of a given set and return direct and indirect TGs. The resulting output network can be loaded and visualized in Cytoscape.

3.2.4. ISMARA

ISMARA is a web tool allowing to analyze gene expression or ChIP-seq datasets to identify the key regulators (TFs and miRNAs) driving TEP or chromatin state changes, with the capability to operate on time course data. *ISMARA* integrates a computational method to reconstruct transcriptional regulatory dynamics by leveraging TF binding site predictions to model gene expression in terms of regulatory motif activities. The output of *ISMARA* analysis includes the inferred activity for each regulatory motif across samples, functional and pathway enrichment analysis, and TF-TF direct interactions prediction [119].

3.2.5. MERLIN + P

The *MERLIN + P* framework extends the expression based GRNs inference algorithm *MERLIN*, a Bayesian framework of learning a probabilistic graphical model integrating additional structure priors such as sequence-motifs, ChIP data or gene knockout experiments. As shown in [120], prior-based methods greatly improved the inferred network structure compared to other methods without integrated prior analysis. Interestingly, *MERLIN + P* enables the possibility to integrate all the three prior networks (coexpression-based, ChIP-Seq-based and motif-based) resulting in higher predictive performances over methods integrating only one prior source.

3.2.6. PANDA

PANDA is an algorithm specifically developed to reconstruct genome-wide GRNs by integrating gene expression, protein-protein interaction and sequence motif data. Different from message-passing models, the *PANDA* method searches for agreement between different data types by using the information from each type to iteratively refine predictions in the others. Originally used to reconstruct a condition-specific network in yeast [121], the method is scalable for higher eukaryotes. *PANDA* is implemented in several programming languages: MATLAB/Octave, C/C++, Python and R.

3.2.7. TF2Network

TF2Network is a web-based tool to predict candidate TF regulators from a set of co-expressed or functionally related genes in *Arabidopsis thaliana*, allowing the integration of PPIs as well as co-expression information [122]. The output of this tool consists of a list of regulators with positively or negatively correlated TGs, together with information on the enrichment statistic score. Interactive GRN visualization in a Cytoscape panel is integrated.

3.3. Single-Cell resource

3.3.1. SCENIC

SCENIC is an R package to infer GRNs by single-cell expression data using *cis*-regulatory sequences in a three-step workflow: in the first step, *GENIE3* is applied to identify sets of genes coexpressed with TFs, then motif analysis is applied to each coexpression. Motif analysis is operated by *RcisTarget*, which is a Bioconductor package identifying regulatory motifs over-represented on a gene list. Only significantly enriched motifs are retained in the model. Finally, the regulatory subnetwork is used to cluster cell types and states. The current version of *SCENIC* supports human, mouse and fly data analysis. A Python implementation of this method is also available [123].

4. ChIP-based resources

Although the presence of a given TF binding site in the promoter or enhancer region of a gene may suggest its transcriptional regulation, the sole information on its binding motif does not prove the existence of a TF-TG interaction in a given context, for example due to tissue-specific chromatin accessibility or tissue-specific availability of a key co-transcription factor required for the functional TF binding. In fact, an essential step in the definition of GRNs is to identify the associations between TFs and the genes they regulate and to establish the correct nature of their relationship (i.e. context-specific activation or repression) [124]. Great advances in this field have been made possible by the development of new laboratory techniques enabling the identification of TF-DNA binding sites on a genome scale, providing a measurement of epigenetic regulation of genes. One of such techniques is Chromatin Immunoprecipitation (ChIP), a method which allows to extracting and isolating specific protein-DNA chromatin complexes from living cells [125]. Coupling ChIP to high-throughput techniques such as NGS sequencing (ChIP-seq) or DNA microarrays (ChIP-chip) allows identifying a map of genome-wide binding sites for the TF under investigation, which can be used as input to infer GRNs [126]. Associating TF binding regions to target gene expression is crucial to build GRNs from ChIP data [127], and several databases are now available collecting hundreds of such datasets that are publicly accessible [126]. A set of tools is available to interpret and annotate peak calls, which are the significantly enriched regions of interest defining TF binding sites from a ChIP experiment [128]. Peak annotation tools mainly focus on the assignment of each peak to the corresponding gene by applying methods to calculate the proximity of the closest TSS in the direct neighborhood.

Most widely used peak annotation tools include HOMER (see above) [117], GREAT [129] and many others, which are reviewed below.

Discovery of gene regulatory modules from DNA occupancy is difficult, because physical interaction between regulators and DNA regions alone does not provide sufficient functional information. Integrating TEPs with chromatin state and TF location analysis provides a complementary method to infer more robust GRNs. While it is difficult to determine direct interactions from transcriptomic analysis alone, the use of ChIP-seq information can help to identify such interactions within a GRN, improving the resulting model [127]. *GRAM* (Genetic Regulatory Modules) was one the first algorithms to reconstruct GRNs by integrating TEPs with ChIP data by using 106 TF ChIP-Chip data and 500 expression profiles in *Saccharomyces cerevisiae* [130]. Since then, many other methods and tools have been developed to integrate ChIP data in GRN inference methods and pipelines.

For the sake of the current review, we identified two classes of resources that can be useful for GRN reconstruction: 1) databases that collect and provide context-specific ChIP-Seq information. 2) Tools that annotate ChIP-Seq data, often associating “peaks” (regions where the ab-targeted TF binds more than observed in a control/input sample) with functional genomic locations (e.g. promoters, enhancers), and interrogate databases in group 1 to infer gene regulation from promoter specific TF binding.

4.1. ChIP-Seq databases and Web tools

Since the advent of the NGS Revolution, an avalanche of ChIP-Seq data has been generated and deposited for thousands of target proteins (mostly TFs), organisms, tissues and experimental setups. The results of these experiments are commonly stored by authors in NGS repositories like the European Nucleotide Archive (ENA) [131] or the Sequence Read Archive SRA (SRA) [132].

4.1.1. ENCODE

The Encyclopedia of DNA Elements (*ENCODE*) is a large project which collected vast amounts of quantitative molecular data from the human genome in different cell types and conditions [133]. *ENCODE* currently provides ChIP-Seq data on 12 histone markers and 187 transcription factors across 118 human cell lines, for a total of 9308 samples. The *ENCODE* Consortium has defined and implemented an “audit” system to classify the quality of experimental data based on flags. The standardized high-quality dataset is available both as raw data (FASTQ sequences or BAM alignments) and as already inferred binding sites, in the form of Browser Extensible Data (BED) files describing peak locations in the genome. These tracks are also available at the UCSC Genome Browser for checking specific genome regions for TF binding [134].

4.1.2. ChIP Atlas

While not as standardized as *ENCODE*, the *ChIP Atlas* [135] currently contains a larger amount of experiments (96,000), essentially all those deposited in the major public repositories, including ENA and SRA. It focuses on six model organisms: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae*. It can be queried using four modes. The first, Peak Browser, allows to download multiple TF-binding tracks (peaks) as a BED file or to visualize them on an online genome visualizer. The second, Target Genes, is the one most immediately useful for GRN inference, as it allows identifying TGs of a specific TF in selected experiments and cellular contexts. It also allows defining the distance of the TF-bound region from the TSS to define putative TGs. The third, Colocalization, allows predicting TFs co-localized with other TFs. The fourth, called “Enrichment Analysis”, allows the user to upload a custom BED file and compare it to available ChIP-Seq tracks.

4.1.3. ChIPBase

ChIPBase is an online collection of ChIP data which can be queried to identify genomic regions targeted by TFs [136]. It contains a manual curation of 10,200 peak datasets from 10 species. On top of being a curated collection of experiments, *ChIPBase* can be used to further investigate ChIP-Based GRN hypotheses with motif coexpression analyses. Its main strength is in the addition of noncoding TGs, whereas most of the other tools focus on coding genes.

4.1.4. Cscan

Cscan is a simple online tool that allows querying online databases with a list of transcript ids for TGs [137]. For this set of TGs' promoters, the tool detects TFs (and DNA-binding proteins) whose binding sites are significantly enriched, using mainly information from ChIP-Seq experimental databases. *Cscan* allows to data mine individual cell lines and perturbation experiments for specific TFs, allowing for the comparison of context-dependent GRN rewiring.

4.1.5. GREAT

GREAT (Genomic Regions Enrichment of Annotations Tool) is a popular online tool to annotate a user-provided ChIP-Seq experiment, in the form of a BED file [129]. For each ChIP-Seq peak set analyzed, *GREAT* finds the collection of genes likely regulated by the investigated TF (proximal and distant TGs are based on nucleotide distances defined by the user). GREAT also provides automatic gene ontology analysis to predict the functional role of the TF GRN and a large series of other ontologies, as well as mouse and human phenotypes associated to the GRN. The only disadvantage of this tool is the small selection of genomes on which it operates: *Homo sapiens* (hg19), *Mus musculus* (mm9 and mm10) and zebrafish (*Danio rerio*) (danRer7). A common workaround for this issue is to use the *hgLiftOver* tool to convert between genome versions (e.g. hg38 to hg19) provided by the UCSC Genome Browser platform [130].

4.1.6. PAVIS

Like *GREAT*, *PAVIS* (Peak Annotation and VISualization tool) is an online resource that can annotate ChIP-Seq data [138]. It supports more genomes and annotations than GREAT and provides less information on disease-related phenotypes associated to the annotated genomic regions. It provides however a constantly updated genome annotation, allowing connecting any genomic peak set (such as a specific TF binding site) to the promoters of both coding and noncoding genes, including alternative TSSs for splice variants.

4.1.7. LOLA

LOLA (Locus Overlap Analysis) is a web tool that allows the characterization of genomic regions in terms of overlap with genomic annotation tracks [139], mostly TF binding sites from *ENCODE* and *JASPAR*, a manually curated repository of GRN information discussed later here [140]. *LOLA* also provides comparative analyses with the *ROADMAP* database for epigenomics annotation in human, focusing on regions with different chromatin accessibility, histone modification or DNase I hypersensitivity [141]. In short, *LOLA* can be used to understand transcriptional regulatory events happening upstream of a list of user provided genomic regions and/or TGs. The code running beneath the *LOLA* web tool is available as an R/Bioconductor package, which can be extended to any organism or genome annotation track set.

4.1.8. ChIP-Array

ChIP-Array is a full-fledged online GRN reconstruction pipeline that integrates user-provided ChIP-Seq data for a specific TF [142] with a specific transcriptional quantification dataset (microarrays or RNA-Seq), making it a hybrid between the ChIP-Seq and coexpression classes of tools. *ChIP-Array* then generates a GRN based on ChIP-Seq peak locations and coexpression derived from the user input. It then performs several coexpression conditioning and data mining steps on motif

databases to remove indirect GRN edges. It allows for the inclusion of long-range (> 10 kb) Chromatin Interactions to justify TF-TG interactions and integrates database information on tissue-specific chromatin accessibility to define active and inactive TF-TG relationships.

4.1.9. ChIP-Enrich

ChIP-Enrich is another peak annotation web tool [143], like *GREAT* or *PAVIS*, with focus on human, mouse, rat and zebrafish genes and pathways. Compared to the other tools, it allows speeding up the analysis by selecting specific annotation resources, to define which classes of TGs are associated to the TF investigated in a ChIP-Seq experiment. Specifically, it can focus the analysis on proximal and distant TF-TG interactions and provide a list of microRNAs potentially influencing these interactions. Importantly, *ChIP-Enrich* provides putative pharmacological targets in the predicted GRN, through the integration with DrugBank information [144], and it can therefore be used as a hypothesis generator tool for pharmacological intervention for specific ChIP-inferred GRNs.

4.2. ChIP-Seq-based tools

4.2.1. ChIPpeakAnno

ChIPpeakAnno is a powerful R/Bioconductor package [145] that can annotate genomic regions defined with the *GenomicRanges* framework [146]. Following the tutorial, it is possible to convert the BED file obtained from a ChIP-Seq experiment into a set of genomic regions, and then annotate them with respect to specific genomic annotations. To reconstruct a TF-centered GRN, for example, the user simply needs to load a BED file from a ChIP-Seq experiment targeting that TF and find, for each ChIP-Seq peak, the closest TSS(s), defined by the user or by genome annotation R packages such as *org.Hs.eg.db*. This will define a list of putative TGs for the given TF.

4.2.2. DROPA

DROPA (DRIP Optimized Peak Annotator) is a new peak-annotation tool optimized for DNA/RNA hybrid regions (R loops) and data generated from DRIP-Seq [147]. *DROPA* is a command line Python tool and it is currently provided with a user-friendly tutorial, together with a companion test dataset. While originally developed for DRIP-Seq tracks, it can be used for NGS data from specific protocols such as Histone marks IP-Seq, DNase-Seq and FAIRE-seq, as its true strength lies on the ability to annotate genomic regions taking into account gene expression information. It can be used as a GRN reconstruction method focusing not only TF-TG interactions, but on any functional genomic feature-TG relationship.

