**Gene regulatory networks in plants: Learning causality from time and perturbation.**

**Gabriel Krouk[3], Jesse Lingeman[1], Amy Marshall Colon[2], Gloria Coruzzi[2], and Dennis Shasha[1]**

**1** Courant Institute of Mathematical Sciences, New York University, New York, NY, 10003, USA

**2** Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, 10003, USA

**3** Biochimie et Physiologie Moléculaire des Plantes (UMR 5004 CNRS-INRA-SupAgro-UM2), Institut Claude Grignon, Place Viala, 34060 Montpellier Cedex 1, France

**Abstract**

The ultimate goal of Systems Biology is to generate models that can predict how a system will react under untested conditions or in response to genetic perturbations. This area of research is particularly relevant to plants where such predictive models can be useful for interventions in agriculture, for example to engineer plants to improve their robustness to environmental change. In the present review we: i) describe experimental approaches to understand such dynamic and causal gene relationships in plants using time series (kinetic) and other data, ii) review the analytical approaches used to infer causality in the Gene Regulatory Network (GRNs) from these types of genomic data, iii) review methods for high throughput validation of gene regulatory networks in plants. This systems biology cycle of experiment-model-experiment cycle can lead to rapid understanding of TF-targeted gene network modules in plants.

**Introduction:**

Due to their sessile mode of life, plants are subject to drastic variations in their environment that lead to rapid adaptation of their gene expression states resulting from their complex gene regulatory networks. The ultimate goal in Plant Systems Biology is to infer how such regulatory networks will respond under untested conditions for both scientific and practical gain. In prokaryotes, models to infer gene regulatory networks have successfully predicted genome-wide variations in untested environmental conditions as well as the causal relationships between genes [1-4]. However, there has been less success in generating predictive network models for multi-cellular organisms including plants. With the increasing availability of high throughput -omic techniques and data, we think it useful to review both experimental and informatic approaches for inferring causal relationships in Gene Regulatory Networks (hereafter, GRNs). This article consists of three parts: 1. A review of efforts to use time series and other –omic data to infer causal regulatory edges, and show the kinds of biological insights that can be obtained. 2. A description and a categorization of the informatic methods that are used to infer causal networks. 3. A discussion of recent high throughput experimental techniques to validate inferred regulatory networks in plants.

**I) Successful case studies of learning Gene Regulatory Networks in Plants**

Different kinds of systems approaches are used to model GRNs in plants. One characterization of the approaches is whether they start either with a significant amount of prior experimental knowledge of the connectivity of the modeled gene regulatory network or not. Thus, we call them here “Strong Prior” and “Weak Prior” approaches, respectively.

**Strong Prior approaches.**

In our terminology, Strong Prior approaches are grounded in extensive previous knowledge about the components involved in the GRNs [5] of well-studied functions, for example: auxin signaling [6-8], the circadian clock [9-11], or flower development [12-14]. This given network connectivity knowledge is extended to determine the strengths of edges. Outputs of the models are then compared to experimental data in order determine their predictive accuracy. When

predictive, the models can be used to explore, *in silico,* regulatory behavior in untested conditions and determine overall system properties/architecture. This kind of investigations has lead to striking results exemplified herein.

For auxin signaling, Vernoux et al., 2001 [6] built a model based on previous knowledge of the AUX/IAA-ARF transcription factors network and Yeast-2-hybrid experiments. This gene regulatory network that was finally built displays a strong buffering capacity that is consistent with the *in planta* behavior of the shoot apical meristem as shown by using DII-VENUS as a reporter of the input of signaling pathway and DR5 reporter gene as an output.

The circadian clock is also a well-studied gene regulatory system (for a comprehensive review see Bujdoso and Davis, 2013 [15]) that consists of interlocked transcription factor feedback loops [16-18]. GRN modeling of the circadian system has been successful in determining its evolution in time and the critical components involved in some key features of the oscillations. For instance, in Pokhilko et al., (2010) [17], the GRN model was central to the discovery of the role of PRR5 as a night inhibitor of the LHY/CCA1 expression, including its role in the control of the phase of the of morning gene expression. In the same work, this GRN-generated hypothesis was validated by matching the behavior of *prr5* mutants to gene expression predicted by the model [17]. Alternatively, Akman et al. (2012) [10] used Boolean logic to describe circadian circuits in a quantitative model. The simplified model with decreased parameterization was able to accurately simulate observed circadian oscillations and identify regulatory structures consistent with experimental data.

