4-11-2011

What does biologically meaningful mean? A perspective on gene regulatory network validation

Albertha J. M. Walhout
University of Massachusetts Medical School, Marian.Walhout@umassmed.edu

Genome **Biology**

OPINION

What does biologically meaningful mean? A perspective on gene regulatory network validation

Albertha JM Walhout*

Abstract

Gene regulatory networks (GRNs) are rapidly being delineated, but their quality and biological meaning are often questioned. Here, I argue that biological meaning is challenging to define and discuss reasons why GRN validation should be interpreted cautiously.

The control of gene expression is pivotal in biology. It is accomplished by a large number of regulators, including transcription factors (TFs), that can modulate mRNA synthesis by directly interacting with regulatory genome sequences. The human genome contains about 20,000 genes and an estimated 1,400 TFs [1]. Although much is known about the basic mechanics of transcription, little is known about how TFs function collectively in the context of intricate gene regulatory networks (GRNs) to achieve complex biological outputs during development and in physiology and disease.

Ideally, one would like to comprehensively map the binding of each TF within the genome and understand the effects that such interactions have on its target genes. Conversely, for each gene, one would like to know which TFs contribute to its expression and in which cells and under which circumstances this contribution occurs.

Over the past decade several high-throughput methodologies have been developed, standardized and implemented to map GRNs, including computational reverse engineering (reviewed in [2,3]), chromatin immuno-precipitation (ChIP) combined with microarrays (ChIP-chip) or next generation sequencing (ChIP-seq) (reviewed in [4,5]), and yeast one-hybrid (Y1H) assays (reviewed in [6,7]). Each of these methods has inherent limitations, and therefore GRNs might miss interactions (false

*Correspondence: marian.walhout@umassmed.edu
Program in Gene Function and Expression, Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

negatives) and contain interactions that are not 'biologically meaningful' (false positives; Table 1). Determining the scale of these limitations has proven difficult, in part because it is challenging to define the term 'biologically meaningful'. For instance, interactions between genes and regulators are often deemed biologically meaningful only if the expression of the gene changes following removal or reduction of the regulator, and/or if mutations in the gene and the regulator confer similar phenotypes. Here, I discuss methods that are used to identify interactions between genes and regulators and illustrate different levels of validation that can be used to obtain further support for these interactions. In addition, I argue that lack of validation does not necessarily signify irrelevance because validation assays each come with their own caveats, and their interpretation can be further complicated by mechanisms such as TF redundancy. Instead, results from many assays should be combined to generate increasingly comprehensive, high-quality network models.

Gene regulatory networks

GRNs are graph diagrams that depict interactions between genes and their regulators (such as TFs). These interactions can indicate a regulatory relationship, and/or can depict a physical interaction between a TF and a genomic DNA region associated with a particular gene. A genome-scale method used for inferring regulatory interactions is to computationally search transcriptomic data for correlations between gene and TF expression. This reverse engineering approach has been pioneered in yeast [8] and has also been applied to mammals [9,10]. ChIP starts with a protein and is an example of a TF-centered (protein-to-DNA) method, whereas Y1H starts with a DNA fragment and can be referred to as a gene-centered (DNA-to-protein) technique (reviewed in [6,7]). Both ChIP (for example, [11-16]) and Y1H assays (for example, [17-21]) have been successfully used in various systems and have each led to a wealth of data. Importantly, inferred regulatory relationships are not necessarily a result of direct physical interactions. Conversely, for

Table 1. Overview of commonly used techniques for gene regulatory network mapping and their advantages and limitations

