

OPINION

# Gene regulatory networks in plants: learning causality from time and perturbation

Gabriel Krouk<sup>3</sup>, Jesse Lingeman<sup>1</sup>, Amy Marshall Colon<sup>2</sup>, Gloria Coruzzi<sup>2</sup> and Dennis Shasha<sup>1\*</sup>

## Abstract

The goal of systems biology is to generate models for predicting how a system will react under untested conditions or in response to genetic perturbations. This paper discusses experimental and analytical approaches to deriving causal relationships in gene regulatory networks.

**Keywords:** Gene regulatory networks, network interference, plant, systems biology

## Introduction

Owing 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, for both scientific and practical gain, how such regulatory networks will respond under untested conditions. In prokaryotes, models to infer gene-regulatory networks (GRNs) 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 multicellular organisms, including plants. With the increasing availability of high-throughput 'omic' techniques and data, we think it useful to summarize both experimental and informatic approaches for inferring causal relationships in GRNs. Here, we use the term GRN to refer to the set of transcriptional interactions between transcription factors (TFs) and their targets, as opposed to a multimodal set of gene-to-gene or gene-to-metabolite interactions.

Here, we have three aims: first, to summarize efforts to use time-series and other -omic data to infer causal

regulatory edges, showing the kinds of biological insights that can be obtained; next to provide a description and a categorization of the informatic methods that are being used to infer causal networks; and finally to discuss recent high-throughput experimental techniques to validate inferred GRNs in plants.

## Successful case studies of learning gene-regulatory networks in plants

Different kinds of systems approaches are used to model GRNs in plants. One way of characterizing these systems approaches is dependent on whether or not they start with a significant amount of prior experimental knowledge of the connectivity of the modeled GRN. Thus, in this article, we call them '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 previous knowledge is paradigmatically derived using differential-equation systems and Boolean models (described below). Outputs of the models are then compared with experimental data to determine their predictive power. When the predictions hold, the models can be used to explore GRN behavior in untested conditions *in silico* and to determine the overall system properties and architecture. These kinds of investigations have led to some striking results, as discussed below.

For auxin signaling, Vernoux and colleagues [6] built a model based on previous knowledge of the auxin/indole-3-acetic acid and auxin response factor (AUX/IAA-ARF) transcription factor network and yeast two-hybrid experiments (taking into account the possibility of interactions between the protein partners). This ordinary differential equation model demonstrated that the resulting GRN shows a strong buffering capacity as the transcriptional induction of auxin-induced genes is stabilized even when auxin inputs display strong variations. This property was experimentally revealed *in planta*, in the shoot apical

\*Correspondence shasha@courant.nyu.edu

<sup>1</sup>Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA

Full list of author information is available at the end of the article

meristem, by using the fluorescent sensor DII-VENUS as a reporter of the input of the signaling pathway and the *DR5* reporter gene as the output.

The circadian clock is also a well-studied gene-regulatory system (for a comprehensive review, see Bujdosó and Davis [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 crucial components involved in some key features of the oscillations. For instance, in studies by Pokhilko and colleagues [17], the GRN model was central to the discovery of the role of *PRR5* as a night inhibitor of the expression of *LHY/CCA1*, including the role for *PRR5* in the control of the phase of morning gene expression. In the same work, this GRN-generated hypothesis was validated by matching the behavior of *prp5* mutants to gene expression predicted by the model [17]. In an alternative approach, Akman *et al.* [10] used Boolean logic to describe circadian circuits in a quantitative model. The simplified model with decreased parameterization was able to simulate observed circadian oscillations accurately and identify regulatory structures consistent with experimental data.

Flower development (described by the ABC model) is a textbook example of a conserved GRN that controls the fate of cells becoming 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 formation of floral organ primordia 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 cell-fate-associated 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 (for example, making 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 primordia [13] as well as for floral transition [14].

These few examples of successful Strong Prior approaches demonstrate that GRNs 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 unearthed.