4.2.3. UROPA

UROPA (Universal RObust Peak Annotator) [148] is a ChIP-Seq annotation method implemented in R and Python, which focuses on advanced aspects of peak annotation. For the purpose of GRN reconstruction, it allows the user to describe genomic elements associated with TF binding sites, in particular strandedness, with specific functions (coding, noncoding, miRNAs) and with a multitude of different gene annotation supports (such as those from ENSEMBL or Gencode).

4.2.4. Goldmine

Goldmine is an R package for annotation of genomic regions (such as those generated via ChIP-Seq) [149]. As other tools described here, Goldmine provides peak annotation with reference to promoter proximity, DNase hypersensitivity regions, overlap with other TF binding sites, in all species and tracks available at the UCSC Genome Browser. Its unique feature is the ability to annotate peaks with high-throughput methylome data, highlighting which methylation patterns may influence the binding of TFs to TG promoters, and therefore yielding a further layer of regulatory complexity in GRN reconstruction.

4.2.5. ChIPMunk

ChIPMunk is a Java command line tool optimized for the analysis of large amounts of ChIP-Seq and DNase footprint data (hundreds of Gigabases of aligned NGS reads) to find specific consensus regions supporting the existence of sequence-based GRNs for an investigated TF or regulatory feature [143].

4.2.6. Genexpi

Genexpi is a ChIP-Based tool for GRN inference that combines user-provided ChIP-Seq profiles with time series expression data to define TF-TG pairs [150]. It starts with defining candidate targets with ChIP-Seq which are then tested for coexpression with the investigated TF. *Genexpi* in particular can work with time series, effectively calculating coexpression between a TF TEP at time t with any TG TEP at time t + 1, t + 2 or more. This provides not only a GRN inference, but also provides a time frame describing the delay in transcriptional responses between the TF and its targets.

4.2.7. VULCAN

The *VULCAN* tool uses ChIP-Seq data from perturbation experiments to predict GRNs and transcriptional co-regulators. Specifically, it has been recently applied on ChIP-Seq data from estrogen-treated cells targeting the Estrogen Receptor (ER) TF, and successfully identifying its TGs and ER's direct interactors [151]. *VULCAN* implements similarities in TEP between TFs and ChIP-Seq overlapping sites to predict the components of transcriptional complexes, thereby focusing on a similar task as the *ChIP-Atlas* or *PlantPAN* colocalization tools described earlier.

5. Orthology-based resources

Gene networks, and in particular transcriptional networks, can benefit from knowledge and discoveries available in other species. The key concept is that a TF-TG relationship proven in one organism can be conserved in another one (Fig. 3). This knowledge transfer, however, requires knowing the specific phylogenetic translation between two organisms in order to identify the correct functional counterpart for any TF-TG pair. A correct definition of orthology is of fundamental importance to reliably transfer GRN models across species [152].

Orthology is an evolutionary relationship between a pair (or multiple forms) of genes that share a common ancestor. This form of relationship contemplates sequence similarity, or, at least, pattern conservation [153]. Because of the involvement of pattern conservation, orthology can also be stated as similarity in function or conservation in function: this definition is rather equivocal since there is not a definite method to assign functional conservation to genes or proteins [154].

In a context of gene network inference, detecting orthology is especially important to maximize information content and accuracy. Identifying the presence of genes in different species that can be traced to a common ancestor and assuming that the orthology is not only sequence-based but also functional, allows for the hypothesis of a complex TF-TG network. The presence of multiple candidate orthologs of both TFs and their TGs enriches the overview on all the possibilities of regulatory interactions between these two elements. Furthermore, the prediction of an interaction in one species can be extended and verified in other species. The extension of the prediction of regulatory interactions among species can be achieved with bioinformatic tools, with the aim to gain more and more detailed knowledge about the regulatory systems of transcription [155].

The prediction of orthology can be obtained with a variety of methods, some of which can be used for the inference of both evolutionary and functional orthology. A complete list of these methods has been described before [156] and is beyond the scope of this review. We provide a summary of these methods in the Supplementary Materials. There are large databases collecting information on orthologous relationships, such as eggNOG [157], inParanoid [158] and orthoDB [159]. We provide a curated example of Orthology-inference of GRN

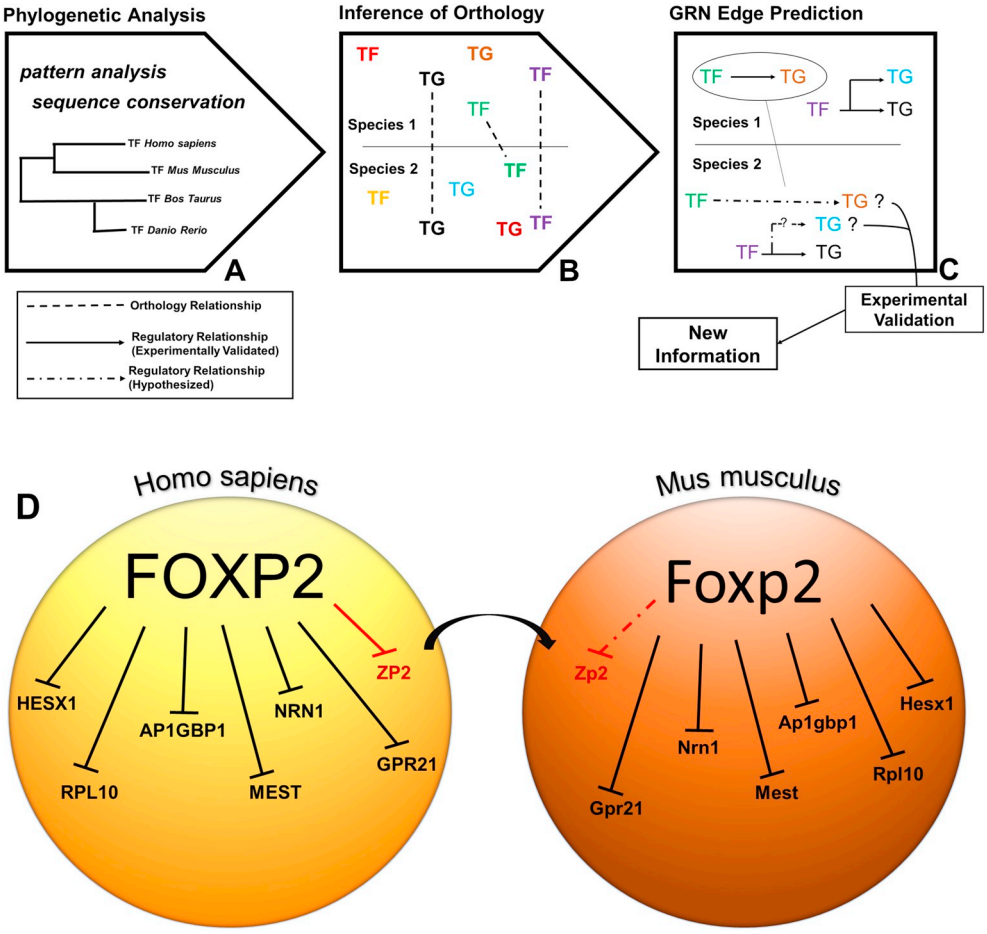

Fig. 3. Schematic representation and examples of the steps to predict new interactions between TFs and TGs using orthology-based approaches. (A) Phylogenetic Analysis is required to define sequence similarity across species. (B) Orthologous relationships are then inferred between two species, identifying proteins that are structurally and/or functionally correlated. (C) GRN TF-TG relationships (edges) are then inferred in a species, leveraging known relationships between those genes in another species. A set of prediction is therefore generated computationally and, if validated, will generate novel GRN information. (D) Example of TF that has orthologs in *Homo sapiens* and *Mus musculus*. FOXP2 is mainly a transcriptional repressor [198], whose targets have been obtained from published ChIP-seq data in *Homo sapiens* [199] and *Mus musculus* [200]. We selected genes that correspond to the 'PROMOTER' location to the nearest gene and we then created an intersection that includes only the common target genes. We then selected a target gene in *Homo sapiens* that was missing in *Mus musculus* (ZP2). Zp2 exists in *Mus musculus* and is an ortholog of the human ZP2, but it's missing the information about the involvement of Foxp2 in its regulation. Since orthology is verified for both the TF and the TG we can make the prediction of interaction between murine Foxp2 and Zp2, that can be then verified experimentally.

edges in Fig. 3B, focusing around the Foxp2 TF.

However, there are also a few tools which adopt orthology-based ideas to automatically infer large scale GRNs, described below.

5.1. Orthology-Based GRN inference tools

5.1.1. MRTLE

MRTLE is a command line tool written in C/C++ that takes as input expression data from multiple species, a phylogenetic tree describing relationships between species, and optionally species-specific regulatory information, such as known TFs and their regulation on gene families [160]. *MRTLE* processes this information and produces a regulatory network for each species included in the phylogenetic tree. It can be considered as a validation tool for GRNs, validating which relationships are conserved across evolution and which are specific to a particular species or *phylum*.

5.1.2. TargetOrtho

TargetOrtho is a Python standalone GUI to identify putative TGs regulated by specific TFs. The user provides TF binding sites motifs (obtained from experimental data) and *TargetOrtho* scans the genomes of interest for matches, extending this analysis on orthologs in other species through motif-based remote homology detection [161]. *TargetOrtho* performs well in compact genomes (such as *C. elegans*, where it has been tested) but not as well in higher eukaryotes, where the regulatory elements are dispersed and often located far away (> 10 kb) from the TSSs.

5.1.3. Phylogene

PhyloGene is an online web service that provides information on significantly co-evolving proteins according to phylogenetic profiles and orthology inference. It contains information on thousands of proteins from model species (Human, mouse, *Drosophila* and nematode) [162]. It can be specifically queried to detect TF-TG pairs that are coevolving across phylogenetic groups.

5.1.4. OrthoClust

OrthoClust is a command line tool written in the Julia programming language which can detect if a specific gene regulatory modules are functionally conserved across species or are sequence-specific [163]. *OrthoClust* requires as input a network file of TF-TG relationships (inferred by any method) in all species considered and a coupling information file, which describes the orthologous relationships between genes the input species and another species. *OrthoClust* finally groups genes in the input species in functional modules that are conserved and provides a conservation score for them. *OrthoClust* can be useful in GRN inference because a TF-centered module highly conserved across species has a higher chance to be correct.

6. Literature-based resources

6.1. JASPAR

JASPAR is a manually curated and open access database of non-redundant TF binding profiles [140]. Its extensive manual curation, based on constant literature searches by experts in the field of genomics, elevates to a full-fledged literature-based resource. *JASPAR* stores TF binding information as position-frequency matrices (PFMs) and TF flexible models (TFFM), based upon TF binding site predictions made by the UCSC Genome Browser track data hub. As of 2018, JASPAR

counted with 1564 position-frequency matrices, including 719 from vertebrates, 501 from plants and 140 from insects.

6.2. TRANSFAC

Similar to *JASPAR*, the *TRANSFAC* (for TRANScription FACtor) database is a manually curated repository of eukaryotic TF binding sites and DNA binding profiles [164]. TRANSFAC is frequently used to computationally predict TF binding sites, to predict genes regulated downstream a target sequence and to predict all TFs that regulate a given set of genes. To these ends, TRANSFAC counts on nearly a dozen algorithms compute its results and to compare them to other sources. Currently, the most up-to-date version of the database must be licensed, whereas older versions can be accessed for free.

6.3. KEGG

A massive data collection effort, the *KEGG* (Kyoto Encyclopedia of Genes and Genomes) database generates high-level maps of complex biological systems, such as biological processes, diseases and chemical substance effect pathways, by integrating numerous large-scale molecular datasets [165]. Included in the array of *KEGG* resources are those generated by genome sequencing, various small molecule and enzyme assays and other high-throughput experimental technologies.

6.4. MSigDB

The *MSigDB*, or Molecular Signatures Database, comprises a collection of annotated gene sets originally formatted for use in Gene Set Enrichment Analysis (GSEA) [166]. The determination of whether a given gene set is significantly up- or down-regulated in response to a certain perturbation can help explain underlying biological mechanisms. The *MSigDB* relies upon published expression profiles to generate gene sets, as the developers believe that this provides unbiased readouts of biological states [167]. Conveniently, *MSigDB* output includes files to easily convert between HUGO and ENTREZ gene symbol nomenclature. One caveat to the *MSigDB*'s gene sets is that many are derived from microarray experiments, which reflect purely transcriptional events.

6.5. Harmonizome

Harmonizome is a vast collector of genomics databases, including *TRANSFAC*, *KEGG*, *MotifMap*, *MsigDB* (all described in this review) and others not specifically connected to GRNs [168]. *Harmonizome* allows downloading entire databases as flat files of interactions and information, which is considered very useful by the bioinformatics community. Furthermore, it allows for gene-centered queries to investigate specific GRN of interest, a feature considered optimal for experimental scientists. Interestingly enough, *Harmonizome* contains a snapshot of the *Biocarta* database, a historical, manually curated, literature-based collection of biomolecular relationships from both signal transduction pathways and GRNs [169]. The *Biocarta* web domain has been inactive for several months and the project seems to be discontinued at the moment of writing this review.

