

MIT Open Access Articles

*Transcriptional Regulatory Circuits:
Predicting Numbers from Alphabets*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Kim, Harold D., Tal Shay, Erin K. O'Shea, and Aviv Regev. "Transcriptional Regulatory Circuits: Predicting Numbers from Alphabets." *Science* 2009 July 24; 325(5939): 429–432.

As Published: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745280/>

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: <http://hdl.handle.net/1721.1/84942>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





Published in final edited form as:

Science. 2009 July 24; 325(5939): 429–432. doi:10.1126/science.1171347.

Transcriptional Regulatory Circuits: Predicting Numbers from Alphabets

Harold D. Kim^{1,*}, Tal Shay^{2,*}, Erin K. O'Shea¹, and Aviv Regev^{2,†}

¹Howard Hughes Medical Institute, Harvard University Faculty of Arts and Sciences Center for Systems Biology, Departments of Molecular and Cellular Biology and Chemistry and Chemical Biology, Cambridge, MA 02138, USA

²Department of Biology, Massachusetts Institute of Technology, and Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Abstract

Transcriptional regulatory circuits govern how cis and trans factors transform signals into messenger RNA (mRNA) expression levels. With advances in quantitative and high-throughput technologies that allow measurement of gene expression state in different conditions, data that can be used to build and test models of transcriptional regulation is being generated at a rapid pace. Here, we review experimental and computational methods used to derive detailed quantitative circuit models on a small scale and cruder, genome-wide models on a large scale. We discuss the potential of combining small- and large-scale approaches to understand the working and wiring of transcriptional regulatory circuits.

The next frontier in genomics is to assemble systematically the functional components of genomes and cells into the circuits that transform signals into cellular responses. These circuits include signal transduction, metabolic, and transcriptional pathways. One major challenge is to predict, in a given cell type and under a given condition, the expression level of every gene. This view focuses on regulatory circuits that take trans regulators and cis sequence determinants as input and yield gene expression level as output (Fig. 1A).

Small-scale approaches have been used to develop detailed quantitative models of the regulatory circuits controlling one or very few genes. Complementary large-scale approaches used computational algorithms to reconstruct genome-scale circuits; these methods typically result in comprehensive, albeit less-detailed, models. These two strategies already inform and influence each other (1,2) and are likely to merge and deliver quantitative and biochemically interpretable genome-wide models of circuit wiring (3).

Small-Scale Approaches: Success Stories from Prokaryotes

Small-scale measurements focus on a single circuit and its cis regulation, typically by using thermodynamics or kinetic models as a framework for interpreting data. Inputs and outputs for several regulatory circuits have been measured in vivo on a small scale in prokaryotes (4–7). Inputs include inducers that modulate the activity of the trans factor (6,7), or the trans factor itself (4,8,9). Such studies rely on the availability of a transcription factor (TF) that binds a well-characterized promoter, with activity that can be continuously tuned in a controlled manner. The output is typically represented by the rate of protein production (4,6), protein

[†]To whom correspondence should be addressed. aregev@broad.mit.edu.

^{*}These authors contributed equally to this work.

activity (6,7), or steady-state protein level (8,9) and is measured by means of a reporter, such as a fluorescent protein gene or gene encoding an enzyme (whose activity serves as the reporter) that is placed under the control of the promoter of interest (4,10). The input-output relation, termed the gene-regulation function (4), is generally measured from a population of cells, although the same approaches apply to single-cell measurements.

Direct measurements, along with prior knowledge about the molecular system, are used to infer a quantitative model of gene regulation (5). The gene-regulation function is often approximated by the Hill function (Fig. 1B) which summarizes threshold, sensitivity, and maximum expression level. To understand the relation between phenomenological Hill parameters and the actual biochemical parameters of gene regulation, one needs to know whether TFs oligomerize, how strongly they interact with cis elements in the promoter region, whether they interact cooperatively with themselves or other factors when binding to the promoter, and how they interact with RNA polymerase (5). Numerous studies of prokaryotic systems, such as the *lac* and lambda promoters (11,12), have provided insights into the biochemical basis for the Hill parameters of single-input gene-regulation functions (7).

