Biochemical Society Transactions (2023) **51** 1–12
https://doi.org/10.1042/BST20210145

Review Article

Computational approaches to understand transcription regulation in development

Maarten van der Sande*, Siebren Frölich* and Simon J. van Heeringen

Radboud University, Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, 6525GA Nijmegen, The Netherlands

Correspondence: Simon J. van Heeringen (s.vanheeringen@science.ru.nl)

Gene regulatory networks (GRNs) serve as useful abstractions to understand transcriptional dynamics in developmental systems. Computational prediction of GRNs has been successfully applied to genome-wide gene expression measurements with the advent of microarrays and RNA-sequencing. However, these inferred networks are inaccurate and mostly based on correlative rather than causative interactions. In this review, we highlight three approaches that significantly impact GRN inference: (1) moving from one genome-wide functional modality, gene expression, to multi-omics, (2) single cell sequencing, to measure cell type-specific signals and predict context-specific GRNs, and (3) neural networks as flexible models. Together, these experimental and computational developments have the potential to significantly impact the quality of inferred GRNs. Ultimately, accurately modeling the regulatory interactions between transcription factors and their target genes will be essential to understand the role of transcription factors in driving developmental gene expression programs and to derive testable hypotheses for validation.

Introduction

Multicellular organisms develop from a single fertilized egg, guided by the genetic information encoded in the genome. Cell lineages diverge and form tissues and organs, based on the interplay between signaling pathways, biomechanical forces [1] and the regulation of gene expression programs [2]. While development is controlled on many levels, transcription regulation is crucial [3]. To better understand these regulatory principles in development and evolution, it is essential to construct informative models of gene regulation.

Transcription is regulated by transcription factors (TFs) within the chromatin context [4]. TFs bind the DNA either directly, mostly in a sequence-specific manner [5], or indirectly via other TFs [6]. They can recruit various other proteins, such as co-activators, RNA polymerase, chromatin remodelers and histone modifying enzymes, to remodel or stabilize the chromatin or to activate or repress transcription [7,8]. In metazoans, TFs form up to 8% of the known proteome [9,10], with DNA binding domains and affinities being highly conserved between metazoans [11–13]. They bind specific DNA motifs that are clustered in relatively short *cis*-regulatory elements (CREs) that can be categorized as promoters, enhancers and insulators [14]. The exact function of an element depends on the combination of bound transcription factors, which is influenced by motif specificity, distance between motifs and motif directionality [15–19]. Core regulatory modules and pathways involved in germ layer and axis formation are deeply conserved in metazoans [20].

A useful abstraction to study transcription regulation is a network of transcription factors and their target genes. This concept of a gene regulatory network (GRN) was introduced in 1969 by Roy Britten and Eric Davidson and later experimentally demonstrated in sea urchin embryos [21,22]. GRNs serve to predict the effect of transcription factor expression on gene transcription and to derive testable hypotheses for validation. More generally, they function to model cell type specification and differentiation in development as well as regulatory perturbations in disease. GRNs have been constructed,

*These authors contributed equally to this work.

Received: 28 October 2022
Revised: 7 January 2023
Accepted: 13 January 2023

Version of Record published: 25 January 2023

Biochemical Society Transactions (2023) **51** 1–12
https://doi.org/10.1042/BST20210145

mostly based on experimental loss-of-function and gain-of-function studies, for a variety of developmental models. Examples include germ layer formation in echinoderms [23–25] and frogs [26–29], neural crest formation [30,31], the *Drosophila* gap gene network [32] and hematopoietic development [33–35]. However, experimental elucidation of a limited number of interactions is hard to scale. Regulatory interactions are highly context-specific [17,36] and most remain unknown [8,37].

Computational inference of genome-wide GRNs was made possible with the advent of expression microarrays. Expression levels between transcription factors and their target genes tend to correlate [38] and genes with similar mRNA expression patterns are more likely to be regulated by a common transcription factor [39,40]. This led to the conception of gene co-expression networks, where functional connections between genes are inferred by expression pattern similarity. WGCNA [41] and ARACNe [42] were among the first gene co-expression-based tools and remain popular. Presently, a multitude of GRN inference methods exists. Reviews on the technical details can be found here [14,43–45]. Recent advances in experimental and computational techniques means that GRN inference has progressed beyond simple co-expression. In this review, we will highlight three approaches that have the potential to significantly impact GRN modeling: (1) moving from one modality, gene expression, to multi-omics, (2) single cell sequencing for cell type-specific signal and (3) neural networks as flexible gene regulatory models (Figure 1).

Multi-omics to capture gene regulation

Gene regulation by TFs is mediated through CREs including promoters and enhancers. By incorporating TF binding at enhancers, regulatory networks can be constrained by direct, causal relationships. Ideally, binding of TFs would be determined experimentally with chromatin immunoprecipitation followed by sequencing (ChIP-seq) [46] or related techniques [47–49]. While large compendia of TF binding profiles in different cell types have been collected for humans [37], this effort remains unfeasible for less well-studied organisms, including most developmental model systems. With sufficient training data, TF binding can be computationally imputed [50–65], however, this does not necessarily generalize across species [66]. As a result, most current approaches use relatively simple models that combine experimentally measured CRE activity with TF binding motifs to computationally predict TF binding.

Putative CREs and their activity can be mapped genome-wide using chromatin accessibility assays, such as DNase I hypersensitive sites sequencing (DNAse-seq) [67] and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) [68]. The number of reads in an element can then be used as a measure for CRE activity in the experimental system [69]. ATAC-seq especially has been widely applied in developmental model systems, as it is experimentally relatively straightforward [31,70–78]. The chromatin environment can supply additional information on CRE location, function and activity. For instance, the transcriptional co-activator p300 is a histone acetyltransferase and can acytelate lysine 27 of histone H3 (H3K27ac). ChIP-seq using antibodies specific to p300 or H3K27ac can therefore identify active enhancers and promoters [79,80]. Other histone modifications that can be linked to CRE activity include H3K4me1 (enhancers) and H3K4me3 (promoters) [81].

CRE activity is determined by (in)direct binding of several TFs [82,83]. Therefore, characterizing TF binding at enhancers can identify their relative importance to the function of an enhancer. One approach to infer TF binding from genome-wide DNA accessibility is digital genomic footprinting [84], which has been used to directly infer GRNs [85,86]. However, sequence bias of the enzymes needs to be taken into account and TFs with more dynamic binding kinetics, such as some nuclear receptors, are not detected by footprint analysis [87–89]. Regardless, footprint analysis using cleavage bias correction can still be informative, especially in differential conditions [90,91]. A more routinely applied approach is to combine TF binding probabilities derived from TF motif scores with DNA accessibility. In some approaches, these are used as priors or constraints on network topology, where the network is inferred from gene expression measurements [92–94]. In alternative approaches, TF motif scores and accessibility are combined with RNA expression using regression models or co-variation of accessibility and expression [95–100].