Method	Characteristics	Advantages	Limitations
Computational inference (reverse engineering)	Infers putative regulatory relationships from gene expression data	Fast; cheap	The interactions predicted could be indirect (they do not have to reflect physical interactions); regulators that themselves do not change in expression will be missed; detection limits of mRNA measurements will affect GRN predictions (regulators or genes expressed at low levels will be missed)
Chromatin immunoprecipitation (ChIP)	Experimentally identifies physical interactions between TFs and DNA; TF-centered (protein-to-DNA)	*In vivo*; can detect TF dimers and complexes	Condition-dependent interactions can be missed; needs high-quality, highly specific ChIP-grade antibodies; when a universal tag is used for immunoprecipitation, TF is usually overexpressed; peak calling required
Yeast one-hybrid (Y1H)	Experimentally identifies physical interactions between DNA and TFs; gene-centered (DNA-to-protein)	Heterologous; condition- and context-independent	TFs that require post-translational modifications before binding DNA will be missed; not yet suitable for heterodimers

physical interactions between TFs and DNA, the regulatory consequence (repression or activation) is usually not known. Therefore, for optimal coverage and information content, both types of approaches need to be applied and integrated.

What is 'biologically meaningful'?

The quality of GRNs depends on the proportion of real interactions that are retrieved and the proportion of retrieved interactions that are real. For many scientists, interactions identified by high-throughput methods are deemed biologically meaningful (real) only if a regulatory and/or functional consequence is demonstrated, after which the interaction is considered 'validated *in vivo*'. However, a lack of *in vivo* validation does not necessarily invalidate an interaction because: (i) *in vivo* validation methods have their own limitations; (ii) a TF binding event might have been attributed to the wrong gene - for instance, when a TF binds an enhancer far from a transcription start site; and (iii) biological safety nets that buffer the loss of individual TFs can mask the effect of genuine DNA-TF interactions on gene expression.
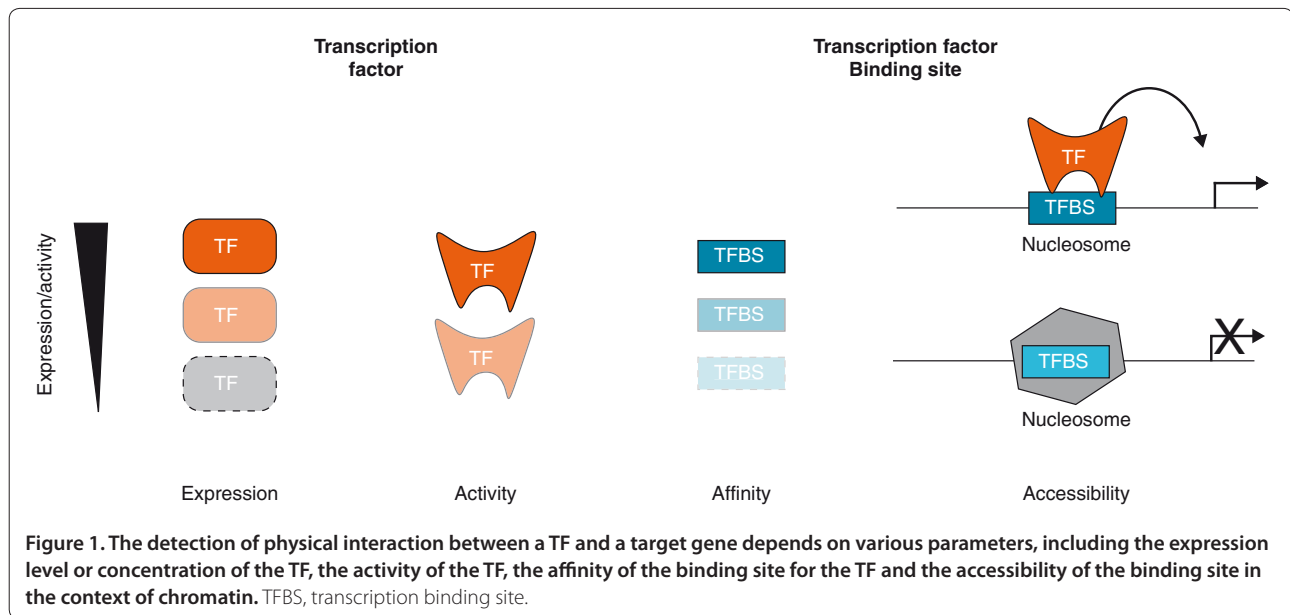
Data quality: false negatives

