Published in final edited form as:
Trends Cell Biol. 2011 August ; 21(8): 442–451. doi:10.1016/j.tcb.2011.05.001.

Reconstructing Regulatory Network Transitions

Jalean J. Petricka[1] and **Philip N. Benfey**[1,2]

[1] Department of Biology and IGSP Center for Systems Biology, Duke University, Durham, NC 27708

Abstract

Cellular responses often involve a transition of cells from one state to another. A transition from a stem cell to differentiated cell state, for example, may occur in response to gene expression changes induced by a transcription factor, or signaling cascades triggered by a hormone or pathogen. Regulatory networks are thought to control such cellular transitions. Thus, many researchers are interested in reconstructing regulatory networks, not only to gain a deeper understanding of cellular transitions, but also with the aim of using networks to predict and potentially manipulate cellular transitions and outcomes. In this review, we highlight approaches to the reconstruction of regulatory networks underlying cellular transitions, with special attention to transcriptional regulatory networks. We describe recent regulatory network reconstructions in a variety of organisms and discuss the success they share in identifying new regulatory components as well as shared relationships and phenotypic outcomes.

Regulatory Networks Underlying Cellular Transitions

Cells respond to various stimuli, such as hormones and pathogens, as well as changes in environmental conditions. A yeast cell, for example, undergoes changes in response to low oxygen to produce ethanol. Cellular responses such as this often involve a transition from one state to another. Other examples include when cells transition between different states during the phases of the cell division cycle and during stages of pathogen infection. Cellular transitions from one state to another can occur over various time frames and are impacted by interactions between many internal and external factors (Figure1). Such transitions are believed to be orchestrated by regulatory networks [1–6; Glossary Box, Fig.1], which are composed of biological molecules, such as proteins, that are involved in the control of a range of biological activities, including signaling cascades and transcriptional activity.

One recent example of a regulatory network underlying a cellular transition is found in human stem cells transitioning to differentiated endoderm, which later produces lung, thyroid, and pancreatic cells [7]. In this example, a regulatory network of the transcription factors NANOG, OCT4, and SOX2 is known to be important for stem cell pluripotency the potential of stem cells to differentiate into different germ layers. The authors showed that these transcription factors directly control expression of EOMESODERMIN, which transitions cells to specify endoderm. In turn, EOMESODERMIN interacts with SMAD2/3 to initiate the subsequent formation of endoderm from stem cells [7]. These findings are significant because they not only describe the regulatory network underlying the cellular

[2]Corresponding Author: Benfey, P.N. (philip.benfey@duke.edu).

transition from stem cells to endoderm, but also point to potential therapeutic uses in the regeneration of human organs derived from endoderm [7]. This example highlights the relevance and importance of understanding regulatory networks controlling cellular transitions.

Regulatory networks can be difficult to characterize and may be represented in a variety of ways. One simplistic view can be obtained from playing Perfection, a game from Milton Bradley. This game requires skill and speed to place the many different shaped pieces (like circles and squares) into the corresponding shaped spaces of the depressed, ordered grid before it pops up and spews out all of the pieces. On a gross level, biological networks are similar in that they are composed of the molecules in a cell. In the simplest sense, these molecules or pieces are localized to specific positions relative to one another. For example, molecules can be found in the nucleus versus the cytoplasm of the cell, relative to each other based on their sequential action (such as enzymes acting in a biochemical pathway), or precisely interacting with a molecule to regulate its action or expression (as in phosphatases regulating kinase receptors, or transcription factors controlling expression of a gene). Many researchers are working hard to correctly place all of the pieces or components into a larger framework, or network, to understand their location and relationships.

Of course, biological networks are much more complex than this simple analogy. The components or pieces consist of an assortment of different molecules, such as DNA, RNA, metabolites, and proteins. While the relationships between molecules within networks have been represented as simple connections [8,9], in reality molecules act dynamically in the cell, sometimes interacting with multiple different partners in less than a second. The partners and/or targets of these molecules can then change a few seconds later. In addition, over time, evolutionary changes occur in the molecules of a network, which influence the relationships between molecules and thereby the architecture of the network. Thus, the study and representation of the components, interactions, and their dynamics within a biological network are quite challenging and complex.

One aim of systems biology [Glossary Box] is to characterize and manipulate these highly dynamic and complex regulatory networks. To accomplish this aim, systems biology utilizes and often combines methods and approaches from a variety of disciplines, including, but not limited to, biology, chemistry, physics, mathematics, statistics, engineering and computer science. A systems approach encompassing molecular, genetic, genomic, mathematical, and computational methods, for example, has been successfully used to discern the cellular response of human cells to influenza virus [10]. The experimental strategy included assays of protein-protein interactions between human and viral proteins (yeast two-hybrid assays), gene expression in human bronchial epithelial cells (HBECs) exposed to virus and virus infection (microarrays), as well as genetic knock-down experiments combined with viral replication and reporter assays in HBECs [10]. Data obtained from these experiments were combined and divided into groups, or clusters , using computational algorithms to identify signaling pathways in human cells involved in the detection and elimination of influenza proteins. Using this systems approach, the authors identified a regulatory network that includes RNA binding, WNT signaling, and viral polymerase subunit proteins that have functional roles in HBECs infected by influenza virus [10].

Systems biology approaches embrace traditional experimental approaches from molecular biology and genetics that are focused on individual molecules, in addition to high-throughput experimental approaches like large-scale analyses of gene expression. While traditional and high-throughput strategies both allow detection and quantification of gene expression based on nucleotide hybridization, for instance, these approaches differ considerably in many aspects such as cost, scale, feasibility, and sensitivity [Box 1]. The

questions that systems biology approaches are being used to address cover a wide range of topics in a variety of organisms. Some examples include: How do components of a cell, tissue, or organ cooperate to recognize signals and coordinate an appropriate response? How is cellular homeostasis maintained when an underlying regulatory network is perturbed by unfavorable environmental conditions or pathogen attack? How stable are cellular transitions in development, for instance, from one cellular identity to another, and can they be induced or reversed by alterations in network components and relationships? Can we predict which drugs and therapies will be most effective in treating human cancers and disease, and specifically target them to affect only certain portions of a larger regulatory network controlling overall human health? Systems biology approaches utilizing and integrating knowledge, techniques, and methodologies from diverse disciplines are therefore necessary to address the complexity of these important biological questions.