6.6. AGRIS

The first of species-specific resources in this paragraph, the Arabidopsis Gene Regulatory Information Server (*AGRIS*) comprises three interlinked databases (*AtTFDB*, *AtcisDB* and *AtRegNet*) providing a broad collection of TFs, computationally and experimentally derived *cis*-regulatory elements, and their interactions in *Arabidopsis thaliana* [170]. Currently, this database comprises around 33,000 upstream regions of annotated *Arabidopsis* genes and 1700 TFs, counting > 1.5 million direct interactions between TFs and TGs. It includes the Grassius Regulatory Grid eXplorer (GRG-X), a web application to visualize and manipulate the regulatory network.

6.7. Flybase and Wormbase

Flybase is the primary repository for genetic and molecular information concerning *Drosophila melanogaster*, one of the most extensively studied model organisms, as well as other fly species. Manually curated, *Flybase* incorporates an integrated collection of genetic, molecular, genomic and developmental information. The *Flybase* team made a significant update to the database in 2018, with “Flybase 2.0”, which features enhanced reference lists and new protein domain graphics, powered by the Pfam and SMART protein databases. Another key feature of the updated *Flybase* is an ‘experimental tools’ function that consolidates information concerning the reagents involved in designing fly strain-specific experiments [171]. The *Flybase* counterpart for nematodes, containing similar information, is called *Wormbase* [172].

6.8. RegulonDB

Curated over a period of 25 years, *RegulonDB* contains the most extensive collection of knowledge concerning the genome organization and transcriptional regulation of *Escherichia coli* K-12 [173]. *RegulonDB* integrates information concerning transcriptional units, operons and regulons and it makes an attempt to incorporate both strong and weak interaction evidence by classifying distinct types of evidence and weighting them appropriately. Although this method risks some false positive interactions, such as when multiple pieces of weak evidence are combined and enhanced to be considered as one piece of strong evidence, it has the advantage of not arbitrarily dismissing potentially useful information. The latest versions of *RegulonDB* also include features to quantify coexpression of all possible gene pairs and to identify upstream TF binding sites, whose activity supports evidence for regulatory interactions.

6.9. Saccharomyces-Genome Database (SGD)

SGD (Saccharomyces Genome Database) [92] is the foremost repository of genetic information pertaining to the budding yeast *Saccharomyces cerevisiae*. Published experimental results are combined with high-throughput results obtained via the Locus Summary genome browser, forming a vast encyclopedia of yeast genomic features. The detailed chromosomal characteristics, functions and interactions contained in the *SGD* are publicly accessible and indispensable for designing yeast experiments and interpreting their results. The S288C reference genome is frequently updated and annotation is performed using GO terms, all of which can be easily downloaded for offline work. Changes and updates to the *SGD*, as well as other relevant announcements are also accessible on social media, via Facebook and Twitter (@yeastgenome).

7. PPI resources for inferring transcriptional complexes

PPI networks describe physical interactions between gene products and are beyond the scope of this review. However, a specific sub-graph of PPI networks is important for understanding regulation of transcription, specifically the interactions between transcriptional regulators. In fact, we have mentioned earlier the VULCAN tool, which is able to infer transcriptional complexes using a combination of coexpression and ChIP-Seq inferences [151]. Another important tool trying to define interactions between TFs is PTHGRN [174], which goes beyond the scope of this review as it focuses on the effects of post-translational modifications (acetylations, phosphorylations) on TF activity.

There are several resources for investigating PPI networks that could be queried in specific ways to predict or visualize known TF-TF

interactions. A recent effort to standardize these resources is currently undergoing under the IMEx consortium [175].

7.1. HPRD

HPRD (Human Protein Reference Database) is a manually curated collection of protein-protein interactions [176]. Each physical interaction is experimentally validated by at least one published study (referenced to for each edge). Furthermore, interactions are grouped among "yeast 2-hybrid" (high throughput interaction assays), "in vitro" and "in vivo" validation classes, so the user may choose to trust only one of them. However, as the name implies, HPRD focuses exclusively on the human proteome. 4.27% of PPIs reported by HPRD (Release 9) are TF-TF interactions (Table 1).

7.2. STRING

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is the largest PPI database available today, currently covering 5090 organisms, 24.6 million proteins and > 2000 million interactions. STRING automatically combines PPI evidences from several sources such as direct experiments, orthology screenings, coexpression and text mining. Unlike the other tools described here, which are mostly based on collecting PPI evidences, STRING provides also predictions, where a protein-protein summary score is computed by combining the probabilities of each evidence [177]. Roughly 1.30% of *STRING* (v9.05) reported interactions draw PPI predictions between TFs (Table 1).

7.3. BioGRID

BioGRID is similar to *HPRD*, as it searches literature and databases for experimental evidences of PPI, but in an automated way [178]. It provides data for 68 organisms (most prominently model ones, such as human, yeast, fruit fly and *Escherichia coli*). Each edge is classified by virtue of the experimental system(s) supporting the PPI evidence, in a more precise way than HPRD such as co-immunoprecipitation, synthetic lethality, fluorescence resonance energy transfer (FRET) and yeast two-hybrid. In the 3.5.172 version of *BioGRID*, 1.84% of the PPIs connect a TF pair (Table 1).

7.4. IntACT

IntACT is the most recent among the short list of PPI databases described here [179]. Its multi-species molecular interactions database is populated by manually curated literature data and by high-throughput data analysis. It does not only contain PPI data but also interactions between proteins, nucleic acids and small compounds. The fraction of TF-TF interactions in the Human *IntACT* interactome is 1.29%, very similar to the percentage predicted by STRING (Table 1).

Table 1
PPI databases, with focus on TF-TF interactions. The percentage of TF-TF interactions is estimated using the human proteins portion of the resource, and the TF list is derived from [7].

Tool	Nr. proteins	Nr. interactions	Nr. interactions (human)	% of TF-TF interactions
HPRD	9617	39,240	39,240	4.27%
STRING	~24,600,000	~2,000,000,000	4,009,084	1.30%
BioGRID	87,223	1,693,097	481,338	1.84%
IntACT	107,320	889,774	379,393	1.29%

8. Other resources

8.1. Graphite

Graphite is a framework that converts pathway topologies into gene networks [180]. It is available as both an R/Bioconductor package and as a web server (*Graphite* Web). *Graphite* provides two main functionalities. In the first mode, "Browse", it allows to navigate manually curated pathways (such as the *KEGG* collection) using a list of genes, to identify those where they are enriched in specific pathways and GRNs. In the second mode, "Analyze", Graphite allows the user to provide an expression dataset with multiple conditions (e.g. control/treatment) to investigate changes in pathway correlation structure, i.e. gene network (and specifically GRN) rewiring.

8.2. TRRUST

TRRUST (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text-mining) [181] is a peculiar method for GRN inference in human and mouse, as it tries to detect co-occurrences between TFs and TGs by large scale literature and text mining. The results are then curated manually and included in an online database, which currently contains 8400 TF-TG relationships for 800 human TFs and 6500 relationships for mouse TFs.

8.3. EnhancerAtlas

Enhancer regions are incompletely characterized functional parts of the genome with transcriptional effects on distal genes. Tools like *EnhancerAtlas* allow for the inclusion of enhancers as entities (nodes) to be represented in GRNs [182]. *EnhancerAtlas* provides a cell line-specific collection of human enhancers, both as FASTA sequences and coordinates on the hg19 genome. It also contains a cell line-specific set of GRNs with proven relationships between enhancers and TGs.

8.4. TF2DNA

TF2DNA is a database of TF-TG relationships in six model species, encompassing hundreds of TFs and thousands of TGs [183]. These relationships are inferred through 3D models of TFs deposited at the Protein Data Bank (PDB) [184]. Based on structural data and computational simulations, *TF2DNA* builds models of TF-DNA binding, identifying binding motifs for TFs with a deposited high-quality 3D structure and defining structure-based models of GRNs. Then, via homology modelling, information is transferred over other TFs to extend the prediction to a larger number of TF-centered GRNs.

8.5. ELMER

ELMER is a R/Bioconductor package and web-tool that infers GRNs by combining genome-wide methylation data with transcriptome-wide TEPs [185]. *ELMER* can be used on publicly available data where methylation and gene expression were measured in the same samples, such as those from TCGA [186], or with user provided data: RNA-Seq and Illumina 450K or EPIC arrays for methylation probes. *ELMER* starts by analyzing which genomic locations have the most significant methylation changes between two groups, and then associates these probes to proximal (10 kb by default) TSSs. *ELMER* performs a motif enrichment analysis to identify a list of putative TFs associated to differentially methylated regions and then identifies a subset of these TFs whose expression correlates with the methylation changes. It finally connects the methylation-guided TF to the genes in proximity of these regions to draw a final GRN. It must be noted that *ELMER* necessarily operates on a perturbation setup (e.g. tumor vs. normal) because it relies on expression and methylation differential analysis.

8.6. 3D Genome Browser

The *3D Genome Browser* is an online repository of 3D chromatin data, generated by HiC and ChIA-Pet [187]. It can be queried by dataset or by gene name to investigate topological distances between specific genomic regions. While specific for human and mouse, it allows to infer GRNs over long range distances (> 100 kb) as it allows to detect real 3D proximity between TF binding regions (e.g. enhancers) and influenced TSSs.

8.7. CrossNet

We conclude our collection of tools for GRN inference with *CrossNet*, a web application for comparative analysis of multiple biological networks [188]. *CrossNet* allows the user to upload weighted and unweighted networks as simple text files and then compare them using several network comparison methodologies and analysis workflows, spanning from a simple edge intersection to a network-wide topological analysis (clustering coefficient distributions, shortest paths, etc.), conserved hubs (nodes with a high number of edges in several networks) and higher-order network structures (such as conserved network motifs). *CrossNet* can compare multiple (more than two) networks and display their similarities as a hierarchical tree, with the possibility to install it on a personal workstation to accommodate the offline requirement of processing large (> 2000 nodes) networks. For the purpose of GRN reconstruction, while not doing any GRN inference per se, *CrossNet* can be used to compare and merge GRNs inferred from other tools, highlighting similarities and differences in both individual transcriptional units and in the whole GRN structure.

9. Best practices and example

The inference of GRNs is a useful approach to handle several biological problems and to derive new functional hypotheses to be experimentally validated. The complementary use of some of the tools reviewed above can help researchers in solving the inference problem. Once approaching the task of inferring a GRN-centered around a specific TF, the investigator should first of all familiarize with the TF itself: is it an activator or a repressor (which could explain positive or negative correlations in coexpression analysis)? Does it have a clearly defined and well characterized ortholog in a more studied organism? Is there knowledge on the DNA motifs it recognizes and binds? Are there fingerprinting, ChIP-chip or ChIP-Seq experiments focusing on such TF? Are there perturbation experiments that perturbed the TF by either knockdown or overexpression?

The following example focuses on a well-known repressor TF, B-cell lymphoma 6 (BCL6), and it is based on a pipeline integrating methods from complementary data sources.

9.1. Characterize TF and context (species and cells)

BCL6 is a transcriptional repressor belonging to the BTB/POZ zinc-finger family of TFs. It plays a critical role during normal B-cell development, and a variety of structural and functional alterations of BCL6 have been associated with lymphomagenesis. In the last years, the deregulation of BCL6 has been linked to several other malignancies, such as acute lymphoblastic and chronic myeloid leukemia, breast, colorectal, and non–small cell lung cancer [189]. As a transcriptional repressor, BCL6 mediates its effects on several TGs by interacting with different chromatin modifying repressor complexes.

The network perspective can broaden the understanding of BCL6 by identifying the full set of TGs under its direct influence, allowing drawing predictions on the effects of targeting BCL6 with targeted drugs, causing rewiring of its GRN. Furthermore, the depiction of its context specificity may represent an interesting tool for drug discovery/drug repurposing by creating an in silico platform to test and predict possible treatment strategies, or to identify new druggable targets in a dysregulated BCL6 context.

In order to investigate the cancer-specific regulatory action of BCL6, we focused the GRN reconstruction in the specific context of human leukemia cells. Our intention is to run user-friendly methods to understand which TGs are regulated by BCL6 and that could be also handled by researchers who are completely new to bioinformatics.

9.2. Coexpression analysis

If large transcriptome-wide expression datasets are available in the desired cellular context, coexpression methods are an excellent starting point as they will provide an initial list of candidate TGs. In our case, we ran the ***ARACNe-AP*** on the TCGA Acute Myeloid Leukemia dataset, as in [43], to reconstruct a BCL6 coexpression network keeping only negatively correlated TGs, consistently with the role of transcriptional repressor of this TF.

9.3. ChIP-Seq analysis

Since correlation does not imply direct causation, coexpression alone cannot be used as a reliable evidence for TF-TG interactions. Ideally, coexpression evidence should be strengthened by ChIP-Seq data for the selected TF in the same context used in the coexpression analysis. ChIP-Seq analysis will provide regions bound by the TF, which can be associated to genes e.g. by proximity to the TSS.