To obtain a complete picture of gene regulation, it is important to detect all physical and regulatory interactions that occur between genes and TFs. However, it is likely that the interaction networks that have been delineated so far are incomplete because not all interactions can be detected by the method(s) used. There are several reasons why DNA-TF interactions can be missed (false negatives; Table 1). In computationally inferred GRNs, relationships can be missed when the required cut-off for correlation was set too high, when TFs do not change in expression in accordance with their target genes, or when the TF or its target is expressed at very low levels, thereby disabling detection of expression

changes. With ChIP, the detection of interactions depends on the expression level, concentration and activity of the TF in the cell or tissue sampled, and the strength and accessibility of its binding sites (Figure 1). With ChIP-Seq, precipitated DNA fragments are sequenced, and bound regions are 'called' by compiling all the reads that correspond to particular genomic DNA regions into 'peaks'. Subsequently, cut-offs are selected somewhat arbitrarily to distinguish bound from unbound regions [22]. This will inevitably cause the strongest and/or most well-represented (robust) interactions to be considered at the expense of weaker interactions that might just as likely be biologically meaningful. Gene-centered Y1H assays also miss DNA-TF interactions. For instance, interactions with obligatory heterodimers cannot be detected with the current configurations of the assay. In addition, TFs that require post-translational modification or a cofactor in order to bind DNA may not be retrieved. When cDNA libraries are screened, low-abundance TFs have a high likelihood of being missed. This disadvantage has been partially alleviated by using directed Y1H assays in which TFs are tested one by one for their ability to bind to a particular DNA fragment [23]. Such clone-based assays, however, depend on clone resources such as the ORFeome [24,25], and TFs for which open reading frame clones are not available will obviously not be represented in these assays.

Data quality: false positives

False positives can be incorporated into GRNs and can be either technical or biological. Technical false positives are interactions that are sporadic in nature and cannot be repeated, even with the same assay with which they were originally retrieved. Obviously, any high-throughput method should avoid detecting spurious interactions by carefully optimizing and evaluating the robustness of the assay. Biological false positives are defined as interactions

Figure 1. The detection of physical interaction between a TF and a target gene depends on various parameters, including the expression level or concentration of the TF, the activity of the TF, the affinity of the binding site for the TF and the accessibility of the binding site in the context of chromatin. TFBS, transcription binding site.

that are robustly detected but that are not biologically meaningful. In computationally derived GRNs spurious edges (regulatory interactions) can arise if both a TF and its inferred target are regulated by another TF that itself does not change in expression. In ChIP experiments, biological false positives might be obtained when the antibody is not exclusively specific for the TF that is being studied, or when a TF is overexpressed (for instance, from a transgene), and starts binding to lower affinity or non-specific sites. Furthermore, selecting a threshold that is too low when 'calling' interactions from a background of non-interacting fragments may result in the inclusion of false interactions. Y1H assays can retrieve interactions that do not occur *in vivo* when a TF binding site is available in the context of yeast chromatin but not in the organism from which the DNA fragment was cloned. TF levels in yeast are controlled by a yeast promoter and by the copy number of the TF-expressing plasmid, and it is possible that lower affinity DNA sequences are bound when TF levels are high.

Five levels of validation
There are five conceptual levels of validation of interactions between genes and their regulators.

The first level is retesting interactions detected with the same experimental approach and reagents to minimize technical false positives. This can be done by retesting individual interactions, or by performing larger, genome-scale experiments multiple times. For example, in ChIP assays, the DNA regions deemed bound by a TF are often confirmed by quantitative PCR of the ChIPped DNA. However, this only confirms that the DNA fragment was precipitated; it is not a retest of the ChIP assay itself. In

Y1H assays, interactions retrieved can be confirmed in freshly grown yeast cells containing the 'DNA bait', using, for instance, a TF-encoding clone [26].