Flower development (the ABC model) is a textbook example of a conserved GRN that controls the fate of cells into sepals, petals, stamens, and carpels [19]. A successful approach using a discrete network model (gene expression is coded into discrete values) has been to simulate the cell-fate determination during floral organ primordial formation in Arabidopsis [12]. This particular GRN dynamically converges towards different steady states in gene expression, each of which defines the different cell fates in flower organs. Plants arrive at these steady states (or

basins of attraction) independently of the initial gene expression values. This shows that this GRN has feedback/buffering capacities that direct gene expression behavior towards a dedicated state (e.g. to make a particular organ) [12]. More recent studies have taken advantage of the wealth of interaction and expression data available in public databases to construct extensive [13] and condensed [14] models of GRNs involved in floral development, resulting in time-evolving molecular regulatory networks for the development of sepal primordium [13] as well as floral transition [14].

These few examples of successful Strong Prior approaches demonstrate that gene regulatory networks confer robust emergent properties supporting developmental or environmental adaptations.

**Weak Prior approaches.**

The Strong Prior approaches described above begin with some physical connection data and then use time series and other experiments to model behavior [5]. However, for many systems - in plants, animals and microbes - this initial knowledge has yet to be discovered.

"Weak Prior" approaches have no knowledge of gene network connectivity. Thus, computational algorithms are necessary to infer potential connections/causality in GRNs from -omics datasets and then to assign weights to the resulting edges. Many techniques are used to infer unknown weighted networks in the field of systems biology (for reviews see [1, 20]), such as correlation networks and machine learning. A striking success story is the model of gene regulatory programs built from a multi-level dataset (including transcriptomic data and CRE inference) to describe the response of *Halobacterium salinarum* to environmental cues [2]. The model was built *de novo* by a machine learning procedure based on 72 transcription factors responding to 9 environmental factors. The same model was able to predict correct gene response (80% of the genome) in 147 untested conditions [2]. This study clearly demonstrates

the feasibility of Weak Prior approaches in prokaryotic systems. However, this type of approach must be scaled up in order to attain the same predictive power in complex multi-cellular systems [21]. In plant science, since this eukaryotic system is far more complex than yeast or bacteria, the field of GRN *de novo* learning is far less advanced. However, Weak Prior approaches have been developed with some success, as described below.

In the plant field of gene regulatory network modeling the three most popular/used top-down approaches methods are i) classical correlations networks, ii) Graphical Gaussian models (based on partial correlation), iii) Machine learning modeling or combinations of the above.

Correlation networks have been used extensively to study GRNs in plants [22]. When combined with other experimental information, correlation networks help to identify key features of plant regulatory networks. For example, an Arabidopsis multi-network was constructed from all available information about putative TF→Cis-Regulatory-Elements (CRE), protein-protein interactions, and miRNA--| mRNA interactions [23]. Significantly, correlation data integrated with the Arabidopsis multi-network has uncovered i) biomodules involved in carbon/nitrogen signal integration [24] and ii) a central role for CCA1, the central component of the circadian clock in nutrient control [25]. Additionally, correlation network approaches were strikingly successful in identifying two genes (a myo-inositol-1-phosphate synthase, and a Kelch-domain protein) correlating with biomass accumulation in plants [26]. The individual role of these two genes was further supported by an association mapping study that demonstrated coherent allelic diversity at their loci [26].

Gaussian Graphical Models (GGMs) attempt to find pairwise relationships between entities, say A and B, after subtracting out the relationships to other entities, say C, D, …. GGMs have been successfully developed [27] and applied to the inference of plant regulatory elements [28, 29]. In Ingkasuwan et al., time series were analyzed to identify genes regulated across diurnal cycle. Then a sub-network of starch metabolism genes together with the diurnally regulated TFs were modeled using GGM. This model was tested and validated by studying regulator mutants that

displayed starch granule defects in plastids [28].