Quantitative models have been built for more complex circuits as well, such as a circuit based on the *lac* promoter that can perform simple computations such as AND and OR logic (6,10). Circuit models can be used in a bottom-up approach to predict the behavior of more-complex regulatory networks, as demonstrated in synthetic gene networks in *Escherichia coli* (3). Overall, the agreement between data and models for prokaryotic gene expression can be attributed to the tractability of the system and the accumulated knowledge of the relevant biochemistry and biophysics.

There are several ways to validate and refine such models. Researchers can compare a biochemical parameter extracted from the model (e.g., the equilibrium dissociation constant between a TF and the DNA) with a corresponding in vitro measurement. Another approach requires perturbing the TF binding equilibrium, by varying the DNA sequence of cis elements and observing whether the data fit the model. Such measurements also offer insights into the plasticity and evolution of the regulatory circuit (10). The model can be further constrained by measuring additional variables, such as noise in gene expression (13).

When a model fails to be validated by the data, we can discover previously unknown events or components that likely are associated with the regulatory circuit, thus correctly identifying the circuit's wiring. For example, a smaller-than-expected maximal expression level suggests the existence of another pathway or factor (7) that can limit binding of the TF to the promoter or modulate its activity post binding. A higher-than-expected sensitivity might point to hidden mechanisms of cooperativity, such as physical interactions between TFs, or to indirect interactions that are mediated by other factors (7,14). Circuits connected in series can also lead to a higher overall sensitivity (15). Moreover, hidden feedback loops can dramatically influence the shape of the gene-regulation function (16).

Small-Scale Approaches to Eukaryotic Gene-Regulation Functions

Eukaryotic systems have regulatory circuits that are more complex than those of prokaryotes. For example, eukaryotic cis elements are not equally accessible to TFs because nucleosomes generally hinder binding of TFs. Moreover, nucleosomes are not passive inhibitors of TFs but are removed from the DNA by adenosine triphosphate-dependent chromatin-remodeling factors. Furthermore, nucleosomes are bound with a different affinity to the DNA, depending on their chemical modification and underlying DNA sequence (17).

The eukaryotic cis regulatory code is also more diverse and has many types of binding sites for multiple TFs (18). Additionally, complexities on the trans side can affect the activity of a

TF as a function of location and posttranslational modification, posing both measurement and modeling challenges. Moreover, sequence-specific TFs interact with various chromatin-remodeling factors and histone-modifying complexes, but we have limited knowledge about the identity of these factors and their quantitative effect on gene regulation. Thus, even for well-studied model eukaryotic promoters, we lack a full understanding of the components, wiring, and biochemical interactions in the circuit.

Gene expression levels have been modeled according to promoter architecture in yeast (8, 19–21), worms (22), and sea urchins (23). Yeast gene expression levels were measured from both synthetic and genomic promoters varying in composition and organization of TF-binding sites among different environments and could be successfully modeled on the basis of the equilibrium binding of TFs to DNA and to each other (19,20). Other studies underscore the importance of nucleosomes in determining the level of gene expression (24). For example, a study testing the gene-regulation functions of phosphate-regulated promoters in yeast, by using different combinations of cis elements, identified a role for nucleosomes in decoupling threshold and maximum expression levels of the gene-regulation function (8).

Large Scale: From Single Gene to Genome-Wide Models

Small-scale models focus on deciphering regulatory functions for circuits where the key components and wiring were previously known. In contrast, large-scale approaches tend to have less precision but can be used to infer circuit components and wiring. Such efforts are particularly critical when studying gene regulation in higher organisms, with largely uncharacterized circuits.

The emergence of large-scale approaches that aim for genome-wide prediction of output has been tightly coupled to an improved ability to measure circuit input, output, and wiring on a genomic scale. For example, output such as mRNA levels can be measured by microarrays or sequencing (25). Also, as an example of cis input, we now can more easily detect promoter sequences because of advances in computation and the availability of whole-genome sequences (26). Trans inputs can be measured as direct interactions of DNA with proteins, including TFs (18,27), histone-modification states (28), and nucleosome positions in vivo (29). TF-promoter interactions also can be inferred from in vitro assays, such as protein-binding microarrays (30).