Box 1

Strengths and weaknesses of traditional and high-throughput approaches
Traditional and high-throughput strategies are both used for regulatory network reconstruction. Since these approaches differ considerably in scale, feasibility, and sensitivity, traditional and high-throughput each have different advantages and disadvantages.

Both strategies allow detection and quantification of gene expression based largely on nucleotide hybridization. Traditional methods of detecting gene expression differences include Northern blots, *in situ* hybridization, and Real Time quantitative PCR (RT-qPCR). Northern blots have the advantage that large probes can be used to detect the full-length transcript of most mRNA sequences. *In situ* hybridization has an edge in that it detects the localization of a specific transcript within a tissue or organism instead of in a test tube, or on a blot or glass slide. RT-qPCR is considered the gold standard for sensitivity, as it can reliably detect as little as a few transcripts in a sample. The disadvantages of these traditional assays mirror the advantages of high-throughput assays quantifying gene expression.

High-throughput methods include microarrays and next-generation or deep sequencing [Glossary Box] of cDNAs corresponding to mRNA. The advantages of these assays are that they are less labor intensive due to the use of robots, and they measure gene expression for all genes in the genome for a reasonable cost, representing a sizable increase in gene number over the 100 s feasible by traditional approaches.

Both strategies can also detect and quantify phenotypes and determine genetic relationships that can then be used for regulatory network reconstruction. Traditional approaches to studying phenotypes involve generation and analysis of single and/or double mutants by procedures including mutagenesis, knock-out by homologous recombination, transformation, crossing, mating, growth measurements, behavior and disease assays, and microscopy. Here again, the advantages and disadvantages are reciprocal for traditional and high-throughput approaches. High-throughput phenotyping platforms use robots and automation of machines, such as microscopes, to standardize and increase the number of mutants or conditions assayed. However, each platform is often optimized for a specific phenotypic assay and platforms are limited by growth habits of organisms and other factors, presenting challenges to high-throughput phenotyping [11]. Nevertheless, it is a key future goal considering that an Escherichia coli network that has 4500 genes and 300 regulators would require 4500X300 = 1,350,000 experiments to test all of the possible connections [12].

In this review, we highlight studies reconstructing regulatory networks underlying biological or cellular transitions, with special emphasis on studies of transcriptional regulatory networks. Recent interest has focused on reconstructing these networks using experimental and theoretical approaches (for a review of theoretical approaches, see [13]). Generally, these approaches have started from molecular or phenotypic data and inferred relationships between molecules and or phenotypes associated with given cellular transitions. The resulting regulatory network reconstructions in a variety of organisms have identified new network components as well as shared relationships and phenotypic outcomes that are involved in cellular transitions.

Network Reconstruction Using 'Guilt-by-Association'

One common approach to reconstructing regulatory networks is to identify and characterize clusters of components and/or connections associated with a given cellular transition. Input from this guilt-by-association approach often comes from high-throughput data sets obtained from large-scale analyses of changes in gene expression (e.g. DNA/RNA microarrays or deep sequencing [Glossary Box]) or from protein interaction analyses (e.g. yeast-two hybrid assays or protein arrays) before, during, and/or after a cellular transition. Connections between genes or proteins are inferred, and then grouped into clusters based on the correlation of gene expression or protein interaction profiles with each other and the cellular transition. A critical assumption used frequently to reconstruct networks with this approach is that statistical relations in the data arise from relationships and interactions between molecular components. For example, in a recent study of Alzheimer s disease in humans, mRNA expression profiles previously obtained from a number of normal tissues and cell-types were used to determine the correlation between genes known to increase susceptibility to, or cause, this neurodegenerative disorder and genes located in chromosomal regions associated with Alzheimer s disease [14]. If a highly significant correlation was observed for an expression profile of a gene located in a chromosomal region associated with the disease and the expression profiles of known Alzheimer s disease genes (i.e., the genes were co-expressed), then the gene was inferred to have a relationship with known Alzheimer s disease genes. Genes identified using a guilt-by-association approach are often referred to as candidate genes to reflect the predictive nature of the inference. The candidate genes resulting from this approach are thus hypothesis-generating in that they are implicated, but not demonstrated, to be involved somehow in the etiology of Alzheimer s. This is one of the major strengths of this approach: new genes are identified that potentially are involved in a given cellular transition, or, as in the example described, in increased susceptibility to Alzheimer s disease. When data generated from many large-scale experiments, such as microarrays are used, the search is unbiased and often a large number of new candidate genes can be identified.

There are, however, drawbacks to the guilt-by-association approach. One major weakness is that while genes can be co-expressed, the relationship between some of these genes may be unclear or even not exist. For instance, in the Alzheimer s study [14], genes may share mRNA expression profiles across a number of tissues and cell-types but the purpose or outcome of the expression profile may be very different. Some genes may be expressed for specific processes like development or metabolism in these tissues and cell-types unrelated to disease, while others are expressed as part of signaling cascades specific to Alzheimer s disease. Approaches to addressing this issue include applying more stringent statistical thresholds for determining correlations, or using and integrating additional data or information, such as gene expression data sets more specific to Alzheimer s disease or the cellular transition of interest. As there are many methods for assessing statistical relationships in data, it is also worth noting that the genes and relationships identified using the guilt-by-association approach may differ depending on the computational and statistical

methods employed. If this occurs, it is possible to obtain a more confident set of genes and relationships by considering only those found by several methods. Additional confidence can also be gained by performing additional genetic and/or molecular experiments to directly test and verify the nature of inferred gene relationships. For instance, the researchers in the Alzheimer s disease study performed pair-wise protein-protein interaction assays of known and candidate genes identified by the guilt-by-association approach [14]. Positive interactions provided further evidence for a role in Alzheimer s for a number of proteins, including Programmed Cell Death 4 (*PDC4*), which could act as a neuronal death regulator in conjunction with PRESENILIN2 and apolipoprotein E, known Alzheimer s genes [14]. This leads to perhaps the largest limitation of this approach, which is that it does not provide mechanistic, or functional , information about the resulting genes and relationships (We discuss functional approaches to network reconstruction later in the review). The guilt-by-association approach does, however, generate interesting hypotheses and candidate genes for further study and also enables reconstruction of regulatory networks.