Enhancers regulate transcription via context-dependent enhancer-promoter interactions [101], usually within a transcriptionally active domain [102]. Combined with TF binding data, these interactions allow for the inference of directed GRNs. Enhancer-promoter interactions can be identified experimentally with Chromatin Conformation Capture techniques [103–105], although this is still uncommon in non-model systems. Inferring interaction between enhancers to target genes is an active field of research. The most commonly used heuristic is to link enhancers to the nearest gene. However, this heuristic is often still incorrect [106,107]. Accuracy can

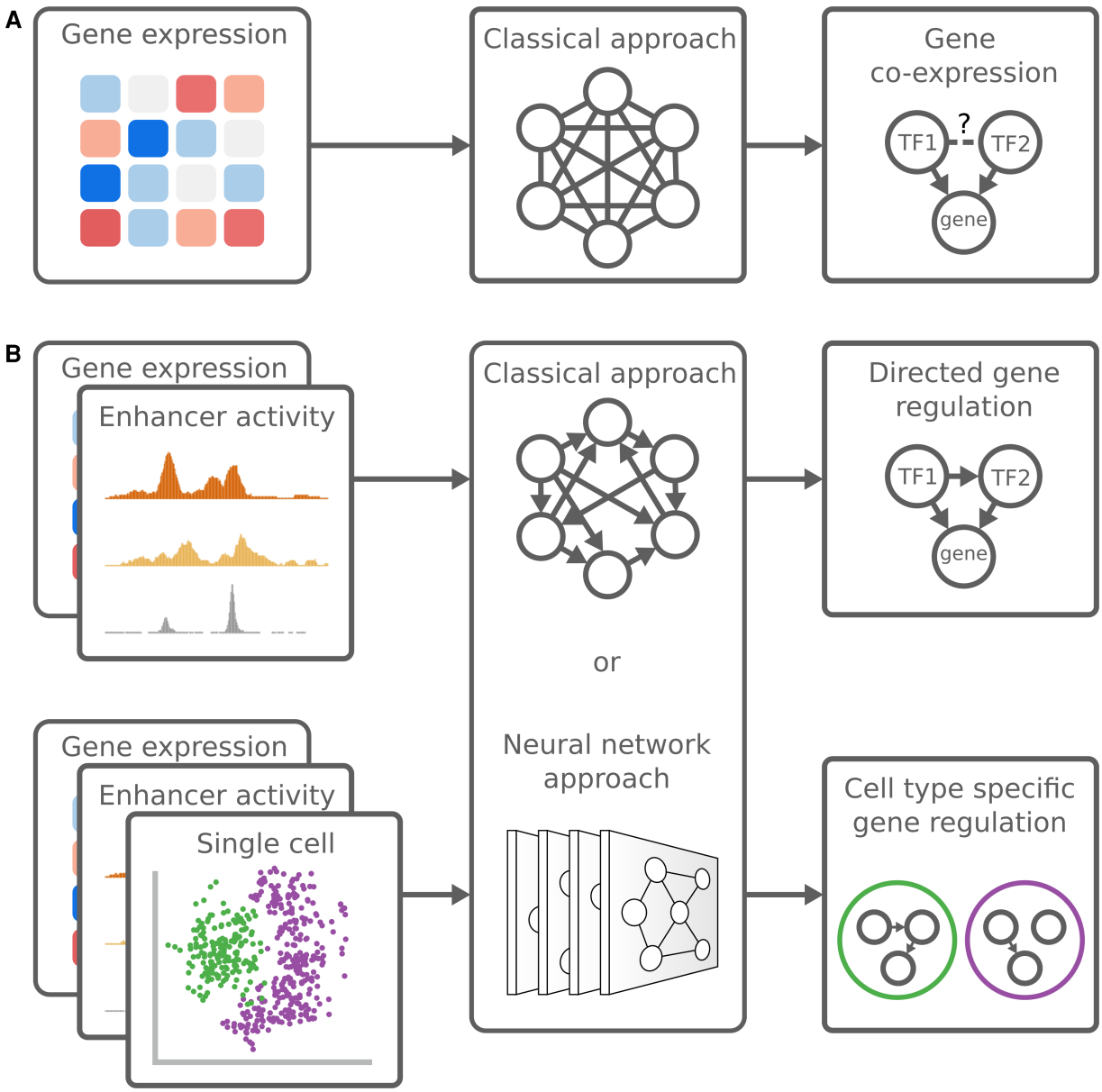

Figure 1. Schematic overview of different gene regulatory network inference approaches.

(**A**) Classical approaches, e.g. correlation, regression or mutual information, can be applied on gene expression data to generate undirected co-expression networks. With prior knowledge about TFs the directionality between TF and target gene can be inferred, however, the directionality between two TFs cannot be established. (**B**) More recent approaches combine multiple types of genome-wide functional data (multi-omics), with either a classical approach or neural networks to identify directed gene regulatory networks. Single cell sequencing allows for the identification of cell type specific regulatory networks.

be improved by combining enhancer to gene distance with TF-target gene co-expression [108]. Finally, Activity-by-Contact based models significantly outperform the nearest gene heuristic by using enhancer to gene distance and enhancer activity [109].

By combining gene expression data with at least one source of enhancer data (e.g. accessibility or interaction data), directed regulatory networks may be inferred with significantly higher accuracy compared with traditional co-expression approaches [98,110,111]. Not only does the combined approach filter out spurious interactions and add causality, but it also reduces the biases introduced by singular approaches. Therefore, we believe that the use of multiple omics will become dominant in all modalities of GRN inference approaches.

Biochemical Society Transactions (2023) **51** 1–12
https://doi.org/10.1042/BST20210145

Single cell sequencing for cell type specific regulation

Developmental transcription regulation has mainly been studied by either *in situ* hybridization [112], which maps the spatial distribution of gene expression of a small set of genes, or bulk gene expression studies [113]. The latter measures the whole transcriptome as a compound signal of all the different cells present in the sample. Single cell sequencing is a fast developing technique to measure the gene expression of individual cells separately, with newer techniques even capable of tagging cells to their spatial coordinates [114,115]. These techniques increase the number of measurements from a handful to several (tens to hundreds of) thousands. This substantial increase in data allows for interesting new ways of GRN inference, but poses new challenges as well.

The output of a single cell experiment generally consists of count tables containing several thousands of cells with low coverage, e.g. only a few thousand of measured transcripts per cell. The low coverage makes the detection of relations between lowly expressed genes difficult. Although it is possible to artificially increase the sequencing depth by simulation (imputation), this does not seem to improve GRN inference [116,117]. Furthermore, it is important to note that cells are repeated measures [118], meaning that the cells come from the same environmental and genetic background, which breaks most statistical assumptions. Computationally clustering related cells, called pseudobulk or meta-cells [119], and using their combined signal solves the issues of low coverage and repeated measures, and still yields cell-type specific signals.

Since fundamentally there are small differences between bulk and pseudobulk data, it is not uncommon to apply bulk GRN inference approaches, such as gene co-expression, ARACNE [42] and GENIE3 [120], to pseudobulk data without much adjustment.