Machine learning methods have also been employed to learn GRNs from time series and other data. State-space modeling is a modern machine learning technique devoted to detecting causality in networks by inferring ordinary differential equations specifying the relationships among genes in those networks while avoiding over-fitting. In plants, this technique has been applied to probe GRNs involved in leaf senescence [30] and GRNs involved in regulating early, time-dependent transcriptional responses to $NO_3^-$ [31]. Breeze et al. (2011) [30] provided a high-resolution temporal picture of the aging leaf transcriptome. Machine learning revealed modules that play various roles at different times, where each module involves particular TF families and CREs. This approach resulted in a GRN model that correctly predicted the influence of the TF ANAC092, and proposed several new regulatory edges that remain to be validated [30]. In another study [31], state-space modeling and machine learning were applied to an Arabidopsis, high-resolution time course of genome-wide transcriptional response to $NO_3^-$ treatments. A subset of TFs and N -transport and -assimilation genes were modeled with transcriptional regulators, in order to propose a GRN that explains $NO_3^-$ signal propagation. *In silico* validation demonstrated that the state space model trained on the early time points was able to predict gene expression modulation in later time points. Experimental validation consisted of studying the effect of over-expressing a predicted hub (SPL9 TF) on the response of other $NO_3^-$ regulated genes. Indeed, SPL9 over-expression modified the regulation of the predicted target genes in the subnetwork (e.g. NIR, NIA2), but also of many other $NO_3^-$ regulated TFs in the GRN [31]. This supports the idea that network relationships adapt to genetic perturbations.

**II) Analytical approaches used to infer causality in the Gene Regulatory Network (a mathematical point of view).**

Inferring a causal link between objects is useful in many applications in plant biology, from genomics to ecology. If some object A can cause some object B to take on a high value (where A could be a gene in our context, a hormone, or a species in ecology), then preventing B from taking such a value can be done by i) removing some B, ii) removing some A, or iii) interfering

with the link from A to B. Conversely, making B achieve a higher value can be done by i) adding more B, ii) adding more A, or iii) enhancing the efficiency of the link from A to B. Commonly, causal relationships in biology may involve several elements A1, ..., Ak influencing some B, sometimes positively and sometimes negatively. The influences can be "linear" in which each element has either a positive or negative weight (or coefficient) or "non-linear" in which case the elements work synergistically. An example of synergy would be a dependency of B on the product of the concentrations of some genes X and Y.

Generally, simpler models scale to larger numbers of genes, but are less informative as summarized by the classes of network inference methods listed in Table 1. Virtually all approaches deteriorate as the size of networks becomes larger, some more than others. Fortunately, biology tends to be modular, so large analyses can be broken down into smaller ones and then recombined [5].

The approaches to network inference fall into the following categories, that can be classidified based on level of information richness (low, medium and high) and scalability of the derived network (large, medium and small networks), as shown in Table 1.

Correlation techniques are scaleable to thousands of genes, but are based on low information richness (see Table 1). **Correlation techniques** are techniques that try to find single source-target relationships. To try to isolate the effects of one gene on another, many researchers make use of partial correlations. Schaefer and Strimmer (2005) [32] and Ingkasuwan et al. (2012) [28] present an analysis of **Graphical Gaussian Models**. These models assume a Gaussian noise distribution and try to infer partial correlations (gene X influences gene Y while holding the effects of other genes constant). **Partial correlations** can be computed indirectly by computing regressions and then computing the correlations among the residuals. Such analyses require heuristic approximations for large networks because the number of experiments (e.g. microarrays) is always far less than the number of genes. Thus, partial correlation approaches can result in medium sized networks (up to 100 genes) (Table 1).

Like correlation, **Mutual Information** [27] seeks pairwise relationships among variables but without assumptions of linear or rank dependencies. Also like correlation, mutual information can be used for large scale networks and does not try to compute the weight of influence of one gene on another in predicting the target's expression value.

Use of **Ordinary Differential Equations** (ODE), often based on mass action, yields equations of the form: change in Gene A concentration = Synthesis rate - Decay rate. Such approaches work especially well for small, information rich networks such as the auxin networks mentioned above [5, 33]. An issue with the mass action approach is that it assumes that different inputs interact in a multiplicative manner (product of concentration of each component), whereas the interaction is likely more complex in biological as opposed to chemical settings.