Reconstruction of Transcriptional Regulatory Networks

A large amount of work over the last few years has gone into the reconstruction of transcriptional regulatory networks (TRNs). In this type of network, the connections represent binding of transcription factor proteins (TFs) to the regulatory region of their target genes (Figure 2). One way to infer these regulatory relationships is from observed changes in mRNA expression levels. For example, from transcriptional profiling data collected before and after a cellular transition, and prior knowledge of the type, number, and organization of specific *cis*-regulatory elements bound by individual TFs in the promoters of TF target genes, network relationships can be drawn. This strategy was successfully used for TRN reconstruction, and impressively, prediction of TF target gene expression patterns in *Drosophila melanogaster* segmentation [15]. Here TRN reconstruction extended from input TF expression to TF binding of cis-regulatory sequences to the output spatiotemporal expression patterns of TF target segmentation genes during embryogenesis. In this TRN reconstruction, the number and position of known transcription factor binding sites (TFBSs) in the promoter of each target gene was recorded for each TF known to regulate patterning of the embryonic segments. The authors then calculated the probability that a given TF would bind to a given site in the promoter of a TF target gene from prior knowledge of the binding affinity of that TF [15]. This was calculated for all sites in all promoters of predicted TF targets; the resulting probabilities allowed the authors to predict the strength of TF binding to sites in target promoters in the TRN. These probabilities were then integrated with information on TF expression levels in each fly segment (i.e. how much TF was available for binding to sequence sites) to predict the output expression level of each target gene for each of the segments [15]. Remarkably, the authors accurately predicted expression patterns of TF target genes in fly segmentation [15]. This was a surprising result for a number of reasons. First, the reconstructed TRN relied on inferences of TF binding events and their contributions to the expression of specific target genes, many of which were not demonstrated experimentally. Second, further assumptions were made that TF binding not only had functional effects on target gene expression, but also that these functional effects led to a specific output level of gene expression. Finally, the impacts of other factors known to affect gene expression levels, such as the accessibility of chromatin to TFs, were not included in the model [15]. This work stands in contrast to most other studies of TRNs in that it demonstrates the predictive power of TRN reconstructions to generate quantified outputs.

Another approach to TRN reconstruction is to use data from direct binding assays of TFs to DNA. TRNs have been reconstructed in this way by using *in vitro* TF binding data. For example, yeast one-hybrid assays or promoter binding microarrays combined with gene

expression information in *Caenorhabditis elegans* and *Arabidopsis thaliana* have been used to reconstruct TRNs [16–18]. However, TRN data are more commonly obtained *in vivo* from Chromatin Immuno-Precipitation followed by quantitative PCR (ChIP-qPCR), genome-wide microarray (ChIP-chip) or deep sequencing (ChIP-seq, [Glossary Box]) experiments because these data are thought to reflect *in vivo*, direct binding events. There are numerous examples of the use of these approaches and we will only highlight a few.

The largest atlas of ChIP-chip experiments has been compiled in yeast [19,20]. TFBSs were determined for most yeast TFs and then mapped onto the yeast genome to generate a TRN. This representation of the network was first reconstructed from standard conditions without incorporating gene expression information (Figure 2). This is a valid approach because this type of TRN represents a large set of inferred regulatory connections. However TRNs, as defined above, indicate regulation by a TF; thus knowledge about whether or not these binding events contribute to gene regulation (i.e. expression changes) must be incorporated to determine if they are regulatory. For example, yeast ChIP-chip results were combined with data from 500 microarray experiments to identify groups of genes that are coordinately bound and expressed, which were called multi-input motifs refined for common expression (MIM-CE) [19]. In brief, the method defines a group of genes that are bound by a set of TFs and then refines the cluster to include only genes that are similarly expressed in all of the microarrays. Next, the algorithm searches for other genes with similar expression profiles, which are also bound by the same TFs. These genes are then added to the group. This process is repeated until all combinations of genes bound by TFs have been queried. Applying this method to the extensive expression data available for the cell cycle, the authors reasoned that MIM-CEs enriched in genes whose expression oscillated through the cell cycle would identify TFs that control these genes. Indeed, the authors accurately identified 9 TFs and correctly assigned them to the corresponding phase of the cell cycle. This included two TF MIM-CEs that had been implicated in the cell cycle, but had ill-defined functions [19]. This method and resulting cell cycle TRN are impressive because direct binding events of various TFs are likely to differ depending on the phase of the cell cycle, and these were not experimentally determined in this study (TF targets were determined by genome-wide ChIP studies performed on non-synchronized cells under standard conditions) [19]. However, the success of using dynamic expression data to achieve an appropriate cell cycle TRN points to the importance of starting with dynamic data for TRN reconstruction of cellular transitions.

Other recent studies have also reported TRNs derived from dynamic transcriptome and ChIP-chip or ChIP-seq data. For instance, dynamic transcriptional data was used in a study of two TFs, SHORT-ROOT (SHR) and SCARECROW, which regulate the asymmetric cell divisions that generate the endodermal and cortex cell lineages of the *Arabidopsis thaliana* root. Transcriptional profiling was performed at multiple time points after induction of these TFs in sorted cells [21]. A previously described algorithm [22] was then used to identify genes directly bound and regulated at the precise time and location of the asymmetric cell division [21]. One of these confirmed targets was a D-type cyclin specifically expressed in these asymmetrically dividing cells, linking cellular patterning and division [21]. Previous transcriptional studies addressing the role of SHR did not identify this D-type cyclin [23], suggesting that this integral link between cellular patterning and division would not have been discovered if dynamic transcriptional data had not been used. Thus, this study emphasizes the importance of incorporating dynamic transcriptional information to determine TRNs underlying cellular transitions.

Another recent study uses dynamic expression data and concentrates on the transition from vegetative to reproductive development in *Arabidopsis*, which is controlled by a set of TFs. When these TFs are mutated, plants have disrupted floral initiation, patterning, and

development. APETALA1, a key TF regulator of floral initiation and development, was chosen in this study to profile expression changes over time with microarrays and TF binding using ChIP-seq during the course of floral initiation [24]. The overlap between the genes found from each of these experiments was then used to reconstruct the TRN. The results indicated that genes involved in vegetative development were repressed by this TF, while genes specifying reproductive development were activated [24]. The dual action of this TF in repressing one fate while activating another is intriguing because a similar dual function has been reported in animals. For example, in embryonic mouse stem cells the TFs OCT4, SOX2, and NANOG have been shown to simultaneously repress targets associated with cellular differentiation while activating genes specifying stem cell fate [25]. This suggests that dual activation and repression by TFs in TRNs may be a general mechanism for controlling cellular transitions.