As BCL6 has previously generated ChIP-Seq data, we used the ***ChIP-Atlas*** "Target Genes" query [135], we retrieved the analysis of a ChIP-seq experiment on a human chronic myelogenous leukemia cell line (GSM640427) showing all ChIP peaks within 5 kb distance from the TSS, and intersected it with the coexpression list to refine for results that are present in both inferences.

9.4. Motif analysis

Coexpression and ChIP-Seq may define high-likelihood TGs for a specific TF, but they do not provide a molecular explanation on why the TF binds and activates/represses those particular genes. Motif-based methods can be used when a known TF binding motif is known, or, in the case of poorly characterized regulators, to perform a de novo discovery of the binding motif (e.g. by *HOMER*) by analyzing the ChIP-Seq regions bound by the TF. Motif-based predictions are to be taken carefully, as often the TF binding is not fully characterized or, in some cases, the TF does not bind the DNA directly to exert its regulatory action, but rather chromatin-associated complexes.

In our example, a BCL6 binding motif is known, so we used ***Motifmap*** [107] to obtain a genome-wide prediction of BCL6 TGs based on the "Motif Search" web-tool. We used default search settings on hg19 reference genome, using a 5 kb distance to the closest TSS. We identified 223 candidate genes, but when crossed with the coexpression list, only 3 common TGs were kept in the final model. In this particular example, the fact that the canonical binding motif of BCL6 explains such a low fraction of the candidate TGs (derived from ChIP-Seq + Coexpression) allows us to venture the opinion that the precise binding of BCL6 to the genome is not yet fully explained by the current scientific knowledge.

9.5. Literature analysis

A key step in GRN reconstruction is to quantify the number of predicted TGs that are already known in literature. As reading each article can be a tedious task, the user may use some of the online repositories of known GRN interactions described in this review.

In our example, we queried the ***TRANSFAC*** Predicted Transcription Factor Targets Dataset for BCL6 available ***Harmonizome*** [168] and compared them with the previously inferred candidate TG list. Some

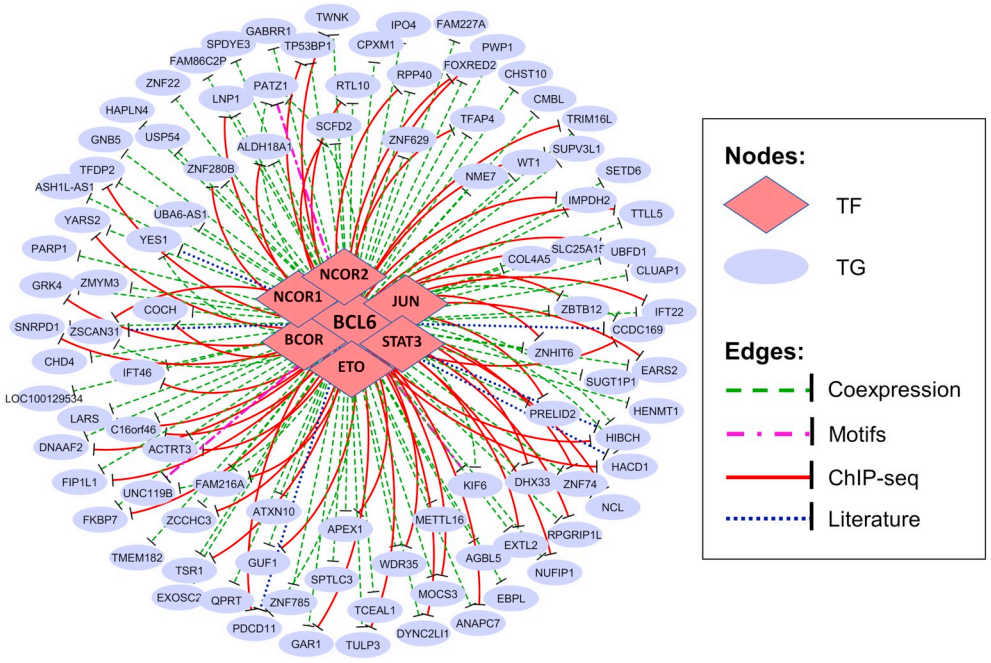

Fig. 4. Partial BCL6 regulon inferred using an integration of GRN reconstruction methods: coexpression (dashed green edges), motif-based (dashed/dotted purple edges), ChIP-based (solid red edges), and literature-based (blue dotted edges). Transcriptional regulators directly interacting with BCL6 are shown as rhombi.

coexpression-inferred BCL6 targets had been already characterized in literature, such as HACD1, ZSCAN31 and HIBCH. Some ChIP-Seq-based inferred TGs were also already known, such as PDCD11 or CCDC169 (also inferred by coexpression).

9.6. PPI Analysis to infer co-transcriptional interactors

TFs are not isolated in the cell and rarely bind the DNA by themselves, let alone drive their regulatory action in solitude. While PPI networks and GRNs should not be confused, it is important to understand which other proteins (especially those with transcriptional roles) physically bind the query TF, to promote or repress its effects on TGs' transcription.

In our example, we used PPI resources to infer the BCL6 transcriptional repression complex. The ***BioGrid*** database readily reports strong interactions between BCL6 and corepressors NCOR1, NCOR2 and BCOR: the disruption of these particular interactions is considered to be a promising pharmacological strategy for B Cell Lymphoma [190]. ***HPRD*** reports experimentally validated physical interactions of BCL6 with JUN TFs [191] and putative corepressor ETO [192].

9.7. Orthology analysis

Due to their vast employment in biology, the amount of raw data generated from model organisms often surpasses that from human. A specific example of GRN knowledge transfer across species is shown in Fig. 3, using orthology-based inference principles.

In our example, an important functional transcriptional interaction for BCL6 is reported in a different organism: the ***STRING*** database reports a BCL6-STAT3 interaction via a PPI network orthology inference, as the interaction has been observed in putative orthologs of these proteins in *Drosophila melanogaster*.

9.8. Aggregation of inferences

We suggest inferring (or at least attempt to infer) TGs using at least one method for each of the six complementary data sources described in this review. An even better approach would be to test several tools, even from the same category, and using different data sources. For example, coexpression analysis can be highly dependent on the dataset used: in some cases, the user may want to identify the generic GRN for a TF, and therefore generate GRNs from all tissues (e.g. all those available in GTEx [193]). In other cases, the user may want to infer a tissue or developmental stage-specific GRN, in that case the analysis should be limited to context-specific datasets. Furthermore, adopting a tissue-wide coexpression analysis can be problematic for TFs that are tissue-specific.

In the end, the user may want to collect all inferred TGs, weighted by the number of tools predicting them, and proceed with an experimental validation of the interaction. As the name implies, one of the most popular and intuitive ways to represent the results of a full pipeline of GRN inference is a network visualization: in our BCL6 example, we collected the inferences from multiple approaches to draw a model GRN with *Cytoscape* (Fig. 4). In order to collectively characterize the TF function by means of inferred TGs, one may want to run an ontological enrichment analysis to find over-represented pathway elements. The TGs inferred by our simple GRN pipeline are predominantly enriched for vesicle transport (PANTHER GO enrichment analysis [194], FDR = 3.28×10^{-4}), consistently with the validated notion that BCL6 represses autophagia and vesicle trafficking in B-cell lymphoma cells [195].

As a final note, GRN inference is sometimes performed full-scale and genome-wide, by calculating all TF-TG interactions at once. This holistic approach, when compared to the single TF one exemplified above, is more time-consuming and difficult to interpret, but it follows the same guidelines and caveats. Specifically, in coexpression analysis, considering all TFs at once is important because it may remove indirect, spurious edges, through solutions like the Data Processing Inequality or the Partial Correlation described before. In the BCL6 example, we actually considered all TFs in the coexpression step, since we performed a full-scale ARACNe-AP inference, and only then extracted the BCL6-specific edges.

10. Conclusions and future directions

GRNs provide both a theoretically sound and a graphically

convenient representation of genome-wide transcriptional dependencies. Understanding GRNs requires tools, methods, datasets and integration to infer the hundreds of thousands of interactions happening in every organism, and how these interactions are rewired in response to external stimuli and during pathophysiological processes. In particular, recent years have seen a shift towards using GRNs and GRN rewiring as biomarkers for stratifying heterogeneous diseases such as cancer [28,33,66,87]. GRNs provide not only biomarkers for survival predictions, but they also directly depend upon one or more TFs acting as master regulators of a specific set of TGs carrying on molecular functions. Prognostic GRNs, and their TF hubs, constitute an ideal target for pharmacological modulation, more likely to succeed the individual gene expression profiles currently used as markers for clinical outcome, as these constitute just one component of the GRN responsible for pathological progression [28].

The tools described in this review provide an array of methods for inferring, interrogating and visualizing GRNs. Because each tool can only reasonably focus on one or a few aspects of a regulatory network, best practices for inferring and interrogating a network of interest will include integrating results from several sources into logical and human-readable graphical outputs. The research objective will inform the best resources to be used. For instance, no method can definitively claim to be the absolute best at inferring putative transcriptional targets, putative post-translational targets, or master regulators that drive certain phenotypes. Despite each method's strengths, careful analysis of the results of multiple methods will provide the researcher with the most complete and useful GRN-based insights into experimental results.

An expanding set of tools, complementary to those for inferring TF-TG relationships, focuses on less characterized GRN elements, such as long noncoding RNAs, miRNAs, enhancer elements and functional chromatin regions [136]. Tools should emerge to describe elements whose contribution to transcriptional regulation has been neglected, such as oxygen availability controlling TF degradation in plants [196] or repetitive DNA acting as a functional sponge for TFs [197].

While the inclusion of more features and datasets will lead to iterative improvements in the algorithms and methods used to infer GRNs, it remains of paramount importance for the scientific community to actively develop accessible tools, such as some of those described here, to allow a broad audience of scientists to investigate the vast amount of species- and context- specific information on transcriptional regulation.

Transparency document

The Transparency document associated this article can be found, in online version.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

First and foremost, we would like to show our honest appreciation to the two anonymous reviewers who contributed in massively improving the quality of our work during the revision process. We also thank our coworkers Prof. Giorgio Milazzo, Dr. Emanuele Valli, Chiara Cabrelle, Erika Gardini and Eleonora Fornasari for the fruitful discussions on gene regulatory networks; Prof. Alberto Danielli for the engaging conversations on DNA Shape-based TF binding; and Prof. Giovanni Perini for his scientific guidance and mentorship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2019.194430.

References

[1] G.W. Li, X.S. Xie, Central dogma at the single-molecule level in living cells, Nature. 475 (2011) 308, https://doi.org/10.1038/nature10315.
[2] H. Kitano, Systems biology: a brief overview, Science. 295 (2002) 1662–1664, https://doi.org/10.1126/science.1069492.
[3] E.R. Gibney, C.M. Nolan, Epigenetics and gene expression, Heredity (Edinb). 105 (2010) 4–13, https://doi.org/10.1038/hdy.2010.54.
[4] I.G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation, Nat Rev Genet. 13 (2012) 505–516, https://doi.org/10.1038/nrg3229.
[5] S. Djebali, C.A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G.K. Marinov, J. Khatun, B.A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R.F. Abdelhamid, T. Alioto, I. Antoshechkin, M.T. Baer, N.S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M.J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O.J. Luo, E. Park, K. Persaud, J.B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A.M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S.E. Antonarakis, G. Hannon, M.C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T.R. Gingeras, Landscape of transcription in human cells, Nature. 489 (2012) 101–108, https://doi.org/10.1038/nature11233.
[6] G. Orphanides, T. Lagrange, D. Reinberg, The general transcription factors of RNA polymerase II, Genes Dev. 10 (1996) 2657–2683.
[7] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The human transcription factors, Cell. 175 (2018) 598–599, https://doi.org/10.1016/j.cell.2018.09.045.
[8] A.C. Wilkinson, H. Nakauchi, B. Goettgens, Mammalian transcription factor networks: recent advances in interrogating biological complexity, Cell Syst. 5 (2017) 319–331, https://doi.org/10.1016/j.cels.2017.07.004.
[9] A. Ishihama, T. Shimada, Y. Yamazaki, Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors, Nucleic Acids Res. 44 (2016) 2058–2074, https://doi.org/10.1093/nar/gkw051.
[10] C.G. de Boer, T.R. Hughes, YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities, Nucleic Acids Res. 40 (2012) D169–D179, https://doi.org/10.1093/nar/gkr993.
[11] S. Shazman, H. Lee, Y. Socol, R.S. Mann, B. Honig, OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites, Nucleic Acids Res. 42 (2014) D167–D171, https://doi.org/10.1093/nar/gkt1165.
[12] P. Brazhnik, A. de la Fuente, P. Mendes, Gene networks: how to put the function in genomics, Trends Biotechnol. 20 (2002) 467–472, https://doi.org/10.1016/S0167-7799(02)02053-X.
[13] J. Zhu, B. Zhang, E.N. Smith, B. Drees, R.B. Brem, L. Kruglyak, R.E. Bumgarner, E.E. Schadt, Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks, Nat. Genet. 40 (2008) 854–861, https://doi.org/10.1038/ng.167.
[14] A.A. Margolin, K. Wang, W.K. Lim, M. Kustagi, I. Nemenman, A. Califano, Reverse engineering cellular networks, Nat. Protoc. 1 (2006) 662–671, https://doi.org/10.1038/nprot.2006.106.
[15] F.M. Lopes, D.C. Martins, R.M. Cesar, Comparative study of GRNS inference methods based on feature selection by mutual information, 2009 IEEE International Workshop on Genomic Signal Processing and Statistics, IEEE, Minneapolis, MN, USA, 2009, pp. 1–4, , https://doi.org/10.1109/GENSIPS.2009.5174334.
[16] E.H. Davidson, Gene regulatory networks and the evolution of animal body plans, Science. 311 (2006) 796–800, https://doi.org/10.1126/science.1113832.
[17] F.M. Delgado, F. Gomez-Vela, Computational methods for gene regulatory networks reconstruction and analysis: a review, Artif. Intell. Med. 95 (2019) 133–145, https://doi.org/10.1016/j.artmed.2018.10.006.
[18] Y. Shimoni, M.Y. Fink, S. Choi, S.C. Sealfon, Plato's cave algorithm: inferring functional signaling networks from early gene expression shadows, PLoS Comput. Biol. 6 (2010) e1000828, , https://doi.org/10.1371/journal.pcbi.1000828.
[19] J. Aldrich, Correlations genuine and spurious in Pearson and Yule, Stat. Sci. 10 (1995) 364–376, https://doi.org/10.1214/ss/1177009870.
[20] D.E. Carlin, E.O. Paull, K. Graim, C.K. Wong, A. Bivol, P. Ryabinin, K. Ellrott, A. Sokolov, J.M. Stuart, Prophetic Granger Causality to infer gene regulatory networks, PLoS One 12 (2017) e0170340, , https://doi.org/10.1371/journal.pone.0170340.
[21] M.H. Maathuis, D. Colombo, M. Kalisch, P. Buhlmann, Predicting causal effects in large-scale systems from observational data, Nat. Methods 7 (2010) 247, https://doi.org/10.1038/nmeth0410-247.
[22] M.J. Alvarez, Y. Shen, F.M. Giorgi, A. Lachmann, B.B. Ding, B.H. Ye, A. Califano, Functional characterization of somatic mutations in cancer using network-based inference of protein activity, Nat. Genet. 48 (2016) 838–847, https://doi.org/10.1038/ng.3593.