Weak Prior approaches infer potential connections in GRNs from -omic datasets. Many techniques are used to infer unknown networks in the field of systems biology (for reviews, see [1,20,21]). These techniques have enjoyed great success in simpler systems, such as for bacteria. For instance, a striking success story is the model of gene-regulatory programs built from a multi-level dataset (including transcriptomic data and *cis*-regulatory element (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 the 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. In plant science, as this eukaryotic system is far more complex than that of yeast or bacteria, the field of GRN *de novo* learning is far less advanced [22]. However, Weak Prior approaches have been developed with some success, as described below.

In the plant field of GRN modeling, the three most popular top-down approaches are: (i) classical correlations networks (in combination with other information to establish causality), (ii) graphical Gaussian models (based on partial correlation) and (iii) machine-learning modeling, or combinations of the above.

Correlation networks have been used extensively to study GRNs in plants even if, by themselves, they do not directly determine causality in networks [23]. 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-to-CRE interactions, protein-protein interactions and microRNA-mRNA interactions [24]. Correlation data integrated with the *Arabidopsis* multi-network have uncovered biomodules involved in carbon/nitrogen signal integration [25] and have also revealed a central role for *CCA1*, the central component of the circadian clock in nutrient control [26]. Additionally, correlation network approaches were strikingly successful in identifying two genes (encoding a myo-inositol-1-phosphate synthase and a Kelch-domain protein) correlating with biomass accumulation in plants [27]. The individual role of these two genes was further supported by an association-mapping study that demonstrated coherent allelic diversity at their loci [27].

Graphical Gaussian models can be viewed as an approximate method to find partial correlation networks. Partial correlation is a measure of correlation between pairs while controlling for other factors. Mathematically,

if A, B and C correlate together, partial correlation correlates A and B by 'subtracting' the correlation due to A and C. Practically, partial correlation is the correlation between the residuals resulting from the linear regression of A with C, and of B with C. Graphical Gaussian models have been successfully developed [28] and applied to plant GRNs [29,30]. Ingkasuwan and colleagues analyzed a time-series to identify genes regulated across the diurnal cycle [29]. Then a sub-network of starch-metabolism genes together with the diurnally regulated TFs were modeled using graphical Gaussian models. This model was tested and validated by studying regulator mutants that displayed starch granule defects in plastids [29].

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 [31] and GRNs involved in regulating early, time-dependent transcriptional responses to  $\text{NO}_3^-$  [32]. Breeze and colleagues [31] provided a high-resolution temporal picture of the transcriptome of the aging leaf. 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 between genes (representing causal relationships) that remain to be validated [31]. In another study [32], state-space modeling and machine learning were applied to an *Arabidopsis* high-resolution time-course of genome-wide transcriptional response to treatments with  $\text{NO}_3^-$ . A subset of TFs and nitrogen transport and assimilation genes has been modeled in order to propose a GRN that explains  $\text{NO}_3^-$  signal propagation. The model has been tested *in silico* as well as experimentally. *In silico* validation demonstrated that the model trained on the early time points of the time-series experiment is able to predict modulation of gene expression at later time-points (not used to train the model). Experimental validation consisted of studying the effect of overexpressing a predicted hub (SPL9 TF) on the  $\text{NO}_3^-$  response of other  $\text{NO}_3^-$ -regulated genes. Indeed, SPL9 overexpression modified the regulation of the nitrate assimilation gene *NIA2* but also of many genes encoding  $\text{NO}_3^-$ -regulated TFs [32].

### **Analytical approaches used to infer causality in the gene-regulatory network (a mathematical point of view)**

Inferring a causal edge between objects is useful in many applications in plant biology, from genomics to ecology.

If some population of objects A can cause an increase in the population of object B (where A could be a gene in our context, a hormone or a species in ecology), then lowering the population of B can be achieved by: (i) removing some members of B, (ii) removing some members of A or (iii) interfering with the edge from A to B. Conversely, making B achieve a higher population can be achieved by: (i) adding more members of B, (ii) adding more members of A or (iii) enhancing the efficiency of the edge from A to B. Commonly, causal relationships in biology can involve several elements, for example A1 to A5, influencing some B, sometimes positively and sometimes negatively. The influences can be 'linear', in which case 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, which can be classified 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. High information richness would, for example, allow the inference of the dynamic behavior of a network [21], whereas low information richness would give some approximation to the connectivity of a causal GRN.