Large-scale genomic studies of regulatory circuits are still limited, most notably in measurements of the abundance and activity of trans input, as is direct manipulation of cis and trans inputs. As a result, most large-scale models do not incorporate detailed regulatory functions and either address cis inputs or trans inputs but not both. Furthermore, the generalizing of the circuit model across the promoters and expression levels of multiple genes comes at the cost of limited power to explain the observed expression of any individual gene (31).

Genome-Wide Reconstruction of Cis Regulatory Functions

Genome-wide sequencing and profiling efforts have spurred the development of numerous approaches to reconstruct cis regulatory functions that explain observed expression levels by the type, number, and organization of cis regulatory elements in promoters. Linear, Bayesian, and thermodynamic approaches have been used and each reflects a different set of assumptions on the biochemical basis of transcriptional regulation (Fig. 1B). Linear models assume that the expression output is a linear combination of cis-element inputs (32–34). Probabilistic Bayesian approaches can handle the noisy nature of large-scale data and are able to capture the combinatorial logic and organization of promoters by modeling combinations of motifs, as well as their relative distance and orientation (35). Some studies cast a linear model in a

probabilistic setting, combining the benefit of both approaches (31). More realistic thermodynamic models have recently emerged. For example, expression patterns in *Drosophila* segmentation were predicted by calculating the probabilities of all possible configurations of trans factors on the cis regulatory sequence and summing their contributions to expression (36). Sequence preferences for nucleosomes and transcription factors were used to predict expression in yeast (37). However, thermodynamic models have not yet been applied on a genomic scale to test their general power. Genome-wide inference of cis regulatory functions critically depends on accurate and comprehensive detection of cis elements in DNA sequences—this is a notoriously difficult problem, especially in higher organisms.

Estimating the success of the models in predicting gene expression is a challenging task. Ideally, large-scale models should be trained on one data set and then tested for their ability to generalize to unseen data. However, most data sets are of limited scale, and systematic experimental follow-up is lacking. Indeed, many studies do not report such objective success rates, whereas others use various measures to assess the quality of their prediction. For example, models that predict module assignment for each gene typically report the percentage of correct assignment of coregulated genes or genes that share the same function with success rate reported from ~30% (38) to 73% (35) in studies in yeast. Other works report the likelihood of the data given the model (39,40), but this measure is hard to compare between models of different complexity. The most quantifiable success level to date is reported for models that predict the actual expression of each gene; they typically report the percentage of the variance in gene expression that is accounted for by the model. Both Bayesian and linear models that predict gene expression from cis regulatory sequences have reported high rates of success in yeast [e.g., 51% in (35), 52 to 72% in (33)]. However, the success of similar approaches in mammalian cells has been much more modest [e.g., 6% in (31), 11 to 24% in (32)]. It is important to consider the amount of expression that we can expect to explain with a particular model. A recent study with synthetic promoter libraries in yeast estimated that cis regulation can explain at most 65% of the variance in expression and that a thermodynamic small-scale model explains 44 to 59% (19). Similar work in the urochordate *Ciona* explains 30 to 89% of the variance at the cis level (22). Establishing standard approaches and data sets on which we can compare the performance of different models is an important goal. The DREAM project aims to achieve such a fair comparison by posting challenges for the community (42).

Discovering Arrows: Inferring Trans Regulation on a Genomic Scale

Cis input alone cannot predict expression, because dynamic expression patterns change with environmental conditions, cell type, and cell cycle stage. These changes are accompanied by corresponding changes in trans factors, which have been studied by three genome-scale strategies.

Genome-scale measurement of physical interactions allows direct incorporation of trans factors into circuit models. For example, TF-DNA interactions, measured either *in vivo* (18,41) or *in vitro* (43), were used to derive a detailed map of TF binding in yeast. Such maps can then be coupled with measurements of expression output to build a model of regulatory circuits (44). Discrepancies between the conditions in which binding and expression are measured can limit such efforts. Conversely, measurement of TF binding before and after a stimulus can help distinguish between direct and indirect targets (1,18,45). Notably, the scale and quality of measurements of trans inputs are still limited by the required effort and cost associated with generating the needed reagents and data (e.g., generating an epitope-tagged version of every TF and using it in chromatin immunoprecipitation across many conditions).