An alternative approach to network inference is to use a **Boolean approach,** which allows other logical relationships among regulators and their targets [5, 10, 12]. Logic gates are based on thresholds, e.g. an "AND gate" will have an effect on target if the minimum input reaches a certain threshold, thus permitting non-linear relationships. These tend to work better on smaller networks than linear equations and better than multiplicative relationships in modeling regulation (Table 1).

Closely related to Boolean approaches are **Decision/Regression Tree** approaches that embody paths of threshold tests (where each path represents a Boolean conjunction of conditions) leading to a prediction (e.g. of expression values). GENIE3 (GEne Network Inference with Ensemble of Trees), is a regression tree algorithm that can be applied to steady state, time series, and/or mutational transcriptome data [34]. This approach has worked particularly well in DREAM3 (Dialogue for Reverse Engineering Assessments and Methods) competitions that use *in silico* data as benchmarks for validating the predictive power of inferred networks [35].

”Integrative genomic” techniques analyze how changes can cause divergent behavior over time [36]. The idea is that genes are in some steady state before some perturbation occurs and the technique follows the genes that change first, that change second, and so on to try to guess causality. This is the qualitative idea behind the differential equation approaches.

**Pipeline Approaches** typically combine different algorthims on different data tyes. For example, the Inferelator is a network inference approach that uses differential equation techniques and mutual information to integrate many different data types including steady state, time series, and mutation/perturbation data [37]. These algorithms treat knowledge in a pipelined fashion. Thus, if physical experiments show that a target gene Z has potential connections from X and Y but not from W, then only X and Y will be considered in the subsequent analysis. The time series-based inference algorithm then might use these potential edges to derive an ordinary differential equation model that may combine linear and non-linear terms. The result of such a pipeline is a set of equations that estimate the change in transcription level of a target gene based on transcriptional levels of other genes using time series data. Fig. 1 illustrates the concept of such pipeline approaches, which refine large, information poor networks into smaller information rich networks with predictive power.

Finally, other work importantly suggests trying many network inference methods in combination (Marbach et al. (2012) [20]) showing empirically that a combination of strategies often lead to the best network resolution supporting the widespread popular use of the **wisdom of crowds** concept.

**III) Validations of inferred GRNs (an experimentalist’s point of view).**

GRN modeling described in the above sections complements genetic studies and generates hypotheses for TF-target interactions to be tested, thus inspiring a new round of the systems biology cycle of high throughput experimentation for model validation and refinement (Fig. 1). A variety of methods have been used to uncover the global structure of gene networks by

inferring regulatory relationships between transcription factors (TFs) and their target genes from genomic data [6, 38-41], in particular transcriptional analysis and chromatin immuno-precipitation.

The most common approach has been TF perturbation in stable over-expression or knock-out/down lines followed by transcriptional analysis [42-45]. However, it remains unclear in such analyses whether changes in transcript levels are a direct consequence of TF manipulation, or whether these changes are caused by indirect or possibly pleiotropic effects. To overcome the limitation of this approach, several other techniques have been used to supplement transcriptional data including yeast-one-hybrid [38], and electrophoretic mobility shift assays [46-48]. However, while these methods can result in a significant enrichment of direct targets, they are often time consuming and not easily applicable to high-throughput analyses.

The introduction of ChIP-X, chromatin immunoprecipitation (ChIP) followed by next generation sequencing (ChIP-seq) or tiling array (ChIP-chip), has greatly improved the genome-wide identification of TF binding sites and has uncovered many potential direct targets [49-51]. Importantly though, ChIP-X reveals the binding of a TF onto a promoter, but does not indicate if this results in activation/repression of gene expression [52]. Therefore, ChIP-X has often been combined with genome-wide transcriptional analysis to characterize the primary targets of a TF [53-55].

Recently, novel combinations of these technologies have yielded vastly improved knowledge about TF→target interaction. For example, whole plant studies using dexamethasone (DEX)-inducible TF translocation into the nucleus followed by separate ChIP-X experiments identified target genes both bound and regulated by a TF of interest [56-58]. Another new technology was recently described by Bargmann et al [59] where a protoplast system combined with fluorescent activated cell-sorting (FACS) has been employed to scale-up validation of GRNs *in vivo*. Briefly, plant protoplasts are transformed with plasmid harboring a fluorescent selection marker together with the over-expression of a studied TF fused to GR (glucocorticoid receptor (from

rat)). Protoplasts co-treated with DEX and the protein synthesis inhibitor cycloheximide (CHX), which blocks secondary target response, results in the identification of only primary TF targets. This rapid technique makes possible high throughput investigations/validations of TFs and the GRNs they regulate in plants [59]. Data from such high throughput TF-target validations can then be fed back into network inference pipelines to refine predicted edges in the derived GRNs, in a true systems biology cycle.