[23] J. Gillis, P. Pavlidis, “Guilt by association” is the exception rather than the rule in gene networks, PLoS Comput. Biol. 8 (2012) e1002444, , https://doi.org/10.1371/journal.pcbi.1002444.
[24] M. Zampieri, N. Soranzo, C. Altafini, Discerning static and causal interactions in genome-wide reverse engineering problems, Bioinformatics. 24 (2008) 1510–1515, https://doi.org/10.1093/bioinformatics/btn220.
[25] M. Saint-Antoine, A. Singh, Evaluating Pruning Methods in Gene Network Inference, IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, vol. 2019, 2019, pp. 1–7, , https://doi.org/10.1109/CIBCB.2019.8791237.
[26] A. de la Fuente, N. Bing, I. Hoeschele, P. Mendes, Discovery of meaningful associations in genomic data using partial correlation coefficients, Bioinformatics. 20 (2004) 3565–3574, https://doi.org/10.1093/bioinformatics/bth445.
[27] T.I. Lee, R.A. Young, Transcriptional regulation and its Misregulation in disease, Cell. 152 (2013) 1237–1251, https://doi.org/10.1016/j.cell.2013.02.014.
[28] F. Emmert-Streib, M. Dehmer, B. Haibe-Kains, Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks, Front Cell Dev Biol. 2 (2014) 38, https://doi.org/10.3389/fcell.2014.00038.
[29] A.J. Singh, S.A. Ramsey, T.M. Filtz, C. Kioussi, Differential gene regulatory networks in development and disease, Cell. Mol. Life Sci. 75 (2018) 1013–1025, https://doi.org/10.1007/s00018-017-2679-6.
[30] D. Mercatelli, F. Ray, F.M. Giorgi, Pan-Cancer and Single-Cell Modeling of Genomic Alterations Through Gene Expression, Front. Genet. 10 (2019). doi:https://doi.org/10.3389/fgene.2019.00671.
[31] B. Usadel, T. Obayashi, M. Mutwil, F.M. Giorgi, G.W. Bassel, M. Tanimoto, A. Chow, D. Steinhauser, S. Persson, N.J. Provart, Co-expression tools for plant biology: opportunities for hypothesis generation and caveats, Plant Cell Environ. 32 (2009) 1633–1651, https://doi.org/10.1111/j.1365-3040.2009.02040.x.
[32] R. Sibout, S. Proost, B.O. Hansen, N. Vaid, F.M. Giorgi, S. Ho-Yue-Kuang, F. Legée, L. Cézart, O. Bouchabké-Coussa, C. Soulhat, N. Provart, A. Pasha, P. Le Bris, D. Roujol, H. Hofte, E. Jamet, C. Lapierre, S. Persson, M. Mutwil, Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon, New Phytol. 215 (2017) 1009–1025, https://doi.org/10.1111/nph.14635.
[33] M. Giulietti, G. Occhipinti, A. Righetti, M. Bracci, A. Conti, A. Ruzzo, E. Cerigioni, T. Cacciamani, G. Principato, F. Piva, Emerging biomarkers in bladder cancer identified by network analysis of transcriptomic data, Front. Oncol. 8 (2018) 450, https://doi.org/10.3389/fonc.2018.00450.
[34] M. Banf, S.Y. Rhee, Computational inference of gene regulatory networks: approaches, limitations and opportunities, Biochim Biophys Acta Gene Regul Mech. 1860 (2017) 41–52, https://doi.org/10.1016/j.bbagrm.2016.09.003.
[35] L.E. Chai, S.K. Loh, S.T. Low, M.S. Mohamad, S. Deris, Z. Zakaria, A review on the computational approaches for gene regulatory network construction, Comput. Biol. Med. 48 (2014) 55–65, https://doi.org/10.1016/j.compbiomed.2014.02.011.
[36] S. Barbosa, B. Niebel, S. Wolf, K. Mauch, R. Takors, A guide to gene regulatory network inference for obtaining predictive solutions: underlying assumptions and fundamental biological and data constraints, Biosystems. 174 (2018) 37–48, https://doi.org/10.1016/j.biosystems.2018.10.008.
[37] S.M. Hill, L.M. Heiser, T. Cokelaer, M. Unger, N.K. Nesser, D.E. Carlin, Y. Zhang, A. Sokolov, E.O. Paull, C.K. Wong, K. Graim, A. Bivol, H. Wang, F. Zhu, B. Afsari, L.V. Danilova, A.V. Favorov, W.S. Lee, D. Taylor, C.W. Hu, B.L. Long, D.P. Noren, A.J. Bisberg, G.B. Mills, J.W. Gray, M. Kellen, T. Norman, S. Friend, A.A. Qutub, E.J. Fertig, Y. Guan, M. Song, J.M. Stuart, P.T. Spellman, H. Koeppl, G. Stolovitzky, J. Saez-Rodriguez, S. Mukherjee, H.-D. Consortium, Inferring causal molecular networks: empirical assessment through a community-based effort, Nat. Methods 13 (2016) 310–8. doi:https://doi.org/10.1038/nmeth.3773.
[38] D. Marbach, J.C. Costello, R. Koffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, M. Kellis, J.J. Collins, G. Stolovitzky, D. Consortium, Wisdom of crowds for robust gene network inference, Nat. Methods 9 (2012) 796–804, https://doi.org/10.1038/nmeth.2016.
[39] D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, Expression profiling using cDNA microarrays, Nat. Genet. 21 (1999) 10–14, https://doi.org/10.1038/4434.
[40] B. Jia, S. Xu, G. Xiao, V. Lamba, F. Liang, Learning gene regulatory networks from next generation sequencing data, Biometrics. 73 (2017) 1221–1230, https://doi.org/10.1111/biom.12682.
[41] E.A. Serin, H. Nijveen, H.W. Hilhorst, W. Ligterink, Learning from co-expression networks: possibilities and challenges, Front. Plant Sci. 7 (2016) 444, https://doi.org/10.3389/fpls.2016.00444.
[42] F.M. Giorgi, C. Del Fabbro, F. Licausi, Comparative study of RNA-seq- and microarray-derived coexpression networks in Arabidopsis thaliana, Bioinformatics. 29 (2013) 717–724, https://doi.org/10.1093/bioinformatics/btt053.
[43] A. Lachmann, F.M. Giorgi, G. Lopez, A. Califano, ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information, Bioinformatics. 32 (2016) 2233–2235, https://doi.org/10.1093/bioinformatics/btw216.
[44] A. Athar, A. Füllgrabe, N. George, H. Iqbal, L. Huerta, A. Ali, C. Snow, N.A. Fonseca, R. Petryszak, I. Papatheodorou, U. Sarkans, A. Brazma, ArrayExpress update – from bulk to single-cell expression data, Nucleic Acids Res. 47 (2019) D711–D715, https://doi.org/10.1093/nar/gky964.
[45] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Res. 41 (2013) D991–D995, https://doi.org/10.1093/nar/gks1193.
[46] J. Mashima, Y. Kodama, T. Fujisawa, T. Katayama, Y. Okuda, E. Kaminuma, O. Ogasawara, K. Okubo, Y. Nakamura, T. Takagi, DNA data Bank of Japan, Nucleic Acids Res. 45 (2017) D25–D31, https://doi.org/10.1093/nar/gkw1001.
[47] J. Rung, A. Brazma, Reuse of public genome-wide gene expression data, Nat Rev Genet. 14 (2013) 89–99, https://doi.org/10.1038/nrg3394.
[48] M. Goldman, B. Craft, T. Swatloski, M. Cline, O. Morozova, M. Diekhans, D. Haussler, J. Zhu, The UCSC cancer genomics browser: update 2015, Nucleic Acids Res. 43 (2015) D812–D817, https://doi.org/10.1093/nar/gku1073.
[49] Y. Aoki, Y. Okamura, H. Ohta, K. Kinoshita, T. Obayashi, ALCOdb: gene coexpression database for microalgae, Plant Cell Physiol. 57 (2016) e3, https://doi.org/10.1093/pcp/pcv190.
[50] T. Obayashi, Y. Aoki, S. Tadaka, Y. Kagaya, K. Kinoshita, ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index, Plant Cell Physiol. 59 (2018) 440, https://doi.org/10.1093/pcp/pcx209.
[51] M. Van Bel, F. Coppens, Exploring plant co-expression and gene-gene interactions with CORNET 3.0, in: A.D.J. van Dijk (Ed.), Plant Genomics Databases, Springer New York, New York, NY, 2017, pp. 201–212, , https://doi.org/10.1007/978-1-4939-6658-5_11.
[52] T. Obayashi, Y. Kagaya, Y. Aoki, S. Tadaka, K. Kinoshita, COXPRESdb v7: a gene coexpression database for 11 animal species supported by 23 coexpression platforms for technical evaluation and evolutionary inference, Nucleic Acids Res. 47 (2019) D55–D62, https://doi.org/10.1093/nar/gky1155.
[53] S. van Dam, T. Craig, J.P. de Magalhães, GeneFriends: a human RNA-seq-based gene and transcript co-expression database, Nucleic Acids Res. 43 (2015) D1124–D1132, https://doi.org/10.1093/nar/gku1042.
[54] S. Yang, C.Y. Kim, S. Hwang, E. Kim, H. Kim, H. Shim, I. Lee, COEXPEDIA: exploring biomedical hypotheses via co-expressions associated with medical subject headings (MeSH), Nucleic Acids Res. 45 (2017) D389–D396, https://doi.org/10.1093/nar/gkw868.
[55] Q. Zhu, A.K. Wong, A. Krishnan, M.R. Aure, A. Tadych, R. Zhang, D.C. Corney, C.S. Greene, L.A. Bongo, V.N. Kristensen, M. Charikar, K. Li, O.G. Troyanskaya, Targeted exploration and analysis of large cross-platform human transcriptomic compendia, Nat. Methods 12 (2015) 211–214, https://doi.org/10.1038/nmeth.3249.
[56] B.C. Ross, Mutual information between discrete and continuous data sets, PLoS One 9 (2014) e87357, , https://doi.org/10.1371/journal.pone.0087357.
[57] K. Basso, A.A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, A. Califano, Reverse engineering of regulatory networks in human B cells, Nat. Genet. 37 (2005) 382–390, https://doi.org/10.1038/ng1532.
[58] R.A. Chavez Montes, G. Coello, K.L. Gonzalez-Aguilera, N. Marsch-Martinez, S. de Folter, E.R. Alvarez-Buylla, ARACNe-based inference, using curated microarray data, of Arabidopsis thaliana root transcriptional regulatory networks, BMC Plant Biol. 14 (2014) 97, https://doi.org/10.1186/1471-2229-14-97.
[59] A. Floratos, K. Smith, Z. Ji, J. Watkinson, A. Califano, geWorkbench: an open source platform for integrative genomics, Bioinformatics. 26 (2010) 1779–1780, https://doi.org/10.1093/bioinformatics/btq282.
[60] P.E. Meyer, F. Lafitte, G. Bontempi, minet: a R/bioconductor package for inferring large transcriptional networks using mutual information, BMC Bioinformatics. 9 (2008) 461, https://doi.org/10.1186/1471-2105-9-461.
[61] M.A. Castro, I. de Santiago, T.M. Campbell, C. Vaughn, T.E. Hickey, E. Ross, W.D. Tilley, F. Markowetz, B.A. Ponder, K.B. Meyer, Regulators of genetic risk of breast cancer identified by integrative network analysis, Nat. Genet. 48 (2016) 12–21, https://doi.org/10.1038/ng.3458.
[62] P. Zoppoli, S. Morganella, M. Ceccarelli, TimeDelay-ARACNE: reverse engineering of gene networks from time-course data by an information theoretic approach, BMC Bioinformatics. 11 (2010) 154, https://doi.org/10.1186/1471-2105-11-154.
[63] I.S. Jang, A. Margolin, A. Califano, hARACNe: improving the accuracy of regulatory model reverse engineering via higher-order data processing inequality tests, Interface Focus. 3 (2013) 20130011, https://doi.org/10.1098/rsfs.2013.0011.
[64] J. He, Z. Zhou, M. Reed, A. Califano, Accelerated parallel algorithm for gene network reverse engineering, BMC Syst. Biol. 11 (2017) 83, https://doi.org/10.1186/s12918-017-0458-5.
[65] A. Khatamian, E.O. Paull, A. Califano, J. Yu, SJARACNe: a scalable software tool for gene network reverse engineering from big data, Bioinformatics. (2018), https://doi.org/10.1093/bioinformatics/bty907.
[66] C.S. Groeneveld, V.S. Chagas, S.J.M. Jones, A.G. Robertson, B.A.J. Ponder, K.B. Meyer, M.A.A. Castro, RTNsurvival: An R/Bioconductor package for regulatory network survival analysis, Bioinformatics. (2019). doi:https://doi.org/10.1093/bioinformatics/btz229.
[67] Y. Shimoni, Association between expression of random gene sets and survival is evident in multiple cancer types and may be explained by sub-classification, PLoS Comput. Biol. 14 (2018) e1006026, , https://doi.org/10.1371/journal.pcbi.1006026.
[68] G. Altay, F. Emmert-Streib, Inferring the conservative causal core of gene regulatory networks, BMC Syst. Biol. 4 (2010) 132, https://doi.org/10.1186/1752-0509-4-132.
[69] B.-E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, F. d'Alche-Buc, Gene networks inference using dynamic Bayesian networks, Bioinformatics. 19 (2003) ii138–ii148. doi:https://doi.org/10.1093/bioinformatics/btg1071.
[70] H. Bae, S. Monti, M. Montano, M.H. Steinberg, T.T. Perls, P. Sebastiani, Learning Bayesian networks from correlated data, Sci. Rep. 6 (2016) 25156, https://doi.org/10.1038/srep25156.
[71] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, R. Guthke, Gene regulatory network inference: data integration in dynamic models—a review, Biosystems. 96 (2009) 86–103, https://doi.org/10.1016/j.biosystems.2008.12.004.

