Building Gene Regulatory Network: experimental approach

Introduction

Developmental gene regulatory networks (dGRNs) plays a crucial role in the development of every single individual, especially multicellular. They regulate developmental processes not only during embryogenesis, but also affect post-embrionic morphogenetic processes like regeneration, asexual reproduction and growth.

GRN can be viewed as highly complicated network of protein-gene interactions. As a proteins in most of the cases act transcription factors (TFs). TFs are the proteins which has special DNA-binding motifs in their structure. This allows them to bind to regulatory elements in adjacent to gene regions (cis-regulatory modules) and regulate gene expression. Depending on type of the interaction (activation/repression) there are can be several outcomes. If TF binds to enchancer region, gene expression increase. Opposite to that, if protein binds to silencer region then gene decrease its expression. Moreover transcription factors may also recruit chromatin remodelers and histone modifiers. Also there can be more complicated types of interaction and complex GRNs can be composed of genes, noncoding RNAs, proteins, metabolites, and signaling components. As a result most of the processes during development can be seen as complex network with lots of different interactions. Such networks usually illustrated using graphs. Vertices show genes, TFs or signal molecules, edges illustrate types of interaction. Such graphs in some cases reach vary big complexity. Therefore such networks are studied by systems biology. Only systems approach allows one to undersytand, not only separate components, but also process as a whole and reveal emergent properties of a system. As a consequense arise a questions: how can we reconstruct such networks? What kind of data should we use? And what are the potencial outcomes from reasearch of such gene interactions?

The purpose of this review was to demonstrate different approaches in building gene regulatory networks. Special attention was paid to developmental GRNs.

Developmental gene regulatory networks

Term ***developmental gene regulatory networks*** (dGRNs) was first introduced by E. H. Davidson. Usually this term refers to a GRN which preferentially consists of genes crucial for the development.

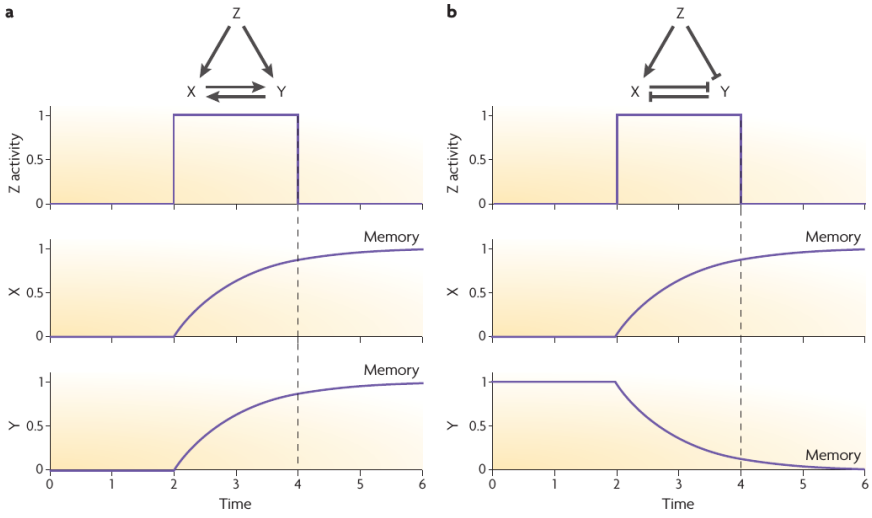

Figure 1: Two types of positive-feedback loops. Graph with curves represent activity (mRNA concentration) of the X,Y and Z genes. a — Network motifs with a double-positive-feedback loop. b — Regulated feedback with a double-negative-feedback loop.