The large number of cells, however, allows for specialized single cell GRN approaches. These include mutual information in combination with partial information decomposition [121], gene coexpression [122], self organizing maps [123], or a combination of single cell RNA-seq and single cell ATAC-seq coexpression and/or bayesian ridge regression [124–126]. Other approaches first order cells by their inferred temporal ordering and then infer the gene-gene relations on this pseudotime, with the assumption that these orderings, also called trajectories, represent cell lineages [127]. Pseudotime can be estimated by simply following the first principal component, or finding the minimal spanning tree between clusters [128], where more advanced methods smoothen the tree [129,130]. A downside of these techniques is that they can not infer the directionality of the relationships. To computationally obtain this directionality, the ratio between spliced and unspliced transcripts per gene can be used as a proxy for whether or not a gene is actively transcribed. By applying this logic across all genes and all cells, one can infer a vector field of velocities of cells which then can be used to get a temporal cell ordering with a start and end [131,132]. These orderings then allow for inferring ordinary differential equations [133,134], Granger causality [135–137], boolean networks [138] or autoregressive models [139]. Most of these methods assume Gaussian noise for gene expression, even though transcription occurs in bursts [140,141], a phenomenon that can only be captured on a single cell level. These dynamics can be modeled as a Markov process including transcriptional bursting and degradation [142]. Theoretically these mechanistic models could be great tools for hypothesis generation, but more work is needed to prove their practical usefulness. Even though the aforementioned GRN inference methods were developed for single cell data specifically, many fail to show consistent improvement over methods that were developed for bulk data, and are seemingly barely any better than purely random models [117,143,144]. Moreover, the added complexity and number of cells leads to computational scaling issues, with some methods taking several days to weeks to finish [117].

Single cell sequencing has the advantage that it disentangles the composite signal present in all biological tissues. The increased number of measurements allows for more complex GRN definitions and inference. Finally, it allows for the inference of fine grained temporal orderings necessary for GRN inference. Even though single cell GRN inference methods have not yet brought the improvements over bulk methods we hoped for, we still expect single cell GRN inference to become the new standard of the field.

Neural networks as flexible gene regulatory models

Computational inference of a GRN depends on a lot of implicit assumptions. For example, a common assumption is that the relationship between genes is additive, which means that the effect on a gene equals the sum of the effects of two regulators separately, but in reality, gene-gene relationships are more complex and for example can include multiplicative effects [145]. A type of model that requires little explicit specification about the possible relationships in the data, but automatically learns these relationships, is an Artificial Neural Network (ANN). ANNs have been successfully applied in a variety of settings, with famously complex

problems such as protein folding [146], image recognition [147], and the board game Go [148]. The successes of ANNs in these unrelated fields shows great promise for application in the field of gene regulatory inference.

Just like GRNs, ANNs consist of nodes and edges. Each edge multiplies the signal from the previous node to the next, and by applying a function to the sum of all the incoming edges the value in the next node is calculated. By adding multiple layers of nodes in between the in- and output nodes (this is where the term *deep* neural network comes from), a network is formed that is capable of learning more and more complex interactions. Learning happens by giving the model examples of input data and expected output, and based on this information the model iteratively updates (learns) its edge weights. After training, hypotheses can easily be tested by systematically querying the model for the predicted effect of certain changes. See [149] for an excellent review on the topic applied to genomics.

ANNs in genomics were first applied to predict the output of a genomic assay, for instance histone modifications in a certain cell type, by using only the DNA sequence as input. Early models showed that convolutional neural networks are capable of predicting functional effects of noncoding variants from short (10–1000 bp) genomic sequences alone [150,151]. These types of models can be used to discover composite motifs and periodic binding [15]. Additionally, these models are capable of learning complex and distal biological relations, as increasing the input sequence to 131 kb still improves accuracy [152].

Whereas ANNs in genomics have mainly been popularized on sequence data, adoption for GRN inference has been relatively slow. Different approaches consist of self-organizing maps [123], variational autoencoders [153], extreme learning machines [154], or graph convolutional neural networks [155,156]. Even though these networks differ in architectural designs, they all report higher levels of accuracy over non-ANN approaches. However, without independent benchmark studies it is hard to verify these results.

The main strength of ANNs is that they can approximate any continuous relationship in the data [157,158], with the downside that large amounts of training data are required. This makes the combination of single cell sequencing and ANNs promising, as current single cell GRN inference approaches have scaling issues [117] and ANNs train relatively fast with the use of GPUs (specialized graphics cards). Fundamentally, understanding how ANNs work is, however, much harder than understanding the classical models typically used for GRN inference. This causes ANNs to be met with skepticism and the persistent misconception that ANNs only function as a black box for predictions and its logic can not be interpreted [159]. We expect ANNs to become commonplace in the field of GRN inference due to their successes in other fields, ease of implementation with high-level programming libraries [160,161], and availability of sufficient training data due to single cell sequencing.

Discussion

Traditional GRNs, mostly based on gene co-expression, have so far served as a useful abstraction to understand regulatory dynamics in developmental systems. However, the way GRNs are currently derived suffers from two fundamental problems. First, the classic GRN that describes TF to target gene relations remains a simplified model and, by design, cannot properly reflect the full complexity of gene regulation. In addition, they are mostly based on mRNA expression as a measure of protein expression, even though this relation is not always linear [162]. In addition, any other types of regulation between transcript and protein product, such as mRNA degradation and post-translational modification, are usually ignored. Second, experiments generally have more features (i.e. genes measured) than samples which is also known as '*the curse of dimensionality*'. In this underdetermined system, many different models can potentially fit to the data, and it is both practically and theoretically impossible to identify the correct model with certainty [163]. It then should not come as a surprise that benchmarks consistently demonstrate that the quality of the inferred GRNs is low [143,144,164–168]. Based on these observations it is clear that our current approach to infer GRN is not sustainable and design changes are needed. Ultimately, we expect the field to move towards GRNs inferred from neural networks trained on single cell multi-omics data.

Having said that, it is not enough to just naively apply single cell multi-omics ANNs. By adding more modalities, and making GRNs more complex, networks become even more underdetermined. This is why most multi-omics approaches use the new modalities to prune the possible TF-target gene relations, which actually reduces the degrees of freedom [98,122,125,126]. Moreover, one can use time-series data to further prune TF-target gene interactions [169], although time-series multi-omics GRN inference tools are still relatively uncommon [170–173]. In addition, computational methods such as regularization [174] and dropout [175] constrain the problem in such a way that you end up with the *simplest* fit out of likely possible fits. In addition, recent developments have made it possible to measure multiple modalities in the same cell, such as combined

Biochemical Society Transactions (2023) **51** 1–12
https://doi.org/10.1042/BST20210145

ATAC-seq and RNA-seq [176–178], which offers new, exciting opportunities for combining single cell sequencing with multi-omics data. ANNs, finally, have been made relatively easy to implement, can learn any type of interaction, and make no assumptions about the data (such as normality), which makes them extremely powerful GRN tools. However, it is not yet clear what the optimal architecture is for these networks, and interpreting the learned network from the ANN remains difficult.

GRN inference has become a data science, and it is time that we start treating it as such. Integrating multiple omics, several thousands of cells, and training complex machine learning models requires specialized knowledge. Common mistakes, such as treating cells from the same sample as independent [118], double dipping [179], and data leakage [180], can be avoided by proper data science training, but are unfortunately still common. Comparing the quality of GRN inference methods requires standardized benchmarks with multiple datasets, preferably a mix of experimental data and simulated data [181–183]. Simulated data has the advantage that the ground truth is known which makes benchmarking straightforward, but has the clear disadvantage that the quality of simulated data depends on its assumptions and may actually not be representative of real biological data. The DREAM challenges [164,184] and BEELINE platform [144] are great examples, with predefined datasets and quality metrics. Only by measuring network accuracy in equal settings will it be possible to properly compare methods. It is however important to note that the goal of GRN inference is to gain mechanistic insights, as opposed to getting an optimal benchmark score, which makes fair comparison between approaches hard.