**Conclusion**

Plant Systems Biology is at the beginning of a new era, in which machine learning techniques and experimental investigations mutually and iteratively reinforce one another. We believe that this experimental-analytical symbiosis will lead plant biologists to better and deeper insights into biological phenomenon and lead computer scientists to develop new algorithms. Together this symbiotic collaboration should accelerate the understanding of plants as system.

**Acknowledgements**: This work is supported by NIH NIGMS GRANT RO1 GM032877, NSF Grants MCB-0929338 & MCB 1158273 to GC and DS; by NIH-NRSA GM095273 to AMC; Grants from ANR (NitroNet: ANR 11 PDOC 020 01) and CNRS (PEPS Bio math Info 2012–2013: SuperRegNet) to G.K. We thank Becca Susko for her outstanding work on the manuscript preparation, and Benoit Lacombe and Sandrine Ruffel for critical reading and help.

**References:**


1. Bonneau R: Learning biological networks: from modules to dynamics. *Nat Chem Biol* 2008, 4:658-664.
2. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, et al: A predictive model for transcriptional control of physiology in a free living cell. *Cell* 2007, 131:1354-1365.
3. Robison K, McGuire AM, Church GM: A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome. *J Mol Biol* 1998, 284:241-254.
4. Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J: RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res* 2004, 32:D303-306.
5. Middleton AM, Farcot E, Owen MR, Vernoux T: Modeling regulatory networks to understand plant development: small is beautiful. *Plant Cell* 2012, 24:3876-3891.
6. Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, et al: The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol* 2011, 7:508.
7. Sankar M, Osmont KS, Rolcik J, Gujas B, Tarkowska D, Strnad M, Xenarios I, Hardtke CS: A qualitative continuous model of cellular auxin and brassinosteroid signaling and their crosstalk. *Bioinformatics* 2011, 27:1404-1412.
8. Havens KA, Guseman JM, Jang SS, Pierre-Jerome E, Bolten N, Klavins E, Nemhauser JL: A synthetic approach reveals extensive tunability of auxin signaling. *Plant Physiol* 2012, 160:135-142.
9. Pokhilko A, Fernandez AP, Edwards KD, Southern MM, Halliday KJ, Millar AJ: The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Mol Syst Biol* 2012, 8:574.
10. Akman OE, Watterson S, Parton A, Binns N, Millar AJ, Ghazal P: Digital clocks: simple Boolean models can quantitatively describe circadian systems. *J R Soc Interface* 2012, 9:2365-2382.
11. Salazar JD, Saithong T, Brown PE, Foreman J, Locke JC, Halliday KJ, Carre IA, Rand DA, Millar AJ: Prediction of photoperiodic regulators from quantitative gene circuit models. *Cell* 2009, 139:1170-1179.
12. Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER: A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 2004, 16:2923-2939.
13. La Rota C, Chopard J, Das P, Paindavoine S, Rozier F, Farcot E, Godin C, Traas J, Moneger F: A data-driven integrative model of sepal primordium polarity in Arabidopsis. *Plant Cell* 2011, 23:4318-4333.
14. Jaeger KE, Pullen N, Lamzin S, Morris RJ, Wigge PA: Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis. *Plant Cell* 2013, 25:820-833.