The second level is confirming an interaction with the same assay but using different reagents. Computationally inferred regulatory interactions can be assessed in an independent dataset, or with a different algorithm. ChIP interactions can be confirmed by using multiple antibodies to the same protein [15,27]. The signal-to-noise ratio can also be improved by including control experiments of samples in which the TF is removed or reduced [12]. In such experiments, DNA regions that are detected by ChIP both in wild-type and TF mutant or RNA interference (RNAi) samples can be considered false positives, and thresholds can be drawn accordingly. Y1H assays use two reporter genes that are integrated into different locations in the yeast genome, and only interactions that result in the activation of both reporters should be considered, as they are basically detected twice, and therefore confirmed. In addition, independent 'DNA bait' strains and 'TF preys' from different clone resources can be used to confirm interactions.

The third level is detecting an interaction with a different assay of the same type. For instance, to validate computationally inferred regulatory interactions one can use RNAi of a TF and examine changes in the expression of its inferred targets *in vivo*. This approach has recently been used for a subset of regulatory relationships inferred to control the pathogenic response of murine dendritic cells [10]. Physical interactions detected by Y1H assays can be confirmed by ChIP and vice versa. For example, we have confirmed multiple Y1H interactions in *Caenorhabditis elegans* and *Arabidopsis thaliana*

interactions by ChIP [17,21]. Additional support for physical interactions identified either by Y1H assays or by ChIP can be obtained by identifying a putative TF binding site within the DNA fragment bound, either using motif prediction algorithms (for example, [28-30]) or by interrogating large TF binding site datasets (for example, [31-34]). In Y1H assays, the putative site can be deleted and interactions with the mutant fragment can be examined. Loss of the DNA-TF interaction with the mutant fragment would confirm that the selected binding site is indeed correct [35].

The fourth level is observing an interaction in a different type of assay, that is, a regulatory interaction is confirmed by a physical interaction or vice versa. An example of how the regulatory effect of physical DNA-TF interactions can be examined in *C. elegans* is the generation of transgenic animals that express green fluorescent protein (GFP) under the control of the DNA fragment with which the physical interaction was detected, and subjection of these transgenic animals to RNAi of the relevant TF or crossing them into TF mutant animals [19,20,36]. An increase in GFP expression following reduction of the TF would indicate that the TF represses gene expression, whereas a decrease in GFP expression following TF reduction would indicate that the TF is an activator. Other methods used to determine target gene expression in mutant/RNAi animals include quantitative RT-PCR and expression profiling (for example, [17,37]).

The final level is observing that a TF and its target share functional roles in a biological process, or confer similar phenotypes when mutated and/or overexpressed. Such functional similarities can be uncovered by performing phenotypic experiments, or alternatively, correlations between functions can be investigated *in silico* using Gene Ontology [38] and KEGG databases [39]. It is likely that a complete correlation between a TF and its target genes in either a biological process or a phenotype occurs only in a minority of cases because networks are highly interconnected, and TFs often have multiple functions according to developmental, physiological or environmental circumstances.