All together, we expect the field of transcription regulation in development to move towards increasingly multimodal GRN inference techniques to identify causal relations between genes. Single cell sequencing adds a cell type-specific precision which bulk sequencing can not provide. Finally, we expect the adoption of artificial neural networks as the field matures in technology and formal training, as these methods are inherently more powerful as previously used techniques.

Perspectives

- Gene regulatory networks have served as powerful models to understand gene regulatory programs in development and disease. Amongst others, these networks have been applied to model developmental patterning, to identify relevant transcription factors for cell fate transitions and to characterize deregulated transcriptional programs in disease.

- We believe three relatively recent developments will impact the computational inference of GRNs. The combination of multiple data modalities, such as RNA expression and DNA accessibility, help to constrain GRN topology and to predict directed networks. Single cell sequencing will become the *de facto* standard, as it allows for cell type-specific models and is able to provide the high number of measurements that are needed. Finally, artificial neural networks have the capability to create flexible and powerful models of gene regulation, which will benefit efficient and accurate GRN inference.

- The developments outlined above have the potential to significantly improve GRN inference. To fully exploit these approaches we have to implement common data science practices, and develop community-driven benchmarks to consistently measure the performance of different techniques.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Netherlands Organization for Scientific Research [NWO grant 016.Vidi.189.081 to S.J.v.H.].

Author Contributions

Maarten van der Sande: writing — original draft and writing — review and editing. **Siebren Frölich:** writing — original draft and writing — review and editing. **Simon J. van Heeringen:** writing — review and editing, funding acquisition, and supervision.

Abbreviations

ANN, artificial neural network; ATAC-seq, assay for Transposase-Accessible Chromatin using sequencing; CREs, *cis*-regulatory elements; DNAse-seq, DNase I hypersensitive sites sequencing; GRN, gene regulatory network; TFs, transcription factors.

References

1 Mammoto, A., Mammoto, T. and Ingber, D.E. (2012) Mechanosensitive mechanisms in transcriptional regulation. *J. Cell Sci.* **125**, 3061–3073 https://doi.org/10.1242/jcs.093005
2 Cameron, R.A., Hough-Evans, B.R., Britten, R.J. and Davidson, E.H. (1987) Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev.* **1**, 75–85 https://doi.org/10.1101/gad.1.1.75
3 Cooper, G.M. (2000) Regulation of Transcription in Eukaryotes. In *The Cell: A Molecular Approach*, 2nd edn, Sinauer Associates, Sunderland, MA https://www.ncbi.nlm.nih.gov/books/NBK9904/
4 Li, Y.J., Fu, X.H., Liu, D.P. and Liang, C.C. (2004) Opening the chromatin for transcription. *Int. J. Biochem. Cell Biol.* **36**, 1411–1423 https://doi.org/10.1016/j.biocel.2003.11.003
5 McMahon, A.P., Novak, T.J., Britten, R.J. and Davidson, E.H. (1984) Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. *Proc. Natl Acad. Sci. U.S.A.* **81**, 7490–7494 https://doi.org/10.1073/pnas.81.23.7490
6 Gordân, R., Hartemink, A.J. and Bulyk, M.L. (2009) Distinguishing direct versus indirect transcription factor–DNA interactions. *Genome Res.* **19**, 2090–2100 https://doi.org/10.1101/gr.094144.109
7 Chen, H. and Pugh, B.F. (2021) What do transcription factors interact with? *J. Mol. Biol.* **433**, 166883 https://doi.org/10.1016/j.jmb.2021.166883
8 Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 https://doi.org/10.1038/nrg2538
9 Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M. et al. (2018) The human transcription factors. *Cell* **172**, 650–665 https://doi.org/10.1016/j.cell.2018.01.029
10 Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z. et al. (2018) Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 https://doi.org/10.1038/s41559-018-0575-6
11 Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J. et al. (2015) Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife* **4**, e04837 https://doi.org/10.7554/eLife.04837
12 Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A. et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 https://doi.org/10.1126/science.1186176
13 Villar, D., Flicek, P. and Odom, D.T. (2014) Evolution of transcription factor binding in metazoans: mechanisms and functional implications. *Nat. Rev. Genet.* **15**, 221–233 https://doi.org/10.1038/nrg3481
14 Levine, M. and Davidson, E.H. (2005) Gene regulatory networks for development. *Proc. Natl Acad. Sci. U.S.A.* **102**, 4936–4942 https://doi.org/10.1073/pnas.0408031102
15 Avsec, Ž, Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K. et al. (2021) Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 https://doi.org/10.1038/s41588-021-00782-6
16 Brown, C.D., Johnson, D.S. and Sidow, A. (2007) Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science* **317**, 1557–1560 https://doi.org/10.1126/science.1145893
17 Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S. and Levine, M.S. (2016) Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl Acad. Sci. U.S.A.* **113**, 6508–6513 https://doi.org/10.1073/pnas.1605085113
18 Wong, E.S., Zheng, D., Tan, S.Z., Bower, N.I., Garside, V., Vanwalleghem, G. et al. (2020) Deep conservation of the enhancer regulatory code in animals. *Science* **370**, eaax8137 https://doi.org/10.1126/science.aax8137
19 Zeitlinger, J. (2020) Seven myths of how transcription factors read the *cis*-regulatory code. *Curr. Opin. Syst. Biol.* **23**, 22–31 https://doi.org/10.1016/j.coisb.2020.08.002
20 Martindale, M.Q. (2005) The evolution of metazoan axial properties. *Nat. Rev. Genet.* **6**, 917–927 https://doi.org/10.1038/nrg1725
21 Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science* **165**, 349–357 https://doi.org/10.1126/science.165.3891.349
22 Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H. et al. (2002) A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev. Biol.* **246**, 162–190 https://doi.org/10.1006/dbio.2002.0635
23 Cary, G.A., McCauley, B.S., Zueva, O., Pattinato, J., Longabaugh, W. and Hinman, V.F. (2020) Systematic comparison of sea urchin and sea star developmental gene regulatory networks explains how novelty is incorporated in early development. *Nat. Commun.* **11**, 6235 https://doi.org/10.1038/s41467-020-20023-4
24 Peter, I.S. and Davidson, E.H. (2011) A gene regulatory network controlling the embryonic specification of endoderm. *Nature.* **474**, 635–639 https://doi.org/10.1038/nature10100
25 Saudemont, A., Haillot, E., Mekpoh, F., Bessodes, N., Quirin, M., Lapraz, F. et al. (2010) Ancestral regulatory circuits governing ectoderm patterning downstream of nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm. *PLoS Genet.* **6**, e1001259 https://doi.org/10.1371/journal.pgen.1001259
26 Charney, R.M., Paraiso, K.D., Blitz, I.L. and Cho, K.W.Y. (2017) A gene regulatory program controlling early Xenopus mesendoderm formation: Network conservation and motifs. *Semin. Cell Dev. Biol.* **66**, 12–24 https://doi.org/10.1016/j.semcdb.2017.03.003