These examples highlight the power of using dynamic expression data to reconstruct TRNs, as key targets (and thus TRN architecture) and mechanisms of transitions can be missed in TRN reconstructions lacking dynamic transcriptional information. Despite the power of these studies, they do not include information about the dynamics of direct TF binding (e.g. dynamic ChIP data). This is a drawback because TRNs reconstructed from dynamic transcriptional data, but single time point ChIP data, still rely on inferences about whether the observed dynamic changes in gene expression are actually dictated by direct TF binding events. Confidence in the resulting TRNs is therefore less than for a TRN reconstructed using ChIP-chip or ChIP-seq data obtained at the same time points profiled in dynamic expression experiments. Future studies incorporating dynamic information of both direct binding and gene expression should add to the confidence in TRN architecture.

The previous two examples of TRN reconstruction in plants used direct TF binding and dynamic gene expression experiments for single TFs. However, larger TRN reconstructions involving a greater number of TFs have been performed. For example, in mouse embryonic stem cells (ESCs), TRNs were reconstructed for 9 and 11 TFs involved in reprogramming of ESCs using ChIP-chip and ChIP-seq, respectively [26,27]. ChIP-chip and transcriptome data were grouped with additional data obtained from ChIP-chip data of methylation patterns in ESCs (H3K4me3 and H3K27me3) using a supervised clustering approach [26, Glossary Box]. What emerged was striking; genes that were active in ESCs and repressed upon differentiation had promoters that were occupied by multiple TFs, whereas a single factor occupied promoters of genes that were inactive in ESCs and activated upon differentiation [26]. Clearly, this feature would not have emerged from a TRN reconstruction based on a single TF. Thus, TRN reconstructions of individual factors overlook combinatorial action of TFs that may be vital to the regulation and control of target genes. For this reason several projects are now trying to systematically analyze binding of all TFs involved in a biological process with the goal of accurately reconstructing TRNs. For example, the *Drosophila* modENCODE project [28, 29] has reconstructed a large TRN using ChIP-based TF binding of 76 TFs, 104 TF conserved binding sequences, and gene expression data [28].

Collectively, TRN reconstructions in many organisms have led to new and important insights about biological transitions, which should allow predictive modeling and future manipulation of transcriptional regulatory events. However, comprehensive TRNs incorporating dynamic direct binding and gene expression for all TFs in a given organism have not yet been achieved. Emerging high-throughput data from large projects in many organisms, such as modENCODE in *C. elegans* [30] and ENCODE in humans [31,32], offer promise for future comprehensive and accurate TRNs. Despite this promise, some TRN studies indicate that reconstruction using direct binding information from transcriptome and genome-wide ChIP data is more complex as TF binding events far exceed the number of genes regulated at the transcriptional level (reviewed in [33]). Many hypotheses have been

proposed for this observed discrepancy, but the significance of direct TF binding without accompanying changes in gene expression remains unclear [33]. A major challenge for the future is to better interpret ChIP data as well as other aspects of TRN reconstruction [Box 2]. It is likely that integration of other data types and/or novel approaches will be needed to improve the accuracy of TRN reconstruction.

Box 2

Considerations for reconstructing TRNs

The goal of TRN reconstruction is to determine the relationship between TFs, their target DNA sequences, and the effects of these relationships on gene expression. Some researchers consider TRNs to be functional if they are able to accurately predict output gene expression levels of TF targets. However, in the context of this review we define functional regulatory networks to be those that make connections to phenotypes.

Much progress has been made in assaying TF binding targets on a genome-wide scale *in vivo* using microarrays and next-generation sequencing, which is allowing reconstruction of more expansive TRN reconstructions. ChIP data are being integrated with gene expression data and *in vitro* assays that determine TF binding site sequences for a number of organisms. Here we mention a few considerations for TRN reconstruction involving direct binding data obtained from ChIP assays.

First, TF expression, as well as chromatin accessibility to TFs, can be different depending on the environment, developmental stage, or spatial location of cells. This presents a challenge for TRNs reconstructed using ChIP, or *in vitro* assays of TF binding sites used to infer TF targets, for multiple TFs. This is because making or inferring connections between different TFs and targets may be complicated if these TFs are not expressed or able to bind accessible chromatin in similar cellular locations or times. This may be why some TRN reconstructions have focused on a core group of TFs that are expressed in the same cells and cellular conditions. Future advances in ChIP allowing the use of small numbers, or even a single cell, may provide the necessary spatial and temporal resolution for reconstructing TRNs with greater cellular relevance for multicellular organisms. However, the utility of TRN reconstructions for whole organs or organisms using TF binding data derived from different spatial and/or temporal conditions remains an open question.

Secondly, many TFs differ in structure, DNA affinity, stability, and turnover. Perhaps as a result, ChIP protocols vary between labs in fixation, shearing of chromatin, and immunoprecipitation conditions and thus produce different results. In addition, different antibodies, epitopes, and tags are utilized for ChIP studies, which may perform differently in ChIP assays. Therefore, standardization of ChIP procedures is a future challenge; surmounting it will likely result in more comparable data sets. These data sets will increase the predictive power of reconstructed TRNs and serve as improved resources for data integration in other regulatory networks.

Functional Approaches to Regulatory Network Reconstruction

In the previous section we described a basic approach to regulatory network construction of a specific type of network, the TRN, which is potentially useful for prediction and manipulation of gene expression levels. However, cellular phenotypes which should constitute the functional outcome of a TRN don t always correspond to the outputs of reconstructed regulatory networks. Predicting phenotypes from the molecular interactions encoded in the genotype is a central goal of systems biology. To achieve this aim, another

common approach to reconstruction of regulatory networks is to start with phenotypes and then connect them to genetic or molecular relationships.