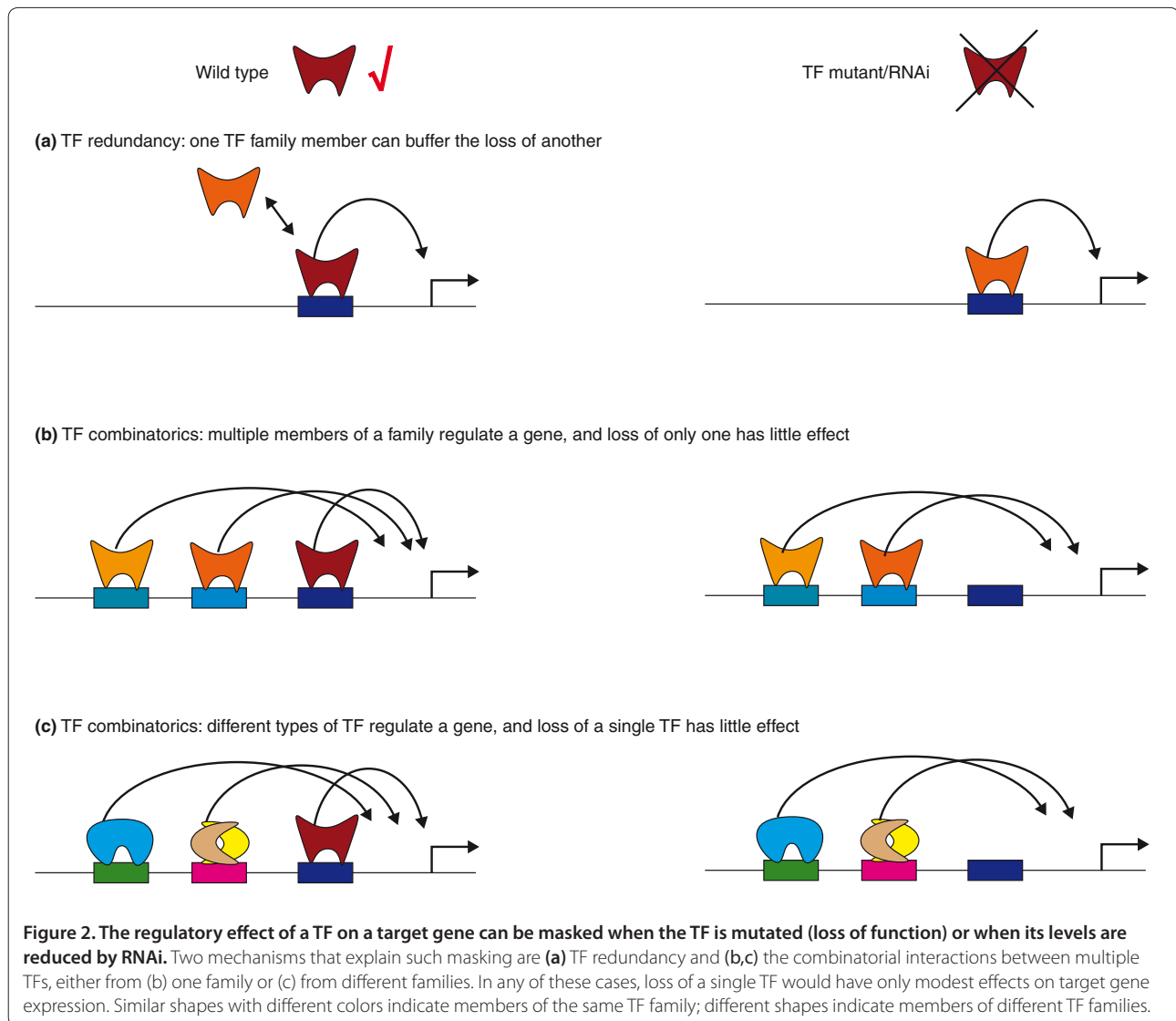## Limitations of DNA-TF validation: detection limits and interpretation

It is important to note that validation experiments are each subject to their own limitations and that a negative result need not invalidate the original interaction observed. For instance, when RNAi is used to examine regulatory relationships between a gene and its putative regulators, off-target effects can complicate the interpretation. Therefore, it is desirable to observe the same effect on target gene expression with multiple small interfering RNAs (siRNAs) designed to target the same regulator. In *C. elegans*, not every tissue is equally amenable

to RNAi; for instance, most neurons are largely refractory, and interactions occurring there would wrongly be deemed not valid when analyzed by RNAi [40,41]. When validating Y1H interactions by ChIP, it is important to perform the experiment using samples that not only express the TF of interest, but also under conditions where the TF is active (Figure 1). An example that illustrates this concept is *C. elegans* DAF-16, a forkhead TF that is located in the cytoplasm in wild-type animals unless they are exposed to nutritional or environmental stress. In *daf-2* mutants, however, DAF-16 aberrantly translocates to the nucleus and is constitutively active, and the *daf-2* mutant background has therefore been used to identify an initial set of DAF-16-bound regions [12]. Some other TFs could truly be bound to DNA sites *in vivo*, but be functionally dormant until activation by a ligand or signaling pathway, and regulatory effects would only be detectable under activation conditions. Prior knowledge of TF expression [42,43] or activity will greatly facilitate the design of an appropriate *in vivo* experiment. For instance, we focused on animals in the dauer stage (in which their development is arrested) to validate interactions between microRNA promoters and DAF-3, whose expression greatly increases in the dauer stage and which confers a dauer-related phenotype [19]. Finally, some TFs may affect the expression of their target genes in a tissue-specific manner, and some TFs can even function as both an activator and a repressor, depending on the circumstances. For instance, using a transgenic GFP approach, we found that the *C. elegans* nuclear hormone receptor NHR-45 activates the promoter of *nhr-178* in some tissues and under some conditions, whereas it represses it in other tissues under other conditions [20]. This example highlights the complexity of gene regulation and the caution one should take when interpreting 'whole animal' validation assays, such as quantitative RT-PCR.