[72] M.A. de Luis Balaguer, R. Sozzani, Inferring gene regulatory networks in the Arabidopsis root using a dynamic Bayesian network approach, Methods Mol. Biol. 1629 (2017) 331–348, https://doi.org/10.1007/978-1-4939-7125-1_21.

[73] V.A. Smith, J. Yu, T.V. Smulders, A.J. Hartemink, E.D. Jarvis, Computational inference of neural information flow networks, PLoS Comput. Biol. 2 (2006) e161, https://doi.org/10.1371/journal.pcbi.0020161.

[74] N.H. Balov, Consistent Model Selection of Discrete Bayesian Networks from Incomplete Data, ArXiv:1105.4507 [Math, Stat], http://arxiv.org/abs/1105.4507, (2011) (accessed July 28, 2019).

[75] S. Lèbre, Inferring Dynamic Genetic Networks With Low Order Independencies, ArXiv:0704.2551 [Math, q-Bio, Stat], http://arxiv.org/abs/0704.2551, (2007) (accessed July 28, 2019).

[76] E.R. Morrissey, M.A. Juárez, K.J. Denby, N.J. Burroughs, On reverse engineering of gene interaction networks using time course data with repeated measurements, Bioinformatics. 26 (2010) 2305–2312, https://doi.org/10.1093/bioinformatics/btq421.

[77] R. Opgen-Rhein, K. Strimmer, From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data, BMC Syst. Biol. 1 (2007) 37, https://doi.org/10.1186/1752-0509-1-37.

[78] J. Cao, X. Qi, H. Zhao, Modeling gene regulation networks using ordinary differential equations, Methods Mol. Biol. 802 (2012) 185–197, https://doi.org/10.1007/978-1-61779-400-1_12.

[79] A. Madar, A. Greenfield, H. Ostrer, E. Vanden-Eijnden, R. Bonneau, The inferelator 2.0: A scalable framework for reconstruction of dynamic regulatory network models, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009: pp. 5448–5451. doi:https://doi.org/10.1109/IEMBS.2009.5334018.

[80] A. Vasilevski, F.M. Giorgi, L. Bertinetti, B. Usadel, LASSO modeling of the Arabidopsis thaliana seed/seedling transcriptome: a model case for detection of novel mucilage and pectin metabolism genes, Mol. BioSyst. 8 (2012) 2566–2574, https://doi.org/10.1039/c2mb25096a.

[81] M. Bansal, G. Della Gatta, D. di Bernardo, Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, Bioinformatics. 22 (2006) 815–822, https://doi.org/10.1093/bioinformatics/btl003.

[82] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics. 9 (2008) 559, https://doi.org/10.1186/1471-2105-9-559.

[83] P. Shannon, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[84] M. Mutwil, B. Usadel, M. Schaette, A. Loraine, O. Ebenhoeh, S. Persson, Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm, Plant Physiol. 152 (2010) 29–43, https://doi.org/10.1104/pp.109.145318.

[85] P.V. Kharchenko, L. Silberstein, D.T. Scadden, Bayesian approach to single-cell differential expression analysis, Nat. Methods 11 (2014) 740–742, https://doi.org/10.1038/nmeth.2967.

[86] M.W.E.J. Fiers, L. Minnoye, S. Aibar, C. Bravo González-Blas, Z. Kalender Atak, S. Aerts, Mapping gene regulatory networks from single-cell omics data, Briefings in Functional Genomics. 17 (2018) 246–254, https://doi.org/10.1093/bfgp/elx046.

[87] S. Mohammadi, V. Ravindra, D.F. Gleich, A. Grama, A geometric approach to characterize the functional identity of single cells, Nat. Commun. 9 (2018) 1516, https://doi.org/10.1038/s41467-018-03933-2.

[88] P. Trairatphisan, A. Mizera, J. Pang, A.A. Tantar, J. Schneider, T. Sauter, Recent development and biomedical applications of probabilistic Boolean networks, Cell Commun Signal. 11 (2013) 46, https://doi.org/10.1186/1478-811X-11-46.

[89] V. Moignard, S. Woodhouse, L. Haghverdi, A.J. Lilly, Y. Tanaka, A.C. Wilkinson, F. Buettner, I.C. Macaulay, W. Jawaid, E. Diamanti, S.-I. Nishikawa, N. Piterman, V. Kouskoff, F.J. Theis, J. Fisher, B. Göttgens, Decoding the regulatory network of early blood development from single-cell gene expression measurements, Nat. Biotechnol. 33 (2015) 269–276, https://doi.org/10.1038/nbt.3154.

[90] C.Y. Lim, H. Wang, S. Woodhouse, N. Piterman, L. Wernisch, J. Fisher, B. Göttgens, BTR: training asynchronous Boolean models using single-cell expression data, BMC Bioinformatics. 17 (2016) 355, https://doi.org/10.1186/s12859-016-1235-y.

[91] C.A. Jackson, D.M. Castro, G.-A. Saldi, R. Bonneau, D. Gresham, Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments, Genomics (2019), https://doi.org/10.1101/581678.

[92] J.M. Cherry, E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K.R. Christie, M.C. Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, K. Karra, C.J. Krieger, S.R. Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simison, S. Weng, E.D. Wong, Saccharomyces Genome Database: the genomics resource of budding yeast, Nucleic Acids Res. 40 (2012) D700–D705, https://doi.org/10.1093/nar/gkr1029.

[93] T. Peng, Q. Zhu, P. Yin, K. Tan, SCRABBLE: single-cell RNA-seq imputation constrained by bulk RNA-seq data, Genome Biol. 20 (2019) 88, https://doi.org/10.1186/s13059-019-1681-8.

[94] H. Liu, P. Li, M. Zhu, X. Wang, J. Lu, T. Yu, Nonlinear network reconstruction from gene expression data using marginal dependencies measured by DCOL, PLoS One 11 (2016) e0158247, , https://doi.org/10.1371/journal.pone.0158247.

[95] H. Matsumoto, H. Kiryu, C. Furusawa, M.S.H. Ko, S.B.H. Ko, N. Gouda, T. Hayashi, I. Nikaido, SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation, Bioinformatics. 33 (2017) 2314–2321, https://doi.org/10.1093/bioinformatics/btx194.

[96] M. Guo, H. Wang, S.S. Potter, J.A. Whitsett, Y. Xu, SINCERA: a pipeline for single-cell RNA-Seq profiling analysis, PLoS Comput. Biol. 11 (2015) e1004575, , https://doi.org/10.1371/journal.pcbi.1004575.

[97] H. Ding, E.F. Douglass, A.M. Sonabend, A. Mela, S. Bose, C. Gonzalez, P.D. Canoll, P.A. Sims, M.J. Alvarez, A. Califano, Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm, Nat. Commun. 9 (2018) 1471, https://doi.org/10.1038/s41467-018-03843-3.

[98] A. Gobbi, F. Iorio, K.J. Dawson, D.C. Wedge, D. Tamborero, L.B. Alexandrov, N. Lopez-Bigas, M.J. Garnett, G. Jurman, J. Saez-Rodriguez, Fast randomization of large genomic datasets while preserving alteration counts, Bioinformatics. 30 (2014) i617–i623, https://doi.org/10.1093/bioinformatics/btu474.

[99] F. Iorio, M. Bernardo-Faura, A. Gobbi, T. Cokelaer, G. Jurman, J. Saez-Rodriguez, Efficient randomization of biological networks while preserving functional characterization of individual nodes, BMC Bioinformatics. 17 (2016) 542, https://doi.org/10.1186/s12859-016-1402-1.

[100] Emma Schwager, The CCREPE (Compositionality Corrected by REnormalizaion and PErmutation) package: detecting statistically significant associations between sparse and high dimensional compositional data, Bioconductor, 2017. doi:10.18129/b9.bioc.ccrepe.

[101] V.A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods, PLoS One 5 (2010), https://doi.org/10.1371/journal.pone.0012776.

[102] V.A. Huynh-Thu, P. Geurts, dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data, Sci. Rep. 8 (2018) 3384, https://doi.org/10.1038/s41598-018-21715-0.

[103] P. Bellot, C. Olsen, P. Salembier, A. Oliveras-Verges, P.E. Meyer, NetBenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference, BMC Bioinformatics. 16 (2015) 312, https://doi.org/10.1186/s12859-015-0728-4.

[104] T. Schaffter, D. Marbach, D. Floreano, GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods, Bioinformatics. 27 (2011) 2263–2270, https://doi.org/10.1093/bioinformatics/btr373.

[105] P. Kheradpour, A. Stark, S. Roy, M. Kellis, Reliable prediction of regulator targets using 12 drosophila genomes, Genome Res. 17 (2007) 1919–1931, https://doi.org/10.1101/gr.7090407.

[106] S. Inukai, K.H. Kock, M.L. Bulyk, Transcription factor–DNA binding: beyond binding site motifs, Curr. Opin. Genet. Dev. 43 (2017) 110–119, https://doi.org/10.1016/j.gde.2017.02.007.

[107] M. Tompa, N. Li, T.L. Bailey, G.M. Church, B. De Moor, E. Eskin, A.V. Favorov, M.C. Frith, Y. Fu, W.J. Kent, V.J. Makeev, A.A. Mironov, W.S. Noble, G. Pavesi, G. Pesole, M. Régnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, Z. Zhu, Assessing computational tools for the discovery of transcription factor binding sites, Nat. Biotechnol. 23 (2005) 137–144, https://doi.org/10.1038/nbt1053.

[108] R. Milo, Network motifs: simple building blocks of complex networks, Science. 298 (2002) 824–827, https://doi.org/10.1126/science.298.5594.824.

[109] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME SUITE: tools for motif discovery and searching, Nucleic Acids Res. 37 (2009) W202–W208, https://doi.org/10.1093/nar/gkp335.