Such a functional approach to regulatory network construction has been taken in yeast by systematically analyzing different pairs of deleted or knocked-down genes for fitness phenotypes, which were measured as yeast colony size [34]. This work involved analysis of yeast mutants in 1712 essential and non-essential genes that were screened against 3885 non-essential mutant strains; a total of over 5 million gene pairs were thus quantitatively scored for fitness [34]. When fitness scores of double and single mutants were compared, about 170,000 genetic interactions were found. A genetic regulatory network was then reconstructed in which connections denoted genes sharing similar genetic interaction profiles or a common set of genetic interactions. The connectivity of the resulting network was then used to successfully predict function of known and uncharacterized genes [34]. The genetic network was also integrated with data from ~4700 deletion mutants exposed to hundreds of chemical compounds. From these data, the authors calculated a “chemical-genetic degree,” the number of chemical perturbations for which a gene deletion mutant exhibits hypersensitivity when exposed. Interestingly, a significant correlation was found between the number of genetic interactions of a given gene and its chemical-genetic degree [34]. Based on this, the authors suggested that the same genes act to protect yeast cells against genetic and chemical perturbations, which may be useful for linking chemical compounds to gene targets and predicting synthetic interactions between drugs [34]. Recently, similar genetic analyses combined with transcriptional profiling of yeast strains with multiple deletions in kinases and phosphatases have allowed reconstruction of a signaling regulatory network [35]. It will be interesting to see if future integration of these data with chemical data will lead to stronger predictions of chemical and genetic targets for drug discovery.

In addition to potential applications in chemical and drug discovery, other advantages of functional approaches using genetic interaction assays include the ability to implicate genes in a genetic pathway or in biological functions. In [34] the authors compared the ability of their genetic interaction network to predict multiple phenotypic functions as compared to networks reconstructed from protein-protein interaction data. They found that genes that have many interactions (i.e. hubs) in their network correlate more highly with multiple biological processes than do hubs from protein-protein interaction networks. This suggested that networks from genetic interaction studies are better at identifying broad phenotypic functions [34]. However, there are some disadvantages to genetic interaction studies. For instance, how the genes act at the molecular level to achieve biological functions must be inferred by annotation or from previous molecular studies. Also, even though certain genes may interact at the genetic level, the nature of this genetic interaction at the molecular level is often unknown and the protein products of interacting genes may not interact at the molecular level to accomplish biological functions. Indeed, the authors suggest that although genetic interaction networks may implicate many phenotypic functions, protein-protein interaction networks elucidate local molecular pathway architecture [34], which may lie within larger genetic interaction hubs. Thus, there are pros and cons to reconstructing networks from genetic versus protein-protein interaction data. One common disadvantage is that the relevance of observed interactions is often unclear. Although a pair of genes or their encoded proteins appear to interact based on genetic or *in vitro* protein-protein assays, *in vivo* this may not be the case. For example, two proteins may not be expressed in the same place or time frame, suggesting *in vitro* interactions do not occur *in vivo*. Integration of gene and protein expression data with genetic and protein-protein interaction data may help to circumvent this problem and provide more solid links between genes, proteins, phenotypes, and biological functions of reconstructed networks.

Reconstruction of functional TRNs has also been accomplished by combining gene expression information of knock-out TF lines and the transcriptome. For example, transcriptional profiling was performed for 263 deletion strains of yeast TFs and then each was compared to wild type under standard conditions [36]. These data were used to reconstruct a functional TRN that showed low, but significant, overlap with direct targets reported for these same TFs in ChIP-chip studies. This small overlap indicated to the authors that some of the targets in their network were indirect. To address this, a regulatory network model was generated using a directed-weighted graph method [Glossary Box] and then refined by removing putative indirect connections. Refinement involved the following logic about whether a given TF A regulates a target gene: if the statistical value of a gene regulated by two TFs (for example a given TF A that targeted another TF B) was higher for TF A than for TF B [36], then the connection between TF A and the gene is indirect. This refinement significantly improved the overlap, as did using higher quality TF binding data sets [36]. Future research may address whether overlap between these regulatory networks can be enhanced with deep sequencing technologies and/or if other levels of regulation, such as post-transcriptional mechanisms, must be incorporated to reconstruct functional TRNs using this approach.

In mammals, a functional approach using knock-down and gene expression data has recently been reported in mouse dendritic cells exposed to various pathogens for reconstruction of an observational regulatory network [37]. The authors first characterized transcriptional responses of dendritic cells to various pathogens at different time points using microarrays. Potential regulators and time points were then selected using an information-theoretic approach [Glossary Box], whereby the regulators and time points selected capture the most gene expression information. Next, this dynamic data set was used to cluster different responses and reconstruct a regulatory network using an Expectation Maximization (EM) approach [Glossary Box]. The expression of selected regulators was then reduced by > 75% using lentiviral shRNAs in dendritic cells. The resulting expression profiles were then determined at a selected time point after exposure to a single treatment activating the majority of pathogen responses. These data were then used to reconstruct a functional regulatory network that largely agreed with their observational model [37]. This suggests to us that dynamic expression data alone may be sufficient to reconstruct a network similar to one generated with functional approaches. However, the remaining false positive interactions of the observational model were attributed to the fact that a correct regulator had gene expression profiles that were indistinguishable from other regulators [37], suggesting this is a shortcoming of models lacking functional approaches.

Collectively from these studies, the main strength of functional approaches is clear and formidable: gene expression changes are directly linked to a cellular state, transition, or perturbation. In the case of [37], gene expression changes were linked to the response of a specific cell type (dendritic cells) to exposure of a specific pathogen, which may facilitate therapeutic targeting of specific pathways to enhance human vaccine efficacy or combat disease. Of course, known challenges also exist for functional approaches employing RNA interference (RNAi) methods to reconstruct functional regulatory networks. One is the difficulty and expense in reproducibly performing large scale RNAi screens. Other problems related to reconstruction of functional regulatory networks include accounting for differences in the levels of RNAi knockdown achieved for different genes and false positive regulatory relationships that arise from off-target effects of RNA interference. Further challenges of network reconstruction using functional approaches can be found in [Box 3]. Another significant hurdle encountered whenever the output is an expression profile is that this does not provide the molecular mechanism responsible, which may be important for effective manipulations and applications of reconstructed regulatory networks. One common tactic to overcome this hurdle is to use other assays to determine and/or validate molecular

relationships implicated in networks reconstructed by functional approaches. These data can then be integrated to refine the original network or to generate a more detailed, and perhaps more predictive, regulatory network reconstruction of phenotypic outputs.

Box 3

Considerations for reconstruction of functional regulatory networks

The goal of reconstructing functional regulatory networks is to connect phenotypic effects to genes or proteins that are relevant to cellular transitions and ultimately development and disease. Although we present reconstructions using data generated from gene deletion or knock-down as a separate approach from approaches starting with molecular quantifications, this distinction may be hazy. It could be more accurate to say this is an alternative starting point for reconstructing regulatory networks. Thus, the two approaches (starting from molecular changes to phenotypes versus phenotypes to molecular changes) may be considered complementary.