Limitations of DNA-TF validation

In the past few years, it has become clear that many physical interactions - for instance, those identified by ChIP - do not convey a detectable regulatory consequence on their predicted target genes (for example, [44,45]). This could be because the regulatory consequences were tested under conditions in which the TF is not active. Alternatively, the wrong gene could have been attributed to the binding event, or the loss or reduction of individual TFs could have been masked by TF combinatorics or redundancy.

Both TF redundancy and combinatorics are prevalent in GRNs (Figure 2). For instance, multiple TFs from the same family can act through a single binding site and function redundantly, so that loss of one TF is compensated by another (Figure 2a). Such redundancy has

Wild type

TF mutant/RNAi

(a) TF redundancy: one TF family member can buffer the loss of another

(b) TF combinatorics: multiple members of a family regulate a gene, and loss of only one has little effect

(c) TF combinatorics: different types of TF regulate a gene, and loss of a single TF has little effect

Figure 2. The regulatory effect of a TF on a target gene can be masked when the TF is mutated (loss of function) or when its levels are reduced by RNAi. Two mechanisms that explain such masking are **(a)** TF redundancy and **(b,c)** the combinatorial interactions between multiple TFs, either from (b) one family or (c) from different families. In any of these cases, loss of a single TF would have only modest effects on target gene expression. Similar shapes with different colors indicate members of the same TF family; different shapes indicate members of different TF families.

been shown in various systems and for various different TF families, including mammalian ETS proteins that were studied by ChIP [46] and *C. elegans* FLYWCH-type zinc fingers that were found by Y1H assays [36]. Loss of a TF can also be masked by combinatorics involving TFs from the same family (Figure 2b) or different (Figure 2c) families, in which each TF contributes only a small regulatory effect, and this can lead to apparent redundancy. Such built-in redundancy may be very useful for critical genes that need to be buffered to avoid detrimental phenotypic consequences of TF loss [47].

Physical interactions between TFs and genomic DNA fragments are often inferred to affect the gene that is in closest linear proximity to the binding event. For interactions that occur close to a transcription start site, either in a gene promoter or just downstream in the transcribed region, this is a reasonable assumption. Interactions

involving more distant DNA elements (such as enhancers), however, may not necessarily affect a nearby gene because the genome is not organized as a linear polymer, but rather is organized in three dimensions [48]. Studies using chromatin conformation capture techniques [49] are providing insights into which genomic regions contact each other to form loops, potentially to bring together enhancers and gene promoters. Integrating physical and regulatory interaction data with structural genome and DNA looping data will facilitate the further dissection of GRNs.

Conclusions

When is an interaction between a genomic DNA fragment and a TF biologically meaningful? When the gene located closest to the binding event changes in expression following loss or reduction of the TF? When

the gene and the TF have the same phenotype? Clearly, the answers are not simple, and the assays used to validate interactions are not foolproof. However, when the methods used for GRN delineation are robust and technically sound, individual interactions should not be discarded when one or more validation methods fail to detect regulatory or phenotypic consequences, particularly in complex, whole organisms. Rather, multiple methods need to be combined into increasingly integrated networks to attribute different degrees of confidence to each of the interactions observed.