[110] A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function, Nucleic Acids Research. 47 (2019) D155–D162. doi:https://doi.org/10.1093/nar/gky1141.

[111] K. Daily, V.R. Patel, P. Rigor, X. Xie, P. Baldi, MotifMap: integrative genome-wide maps of regulatory motif sites for model species, BMC Bioinformatics. 12 (2011) 495, https://doi.org/10.1186/1471-2105-12-495.

[112] C.-N. Chow, T.-Y. Lee, Y.-C. Hung, G.-Z. Li, K.-C. Tseng, Y.-H. Liu, P.-L. Kuo, H.-Q. Zheng, W.-C. Chang, PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants, Nucleic Acids Research. 47 (2019) D1155–D1163. doi:https://doi.org/10.1093/nar/gky1081.

[113] V. Agarwal, G.W. Bell, J.-W. Nam, D.P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, ELife. 4 (2015) e05005, , https://doi.org/10.7554/eLife.05005.

[114] M.T. Weirauch, A. Yang, M. Albu, A.G. Cote, A. Montenegro-Montero, P. Drewe, H.S. Najafabadi, S.A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M.G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J.S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A.J.M. Walhout, F.-Y. Bouget, G. Ratsch, L.F. Larrondo, J.R. Ecker, T.R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity, Cell. 158 (2014) 1431–1443, https://doi.org/10.1016/j.cell.2014.08.009.

[115] T.-P. Chiu, F. Comoglio, T. Zhou, L. Yang, R. Paro, R. Rohs, DNAshapeR: an R/bioconductor package for DNA shape prediction and feature encoding, Bioinformatics. 32 (2016) 1211–1213, https://doi.org/10.1093/bioinformatics/btv735.

[116] L. Yang, T. Zhou, I. Dror, A. Mathelier, W.W. Wasserman, R. Gordân, R. Rohs, TFBSshape: a motif database for DNA shape features of transcription factor binding sites, Nucl. Acids Res. 42 (2014) D148–D155, https://doi.org/10.1093/nar/gkt1087.

[117] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, C.K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol. Cell 38 (2010) 576–589, https://doi.org/10.1016/j.molcel.2010.05.004.

[118] A. Verfaillie, H. Imrichova, R. Janky, S. Aerts, iRegulon and i-cisTarget: Reconstructing regulatory networks using motif and track enrichment: iRegulon

and i-cisTarget: Reconstructing regulatory networks, in: A. Bateman, W.R. Pearson, L.D. Stein, G.D. Stormo, J.R. Yates (Eds.), Current Protocols in Bioinformatics, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2015, pp. 2.16.1–2.16.39, , https://doi.org/10.1002/0471250953.bi0216s52.
[119] P.J. Balwierz, M. Pachkov, P. Arnold, A.J. Gruber, M. Zavolan, E. van Nimwegen, ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs, Genome Res. 24 (2014) 869–884, https://doi.org/10.1101/gr.169508.113.
[120] A.F. Siahpirani, S. Roy, A prior-based integrative framework for functional transcriptional regulatory network inference, Nucleic Acids Res. (2016) gkw963. doi:https://doi.org/10.1093/nar/gkw963.
[121] K. Glass, C. Huttenhower, J. Quackenbush, G.-C. Yuan, Passing messages between biological networks to refine predicted interactions, PLoS One 8 (2013) e64832, , https://doi.org/10.1371/journal.pone.0064832.
[122] S.R. Kulkarni, D. Vaneechoutte, J. Van de Velde, K. Vandepoele, TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information, Nucleic Acids Res. 46 (2018) e31, https://doi.org/10.1093/nar/gkx1279.
[123] S. Aibar, C.B. González-Blas, T. Moerman, V.A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z.K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering, Nat. Methods 14 (2017) 1083–1086, https://doi.org/10.1038/nmeth.4463.
[124] A. Blais, Constructing transcriptional regulatory networks, Genes Dev. 19 (2005) 1499–1511, https://doi.org/10.1101/gad.1325605.
[125] R. Nakato, K. Shirahige, Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation, Brief Bioinform. (2016) bbw023. doi:https://doi.org/10.1093/bib/bbw023.
[126] G. Pavesi, ChIP-Seq data analysis to define transcriptional regulatory networks, Adv. Biochem. Eng. Biotechnol. 160 (2017) 1–14, https://doi.org/10.1007/10_2016_43.
[127] C. Angelini, V. Costa, Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems, Front, Cell Dev. Biol. 2 (2014), https://doi.org/10.3389/fcell.2014.00051.
[128] R. Thomas, S. Thomas, A.K. Holloway, K.S. Pollard, Features that define the best ChIP-seq peak calling algorithms, Brief Bioinform. (2016) bbw035. doi:https://doi.org/10.1093/bib/bbw035.
[129] C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, C.B. Lowe, A.M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions, Nat. Biotechnol. 28 (2010) 495–501, https://doi.org/10.1038/nbt.1630.
[130] Z. Bar-Joseph, G.K. Gerber, T.I. Lee, N.J. Rinaldi, J.Y. Yoo, F. Robert, D.B. Gordon, E. Fraenkel, T.S. Jaakkola, R.A. Young, D.K. Gifford, Computational discovery of gene modules and regulatory networks, Nat. Biotechnol. 21 (2003) 1337–1342, https://doi.org/10.1038/nbt890.
[131] R. Leinonen, R. Akhtar, E. Birney, L. Bower, A. Cerdeno-Tarraga, Y. Cheng, I. Cleland, N. Faruque, N. Goodgame, R. Gibson, G. Hoad, M. Jang, N. Pakseresht, S. Plaister, R. Radhakrishnan, K. Reddy, S. Sobhany, P. Ten Hoopen, R. Vaughan, V. Zalunin, G. Cochrane, The European nucleotide archive, Nucleic Acids Res. 39 (2011) D28–D31, https://doi.org/10.1093/nar/gkq967.
[132] Y. Kodama, M. Shumway, R. Leinonen, On behalf of the international nucleotide sequence database collaboration, the sequence read archive: explosive growth of sequencing data, Nucleic Acids Res. 40 (2012) D54–D56, https://doi.org/10.1093/nar/gkr854.
[133] C.A. Davis, B.C. Hitz, C.A. Sloan, E.T. Chan, J.M. Davidson, I. Gabdank, J.A. Hilton, K. Jain, U.K. Baymuradov, A.K. Narayanan, K.C. Onate, K. Graham, S.R. Miyasato, T.R. Dreszer, J.S. Strattan, O. Jolanki, F.Y. Tanaka, J.M. Cherry, The encyclopedia of DNA elements (ENCODE): data portal update, Nucleic Acids Res. 46 (2018) D794–D801, https://doi.org/10.1093/nar/gkx1081.
[134] D. Karolchik, The UCSC genome browser database, Nucleic Acids Res. 31 (2003) 51–54, https://doi.org/10.1093/nar/gkg129.
[135] S. Oki, T. Ohta, G. Shioi, H. Hatanaka, O. Ogasawara, Y. Okuda, H. Kawaji, R. Nakaki, J. Sese, C. Meno, ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data, EMBO Rep. 19 (2018) e46255. doi:10.15252/embr.201846255.
[136] K.-R. Zhou, S. Liu, W.-J. Sun, L.-L. Zheng, H. Zhou, J.-H. Yang, L.-H. Qu, ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data, Nucleic Acids Res. 45 (2017) D43–D50. doi:https://doi.org/10.1093/nar/gkw965.
[137] F. Zambelli, G.M. Prazzoli, 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 Res. 40 (2012) W510–W515, https://doi.org/10.1093/nar/gks483.
[138] W. Huang, R. Loganantharaj, B. Schroeder, D. Fargo, L. Li, PAVIS: a tool for peak annotation and visualization, Bioinformatics. 29 (2013) 3097–3099, https://doi.org/10.1093/bioinformatics/btt520.
[139] N.C. Sheffield, C. Bock, LOLA: enrichment analysis for genomic region sets and regulatory elements in R and bioconductor, Bioinformatics. 32 (2016) 587–589, https://doi.org/10.1093/bioinformatics/btv612.
[140] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J.A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S.R. Kulkarni, G. Tan, D. Baranasic, D.J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W.W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework, Nucleic Acids Res. 46 (2018) D260–D266, https://doi.org/10.1093/nar/gkx1126.
[141] M.J. Ziller, R. Edri, Y. Yaffe, J. Donaghey, R. Pop, W. Mallard, R. Issner, C.A. Gifford, A. Goren, J. Xing, H. Gu, D. Cacchiarelli, A.M. Tsankov, C. Epstein, J.L. Rinn, T.S. Mikkelsen, O. Kohlbacher, A. Gnirke, B.E. Bernstein, Y. Elkabetz, A. Meissner, Dissecting neural differentiation regulatory networks through epigenetic footprinting, Nature. 518 (2015) 355–359, https://doi.org/10.1038/nature13990.
[142] P. Wang, J. Qin, Y. Qin, Y. Zhu, L.Y. Wang, M.J. Li, M.Q. Zhang, J. Wang, ChIP-Array 2: integrating multiple omics data to construct gene regulatory networks, Nucleic Acids Res. 43 (2015) W264–W269, https://doi.org/10.1093/nar/gkv398.
[143] V.G. Levitsky, I.V. Kulakovskiy, N.I. Ershov, D. Oshchepkov, V.J. Makeev, T.C. Hodgman, T.I. Merkulova, Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data, BMC Genomics 15 (2014) 80, https://doi.org/10.1186/1471-2164-15-80.
[144] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Research. 46 (2018) D1074–D1082. doi:https://doi.org/10.1093/nar/gkx1037.
[145] L.J. Zhu, C. Gazin, N.D. Lawson, H. Pagès, S.M. Lin, D.S. Lapointe, M.R. Green, ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data, BMC Bioinformatics. 11 (2010) 237, https://doi.org/10.1186/1471-2105-11-237.
[146] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M.T. Morgan, V.J. Carey, Software for computing and annotating genomic ranges, PLoS Comput. Biol. 9 (2013) e1003118, , https://doi.org/10.1371/journal.pcbi.1003118.
[147] M. Russo, B. De Lucca, T. Flati, S. Gioiosa, G. Chillemi, G. Capranico, DROPA: DRIP-seq optimized peak annotator, BMC Bioinformatics. 20 (2019) 414. doi:https://doi.org/10.1186/s12859-019-3009-9.
[148] M. Kondili, A. Fust, J. Preussner, C. Kuenne, T. Braun, M. Looso, UROPA: a tool for universal RObust peak annotation, Sci. Rep. 7 (2017) 2593, https://doi.org/10.1038/s41598-017-02464-y.
[149] J.M. Bhasin, A.H. Ting, Goldmine integrates information placing genomic ranges into meaningful biological contexts, Nucleic Acids Res. 44 (2016) 5550–5556, https://doi.org/10.1093/nar/gkw477.
[150] M. Modrák, J. Vohradský, Genexpi: a toolset for identifying regulons and validating gene regulatory networks using time-course expression data, BMC Bioinformatics. 19 (2018) 137, https://doi.org/10.1186/s12859-018-2138-x.
[151] A.N. Holding, F.M. Giorgi, A. Donnelly, A.E. Cullen, S. Nagarajan, L.A. Selth, F. Markowetz, VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERa at enhancers in breast cancer, Genome Biol. 20 (2019) 91, https://doi.org/10.1186/s13059-019-1698-z.
[152] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, S.A. Teichmann, Structure and evolution of transcriptional regulatory networks, Curr. Opin. Struct. Biol. 14 (2004) 283–291, https://doi.org/10.1016/j.sbi.2004.05.004.
[153] T. Gabaldón, E.V. Koonin, Functional and evolutionary implications of gene orthology, Nat Rev Genet. 14 (2013) 360–366, https://doi.org/10.1038/nrg3456.
[154] A.M. Altenhoff, C. Dessimoz, Phylogenetic and functional assessment of Orthologs inference projects and methods, PLoS Comput. Biol. 5 (2009) e1000262, , https://doi.org/10.1371/journal.pcbi.1000262.
[155] K. Tan, T. Shlomi, H. Feizi, T. Ideker, R. Sharan, Transcriptional regulation of protein complexes within and across species, Proc. Natl. Acad. Sci. 104 (2007) 1283–1288, https://doi.org/10.1073/pnas.0606914104.
[156] B.T.L. Nichio, J.N. Marchaukoski, R.T. Raittz, New tools in orthology analysis: a brief review of promising perspectives, Front. Genet. 8 (2017) 165, https://doi.org/10.3389/fgene.2017.00165.
[157] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Research. 47 (2019) D309–D314. doi:https://doi.org/10.1093/nar/gky1085.
[158] K.P. O'Brien, Inparanoid: a comprehensive database of eukaryotic orthologs, Nucleic Acids Res. 33 (2004) D476–D480, https://doi.org/10.1093/nar/gki107.
[159] E.V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F.A. Simão, E.M. Zdobnov, OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs, Nucleic Acids Res. 47 (2019) D807–D811, https://doi.org/10.1093/nar/gky1053.
[160] C. Koch, J. Konieczka, T. Delorey, A. Lyons, A. Socha, K. Davis, S.A. Knaack, D. Thompson, E.K. O'Shea, A. Regev, S. Roy, Inference and Evolutionary Analysis of Genome-Scale Regulatory Networks in Large Phylogenies, Cell Systems. 4 (2017) 543–558.e8. doi:https://doi.org/10.1016/j.cels.2017.04.010.
[161] L. Glenwinkel, D. Wu, G. Minevich, O. Hobert, TargetOrtho: a phylogenetic footprinting tool to identify transcription factor targets, Genetics. 197 (2014) 61–76, https://doi.org/10.1534/genetics.113.160721.
[162] I.R. Sadreyev, F. Ji, E. Cohen, G. Ruvkun, Y. Tabach, PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles, Nucleic Acids Res. 43 (2015) W154–W159, https://doi.org/10.1093/nar/gkv452.
[163] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, M. Gerstein, OrthoClust: an orthology-based network framework for clustering data across multiple species, Genome Biol. 15 (2014) R100, https://doi.org/10.1186/gb-2014-15-8-r100.
[164] E. Wingender, TRANSFAC: an integrated system for gene expression regulation, Nucleic Acids Res. 28 (2000) 316–319, https://doi.org/10.1093/nar/28.1.316.
[165] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (2016) D457–D462, https://doi.org/10.1093/nar/gkv1070.
[166] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdottir, P. Tamayo, J.P. Mesirov, Molecular signatures database (MSigDB) 3, 0, Bioinformatics. 27