Some of the challenges for identifying and measuring molecules and phenotypes are similar. Are the assays sensitive enough? Do they measure enough features? Are they variable between researchers and/or labs?

Platforms capable of monitoring and measuring phenotypic changes in cells and organisms in high-throughput formats are rapidly emerging. These platforms are becoming more flexible to provide dynamic imaging and measurements under a variety of conditions. Since many of these platforms use automated phenotypic quantification, the continuous and multivariate characteristics of phenotypes are being captured rather than just categorical (dead or alive) ones that discard information [11]. Concurrently, techniques for deleting, silencing, and/or inducing genes are improving steadily. Taken together, phenotypic quantification combined with collections of single or double mutants provides promising high-throughput assays for associating genes and phenotypes quickly, reproducibly, and inexpensively.

However, a number of challenges remain for analyses linking phenotypes and genes. First, many phenotypic measurements and assays are not standardized between laboratories. Secondly, mutational strategies are often plagued by redundancy in gene function, off-target effects of RNAi, and indirect phenotypic effects of mutated genes. Thirdly, even though high-throughput phenotyping assays are improving, currently these assays often produce low-dimension measurements of small sample sizes.

While these challenges may be tackled with community efforts and technical advances, the challenges posed by mutational strategies are more problematic. Indirect or unexpected phenotypic effects or genetic interactions may exist for a multitude of reasons, including gene redundancy, buffering, variation between individuals with the same genotype, or modifications at the post-transcriptional or post translational levels. These may be accounted for by controls or multiple replicates, but incorporation of further molecular data about molecular mechanism or multiple mutants (>2) and alleles may be needed to address the gap between genotype and phenotype.

Data Integration and Regulatory Network Reconstruction

“Perform additional experiments” is a comment many researchers have received upon review of a manuscript or presentation regardless if the work was focused on a single gene or a cellular network reconstructed from multiple genes. One reason for this response may be that scientists generally place more confidence in multiple different assays pointing to the

same biological results and conclusions. This may be why integrating multiple different data sets is appealing to those reconstructing regulatory networks of cellular transitions.

Another related potential benefit to integrating data for regulatory network reconstruction, besides increased confidence, is more detail about the molecular nature of inferred relationships. For instance, suppose two TFs in a reconstructed TRN share direct binding profiles to the same sites in the promoters of many target genes. Some TRN reconstructions would denote a relationship between these TFs; however, the nature of this relationship is unclear. Do these TFs bind at sites very close by, or do they interact in or are part of a protein complex? Data from protein-protein interaction studies could address this question. This example also illustrates how integrated regulatory networks are often a source of hypotheses that drive future research. For example, if the TFs were found to interact, then further studies could be performed to examine this relationship, such as single or double knock-down experiments of the TFs evaluating their phenotypic effects on a cellular process inferred from identities of joint targets.

One recent large-scale effort to determine if data integration improves reconstructions of regulatory networks comes from the *Drosophila* modENCODE project [28,29]. In [28] the authors reconstructed a large TRN by integrating ChIP-based TF binding of 76 TFs, 104 TF conserved binding sequences, two chromatin data sets specifying chromatin marks, and three large gene expression data sets using an unsupervised machine learning algorithm [Glossary Box]. For the reconstruction, the possible set of interactions came from 707 TFs and 14,444 target genes, which correspond to the number of genes that had at least one measurement in one data set. The reconstructed functional TRN represents the relationships with the highest confidence between 576 TFs and 9,436 target genes [28]. The resulting TRN was then compared to TRNs reconstructed using known TF binding motifs, direct TF binding information alone, or REDfly literature curation. A few notable findings emerged. First, the integrated reconstructed network had increased biological relevance, as co-targeted genes had increased functional similarity, expression correlation, and protein-protein interactions [28]. Second, although the algorithm used to reconstruct the TRN was not trained on the REDfly literature-curated network, the highest similarity in connectivity to the authors integrated network was found in the REDfly literature-curated network, suggesting to the authors that they had reconstructed a functional TRN [28]. The functionality of the network was further supported by validation metrics (enrichment above random networks of gene pairs that share common gene ontology terms, expression values across time courses, Imago tissue terms, and protein-protein interactions) that were comparable in the authors reconstructed network and the REDfly gold standard network. However, the authors TRN also had 100 and 1000 times more components and connections, respectively, than the REDfly literature curated network [28]. These additional novel relationships remain hypotheses that are valuable for future research. Taken together, the findings in [28] suggest that data integration has significant advantages over reconstructions with fewer data types. Nevertheless, the authors report that further data sets as well as predictive models are needed because the expression of only one-quarter of the genes could be accurately predicted using their reconstructed regulatory network as compared to random networks [28]. However, networks reconstructed from fewer data types had no predictive value over random networks [28], again supporting the value of integrating multiple data types for network reconstruction.

From the above example, we see that integrating different data types may have advantages over regulatory network reconstructions involving fewer data. But how does one know which types of experimental data are the most informative for integration? The answer depends on many variables such as the type of network (signaling cascade, transcriptional, etc.), the desired output of the network, and the experimental strategies available in the cell

or organism of interest. Nevertheless, one answer seems to be integration of any available data previously published. It seems wise to approach this answer with caution as integration of data sets that differ in resolution (i.e. cellular versus organismal or nucleotide versus gene sequence), platform, lab conditions, and/or quality may not improve a regulatory network reconstruction. As few studies have attempted to assess whether integration enhances the confidence or predictive power of a reconstructed regulatory network, future work is needed in this area.