Acknowledgements

References

1. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution.** *Nat Rev Genet* 2009, **10**:252-263.
2. Margolin AA, Califano A: **Theory and limitations of genetic network inference from microarray data.** *Ann NY Acad Sci* 2007, **1115**:51-72.
3. He F, Balling R, Zeng A-P: **Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives.** *J Biotechnol* 2009, **144**:190-203.
4. Park PJ: **ChIP-seq: advantages and challenges of a maturing technology.** *Nat Rev Genet* 2009, **10**:669-680.
5. Collas P: **The current state of chromatin immunoprecipitation.** *Mol Biotechnol* 2010, **45**:87-100.
6. Arda HE, Walhout AJ: **Gene-centered regulatory networks.** *Brief Funct Genomics* 2009, **9**:4-12.
7. Walhout AJM: **Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping.** *Genome Res* 2006, **16**:1445-1454.
8. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.** *Nat Genet* 2003, **34**:166-176.
9. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering of regulatory networks in human B cells.** *Nat Genet* 2005, **37**:382-390.
10. Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, Schubert LA, Birditt B, Shay T, Goren A, Zhang X, Smith Z, Deering R, McDonald RC, Cabili M, Bernstein BE, Rinn JL, Meissner A, Root DE, Hacohen N, Regev A: **Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses.** *Science* 2009, **326**:257-263.
11. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**:99-104.
12. Oh SW, Mukhopadhyay A, Dixit BL, Raha T, Green MR, Tissenbaum HA: **Identification of direct targets of DAF-16 controlling longevity, metabolism and diapause by chromatin immunoprecipitation.** *Nat Genet* 2006, **38**:251-257.
13. Odom DT, Dowell RD, Jacobsen ES, Nekl.udova L, Rolfe PA, Danford TW, Gifford DK, Fraenkel E, Bell GI, Young RA: **Core transcriptional regulatory circuitry in human hepatocytes.** *Mol Syst Biol* 2006, **2**:2006.0017.
14. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M: **Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo.** *Genes Dev* 2007, **21**:385-390.
15. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE: **Combinatorial

16. Tsukagoshi H, Busch W, Benfey PN: **Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root.** *Cell* 2010, **143**:606-616.
17. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, Tissenbaum HA, Mango SE, Walhout AJ: **A gene-centered *C. elegans* protein-DNA interaction network.** *Cell* 2006, **125**:1193-1205.
18. Vermeirssen V, Barrasa MI, Hidalgo C, Babon JAB, Sequerra R, Doucette-Stam L, Barabasi AL, Walhout AJM: **Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network.** *Genome Res* 2007, **17**:1061-1071.
19. Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros V, Walhout AJM: **A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity.** *Genes Dev* 2008, **22**:2535-2549.
20. Arda HE, Taubert S, Conine C, Tsuda B, Van Gilst MR, Sequerra R, Doucette-Stam L, Yamamoto KR, Walhout AJM: **Functional modularity of nuclear hormone receptors in a *C. elegans* gene regulatory network.** *Mol Syst Biol* 2010, **6**:367.
21. Brady SM, Zhang L, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeples M, Kim D, Ahnert S, Ohler U, Ware D, Walhout AJ, Benfey PN: **A stele-enriched gene regulatory network in the *Arabidopsis* root.** *Mol Syst Biol* 2011, **7**:459.
22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
23. Vermeirssen V, Deplancke B, Barrasa MI, Reece-Hoyes JS, Arda HE, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Brent MR, Walhout AJ: **Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping.** *Nat Methods* 2007, **4**:659-664.
24. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, Moore T, Hudson JR Jr, Hartley JL, Brasch MA, Vandenhaute J, Boulton S, Endress GA, Jenna S, Chevet E, Papasotiropoulos V, Tolias PP, Ptacek J, Snyder M, Huang R, Chance MR, Lee H, Doucette-Stamm L, Hill DE, Vidal M: ***C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression.** *Nat Genet* 2003, **34**:35-41.
25. Rual JF, Hirozane-Kishikawa T, Hao T, Bertin N, Li S, Dricot A, Li N, Rosenberg J, Lamesch P, Vidalain PO, Clingingsmith TR, Hartley JL, Esposito D, Cheo D, Moore T, Simmons B, Sequerra R, Bosak S, Doucette-Stamm L, Le Peuch C, Vandenhaute J, Cusick ME, Albala JS, Hill DE, Vidal M: **Human ORFeome version 1.1: a platform for reverse proteomics.** *Genome Res* 2004, **14**:2128-2135.
26. Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJ: **Gateway-compatible yeast one-hybrid screens.** *Cold Spring Harb Protoc* 2006, **14**:2093-2101.
27. Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, Furlong EE: **A core transcriptional network for early mesoderm development in *Drosophila melanogaster*.** *Genes Dev* 2007, **21**:436-449.
28. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z: **Detection of functional DNA motifs via statistical over-representation.** *Nucleic Acids Res* 2004, **32**:1372-1381.
29. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic Acids Res* 2006, **34**:W369-W373.
30. Stormo GD: **Motif discovery using expectation maximization and Gibbs' sampling.** *Methods Mol Biol* 2010, **674**:85-95.
31. Newburger DE, Bulyk ML: **UniPROBE: an online database of protein binding microarray data on protein-DNA interactions.** *Nucleic Acids Res* 2009, **37**:D77-D82.
32. Grove CA, deMasi F, Barrasa MI, Newburger D, Alkema MJ, Bulyk ML, Walhout AJ: **A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors.** *Cell* 2009, **138**:314-327.
33. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML: **Diversity and complexity in DNA recognition by transcription factors.** *Science* 2009, **324**:1720-1723.
34. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: **TRANSFAC and its module TRANSCompel:**