Correlation techniques are techniques that try to find single source-target relationships. To try to isolate the possibly mutual influence of one gene on another, many researchers make use of partial correlations. Schaefer and Strimmer [33] and Ingkasuwan *et al.* [29] have presented 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 calculating regressions and correlations among the residuals. Such analyses require heuristic approximations for large networks because the number of experiments (for example, microarrays) is always far fewer 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' [28] seeks pairwise relationships among variables without assumptions of linear or rank dependencies. Also, like correlation, mutual

**Table 1. Methods for network inference**

| Methods                        | Information richness | Scalability                               | References   |
|--------------------------------|----------------------|---|--------------|
| Correlation/mutual information | Low                  | High (thousands of genes)                 | [20,28]      |
| Partial correlation            | Medium               | Medium (up to 100 genes using heuristics) | [29,33]      |
| Differential equations         | Medium               | Medium                                    | [2,32,34,36] |
| Linear regression              | Medium               | Medium                                    | [38]         |
| Non-linear regression          | High                 | Low (up to 25 genes)                      | [38]         |
| Boolean                        | High                 | Low (up to 25 genes)                      | [11,35]      |

It is clear that there is a 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], although this might not always work.

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 expression value of the target.

Use of differential equations, often based on mass action, yields equations of the form: Rate of 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 [5,34]. 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 to be 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 - for example, 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 (for example, of expression values). 'Gene network inference with ensemble of trees 3' (GENIE3) is a regression tree algorithm that can be applied to steady-state, time-series and/or mutational transcriptome data [35]. This approach has worked particularly well in 'dialogue for reverse engineering assessments and methods 3' (DREAM3) competitions that use *in silico* data as benchmarks for validating the predictive power of inferred networks [36].

'Integrative genomic' techniques analyze how changes can cause divergent behavior over time [37]. 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

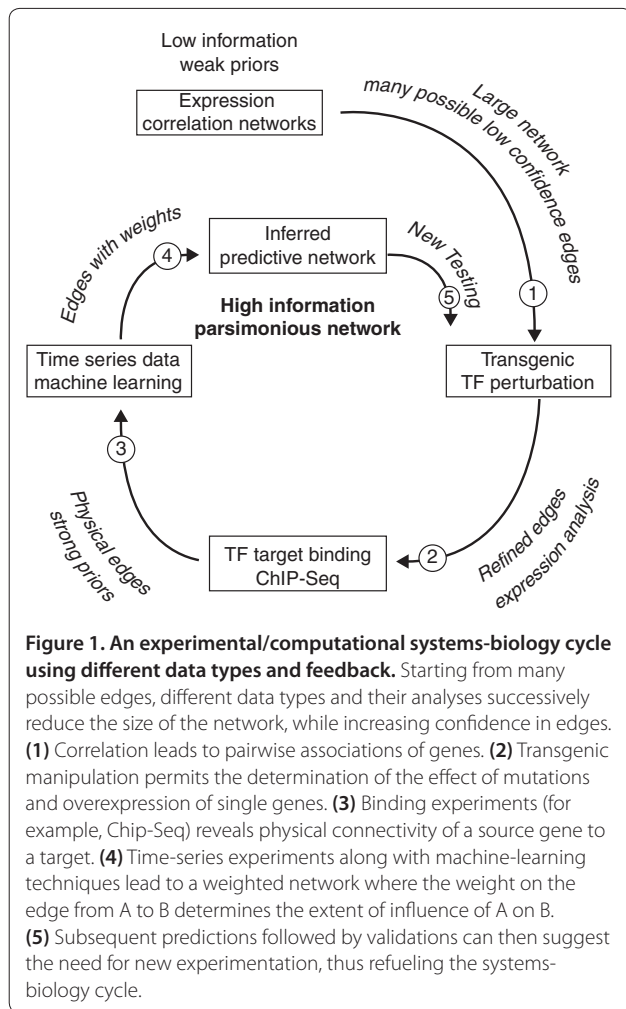
causality. This is the qualitative idea behind the differential-equation approaches.

Pipeline approaches typically combine different algorithms on different data types. For example, the Inferator 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 [38,39]. 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 can 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. Figure 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 [20], showing empirically that a combination of strategies often leads to the best network resolution and supporting the widespread popular use of the 'wisdom of crowds' concept.