In a recent study, the question of which data type is optimal for the study of cellular transitions and network reconstruction was avoided entirely. An unbiased approach involving the generation of as many different data types as possible was taken, as the type or level of regulation (DNA, RNA, protein, etc.) responsible for a cellular transition was largely unknown. In this study of murine embryonic stem cells (ESCs), omics approaches were used to determine the temporal changes in histone acetylation, chromatin-bound RNA polymerase II, mRNA, and nuclear protein levels resulting from the loss of the key pluripotency regulator, Nanog [38]. With the goal of identifying the regulatory layer primarily responsible for changes in protein levels directing cellular phenotypes due to Nanog loss, the authors examined transcriptional, post-transcriptional, translational, and post-translational changes with genome-wide assays. They found that transcriptional changes mediated by transcription factors preceded chromatin reconfiguration and that many of these changes were discordant with nuclear protein levels [38]. The authors concluded that the translational and post-translational steps constituted the majority (43–52%) of the changes involved in Nanog-regulated ESC fate decisions [38], suggesting that proteomic data are the most informative for understanding ESC fate transitions. It is still unclear whether this is true of cellular transitions in other systems [38] and also unknown whether differences in the types of technologies and/or their sensitivity contribute to these differences. Many studies, such as [39–44], evaluating the correlation between RNA and protein expression, for instance, have found that a large proportion (~30–60%) of RNA and protein profiles using current technologies do not correlate. Although a single reason for this has not been established, it may suggest that conclusions about the appropriate layer of regulation that are drawn from comparisons between RNA and protein levels (as in [38]) are not straightforward. Future studies of other cellular transitions are needed to elucidate the most important molecular layers and experiments for regulatory network reconstruction.

Concluding remarks

Cellular responses often involve a transition of cells from one state to another, such as a transition from a stem cell to differentiated cell fate in response to a stimulus. Regulatory networks are thought to control these cellular transitions. From studies of reconstructed regulatory networks featured in this review, a few general themes emerge. First, guilt-by-association and functional approaches have been successful in linking genes to biological processes and phenotypic effects to gene expression changes, respectively, in order to reconstruct regulatory networks. Second, new candidates as well as shared relationships and phenotypic outcomes involved in cellular transitions are valuable deliverables that come out of regulatory network reconstruction. Dynamic data obtained before, during, and/or after a transition have been particularly instrumental in uncovering novel factors involved in cellular transitions. Thirdly, regulatory networks have had limited success in accurately predicting gene expression and phenotypic outcomes of cellular transitions, which restricts the utility of such networks for the purpose of manipulations and applications aimed to control cellular response and disease. Research suggests that the predictive power of networks will improve with further advances in data acquisition and integration, as well as the development of new approaches to regulatory network reconstruction. It is also possible that a large amount of uncertainty will remain and that the primary outcome will be the

generation of hypotheses. Therefore, future research will determine not only if regulatory networks are useful for revealing the mechanisms of cellular transitions, but also if truly predictive regulatory networks can be realized and serve as effective tools in translational research.

Glossary Box

Regulatory network	the molecular components, interactions, and/or relationships of a cell, tissue, organ, or organism that regulate a given biological process. In the context of this review, we use this broad definition as it encompasses more specific regulatory networks, such as those in metabolism, protein signaling, and transcription
Systems biology	a field of biology that studies all components of a biological system, rather than individual ones, to comprehensively and quantitatively describe and understand their functional interactions, dynamics, and contributions to the system using integrative methods from diverse scientific fields, including physics, mathematics, biology, and chemistry
Next-generation or deep sequencing	massive parallel sequencing of millions of small ~35–250 nucleotide fragments from a single sample
Transcriptional Regulatory Network (TRN)	a representation of connections between transcription factors (TF) and the target genes to which they bind and whose expression they control
Chromatin-immunoprecipitation (ChIP)-seq	an experimental approach that combines isolation of protein-chromatin complexes by immunoprecipitation using protein-specific antibodies or epitopes, combined with next-generation sequencing
Supervised clustering	grouping similar genes or expression profiles using a method that is constrained by other data or information
Information-theoretic Approach	an approach that involves quantification of information in terms of entropy (i.e. the uncertainty in the value of a random variable)
Expectation Maximization (EM) Approach	an iterative statistical method for finding the expected (E) log-likelihood and then maximization (M) of it to determine the maximum likelihood of parameters in statistical models that depend on unobserved variables. In biology, this model is used in medical image reconstruction, but also is frequently used for data clustering in machine learning
Directed Weighted Graph Approach	a method of modeling pairwise relationships in a group of objects that are represented by a graph (i.e. network). When the connection between objects (i.e. edge) is given a numerical/statistical value or oriented between nodes, it is said to be to be weighted or directed, respectively. In the case of [54], these objects or nodes are TFs and/or target genes and the relationship between them is given a value, while the regulation is given a direction between them (i.e. the network represents which node is acting upon another)

Unsupervised machine learning	a computational method of grouping data that does not include *a priori* imposed criteria or information. The machine (a computer) first "learns the appropriate similarities from a training data set, for instance from genes or expression data, and then applies the learned similarities to a larger set of unknown data in order to group information

References

1. Nurse P. Systems biology: Understanding cells. Nature. 2003; 424:883. [PubMed: 12931164]
2. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010; 468:911–920. [PubMed: 21164479]
3. Spellman PT, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell. 1998; 9:3272–3297.
4. Simon I, et al. Serial regulation of transcriptional regulators in the yeast cell cycle. Cell. 2001; 106:697–708. [PubMed: 11572776]
5. Santos SDM, Ferrell JE. Systems biology: On the cell cycle and its switches. Nature. 2008; 454:288–289. [PubMed: 18633407]
6. Orlando DA, et al. Global control of cell-cycle transcription by coupled CDK and network oscillators. Nature. 2008; 453:944–947. [PubMed: 18463633]
7. Teo AK, et al. Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. 2011; 25:238–250. [PubMed: 21245162]
8. Alon, U. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapmann & Hall/CRC an imprint of the Taylor & Francis Group; Boca Raton: 2007.
9. Klipp, E., et al. Systems Biology: A Textbook. Wiley-Blackwell; 2009.
10. Shapira SD, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell. 2009; 139:1255–1267. [PubMed: 20064372]
11. Houle, et al. Phenomics: the next challenge. Nature Rev Genet. 2010; 11:855–866. [PubMed: 21085204]
12. Babu MM, Shamir R. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res. 2003; 31:1234–1244. [PubMed: 12582243]
13. De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nat Rev Microbiol. 2010; 8:717–729. [PubMed: 20805835]
14. Soler-López M, et al. Interactome mapping suggests new mechanistic details underlying Alzheimer s disease. Genome Res. 2011; 21:364–376. [PubMed: 21163940]
15. Segal E, et al. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. Nature. 2008; 451:535–540. [PubMed: 18172436]
16. Deplancke B, et al. A gene-centered *C. elegans* protein-DNA interaction network. Cell. 2006; 125:1193–1205. [PubMed: 16777607]
17. Grove CA, et al. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell. 2009; 138:314–327. [PubMed: 19632181]
18. Brady SM, et al. A stele-enriched gene regulatory network in the Arabidopsis root. Mol Syst Biol. 2011; 7:459. [PubMed: 21245844]
19. Lee TI, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae.* Science. 2002; 298:799–80429. [PubMed: 12399584]
20. Harbison CT, et al. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004; 431:99–104. [PubMed: 15343339]
21. Sozzani R, et al. Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. Nature. 2010; 466:128–132. [PubMed: 20596025]
22. Busch W, et al. Transcriptional control of a plant stem cell niche. Dev Cell. 2010; 18:849–861. [PubMed: 20493817]