Alternatively, the level, activity, and wiring of trans factors can be inferred from mRNA levels because many trans regulators are embedded within transcriptional feedback regulatory loops

(46,47); this results in detectable changes in their mRNA levels. Such inference is needed because we currently cannot assay the activities of a large number of regulators in parallel. Translational and post-translational regulations can create a substantial gap between the mRNA level of a trans factor and the level of active protein, but nonetheless, such approaches have been successful in *E. coli*, yeast, mouse, and human (40,48–50). In one approach, temporal expression data were collected to characterize processes such as cell differentiation and responses to environmental stimuli in mammalian systems (46,51), which showed that the transcriptional program was propagated by sequential waves of transcription controlled by different TFs. However, inferring regulatory activity from output levels limits the model's ability to distinguish between causality and correlation.

The perturbation of regulators can also be used to decipher and validate trans-regulation function. For example, 263 TFs were systematically deleted in yeast, and each deletion strain was compared with the wild type for genome-wide expression (52). RNA interference technology, which can target expression of specific genes and RNAs, has enabled similar approaches in mammalian cells, albeit at a more limited scale (31). Furthermore, epistasis analysis can be used to infer fine details of a multi-input circuit, as demonstrated by comparing strains from which one, two, or three trans regulators were deleted for their genome-wide effect on the yeast osmotic stress response (1).

Integrating Cis and Trans Models at a Genomic Scale

The obvious next direction is to systematically measure and model both trans factors and cis regulatory components simultaneously in different conditions (51). However, we are limited in our ability to finely perturb or measure cis and trans inputs to the circuits because of cost, a relative paucity of genetic tools in higher organisms, and the lack of synthetic approaches required to generate promoter-sequence variants (19).

An alternative strategy to engineered perturbations is to focus on the natural genetic polymorphisms underlying variation in gene expression between strains and species. Regulatory circuits that explain expression differences between natural genetic variants in a population or expression quantitative trait loci have been reconstructed on the basis of genetic variation in cis and trans factors (53,54). This approach is particularly appealing as it incorporates the collective effects of subtle perturbations to multiple circuit components and uses the power of genetic linkage and association to determine causality. Probabilistic and linear methods, in combination with linkage analysis, have incorporated genotypes as predictors to infer trans factors and regulatory networks in yeast, mouse, and human (55–57).

Similarly, studying expression in hybrids can exploit the genetic differences between two related species to estimate the relative cis and trans contributions to regulation as observed in yeast and fly (54,58). In yeast, basal differences in expression between species are explained by cis changes, whereas changes in expression regulation between different environmental conditions are associated with trans changes (58). Related studies have also been performed in mouse cells into which a human chromosome was introduced, showing that cis effects are stronger than trans effects (59).

Validation poses a particular challenge for all large-scale studies. Specific predictions can be tested by small-scale approaches, such as manipulation of individual trans factors or cis elements followed by genome-wide profiling [e.g., (31,47)]. However, these strategies are typically used to validate only a small portion of the model (“cherry picks”) and focus on the system's wiring rather than its quantitative aspects. In rare examples, large-scale validation is conducted by comparing a model's prediction (e.g., inferred regulation from cis data) with independent genome-wide data [e.g., TF-DNA binding for all TFs (18)]. In most cases, however, model validation is modest, and few of the model's components are tested. The two

key challenges are, thus, how to experimentally test predictions on a genomic scale and how to incorporate the results of these experiments once collected. One immediate use is as test data, on which to check how well the model generalizes. A more ambitious goal is to use the data for the iterative refinement of the model.

The detailed genetic and molecular manipulation needed to study circuits by small-scale approaches is extremely challenging in higher organisms. For example, genetic deletion of trans factors is about 100 times as expensive and time-consuming in mammals as in yeast and bacteria, and promoter manipulation is similarly challenging. Because many trans factors are pleiotropic, their deletion is lethal, and hence, requires more sophisticated approaches. Small interfering RNA has opened the possibility of manipulation by silencing (knockdown), but it is prone to nonspecific off-target effects that are not yet fully understood. Future characterization of these nonspecific effects will greatly promote the study of mammalian transcriptional circuits. The ability to perturb multiple proteins (combinatorial knockdown) is limited in mammalian cells. The introduction of morpholino oligonucleotides microinjection in developing embryos of zebrafish, sea urchin, and *Xenopus* (60) has been very successful, but this approach has not been established in mammalian systems.