### 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 (Figure 1). A variety of methods have been used to uncover the global structure of gene networks by inferring regulatory relationships between TFs and their target genes from genomic data [6,40-43], in particular transcriptional analysis and chromatin immuno-precipitation.



The most common approach has been TF perturbation in stable overexpression or knockout/knockdown lines, followed by transcriptional analysis [44-47]. 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 assays [40] and electrophoretic mobility-shift assays [48-50]. 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 [51-53]. Importantly, although ChIP-X reveals the binding of a TF onto a promoter, it does not indicate whether this

results in activation/repression of gene expression [54]. Therefore, ChIP-X has often been combined with genome-wide transcriptional analysis to characterize the primary targets of a TF [55-57].

Recently, novel combinations of these technologies have yielded vastly improved knowledge about interactions between TFs and their targets. 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 [58-60]. Another new technology was recently described by Bargmann and colleagues [61] in which a protoplast system combined with fluorescence-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 overexpression of a TF of interest fused to a glucocorticoid receptor from rat. Co-treatment of protoplasts with DEX and the protein synthesis inhibitor cycloheximide, which blocks secondary-target responses, results in the identification of only primary TF targets. This rapid technique makes it possible to perform high-throughput investigations/validations of TFs and the GRNs they regulate in plants [61]. 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 (Figure 1).

## Perspectives

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 phenomena and will encourage computer scientists to develop new algorithms. Together, this symbiotic collaboration should accelerate the understanding of plants as systems.

## Abbreviations

ChIP, chromatin immunoprecipitation; CRE, *cis*-regulatory element; GRN, gene-regulatory network; ODE, ordinary differential equation; TF, transcription factor.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

This work is supported by NIH NIGMS GRANT RO1 GM032877, NSF Grants MCB-0929338 and 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 GK. We thank Becca Susko for her outstanding work on the manuscript preparation, and Benoit Lacombe and Sandrine Ruffel for critical reading and help.

## Author details

<sup>1</sup>Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA. <sup>2</sup>Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, NY 10003, USA.