15. Bujdoso N, Davis SJ: Mathematical modeling of an oscillating gene circuit to unravel the circadian clock network of Arabidopsis thaliana. *Front Plant Sci* 2013, 4:3.
16. Locke JC, Millar AJ, Turner MS: Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. *Journal of theoretical biology* 2005, 234:383-393.
17. Pokhilko A, Hodge SK, Stratford K, Knox K, Edwards KD, Thomson AW, Mizuno T, Millar AJ: Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Mol Syst Biol* 2010, 6:416.
18. Pruneda-Paz JL, Kay SA: An expanding universe of circadian networks in higher plants. *Trends Plant Sci* 2010, 15:259-265.
19. Coen ES, Meyerowitz EM: The war of the whorls: genetic interactions controlling flower development. *Nature* 1991, 353:31-37.
20. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Consortium D, Kellis M, Collins JJ, Stolovitzky G: Wisdom of crowds for robust gene network inference. *Nat Methods* 2012, 9:796-804.
21. Ruffel S, Krouk G, Coruzzi GM: A systems view of responses to nutritional cues in Arabidopsis: toward a paradigm shift for predictive network modeling. *Plant Physiol* 2010, 152:445-452.
22. Mao L, Van Hemert JL, Dash S, Dickerson JA: Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* 2009, 10:346.
23. Gutierrez RA, Lejay LV, Dean A, Chiaromonte F, Shasha DE, Coruzzi GM: Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis. *Genome Biol* 2007, 8:R7.
24. Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE, et al: VirtualPlant: a software platform to support systems biology research. *Plant Physiol* 2010, 152:500-515.
25. Gutierrez RA, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR, Coruzzi GM: Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. *Proc Natl Acad Sci U S A* 2008, 105:4939-4944.
26. Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, et al: Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences* 2009, 106:10348-10353.
27. Carrera J, Rodrigo G, Jaramillo A: Model-based redesign of global transcription regulation. *Nucleic Acids Res* 2009, 37:e38.
28. Ingkasuwan P, Netrphan S, Prasitwattanaseree S, Tanticharoen M, Bhumiratana S, Meechai A, Chaijaruwanich J, Takahashi H, Cheevadhanarak S: Inferring transcriptional gene regulation network of starch metabolism in Arabidopsis thaliana leaves using graphical Gaussian model. *BMC Syst Biol* 2012, 6:100.
29. Ma S, Gong Q, Bohnert HJ: An Arabidopsis gene network based on the graphical Gaussian model. *Genome Res* 2007, 17:1614-1625.
30. Breeze E, Harrison E, McHattie S, Hughes L, Hickman R, Hill C, Kiddle S, Kim YS, Penfold CA, Jenkins D, et al: High-resolution temporal profiling of transcripts during Arabidopsis

leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* 2011, 23:873-894.
31. Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM: Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genome Biol* 2010, 11:R123.
32. Schäfer J, Strimmer K: An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 2005, 21:754-764.
33. Yuan J, Doucette CD, Fowler WU, Feng XJ, Piazza M, Rabitz HA, Wingreen NS, Rabinowitz JD: Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli. *Mol Syst Biol* 2009, 5:302.
34. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 2010, 5.
35. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, Clarke ND, Altan-Bonnet G, Stolovitzky G: Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS One* 2010, 5:e9202.
36. Mendoza-Parra MA, Walia M, Sankar M, Gronemeyer H: Dissecting the retinoid-induced differentiation of F9 embryonal stem cells by integrative genomics. *Mol Syst Biol* 2011, 7:538.
37. Greenfield A, Hafemeister C, Bonneau R: Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* 2013.
38. Brady SM, Zhang L, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeples M, Kim D, et al: A stele-enriched gene regulatory network in the Arabidopsis root. *Mol Syst Biol* 2011, 7:459.
39. Chew YH, Halliday KJ: A stress-free walk from Arabidopsis to crops. *Curr Opin Biotechnol* 2011, 22:281-286.
40. Edwards MA, Whitworth AL, Unwin PR: Quantitative analysis and application of tip position modulation-scanning electrochemical microscopy. *Anal Chem* 2011, 83:1977-1984.
41. Petricka JJ, Benfey PN: Reconstructing regulatory network transitions. *Trends Cell Biol* 2011, 21:442-451.
42. Suzuki M, Ketterling MG, Li QB, McCarty DR: Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol* 2003, 132:1664-1677.
43. Nakabayashi K, Okamoto M, Koshiba T, Kamiya Y, Nambara E: Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J* 2005, 41:697-709.
44. Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis. *Plant Mol Biol* 2006, 60:51-68.
45. Carrera E, Holman T, Medhurst A, Dietrich D, Footitt S, Theodoulou FL, Holdsworth MJ: Seed after-ripening is a discrete developmental pathway associated with specific gene networks in Arabidopsis. *Plant J* 2008, 53:214-224.
46. Ryu KH: The WEREWOLF MYB protein directly regulates CAPRICE transcription during cell fate specification in the Arabidopsis root epidermis. *Development* 2005, 132:4765-4775.