27 Koide, T., Hayata, T. and Cho, K.W.Y. (2005) Xenopus as a model system to study transcriptional regulatory networks. *Proc. Natl Acad. Sci. U.S.A.* **102**, 4943–4948 https://doi.org/10.1073/pnas.0408125102
28 Rankin, S.A., Kormish, J., Kofron, M., Jegga, A. and Zorn, A.M. (2011) A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer. *Dev. Biol.* **351**, 297–310 https://doi.org/10.1016/j.ydbio.2010.11.037
29 Sinner, D., Kirilenko, P., Rankin, S., Wei, E., Howard, L., Kofron, M. et al. (2006) Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development* **133**, 1955–1966 https://doi.org/10.1242/dev.02358
30 Lukoseviciute, M., Gavriouchkina, D., Williams, R.M., Hochgreb-Hagele, T., Senanayake, U., Chong-Morrison, V. et al. (2018) From pioneer to repressor: bimodal foxd3 activity dynamically remodels neural crest regulatory landscape *in vivo*. *Dev. Cell* **47**, 608–628.e6 https://doi.org/10.1016/j.devcel.2018.11.009
31 Williams, R.M., Candido-Ferreira, I., Repapi, E., Gavriouchkina, D., Senanayake, U., Ling, I.T.C. et al. (2019) Reconstruction of the global neural crest gene regulatory network *in vivo*. *Dev. Cell* **51**, 255–276.e7 https://doi.org/10.1016/j.devcel.2019.10.003
32 Jaeger, J. (2011) The gap gene network. *Cell. Mol. Life Sci.* **68**, 243–274 https://doi.org/10.1007/s00018-010-0536-y
33 Kueh, H.Y. and Rothenberg, E.V. (2012) Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 79–102 https://doi.org/10.1002/wsbm.162
34 Pimanda, J.E. and Göttgens, B. (2010) Gene regulatory networks governing haematopoietic stem cell development and identity. *Int. J. Dev. Biol.* **54**, 1201–1211 https://doi.org/10.1387/ijdb.093038jp
35 Singh, H., Khan, A.A. and Dinner, A.R. (2014) Gene regulatory networks in the immune system. *Trends Immunol.* **35**, 211–218 https://doi.org/10.1016/j.it.2014.03.006
36 Ryan, G.E. and Farley, E.K. (2020) Functional genomic approaches to elucidate the role of enhancers during development. *WIREs Syst. Biol. Med.* **12**, e1467 https://doi.org/10.1002/wsbm.1467
37 Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F. et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 https://doi.org/10.1038/nature11247
38 Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K. et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 https://doi.org/10.1126/science.292.5518.929
39 Allocco, D.J., Kohane, I.S. and Butte, A.J. (2004) Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* **5**, 18 https://doi.org/10.1186/1471-2105-5-18
40 Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. U.S.A.* **95**, 14863–8 https://doi.org/10.1073/pnas.95.25.14863
41 Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article 17 https://doi.org/10.2202/1544-6115.1128
42 Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R.D. et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 https://doi.org/10.1186/1471-2105-7-S1-S7
43 Chasman, D. and Roy, S. (2017) Inference of cell type specific regulatory networks on mammalian lineages. *Curr. Opin. Syst. Biol.* **2**, 130–139 https://doi.org/10.1016/j.coisb.2017.04.001
44 Delgado, F.M. and Gómez-Vela, F. (2019) Computational methods for gene regulatory networks reconstruction and analysis: a review. *Artif. Intell. Med.* **95**, 133–145 https://doi.org/10.1016/j.artmed.2018.10.006
45 Mercatelli, D., Scalambra, L., Triboli, L., Ray, F. and Giorgi, F.M. (2020) Gene regulatory network inference resources: a practical overview. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194430 https://doi.org/10.1016/j.bbagrm.2019.194430
46 Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T. et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 https://doi.org/10.1038/nmeth1068
47 He, Q., Johnston, J. and Zeitlinger, J. (2015) ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33**, 395–401 https://doi.org/10.1038/nbt.3121
48 Rhee, H.S. and Pugh, B.F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 https://doi.org/10.1016/j.cell.2011.11.013
49 Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G. et al. (2019) CUT&tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 https://doi.org/10.1038/s41467-019-09982-5
50 Keilwagen, J., Posch, S. and Grau, J. (2019) Accurate prediction of cell type-specific transcription factor binding. *Genome Biol.* **20**, 9 https://doi.org/10.1186/s13059-018-1614-y
51 Li, H., Quang, D. and Guan, Y. (2019) Anchor: trans-cell type prediction of transcription factor binding sites. *Genome Res.* **29**, 281–292 https://doi.org/10.1101/gr.237156.118
52 Quang, D. and Xie, X. (2019) Factornet: A deep learning framework for predicting cell type specific transcription factor binding from nucleotide-resolution sequential data. *Methods* **166**, 40–47 https://doi.org/10.1016/j.ymeth.2019.03.020
53 Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J.A. and Cheng, J. (2021) DeepGRN: prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC Bioinformatics* **22**, 38 https://doi.org/10.1186/s12859-020-03952-1
54 Mariani, L., Weinand, K., Gisselbrecht, S.S. and Bulyk, M.L. (2020) MEDEA: analysis of transcription factor binding motifs in accessible chromatin. *Genome Res.* **30**, 736–748 https://doi.org/10.1101/gr.260877.120
55 Li, H. and Guan, Y. (2021) Fast decoding cell type–specific transcription factor binding landscape at single-nucleotide resolution. *Genome Res.* **31**, 721–731 https://doi.org/10.1101/gr.269613.120
56 Bruse, N. and van Heeringen, S.J. (2018) Gimmemotifs: an analysis framework for transcription factor motif analysis. *bioRxiv* https://doi.org/10.1101/474403
57 Schreiber, J., Bilmes, J. and Noble, W.S. (2020) Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples. *Genome Biol.* **21**, 82 https://doi.org/10.1186/s13059-020-01978-5
58 Cheng, J., Xu, M., Liu, Y. and Huang, W. (2022) *AttBind: Prediction of Transcription Factor Binding Sites Across Cell-types Based on Attention Mechanism*. In *2022 7th International Conference on Computer and Communication Systems (ICCCS)*, pp. 135–139