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

Published: 27 June 2013

## References

- Bonneau R: **Learning biological networks: from modules to dynamics.** *Nat Chem Biol* 2008, **4**:658-664.
- Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, Longabaugh W, Vuthoori M, Whitehead K, Madar A, Suzuki L, Mori T, Chang DE, Diruggiero J, Johnson CH, Hood L, Baliga NS: **A predictive model for transcriptional control of physiology in a free living cell.** *Cell* 2007, **131**:1354-1365.
- 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.
- 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.
- Middleton AM, Farcot E, Owen MR, Vernoux T: **Modeling regulatory networks to understand plant development: small is beautiful.** *Plant Cell* 2012, **24**:3876-3891.
- Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, Guédou Y, Armitage L, Picard F, Guyomarc'h S, Cellier C, Parry G, Koumproglou R, Doonan JH, Estelle M, Godin C, Kepinski S, Bennett M, De Veylder L, Traas J: **The auxin signalling network translates dynamic input into robust patterning at the shoot apex.** *Mol Syst Biol* 2011, **7**:508.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- La Rota C, Chopard J, Das P, Païndavoine 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.
- 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.
- Bujdosó 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.
- Locke JC, Millar AJ, Turner MS: **Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*.** *J Theor Biol* 2005, **234**:383-393.
- 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.
- Pruneda-Paz JL, Kay SA: **An expanding universe of circadian networks in higher plants.** *Trends Plant Sci* 2010, **15**:259-265.
- Coen ES, Meyerowitz EM: **The war of the whorls: genetic interactions controlling flower development.** *Nature* 1991, **353**:31-37.
- 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.
- Albert R: **Network inference, analysis, and modeling in systems biology.** *Plant Cell* 2007, **19**:3327-3338.
- 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.
- Mao L, Van Hemert JL, Dash S, Dickerson JA: ***Arabidopsis* gene co-expression network and its functional modules.** *BMC Bioinformatics* 2009, **10**:346.
- 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.
- Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE, Coruzzi GM, Gutiérrez RA: **VirtualPlant: a software platform to support systems biology research.** *Plant Physiol* 2010, **152**:500-515.
- 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.
- Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, Von Korff M, Steinhauser MC, Keurentjes JJ, Guenther M, Hoehne M, Selbig J, Fernie AR, Altmann T, Stitt M: **Starch as a major integrator in the regulation of plant growth.** *Proc Natl Acad Sci U S A* 2009, **106**:10348-10353.
- Carrera J, Rodrigo G, Jaramillo A: **Model-based redesign of global transcription regulation.** *Nucleic Acids Res* 2009, **37**:e38.
- Ingkasuwan P, Netphan S, Prasitwattanaseree S, Tanticharoen M, Bhumiratanas S, Meechai A, Chajjaruwanich 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.
- Ma S, Gong Q, Bohnert HJ: **An *Arabidopsis* gene network based on the graphical Gaussian model.** *Genome Res* 2007, **17**:1614-1625.
- Breeze E, Harrison E, McHattie S, Hughes L, Hickman R, Hill C, Kiddle S, Kim YS, Penfold CA, Jenkins D, Zhang C, Morris K, Jenner C, Jackson S, Thomas B, Tabrett A, Legaie R, Moore JD, Wild DL, Ott S, Rand D, Beynon J, Denby K, Mead A, Buchanan-Wollaston V: **High-resolution temporal profiling of transcripts during *Arabidopsis* leaf senescence reveals a distinct chronology of processes and regulation.** *Plant Cell* 2011, **23**:873-894.
- 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.
- Schäfer J, Strimmer K: **An empirical Bayes approach to inferring large-scale gene association networks.** *Bioinformatics* 2005, **21**:754-764.
- 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.
- Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods.** *PLoS One* 2010, **5**:e12776.
- 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.
- 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.
- Greenfield A, Hafemeister C, Bonneau R: **Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks.** *Bioinformatics* 2013, **29**:1060-1067.
- Lingeman JM, Shasha D: *Network Inference in Molecular Biology: A Hands-on Framework.* New York: Springer; 2012.
- Brady SM, Zhang L, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeples M, Kim D, Ahnert S, Ohler U, Ware D, Walhout AJ, Benfey PN: **A stele-enriched gene regulatory network in the *Arabidopsis* root.** *Mol Syst Biol* 2011, **7**:459.
- Chew YH, Halliday KJ: **A stress-free walk from *Arabidopsis* to crops.** *Curr Opin Biotechnol* 2011, **22**:281-286.
- Edwards MA, Whitworth AL, Unwin PR: **Quantitative analysis and**

- application of tip position modulation-scanning electrochemical microscopy. *Anal Chem* 2011, **83**:1977-1984.
43. Petricka JJ, Benfey PN: **Reconstructing regulatory network transitions.** *Trends Cell Biol* 2011, **21**:442-451.
  44. 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.
  45. 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.
  46. 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.
  47. 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.
  48. Ryu KH: **The WEREWOLF MYB protein directly regulates CAPRICE transcription during cell fate specification in the *Arabidopsis* root epidermis.** *Development* 2005, **132**:4765-4775.
  49. 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 Mol Biol* 2011, **75**:347-363.
  50. 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 Genet* 2010, **6**.
  51. 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.
  52. 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.
  53. 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.
  54. 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.
  55. 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*.** *Plant Cell Online* 2009, **21**:403-419.
  56. Wang F, Perry SE: **Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development.** *Plant Physiol* 2013, **161**:1251-1264.
  57. Hsieh WP, Hsieh HL, Wu SH: ***Arabidopsis* bZIP16 transcription factor integrates light and hormone signaling pathways to regulate early seedling development.** *Plant Cell* 2012, **24**:3997-4011.
  58. 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.
  59. 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.
  60. 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.
  61. Bargmann BO, Marshall-Colon A, Efroni I, Ruffel S, Birnbaum KD, Coruzzi GM, Krouk G: **TARGET: A transient transcription system for genome-wide transcription factor target discovery.** *Mol Plant* 2013, **6**:978-980.

doi:10.1186/gb-2013-14-6-123

Cite this article as: Krouk G, *et al.*: Gene regulatory networks in plants: learning causality from time and perturbation. *Genome Biology* 2013, **14**:123.