47. Reeves WM, Lynch TJ, Mobin R, Finkelstein RR: Direct targets of the transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors. *Plant Molecular Biology* 2011, 75:347-363.
48. Bustos R, Castrillo G, Linhares F, Puga MI, Rubio V, Perez-Perez J, Solano R, Leyva A, Paz-Ares J: A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in Arabidopsis. *PLoS genetics* 2010, 6.
49. Kuo MH, Allis CD: In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 1999, 19:425-433.
50. de Folter S, Urbanus SL, van Zuijlen LG, Kaufmann K, Angenent GC: Tagging of MADS domain proteins for chromatin immunoprecipitation. *BMC Plant Biol* 2007, 7:47.
51. Zhu JY, Sun Y, Wang ZY: Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq). *Methods Mol Biol* 2012, 876:173-188.
52. Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD: Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* 2012, 484:251-255.
53. Oh E, Kang H, Yamaguchi S, Park J, Lee D, Kamiya Y, Choi G: Genome-Wide Analysis of Genes Targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during Seed Germination in Arabidopsis. *THE PLANT CELL ONLINE* 2009, 21:403-419.
54. Wang F, Perry SE: Identification of Direct Targets of FUSCA3, a Key Regulator of Arabidopsis Seed Development. *PLANT PHYSIOLOGY* 2013, 161:1251-1264.
55. Hsieh WP, Hsieh HL, Wu SH: Arabidopsis bZIP16 Transcription Factor Integrates Light and Hormone Signaling Pathways to Regulate Early Seedling Development. *The Plant Cell* 2012, 24:3997-4011.
56. Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U, Junker A, Weisshaar B, Conrad U, Baumlein H, Altschmied L: Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Res* 2012, 40:8240-8254.
57. Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE: Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell* 2009, 21:2563-2577.
58. Gorte M, Horstman A, Page RB, Heidstra R, Stromberg A, Boutilier K: Microarray-Based Identification of Transcription Factor Target Genes. In *Plant Transcription Factors. Volume* 754. Edited by Yuan L, Perry SE. Totowa, NJ: Humana Press; 2011: 119-141
59. Bargmann BO, Marshall-Colon A, Efroni I, Ruffel S, Birnbaum KD, Coruzzi GM, Krouk G: TARGET: A Transient Transformation System for Genome-wide Transcription Factor Target Discovery. *Mol Plant* 2013.