59 Behjati Ardakani, F., Schmidt, F. and Schulz, M.H. (2019) Predicting transcription factor binding using ensemble random forest models. *F1000Res.* **7**, 1603 https://doi.org/10.12688/f1000research.16200.2
60 Trotter, M.V., Nguyen, C.Q., Young, S., Woodruff, R.T. and Branson, K.M. (2021) Epigenomic language models powered by cerebras. *arXiv* https://doi.org/10.48550/arXiv.2112.07571
61 Yi, R., Cho, K. and Bonneau, R. (2022) NetTIME: a multitask and base-pair resolution framework for improved transcription factor binding site prediction. *Bioinformatics* **38**, 4762–4770 https://doi.org/10.1093/bioinformatics/btac569
62 Pap, G., Zoltán, G., Ádám, K., Tóth, L. and Hegedus, Z. (2021) Transcription factor binding site detection using convolutional neural networks with a functional group-based data representation. *J. Phys. Conf. Ser.* **1824**, 012001 https://doi.org/10.1088/1742-6596/1824/1/012001
63 Karimzadeh, M. and Hoffman, M.M. (2022) Virtual ChIP-seq: predicting transcription factor binding by learning from the transcriptome. *Genome Biol.* **23**, 126 https://doi.org/10.1186/s13059-022-02690-2
64 Koo, P.K. and Ploenzke, M. (2020) Deep learning for inferring transcription factor binding sites. *Curr. Opin. Syst. Biol.* **19**, 16–23 https://doi.org/10.1016/j.coisb.2020.04.001
65 Kundaje, A., Boley, N., Kuffner, R., Heiser, L., Costello, J., Stolovitzky, G. et al. (2017) ENCODE-DREAM *in vivo* transcription factor binding site prediction challenge. *Synapse* https://doi.org/10.7303/syn6131484
66 Cochran, K., Srivastava, D., Shrikumar, A., Balsubramani, A., Hardison, R.C., Kundaje, A. et al. (2022) Domain-adaptive neural networks improve cross-species prediction of transcription factor binding. *Genome Res.* **32**, 512–523 https://doi.org/10.1101/gr.275394.121
67 Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z. et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 https://doi.org/10.1016/j.cell.2007.12.014
68 Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 https://doi.org/10.1038/nmeth.2688
69 Buenrostro, J., Wu, B., Chang, H. and Greenleaf, W. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 https://doi.org/10.1002/0471142727.mb2129s109
70 Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.P., Renno, J. et al. (2018) Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534.e20 https://doi.org/10.1016/j.cell.2018.05.019
71 Marlétaz, F., Firbas, P.N., Maeso, I., Tena, J.J., Bogdanovic, O., Perry, M. et al. (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 https://doi.org/10.1038/s41586-018-0734-6
72 Uesaka, M., Kuratani, S., Takeda, H. and Irie, N. (2019) Recapitulation-like developmental transitions of chromatin accessibility in vertebrates. *Zool. Lett.* **5**, 33 https://doi.org/10.1186/s40851-019-0148-9
73 Pálfy, M., Schulze, G., Valen, E. and Vastenhouw, N.L. (2020) Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet.* **16**, e1008546 https://doi.org/10.1371/journal.pgen.1008546
74 Shashikant, T., Khor, J.M. and Ettensohn, C.A. (2018) Global analysis of primary mesenchyme cell cis-regulatory modules by chromatin accessibility profiling. *BMC Genomics* **19**, 206 https://doi.org/10.1186/s12864-018-4542-z
75 Madgwick, A., Magri, M.S., Dantec, C., Gailly, D., Fiuza, U.M., Guignard, L. et al. (2019) Evolution of embryonic cis-regulatory landscapes between divergent *phallusia* and *ciona ascidians*. *Dev. Biol.* **448**, 71–87 https://doi.org/10.1016/j.ydbio.2019.01.003
76 Esmaeili, M., Blythe, S.A., Tobias, J.W., Zhang, K., Yang, J. and Klein, P.S. (2020) Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Dev. Biol.* **462**, 20–35 https://doi.org/10.1016/j.ydbio.2020.02.013
77 Bright, A.R., van Genesen, S., Li, Q., Grasso, A., Frölich, S., van der Sande, M. et al. (2021) Combinatorial transcription factor activities on open chromatin induce embryonic heterogeneity in vertebrates. *EMBO J.* **40**, e104913 https://doi.org/10.15252/embj.2020104913
78 Yang, H., Luan, Y., Liu, T., Lee, H.J., Fang, L., Wang, Y. et al. (2020) A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* **588**, 337–343 https://doi.org/10.1038/s41586-020-2962-9
79 Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A. et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 https://doi.org/10.1038/nature07730
80 Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J. et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. U.S.A.* **107**, 21931–6 https://doi.org/10.1073/pnas.1016071107
81 Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D. et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 https://doi.org/10.1038/ng1966
82 Simeone, A., Pannese, M., Acampora, D., D'Esposito, M. and Boncinelli, E. (1988) At least three human homeoboxes on chromosome 12 belong to the same transcription unit. *Nucleic Acids Res.* **16**, 5379–5390 https://doi.org/10.1093/nar/16.12.5379
83 Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371–378 https://doi.org/10.1126/science.2667136
84 Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P. et al. (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 https://doi.org/10.1038/nmeth.1313
85 Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B. et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 https://doi.org/10.1038/nature11212
86 Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E. and Stamatoyannopoulos, J.A. (2012) Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150**, 1274–1286 https://doi.org/10.1016/j.cell.2012.04.040
87 He, H.H., Meyer, C.A., Hu, S.S., Chen, M.W., Zang, C., Liu, Y. et al. (2014) Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73–78 https://doi.org/10.1038/nmeth.2762
88 Sung, M.H., Baek, S. and Hager, G.L. (2016) Genome-wide footprinting: ready for prime time? *Nat. Methods* **13**, 222–228 https://doi.org/10.1038/nmeth.3766
89 Sung, M.H., Guertin, M.J., Baek, S. and Hager, G.L. (2014) DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol. Cell* **56**, 275–285 https://doi.org/10.1016/j.molcel.2014.08.016
90 Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M. and Costa, I.G. (2019) Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 https://doi.org/10.1186/s13059-019-1642-2