The most comprehensive and successful model in animals, thus far, has been reconstructed in the sea urchin, including a complete model of the gene regulatory network that specifies the skeletogenic micromere lineage (61). In this model organism, systematic manipulation and measurement of trans factors, cis elements in promoters, and mRNA output measures were combined to devise a detailed validated model of gene regulation. The ability to use morpholine-based knockdown, promoter engineering methods, and mRNA in situ measurements substantially contributed to this success.

Combining Lessons from Small- and Large-Scale Approaches

Despite advances from small-scale and large-scale analysis of gene regulation, most studies do not yet bridge the gap between these approaches. As previously noted, small-scale approaches can generate fine, realistic details and extensive validation, but are limited to a few genes (often one). However, large-scale approaches examine many genes, but often rely on regulation functions that are biologically unrealistic (e.g., Boolean logic or linearity) and lack validation.

Recent studies highlight the promise of combining the power of both approaches. For example, genetic approaches typical of smaller-scale studies combined with linear modeling and mRNA profiling have allowed detailed quantitative reconstruction of the regulatory circuits controlling the response to high osmolarity in yeast (1). In addition, findings from large-scale studies can be followed up in detail with small-scale approaches. For instance, the behavior of a yeast regulon observed in mRNA profiles was explained by small-scale studies on a single input circuit controlling a representative target gene (2). Bayesian network approaches, originally developed to study mRNA profiles (48), have been successfully applied to single-cell measurements of signaling pathways (62).

Indeed, the distinction between small- and large-scale approaches is rapidly blurring, with recent technological advances enabling the experimental toolbox of small-scale approaches to be applied to the genomic level. These include the ability to profile multiple genes in single cells even during multicellular organism development (63), large-scale manipulation of sequences (allowing engineering and perturbation of promoters and TFs), and large-scale tunable trans perturbation, including that in higher organisms (31). As such large-scale perturbations of cellular components become increasingly feasible, the number of samples that need to be genomically profiled becomes exceptionally high. This raises the need for genome-scale profiling methods that are at least two orders of magnitude cheaper and faster and require

substantially lower cell numbers and per-sample effort than, for example, a microarray. In this respect, application of very recent advances in assays for measurement of smaller gene signatures (of several hundred genes) at low cost and high scale (64,65) will be essential for both the generation and validation of large-scale models of gene regulation. Overall, such improved methods will allow construction of large-scale models of gene-regulation functions that are closer to the realistic models used in small-scale systems.

Another direction for a major advance is coupling such approaches with time-series experiments, which will allow the study of dynamic gene-regulation functions. Many current studies rely on steady-state measurements, but transcriptional responses unfold over time in response to environmental changes, during differentiation, and in disease. Furthermore, temporal relations can help distinguish correlation from causality. Most computational methods have not leveraged this power, but recent studies (31,51) have demonstrated its potential. A useful stepping stone will be to construct well-defined subnetworks (“toy-models”), and then to study their regulation and dynamics (20,66) as benchmarks to compare models' performance.

The fundamental challenge for all approaches may be building models that truly generalize to novel states. Almost all current models predict expression in the studied conditions or very similar ones. As the space of possible states may be very large, how successful we can be may depend on our understanding of the underlying features of nonlinearity and combinatorial effects in gene-regulation functions.

In this review we only considered the circuits regulating transcript levels, where the inputs are cis sequences and trans factors that affect transcription directly (e.g., TFs, nucleosomes, or chromatin modifications), and the output is transcript level. However, these circuits are part of a much larger and more complex cellular network composed of many other interacting components. For example, transcripts alone are directly affected by editing, chemical modification, RNA binding proteins, and other noncoding RNAs, which affect their rate of degradation, accessibility for translation, and rate of translation to proteins. More generally, the transcriptional circuits are tightly coupled to signaling, metabolic, and localization systems that are part of the complex three-dimensional organization of cells and organisms. It is the way in which this complex system processes information and executes functions that ultimately determines the phenotype.

Acknowledgments

We gratefully acknowledge the assistance of I. Amit. H.D.K. was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface. E.K.O. was supported by HHMI and NIH GM51377. A.R. was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface and by NIH DP1-OD003958-01.