**TABLE & FIGURE LEGENDS**

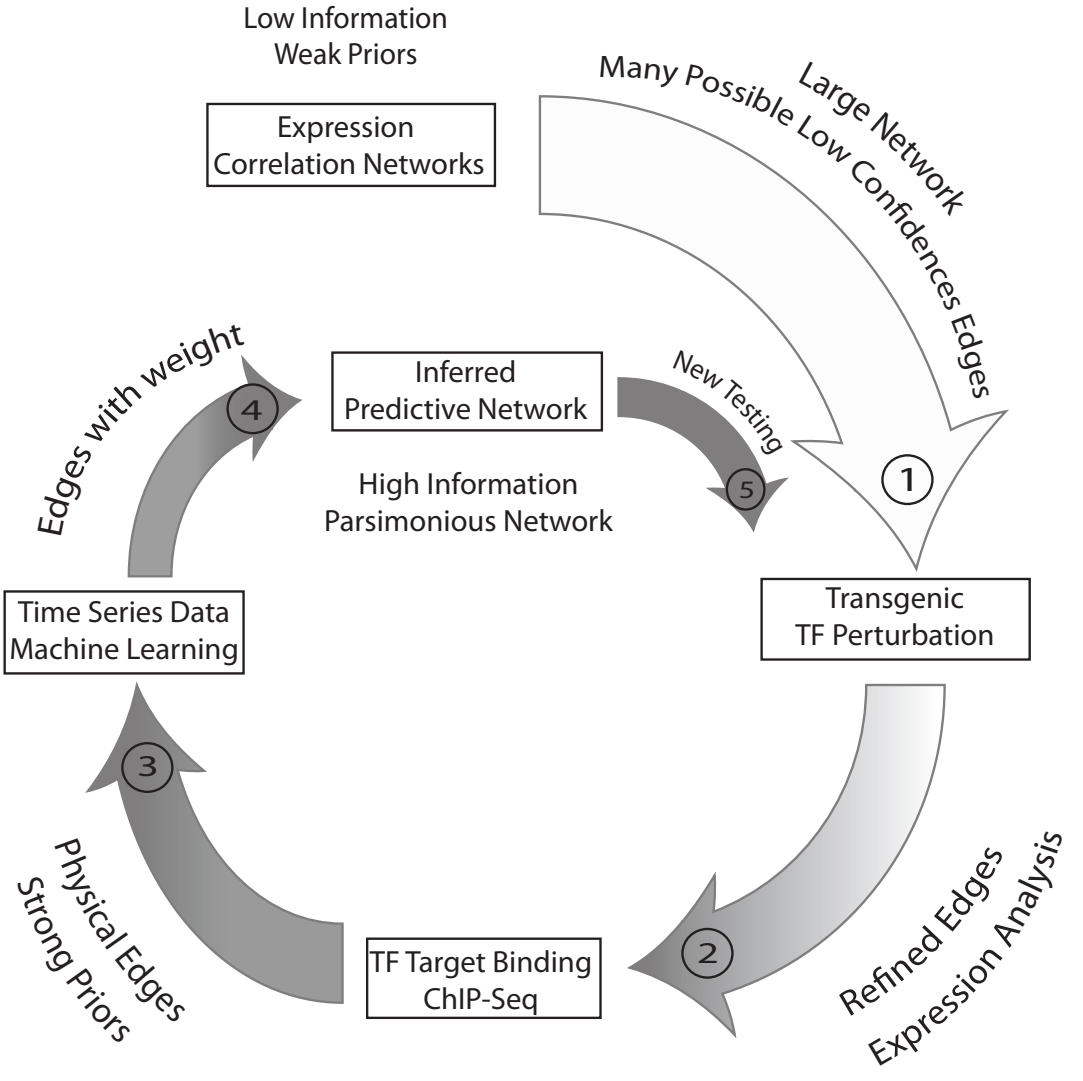
| Methods | Information Richness | Scalability | References |
|---|---|---|---|
| ***Correlation /Mutual Information*** | Low | High (thousands of genes) | [20, 27] |
| ***Partial Correlation*** | Medium | Medium (up to 100 genes using heuristics) | [28, 32] |
| ***Differential equations*** | Medium | Medium | [2, 31, 33, 35] |
| ***Linear regression*** | Medium | Medium | [37] |
| ***Non-linear regression*** | High | Low (up to 25 genes) | [37] |
| ***Boolean*** | High | Low (up to 25 genes) | [11, 34] |

**Table 1. Methods for Network Inference.**

Trade-off between information richness (the number of factors that can be applied to predict gene expression) and the size of the analyzed network. Small networks can be handled by methods that are highly complex and information rich (many linear and non-linear factors can influence a gene within the method). Combining several small network modules holds the potential to analyze a large network [5], though this may not always work.

**Figure 1. An experimental/computational Systems Biology cycle using different data types and feedback.**

Starting from many possible edges, different data types and their analyses successively reduce the size of the network while increasing confidence in edges. 1. Correlation leads to pairwise associations of genes. 2. Transgenics permits the determination of the effect of mutations and over-expression of single genes. 3. Binding experiments (e.g. Chip-SEQ) reveals physical connectivity of a source gene to a target. 4. Time series experiments along with machine learning techniques lead to a weighted network where the weight on the edge from A to B determines the extent of influence of A on B. 5. Subsequent predictions followed by validations may then suggest the need for new experimentation, refueling the systems biology cycle.


Low Information
Weak Priors
Expression
Correlation Networks
Many Possible Low Confidences Edges
Large Network
1
Transgenic
TF Perturbation
2
Refined Edges
Expression Analysis
TF Target Binding
ChIP-Seq
3
Physical Edges
Strong Priors
Time Series Data
Machine Learning
4
Edges with weight
Inferred
Predictive Network
High Information
Parsimonious Network
5
New Testing