23. Levesque MP, et al. Whole-genome analysis of the SHORT-ROOT development pathway in Arabidopsis. PLoS Biol. 2006; 4:e143. [PubMed: 16640459]
24. Kaufmann K, et al. Orchestration of floral initiation by APETALA1. Science. 2010; 328:85–89. [PubMed: 20360106]
25. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. [PubMed: 16153702]
26. Kim J, et al. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–1061. [PubMed: 18358816]
27. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008; 133:1106–1117. [PubMed: 18555785]
28. modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science. 2010; 330:1787–1797. [PubMed: 21177974]
29. Nègre N, et al. A *cis*-regulatory map of the *Drosophila* genome. Nature. 2011; 471:527–531. [PubMed: 21430782]
30. Gernstein MB, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science. 2010; 330:1775–1787. [PubMed: 21177976]
31. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]
32. Raney BJ, et al. ENCODE whole-genome data in the UCSC genome browser (2011 update). Nucleic Acids Res. 2011; 39(Database issue):D871–875. [PubMed: 21037257]
33. MacQuarrie KL, et al. Genome-wide transcription factor binding: beyond direct target regulation. Trends Cell Biol. 2011; 27:141–148.
34. Costanzo M, et al. The genetic landscape of a cell. Science. 2010; 327:425–431. [PubMed: 20093466]
35. van Wageningen S, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell. 2010; 143:991–1004. [PubMed: 21145464]
36. Hu Z, et al. Genetic reconstruction of a functional transcriptional regulatory network. Nat Genet. 2007; 39:683–687. [PubMed: 17417638]
37. Amit I, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen response. Science. 2009; 326:257–263. [PubMed: 19729616]
38. Lu R, et al. Systems level dynamic analyses of fate change in murine embryonic stem cells. Nature. 2009; 462:358–364. [PubMed: 19924215]
39. Gygi SP, et al. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999; 19:1720–1730. [PubMed: 10022859]
40. Chen, et al. Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics. 2002; 4:304–313. [PubMed: 12096112]
41. Ghaemmaghami S, et al. Global analysis of protein expression in yeast. Nature. 2003; 425:737–741. [PubMed: 14562106]
42. Williamson AJ. Quantitative proteomics analysis demonstrates post-transcriptional regulation of embryonic stem cell differentiation to hematopoiesis. Mol Cell Proteomics. 2008; 7:459–472. [PubMed: 18045800]
43. Baerenfaller K, et al. Genome-scale proteomics reveals *Arabidopsis thaliana* gene model and proteome dynamics. Science. 2008; 320:938–941. [PubMed: 18436743]
44. Gry, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics. 2009; 10:365. [PubMed: 19660143]

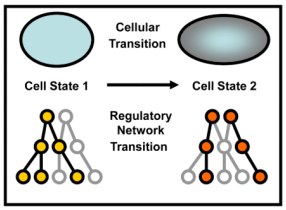

Figure 1.

Schematic drawing of a cellular transition and a regulatory network. At the most basic level, a regulatory network is made up of components (circles) and connections (lines between circles) that may change as the result of a cellular transition. For example, in cell state 1 (blue) only some components (yellow) and connections (black) are present. In response to a stimulus, the cell undergoes a cellular transition from cell state 1 (blue) to cell state 2 (blue-gray). A corresponding transition also occurs in the regulatory network. While some components and connections are unchanged, others are now present (orange circles) or lost.

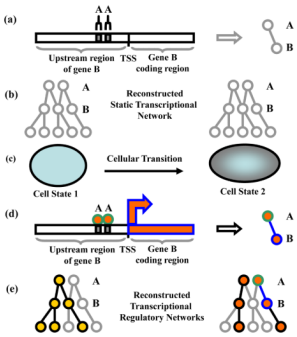

Figure 2.

Example of a Transcriptional Regulatory Network Reconstruction Components of transcriptional networks include transcription factors (TFs) and their target genes (circles); connections between them (lines between circles) denote binding of a TF to the regulatory region of a target gene. If this TF binding results in gene expression changes, then this connection is a regulatory connection; reconstruction of this results in a transcriptional regulatory network (TRN). (a, left) Schematic drawing of the upstream and coding region of a gene (gene B). Predicted or experimentally determined TF binding sites (gray boxes) of TF A are found in the upstream region of gene B. (a, right) This information can be used to reconstruct a representation of the connection (gray line) between a TF A (gray outlined circle A) and its target gene B (gray outlined circle B). (b) The process in (a) can be performed iteratively for each TF and gene in the genome. These can then be assembled to reconstruct a TRN representation. This TRN reconstruction does not include gene expression information, thus this network looks the same in cell states 1 and 2. (c) From Figure 1, a schematic drawing of a cellular transition. It serves as a reference for the networks in (b) and (e). (d, left) Same representation and case as in (a), except there is also gene expression data supporting the expression of TF A (orange circle with green outline) and corresponding changes in the expression levels of its target (gene B, orange circle with blue outline). The orange arrow with blue outline indicates activation of gene B due to binding and regulation by TF A. (d, right) A representation of the relationship between TF A and gene B. TF A action (orange circle with a green outline) results in expression of target gene B (orange circle with a blue outline). The blue line represents the relationship between them, which is a regulatory connection. (e) The process in (d) can be performed iteratively for each TF and gene in the genome for each cell state. These can be assembled to reconstruct a TRN representation for cell states 1 (left) and 2 (right). Note that components (A and B) and regulatory connections in (d) are absent in cell state 1 (left), but present after the cellular transition to cell state 2 (right). The black vertical line labeled TSS in (a) and (d) denotes the Transcriptional Start Site of gene B.