References and Notes

1. Capaldi AP, et al. *Nat Genet* 2008;40:1300. [PubMed: 18931682]
2. Cai L, Dalal CK, Elowitz MB. *Nature* 2008;455:485. [PubMed: 18818649]
3. Guido NJ, et al. *Nature* 2006;439:856. [PubMed: 16482159]
4. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. *Science* 2005;307:1962. [PubMed: 15790856]
5. Bintu L, et al. *Curr Opin Genet Dev* 2005;15:125. [PubMed: 15797195]
6. Setty Y, Mayo AE, Surette MG, Alon U. *Proc Natl Acad Sci U S A* 2003;100:7702. [PubMed: 12805558]
7. Kuhlman T, Zhang ZG, Saier MH, Hwa T. *Proc Natl Acad Sci U S A* 2007;104:6043. [PubMed: 17376875]
8. Kim HD, O'Shea EK. *Nat Struct Mol Biol* 2008;15:1192. [PubMed: 18849996]

9. Becskei A, Kaufmann BB, van Oudenaarden A. *Nat Genet* 2005;37:937. [PubMed: 16086016]
10. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. *PLoS Biol* 2006;4:555.
11. Wall ME, Hlavacek WS, Savageau MA. *Nat Rev Genet* 2004;5:34. [PubMed: 14708014]
12. Johnson AD, et al. *Nature* 1981;294:217. [PubMed: 6457992]
13. Thattai M, van Oudenaarden A. *Proc Natl Acad Sci U S A* 2001;98:8614. [PubMed: 11438714]
14. Wade JT, Belyaeva TA, Hyde EI, Busby SJW. *EMBO J* 2001;20:7160. [PubMed: 11742992]
15. Hooshangi S, Thiberge S, Weiss R. *Proc Natl Acad Sci U S A* 2005;102:3581. [PubMed: 15738412]
16. Alon U. *Nat Rev Genet* 2007;8:450. [PubMed: 17510665]
17. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. *Nat Struct Mol Biol* 2007;14:1025. [PubMed: 17984965]
18. Harbison CT, et al. *Nature* 2004;431:99. [PubMed: 15343339]
19. Gertz J, Siggia ED, Cohen BA. *Nature* 2009;457:215. [PubMed: 19029883]
20. Ellis T, Wang X, Collins JJ. *Nat Biotechnol* 2009;27:465. [PubMed: 19377462]
21. Gertz J, Cohen BA. *Mol Syst Biol* 2009;5:244. [PubMed: 19225457]
22. Brown CD, Johnson DS, Sidow A. *Science* 2007;317:1557. [PubMed: 17872446]
23. Yuh CH, Bolouri H, Davidson EH. *Science* 1998;279:1896. [PubMed: 9506933]
24. Lam FH, Steger DJ, O'Shea EK. *Nature* 2008;453:246. [PubMed: 18418379]
25. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. *Nat Methods* 2008;5:621. [PubMed: 18516045]
26. D'haeseleer P. *Nat Biotechnol* 2006;24:959. [PubMed: 16900144]
27. Johnson DS, Mortazavi A, Myers RM, Wold B. *Science* 2007;316:1497. [PubMed: 17540862]
28. Mikkelsen TS, et al. *Nature* 2007;448:553. [PubMed: 17603471]
29. Kaplan N, et al. *Nature* 2009;458:362. [PubMed: 19092803]
30. Berger MF, Bulyk ML. *Methods Mol Biol* 2006;338:245. [PubMed: 16888363]
31. The FANTOM Consortium and Riken Omics Science Center. *Nat Genet* 2009;41:553. [PubMed: 19377474]
32. Das D, Nahlé Z, Zhang MQ. *Mol Syst Biol* 2006;2:2006.0029. [PubMed: 16760900]
33. Nguyen DH, D'haeseleer P. *Mol Syst Biol* 2006;2:2006.0012. [PubMed: 16738557]
34. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. *Science* 2003;301:102. [PubMed: 12843395]
35. Beer MA, Tavazoie S. *Cell* 2004;117:185. [PubMed: 15084257]
36. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. *Nature* 2008;451:535. [PubMed: 18172436]
37. Raveh-Sadka T, Levo M, Segal E. *Genome Res.* 10.1101/gr.088260.108 published online 18 May 2009, in press
38. Segal E, Yelensky R, Koller D. *Bioinformatics* 2003;19:i273. [PubMed: 12855470]
39. Segal, E.; Barash, Y.; Simon, I.; Friedman, N.; Koller, D. *RECOMB '02: Proceedings of the Sixth Annual International Conference on Computational Biology*; Washington, DC. 18 to 21 April 2002; New York: Association for Computing Machinery; 2002. p. 263-272.
40. Pe'er D, Tanay A, Regev A. *J Mach Learn Res* 2006;7:167.
41. Gao F, Foat B, Bussemaker H. *BMC Bioinformatics* 2004;5:31. [PubMed: 15113405]
42. Stolovitzky G, Monroe D, Califano A. *Ann N Y Acad Sci* 2007;1115:1. [PubMed: 17925349]
43. Zhu C, et al. *Genome Res* 2009;19:556. [PubMed: 19158363]
44. Ye C, Galbraith SJ, Liao JC, Eskin E. *PLOS Comput Biol* 2009;5:e1000311. [PubMed: 19300475]
45. Workman CT, et al. *Science* 2006;312:1054. [PubMed: 16709784]
46. Amit I, et al. *Nat Genet* 2007;39:503. [PubMed: 17322878]
47. Segal E, et al. *Nat Genet* 2003;34:166. [PubMed: 12740579]
48. Pe'er D, Regev A, Elidan G, Friedman N. *Bioinformatics* 2001;17:S215. [PubMed: 11473012]
49. Faith JJ, et al. *PLoS Biol* 2007;5:e8. [PubMed: 17214507]
50. Basso K, et al. *Nat Genet* 2005;37:382. [PubMed: 15778709]
51. Ramsey SA, et al. *PLOS Comput Biol* 2008;4:e1000021. [PubMed: 18369420]