binding predicts spatio-temporal *cis*-regulatory activity. *Nature* 2009, **462**:65-70.

transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006, **34**:D108-D110.

35. Reece-Hoyes JS, Deplancke B, Barrasa MI, Hatzold J, Smit RB, Arda HE, Pope PA, Gaudet J, Conradt B, Walhout AJ: **The** *C. elegans* **Snail homolog CES-1 can activate gene expression** *in vivo* **and share targets with bHLH transcription factors.** *Nucleic Acids Res* 2009, **37**:3689-3698.

36. Ow MC, Martinez NJ, Olsen P, Silverman S, Barrasa MI, Conradt B, Walhout AJM, Ambros VR: **The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress embryonic expression of microRNA genes in** *C. elegans*. *Genes Dev* 2008, **22**:2520-2534.

37. Murphy CT, McCarroll SA, Bargmann CI, Fraser A, Kamath RS, Ahringer J, Li H, Kenyon C: **Genes that act downstream of DAF-16 to influence the lifespan of** *Caenorhabditis elegans*. *Nature* 2003, **424**:277-283.

38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology.** The Gene Ontology Consortium. *Nat Genet* 2000, **25**:25-29.

39. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M: **KEGG for representation and analysis of molecular networks involving diseases and drugs.** *Nucleic Acids Res* 2010, **38**:D355-360.

40. Timmons L, Court DL, Fire A: **Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in** *Caenorhabditis elegans*. *Gene* 2001, **263**:103-112.

41. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J: **Systematic functional analysis of the** *Caenorhabditis elegans* **genome using RNAi.** *Nature* 2003, **421**:231-237.

42. Reece-Hoyes JS, Shingles J, Dupuy D, Grove CA, Walhout AJ, Vidal M, Hope IA: **Insight into transcription factor gene duplication from** *Caenorhabditis elegans* **promoterome-driven expression patterns.** *BMC Genomics* 2007, **8**:27.

43. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, *et al*: **An atlas of combinatorial transcriptional regulation in mouse and man.** *Cell* 2010, **140**:744-752.

44. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD: **Transcription factors bind thousands of active an inactive regions in the** *Drosophila* **blastoderm.** *PLoS Biol* 2008, **6**:e27.

45. Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, Tapscott SJ: **Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming.** *Dev Cell* 2010, **18**:662-674.

46. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ: **Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the** *ETS* **gene family.** *Genes Dev* 2007, **21**:1882-1894.

47. MacNeil L, Walhout AJ: **Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression.** *Genome Res* 2011, in press 10.1101/gr.097378.109.

48. Dekker J: **Gene regulation in the third dimension.** *Science* 2008, **319**:1793-1794.

49. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation.** *Science* 2002, **295**:1306-1311.