91 Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R. et al. (2020) ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 https://doi.org/10.1038/s41467-020-18035-1
92 Miraldi, E.R., Pokrovskii, M., Watters, A., Castro, D.M., De Veaux, N., Hall, J.A. et al. (2019) Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 cells. *Genome Res.* **29**, 449–463 https://doi.org/10.1101/gr.238253.118
93 Siahpirani, A.F. and Roy, S. (2017) A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Res.* **45**, e21 https://doi.org/10.1093/nar/gkw963
94 Sonawane, A.R., DeMeo, D.L., Quackenbush, J. and Glass, K. (2021) Constructing gene regulatory networks using epigenetic data. *NPJ Syst. Biol. Appl.* **7**, 45 https://doi.org/10.1038/s41540-021-00208-3
95 Madsen, J.G.S., Rauch, A., Van Hauwaert, E.L., Schmidt, S.F., Winnefeld, M. and Mandrup, S. (2018) Integrated analysis of motif activity and gene expression changes of transcription factors. *Genome Res.* **28**, 243–255 https://doi.org/10.1101/gr.227231.117
96 Kamal, A., Arnold, C., Claringbould, A., Moussa, R., Servaas, N., Kholmatov, M. et al. (2022) GRaNIE and GRaNPA: Inference and evaluation of enhancer-mediated gene regulatory networks applied to study macrophages. *bioRxiv* https://doi.org/10.1101/2021.12.18.473290
97 Schmidt, F., Kern, F., Ebert, P., Baumgarten, N. and Schulz, M.H. (2019) TEPIC 2—an extended framework for transcription factor binding prediction and integrative epigenomic analysis. *Bioinformatics* **35**, 1608–1609 https://doi.org/10.1093/bioinformatics/bty856
98 Xu, Q., Georgiou, G., Frölich, S., van der Sande, M., Veenstra, G.J.C., Zhou, H. et al. (2021) ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *bioRxiv* https://doi.org/10.1101/2020.06.05.135798
99 Ghaffari, S., Hanson, C., Schmidt, R.E., Bouchonville, K.J., Offer, S.M. and Sinha, S. (2021) An integrated multi-omics approach to identify regulatory mechanisms in cancer metastatic processes. *Genome Biol.* **22**, 19 https://doi.org/10.1186/s13059-020-02213-x
100 Vijayabaskar, M.S., Goode, D.K., Obier, N., Lichtinger, M., Emmett, A.M.L., Abidin, F.N.Z. et al. (2019) Identification of gene specific cis-regulatory elements during differentiation of mouse embryonic stem cells: an integrative approach using high-throughput datasets. *PLoS Comput. Biol.* **15**, e1007337 https://doi.org/10.1371/journal.pcbi.1007337
101 Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 https://doi.org/10.1016/j.cub.2010.06.070
102 Nora, E.P., Lajoie, B.R., Schulz, E.G. Giorgetti, L., Okamoto, I., Servant, N. et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 https://doi.org/10.1038/nature11049
103 Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science* **295**, 1306–1311 https://doi.org/10.1126/science.1067799
104 Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A. et al. (2009) Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 https://doi.org/10.1126/science.1181369
105 Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L. et al. (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 https://doi.org/10.1038/ng.3286
106 Sanyal, A., Lajoie, B.R., Jain, G. and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 https://doi.org/10.1038/nature11279
107 Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P. et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 https://doi.org/10.1016/j.cell.2011.12.014
108 Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. and Bergmann, S. (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 https://doi.org/10.1038/nmeth.3799
109 Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V. et al. (2019) Activity-by-Contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 https://doi.org/10.1038/s41588-019-0538-0
110 Mercatelli, D. and Lopez-Garcia, G. (2020) Giorgi FM. corto: a lightweight R package for gene network inference and master regulator analysis. *Bioinformatics* **36**, 3916–3917 https://doi.org/10.1093/bioinformatics/btaa223
111 Glass, K., Huttenhower, C., Quackenbush, J. and Yuan, G.C. (2013) Passing messages between biological networks to refine predicted interactions. *PLoS ONE* **8**, e64832 https://doi.org/10.1371/journal.pone.0064832
112 Jensen, E. (2014) Technical review: *in situ* hybridization. *Anat. Rec.* **297**, 1349–1353 https://doi.org/10.1002/ar.22944
113 Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 https://doi.org/10.1038/nrg2484
114 Longo, S.K., Guo, M.G., Ji, A.L. and Khavari, P.A. (2021) Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**, 627–644 https://doi.org/10.1038/s41576-021-00370-8
115 Borm, L.E., Mossi Albiach, A., Mannens, C.C.A., Janusauskas, J., Özgün, C., Fernández-García, D. et al. (2022) Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat. Biotechnol.*, 1–10 https://doi.org/10.1038/s41587-022-01455-3
116 Ly, L.H. and Vingron, M. (2022) Effect of imputation on gene network reconstruction from single-cell RNA-seq data. *Patterns* **3**, 100414 https://doi.org/10.1016/j.patter.2021.100414
117 Stone, M., Li, J., McCalla, S.G., Siahpirani, A.F., Periyasamy, V., Shin, J. et al. (2022) Identifying strengths and weaknesses of methods for computational network inference from single cell RNA-seq data. *bioRxiv* https://doi.org/10.1101/2021.06.01.446671
118 Zimmerman, K.D., Espeland, M.A. and Langefeld, C.D. (2021) A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 https://doi.org/10.1038/s41467-021-21038-1
119 Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E. et al. (2019) Metacell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 https://doi.org/10.1186/s13059-019-1812-2
120 Huynh-Thu, V.A., Irrthum, A., Wehenkel, L. and Geurts, P. (2010) Inferring regulatory networks from expression data using tree-Based methods. *PLoS ONE* **5**, e12776 https://doi.org/10.1371/journal.pone.0012776
121 Chan, T.E., Stumpf, M.P.H. and Babtie, A.C. (2017) Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5**, 251–267.e3 https://doi.org/10.1016/j.cels.2017.08.014
122 Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G. et al. (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 https://doi.org/10.1038/nmeth.4463

123 Jansen, C., Ramirez, R.N., El-Ali, N.C., Gomez-Cabrero, D., Tegner, J., Merkenschlager, M. et al. (2019) Building gene regulatory networks from scATAC-seq and scRNA-seq using linked self organizing maps. *PLoS Comput. Biol.* **15**, e1006555 https://doi.org/10.1371/journal.pcbi.1006555
124 González-Blas, C.B., Winter, S.D., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V. et al. (2022) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *bioRxiv* https://doi.org/10.1101/2022.08.19.504505
125 Jiang, J., Lyu, P., Li, J., Huang, S., Blackshaw, S., Qian, J. et al. (2022) IReNA: integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles. *bioRxiv* https://doi.org/10.1101/2021.11.22.469628
126 Kamimoto, K., Hoffmann, C.M. and Morris, S.A. (2020) Celloracle: dissecting cell identity via network inference and in silico gene perturbation. *bioRxiv* https://doi.org/10.1101/2020.02.17.947416
127 Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H. et al. (2019) A lineage-resolved molecular atlas of C. elegans embryogenesis at single cell resolution. *Science* **365**, eaax1971 https://doi.org/10.1126/science.aax1971
128 Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B. et al. (2019) PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 https://doi.org/10.1186/s13059-019-1663-x
129 Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. et al. (2017) Reversed graph embedding resolves complex single-cell developmental trajectories. *Nat. Methods* **14**, 979–982 https://doi.org/10.1038/nmeth.4402
130 Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N. et al. (2018) Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 https://doi.org/10.1186/s12864-018-4772-0
131 Bergen, V., Lange, M., Peidli, S., Wolf, F.A. and Theis, F.J. (2020) Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 https://doi.org/10.1038/s41587-020-0591-3
132 La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V. et al. (2018) RNA velocity of single cells. *Nature* **560**, 494–498 https://doi.org/10.1038/s41586-018-0414-6
133 Aubin-Frankowski, P.C. and Vert, J.P. (2020) Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. *Bioinformatics* **36**, 4774–4780 https://doi.org/10.1093/bioinformatics/btaa576
134 Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N. et al. (2017) SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321 https://doi.org/10.1093/bioinformatics/btx194
135 Deshpande, A., Chu, L.F., Stewart, R. and Gitter, A. (2022) Network inference with granger causality ensembles on single-cell transcriptomic data. *Cell Rep.* **38**, 110333 https://doi.org/10.1016/j.celrep.2022.110333
136 Papili Gao, N., Ud-Dean, S.M.M., Gandrillon, O. and Gunawan, R. (2018) SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* **34**, 258–266 https://doi.org/10.1093/bioinformatics/btx575
137 Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T. et al. (2020) Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst.* **10**, 265–274.e11 https://doi.org/10.1016/j.cels.2020.02.003
138 Woodhouse, S., Piterman, N., Wintersteiger, C.M., Göttgens, B. and Fisher, J. (2018) SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* **12**, 59 https://doi.org/10.1186/s12918-018-0581-y
139 Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I.M., Carrion, M.C. and Huang, Y. (2018) A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics* **34**, 964–970 https://doi.org/10.1093/bioinformatics/btx605
140 Chubb, J.R., Trcek, T., Shenoy, S.M. and Singer, R.H. (2006) Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 https://doi.org/10.1016/j.cub.2006.03.092
141 Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B. et al. (2019) Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 https://doi.org/10.1038/s41586-018-0836-1
142 Ventre, E., Herbach, U., Espinasse, T., Benoit, G. and Gandrillon, O. (2022) One model fits all: combining inference and simulation of gene regulatory networks. *bioRxiv* https://doi.org/10.1101/2022.06.19.496754
143 Chen, S. and Mar, J.C. (2018) Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* **19**, 232 https://doi.org/10.1186/s12859-018-2217-z
144 Pratapa, A., Jalihal, A.P., Law, J.N., Bharadwaj, A. and Murali, T.M. (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 https://doi.org/10.1038/s41592-019-0690-6
145 Kim, D., Risca, V., Reynolds, D., Chappell, J., Rubin, A., Jung, N. et al. (2021) The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. *Nat. Genet.* **53**, 1564–1576 https://doi.org/10.1038/s41588-021-00947-3
146 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 https://doi.org/10.1038/s41586-021-03819-2
147 Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems* (Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., eds), Curran Associates, Inc., Red Hook/New York/United States https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
148 Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G. et al. (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 https://doi.org/10.1038/nature16961
149 Eraslan, G., Avsec, Ž, Gagneur, J. and Theis, F.J. (2019) Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* **20**, 389–403 https://doi.org/10.1038/s41576-019-0122-6
150 Alipanahi, B., Delong, A., Weirauch, M.T. and Frey, B.J. (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 https://doi.org/10.1038/nbt.3300
151 Zhou, J. and Troyanskaya, O.G. (2015) Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* **12**, 931–934 https://doi.org/10.1038/nmeth.3547
152 Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y. and Snoek, J. (2018) Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.* **28**, 739–750 https://doi.org/10.1101/gr.227819.117
153 Shu, H., Zhou, J., Lian, Q., Li, H., Zhao, D., Zeng, J. et al. (2021) Modeling gene regulatory networks using neural network architectures. *Nat. Comput. Sci.* **1**, 491–501 https://doi.org/10.1038/s43588-021-00099-8