52. Hu Z, Killion PJ, Iyer VR. *Nat Genet* 2007;39:683. [PubMed: 17417638]
53. Brem RB, Yvert G, Clinton R, Kruglyak L. *Science* 2002;296:752. [PubMed: 11923494]
54. Wittkopp PJ, Haerum BK, Clark AG. *Nat Genet* 2008;40:346. [PubMed: 18278046]
55. Lee SI, et al. *PLoS Genet* 2009;5:e1000358. [PubMed: 19180192]
56. Emilsson V, et al. *Nature* 2008;452:423. [PubMed: 18344981]
57. Chen Y, et al. *Nature* 2008;452:429. [PubMed: 18344982]
58. Tirosh I, Reikhav S, Levy AA, Barkai N. *Science* 2009;324:659. [PubMed: 19407207]
59. Wilson MD, et al. *Science* 2008;322:434. [PubMed: 18787134]
60. Corey DR, Abrams JM. *Genome Biol* 2001;2reviews1015.1
61. Oliveri P, Tu Q, Davidson EH. *Proc Natl Acad Sci U S A* 2008;105:5955. [PubMed: 18413610]
62. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. *Science* 2005;308:523. [PubMed: 15845847]
63. Dupuy D, et al. *Nat Biotechnol* 2007;25:663. [PubMed: 17486083]
64. Du J, et al. *Nat Biotechnol* 2009;27:77. [PubMed: 19098899]
65. Geiss GK, et al. *Nat Biotechnol* 2008;26:317. [PubMed: 18278033]
66. Cantone I, et al. *Cell* 2009;137:172. [PubMed: 19327819]