(2011) 1739–1740, https://doi.org/10.1093/bioinformatics/btr260.
[167] A. Lachmann, F.M. Giorgi, M.J. Alvarez, A. Califano, Detection and removal of spatial bias in multiwell assays, Bioinformatics. 32 (2016) 1959–1965, https://doi.org/10.1093/bioinformatics/btw092.
[168] A.D. Rouillard, G.W. Gundersen, N.F. Fernandez, Z. Wang, C.D. Monteiro, M.G. McDermott, A. Ma'ayan, The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins, Database. 2016 (2016) baw100. doi:https://doi.org/10.1093/database/baw100.
[169] D. Nishimura, BioCarta, Biotech Software & Internet Report: The Computer Software Journal for Scient. 2 (2001) 117–120.
[170] A. Yilmaz, M.K. Mejia-Guerra, K. Kurz, X. Liang, L. Welch, E. Grotewold, AGRIS: the Arabidopsis gene regulatory information server, an update, Nucleic Acids Res. 39 (2011) D1118–D1122, https://doi.org/10.1093/nar/gkq1120.
[171] J. Thurmond, J.L. Goodman, V.B. Strelets, H. Attrill, L.S. Gramates, S.J. Marygold, B.B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T.C. Kaufman, B.R. Calvi, the FlyBase Consortium, N. Perrimon, S.R. Gelbart, J. Agapite, K. Broll, L. Crosby, G. dos Santos, D. Emmert, L.S. Gramates, K. Falls, V. Jenkins, B. Matthews, C. Sutherland, C. Tabone, P. Zhou, M. Zytkovicz, N. Brown, G. Antonazzo, H. Attrill, P. Garapati, A. Holmes, A. Larkin, S. Marygold, G. Millburn, C. Pilgrim, V. Trovisco, P. Urbano, T. Kaufman, B. Calvi, B. Czoch, J. Goodman, V. Strelets, J. Thurmond, R. Cripps, P. Baker, FlyBase 2.0: the next generation, Nucleic Acids Research. 47 (2019) D759–D765. doi:https://doi.org/10.1093/nar/gky1003.
[172] T.W. Harris, I. Antoshechkin, T. Bieri, D. Blasiar, J. Chan, W.J. Chen, N. De La Cruz, P. Davis, M. Duesbury, R. Fang, J. Fernandes, M. Han, R. Kishore, R. Lee, H.-M. Müller, C. Nakamura, P. Ozersky, A. Petcherski, A. Rangarajan, A. Rogers, G. Schindelman, E.M. Schwarz, M.A. Tuli, K. Van Auken, D. Wang, X. Wang, G. Williams, K. Yook, R. Durbin, L.D. Stein, J. Spieth, P.W. Sternberg, WormBase: a comprehensive resource for nematode research, Nucleic Acids Res. 38 (2010) D463–D467, https://doi.org/10.1093/nar/gkp952.
[173] S. Gama-Castro, H. Salgado, A. Santos-Zavaleta, D. Ledezma-Tejeida, L. Muñiz-Rascado, J.S. García-Sotelo, K. Alquicira-Hernández, I. Martínez-Flores, L. Pannier, J.A. Castro-Mondragón, A. Medina-Rivera, H. Solano-Lira, C. Bonavides-Martínez, E. Pérez-Rueda, S. Alquicira-Hernández, L. Porrón-Sotelo, A. López-Fuentes, A. Hernández-Koutoucheva, V.D. Moral-Chávez, F. Rinaldi, J. Collado-Vides, RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond, Nucleic Acids Res. 44 (2016) D133–D143, https://doi.org/10.1093/nar/gkv1156.
[174] D. Guan, J. Shao, 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 Res. 42 (2014) W130–W136, https://doi.org/10.1093/nar/gku471.
[175] S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F.S.L. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R.E.W. Hancock, L.I. Hannick, I. Jurisica, J. Khadake, D.J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B. Roechert, L. Salwinski, V. Stümpflen, M. Tyers, P. Uetz, I. Xenarios, H. Hermjakob, Protein interaction data curation: the international molecular exchange (IMEx) consortium, Nat. Methods 9 (2012) 345–350, https://doi.org/10.1038/nmeth.1931.
[176] T.S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D.S. Somanathan, A. Sebastian, S. Rani, S. Ray, C.J. Harrys Kishore, S. Kanth, M. Ahmed, M.K. Kashyap, R. Mohmood, Y.L. Ramachandra, V. Krishna, B.A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey, Human Protein Reference Database–2009 update, Nucleic Acids Research. 37 (2009) D767–D772. doi:https://doi.org/10.1093/nar/gkn892.
[177] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, Nucleic Acids Res. 45 (2017) D362–D368, https://doi.org/10.1093/nar/gkw937.
[178] R. Oughtred, C. Stark, B.-J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam, F. Zhang, S. Dolma, A. Willems, J. Coulombe-Huntington, A. Chatr-aryamontri, K. Dolinski, M. Tyers, The BioGRID interaction database: 2019 update, Nucleic Acids Res. 47 (2019) D529–D541, https://doi.org/10.1093/nar/gky1079.
[179] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N.H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, H. Hermjakob, The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases, Nucl. Acids Res. 42 (2014) D358–D363. doi:https://doi.org/10.1093/nar/gkt1115.
[180] G. Sales, E. Calura, D. Cavalieri, C. Romualdi, Graphite - a bioconductor package to convert pathway topology to gene network, BMC Bioinformatics. 13 (2012) 20, https://doi.org/10.1186/1471-2105-13-20.
[181] H. Han, H. Shim, D. Shin, J.E. Shim, Y. Ko, J. Shin, H. Kim, A. Cho, E. Kim, T. Lee, H. Kim, K. Kim, S. Yang, D. Bae, A. Yun, S. Kim, C.Y. Kim, H.J. Cho, B. Kang, S. Shin, I. Lee, TRRUST: a reference database of human transcriptional regulatory interactions, Sci. Rep. 5 (2015) 11432, https://doi.org/10.1038/srep11432.
[182] T. Gao, B. He, S. Liu, H. Zhu, K. Tan, J. Qian, EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types, Bioinformatics. (2016) btw495. doi:https://doi.org/10.1093/bioinformatics/btw495.
[183] M. Pujato, F. Kieken, A.A. Skiles, N. Tapinos, A. Fiser, Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes, Nucleic Acids Res. 42 (2014) 13500–13512, https://doi.org/10.1093/nar/gku1228.
[184] C. Markosian, L. Di Costanzo, M. Sekharan, C. Shao, S.K. Burley, C. Zardecki, Analysis of impact metrics for the Protein Data Bank, Sci Data. 5 (2018) 180212. doi:https://doi.org/10.1038/sdata.2018.212.
[185] T.C. Silva, S.G. Coetzee, N. Gull, L. Yao, D.J. Hazelett, H. Noushmehr, D.-C. Lin, B.P. Berman, ELMER v.2: an R/bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles, Bioinformatics. (2018), https://doi.org/10.1093/bioinformatics/bty902.
[186] Z. Wang, M.A. Jensen, J.C. Zenklusen, A practical guide to the cancer genome atlas (TCGA), in: E. Mathé, S. Davis (Eds.), Statistical Genomics, Springer New York, New York, NY, 2016, pp. 111–141, , https://doi.org/10.1007/978-1-4939-3578-9_6.
[187] Y. Wang, F. Song, B. Zhang, L. Zhang, J. Xu, D. Kuang, D. Li, M.N. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, F. Yue, The 3D Genome Browser: A Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions, (2018), https://doi.org/10.5281/zenodo.1402785.
[188] Nagpal, Sunil, Das Baksi, Krishanu, Kuntal, Bhusan, Mande, Sharmila, CrossNet: A Web Application for Comparative Analysis of Multiple Biological Networks, (n.d.). https://web.rniapps.net/crossnet (accessed September 4, 2019).
[189] M.G. Cardenas, E. Oswald, W. Yu, F. Xue, A.D. MacKerell, A.M. Melnick, The expanding role of the BCL6 Oncoprotein as a cancer therapeutic target, Clin. Cancer Res. 23 (2017) 885–893, https://doi.org/10.1158/1078-0432.CCR-16-2071.
[190] J.M. Granadino-Roldán, C. Obiol-Pardo, M. Pinto, A. Garzón, J. Rubio-Martínez, Molecular dynamics analysis of the interaction between the human BCL6 BTB domain and its SMRT, NcoR and BCOR corepressors: the quest for a consensus dynamic pharmacophore, J. Mol. Graph. Model. 50 (2014) 142–151, https://doi.org/10.1016/j.jmgm.2014.04.003.
[191] F.H. Vasanwala, S. Kusam, L.M. Toney, A.L. Dent, Repression of AP-1 function: a mechanism for the regulation of blimp-1 expression and B lymphocyte differentiation by the B cell lymphoma-6 protooncogene, J. Immunol. 169 (2002) 1922–1929, https://doi.org/10.4049/jimmunol.169.4.1922.
[192] N. Chevallier, ETO protein of t(8;21) AML is a corepressor for Bcl-6 B-cell lymphoma oncoprotein, Blood. 103 (2003) 1454–1463, https://doi.org/10.1182/blood-2003-06-2081.
[193] T. Gte, Consortium, the genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans, Science. 348 (2015) 648–660, https://doi.org/10.1126/science.1262110.
[194] H. Mi, A. Muruganujan, D. Ebert, X. Huang, P.D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, Nucleic Acids Res. 47 (2019) D419–D426, https://doi.org/10.1093/nar/gky1038.
[195] C. Bertolo, R. Malumbres, A. Sagardoy, E.F. Robles, J.I. Martinez-Ferrandis, S. Roa, I.S. Lossos, X. Sagaert, A. Melnick, S. Amar, J.A. Martinez-Climent, LITAF, a BCL6 target gene, regulates Autophagia in B cells and is essential for T-cell dependent humoral responses, Blood. 118 (2011) 1391.
[196] F. Licausi, M. Kosmacz, D.A. Weits, B. Giuntoli, F.M. Giorgi, L.A.C.J. Voesenek, P. Perata, J.T. van Dongen, Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization, Nature. 479 (2011) 419–422, https://doi.org/10.1038/nature10536.
[197] X. Liu, B. Wu, J. Szary, E.M. Kofoed, F. Schaufele, Functional sequestration of transcription factor activity by repetitive DNA, J. Biol. Chem. 282 (2007) 20868–20876, https://doi.org/10.1074/jbc.M702547200.
[198] S. Li, J. Weidenfeld, E.E. Morrisey, Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and Homotypic protein interactions, Mol. Cell. Biol. 24 (2004) 809–822, https://doi.org/10.1128/MCB.24.2.809-822.2004.
[199] E. Spiteri, G. Konopka, G. Coppola, J. Bomar, M. Oldham, J. Ou, S.C. Vernes, S.E. Fisher, B. Ren, D.H. Geschwind, Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain, Am. J. Hum. Genet. 81 (2007) 1144–1157, https://doi.org/10.1086/522237.
[200] S.C. Vernes, P.L. Oliver, E. Spiteri, H.E. Lockstone, R. Puliyadi, J.M. Taylor, J. Ho, C. Mombereau, A. Brewer, E. Lowy, J. Nicod, M. Groszer, D. Baban, N. Sahgal, J.-B. Cazier, J. Ragoussis, K.E. Davies, D.H. Geschwind, S.E. Fisher, Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain, PLoS Genet. 7 (2011) e1002145, , https://doi.org/10.1371/journal.pgen.1002145.