154 Rubiolo, M., Milone, D.H. and Stegmayer, G. (2018) Extreme learning machines for reverse engineering of gene regulatory networks from expression time series. *Bioinformatics* **34**, 1253–1260 https://doi.org/10.1093/bioinformatics/btx730
155 Dutil, F., Cohen, J.P., Weiss, M., Derevyanko, G. and Bengio, Y. (2018) Towards gene expression convolutions using gene interaction graphs. *ArXiv* https://doi.org/10.48550/arXiv.1806.06975
156 Wang, J., Ma, A., Ma, Q., Xu, D. and Joshi, T. (2020) Inductive inference of gene regulatory network using supervised and semi-supervised graph neural networks. *Comput. Struct. Biotechnol. J.* **18**, 3335–3343 https://doi.org/10.1016/j.csbj.2020.10.022
157 Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **2**, 303–314 https://doi.org/10.1007/BF02551274
158 Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 https://doi.org/10.1016/0893-6080(89)90020-8
159 Zhang, Y., Tiňo, P., Leonardis, A. and Tang, K. (2021) A survey on neural network interpretability. *IEEE Trans. Emerg. Top. Comput. Intell.* **5**, 726–742 https://doi.org/10.1109/TETCI.2021.3100641
160 Francois C. Keras. 2015
161 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G. et al. (2019) Pytorch: an imperative style, high-performance deep learning library. *ArXiv* https://doi.org/10.48550/arXiv.1912.01703
162 de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. and Vogel, C. (2009) Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512–1526 https://doi.org/10.1039/B908315D
163 Krishnan, A., Giuliani, A. and Tomita, M. (2007) Indeterminacy of reverse engineering of gene regulatory networks: the curse of gene elasticity. *PLoS ONE* **2**, e562 https://doi.org/10.1371/journal.pone.0000562
164 Cokelaer, T., Bansal, M., Bare, C., Bilal, E., Bot, B.M., Chaibub Neto, E. et al. (2016) DREAMTools: a Python package for scoring collaborative challenges. *F1000Res.* **4**, 1030 https://doi.org/10.12688/f1000research.7118.2
165 (DREAM Challenge) C-Path Analytics - syn21760283 - Wiki [Internet]. [cited 2021 Jun 8]. Available from: https://www.synapse.org/#!Synapse:syn21760283/wiki/603540
166 Guo, W., Calixto, C.P.G., Tzioutziou, N., Lin, P., Waugh, R., Brown, J.W.S. et al. (2017) Evaluation and improvement of the regulatory inference for large co-expression networks with limited sample size. *BMC Syst. Biol.* **11**, 62 https://doi.org/10.1186/s12918-017-0440-2
167 Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M. et al. (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 https://doi.org/10.1038/nmeth.2016
168 Stolovitzky, G., Monroe, D. and Califano, A. (2007) Dialogue on reverse-engineering assessment and methods. *Ann. N. Y. Acad. Sci.* **1115**, 1–22 https://doi.org/10.1196/annals.1407.021
169 Zoppoli, P., Morganella, S. and Ceccarelli, M. (2010) TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* **11**, 154 https://doi.org/10.1186/1471-2105-11-154
170 Ernst, J., Vainas, O., Harbison, C.T., Simon, I. and Bar-Joseph, Z. (2007) Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.* **3**, 74 https://doi.org/10.1038/msb4100115
171 Schulz, M.H., Devanny, W.E., Gitter, A., Zhong, S., Ernst, J. and Bar-Joseph, Z. (2012) DREM 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst. Biol.* **6**, 104 https://doi.org/10.1186/1752-0509-6-104
172 Ding, J., Hagood, J.S., Ambalavanan, N., Kaminski, N. and Bar-Joseph, Z. (2018) iDREM: interactive visualization of dynamic regulatory networks. *PLoS Comput. Biol.* **14**, e1006019 https://doi.org/10.1371/journal.pcbi.1006019
173 Conard, A.M., Goodman, N., Hu, Y., Perrimon, N., Singh, R., Lawrence, C. et al. (2021) TIMEOR: a web-based tool to uncover temporal regulatory mechanisms from multi-omics data. *Nucleic Acids Res.* **49**, W641–W653 https://doi.org/10.1093/nar/gkab384
174 Bühlmann, P. and van de Geer, S. (2011) SpringerLink (Online service). In *Statistics for High-Dimensional Data [electronic resource]: Methods, Theory and Applications*. pp. 575, Berlin, Heidelberg, Springer-Verlag Berlin Heidelberg http://archive.org/details/statisticsforhig00bhlm
175 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 https://dl.acm.org/doi/10.5555/2627435.2670313
176 Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y. et al. (2020) Chromatin potential identified by shared single-Cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 https://doi.org/10.1016/j.cell.2020.09.056
177 Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J. et al. (2018) Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 https://doi.org/10.1126/science.aau0730
178 Chen, S., Lake, B.B. and Zhang, K. (2019) High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 https://doi.org/10.1038/s41587-019-0290-0
179 Gao, L.L., Bien, J. and Witten, D. (2021) Selective inference for hierarchical clustering. *arXiv* https://doi.org/10.48550/arXiv.2012.02936
180 Schreiber, J., Singh, R., Bilmes, J. and Noble, W.S. (2020) A pitfall for machine learning methods aiming to predict across cell types. *Genome Biol.* **21**, 282 https://doi.org/10.1186/s13059-020-02177-y
181 Dibaeinia, P. and Sinha, S. (2020) SERGIO: a single-cell expression simulator guided by gene regulatory networks. *Cell Syst.* **11**, 252–271.e11 https://doi.org/10.1016/j.cels.2020.08.003
182 Li, H., Zhang, Z., Squires, M., Chen, X. and Zhang, X. (2022) Scmultisim: simulation of multi-modality single cell data guided by cell-cell interactions and gene regulatory networks. *bioRxiv* **11**, 252–271 https://doi.org/10.1101/2022.10.15.512320
183 Ventre, E. (2021) Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics. *In Silico Biol.* **14**, 89–113 https://doi.org/10.3233/ISB-210226
184 ENCODE-DREAM in vivo Transcription Factor Binding Site Prediction Challenge [Internet]. DREAM Challenges. [cited 2021 Jun 8]. Available from: https://dreamchallenges.org/encode-dream-in-vivo-transcription-factor-binding-site-prediction-challenge/