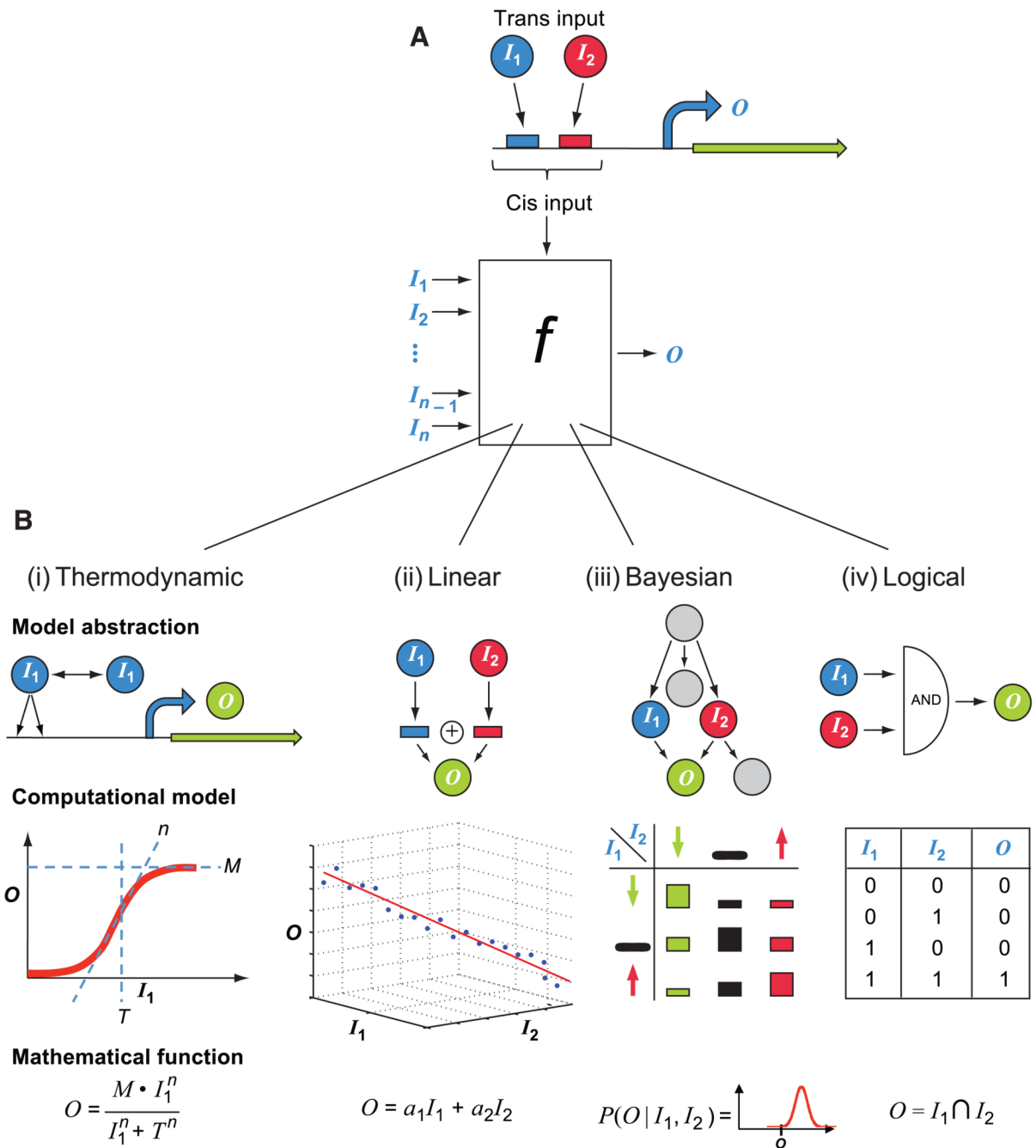


Fig. 1. Gene-regulation functions. **(A)** A gene-regulation function describes how trans inputs, such as transcription factors (I_1, I_2, \dots, I_n), and cis inputs, such as regulatory elements, are transformed into a gene's mRNA level (O). **(B)** Different functions describe distinct scenarios, using distinct mathematical approaches. Most small-scale approaches rely on a thermodynamically motivated model (i), which explains the dependency on a single tunable trans input and fits the response to a Hill function. The extracted Hill parameters, maximum expression level (M), threshold (T), and sensitivity (n), become functions of the remaining trans- and cis-input variables. Large-scale approaches utilize a host of approaches. (ii) Linear models assume that the output is a linear combination of the cis or trans inputs. (iii) Bayesian networks give the

probability distribution of the output given the input values. (iv) Logical (Boolean) circuits consider the output as a result of applying logical operations on the inputs. (Top) The assumption underlying each modeling approach; (middle) the gene-regulation function in this modeling approach; and (bottom) an equivalent mathematical formulation. All modeling approaches are applicable in both small- and large-scale studies.