Developmental GRNs share lots of common features with other types of gene networks [1]. For example lots of networks typically has the same recurring regulation patterns as positive autoregulation (PAR), negative autoregulation (NAP), feedforward loops (FFLs) etc. This common patterns called ***network motifs***. It is recurring circuits of interactions from which the gene networks are built. Also usually GRNs have hierarchical structure [2]. It means that those genes that control the initial stages of development are at the top of the hierarchy and those genes that execute the detailed functions of cell differentiation are at the periphery. Consequently, changes in upstream genes have far larger effect on the final body plan than in downstream. For example changes in gap genes in early development of *Drosofila melanogaster* are more crutial for the develpment than changes in genes which regulate eyes pigmentation. There are lots of others similarities. But at the same time developmental GRNs have their own characteristic features.

First of all, they tend to have longer cascades than some other GRNs. It can be explained by the role of pathways. Often they require transduce signal through the whole process of development, through lots of cell generations. That is why timescale of such process is very slow. Sometimes it is one cell generation at each cascade step. And that brings us to the second characteristic feature of developmental GRNs. They can functioning even after input signal has vanished. The signal can be inherited during cell division and subsequently

be transduced to next cell generations. In this case often used repressor cascades, since they are more tolerant to concentration biases of signal molecule. Such signal inheritance occures because of ***positive-feedback loops*** (PFLs). PFL is a process in which two events are mutually reinforcing each other. There are two kind of PFLs: double-positive loop and a double-negative loop (fig. 1).

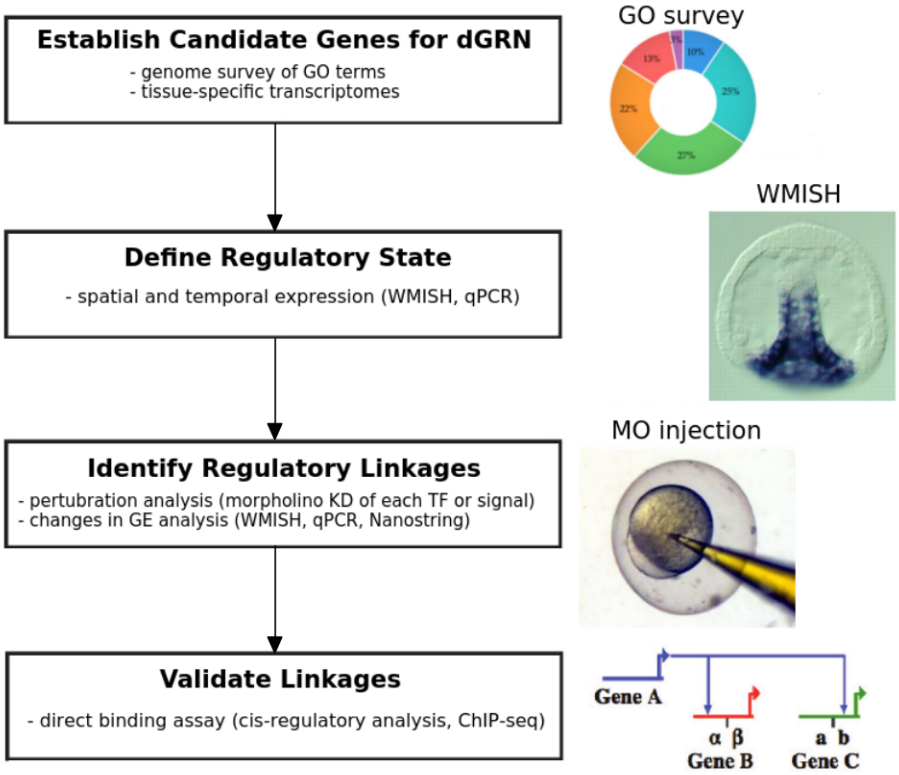

Figure 2: Steps in construction and validation of developmental gene regulatory networks (dGRNs).

In doble-positive-feedback loop (fig.1a) when signal molecule Z is activated, proteins X and Y begin to be produced. After that X and Y develop self-sustained loop and mutually activate each other. So than even if Z is discarded (dashed line), loop A-B continue existinig.

In doble-negative-feedback loop (fig.2a), initially, there is a high concentration of Y and it represes X. After Z is activated, X is produced and Y is repressed. As a result after deactivation of Z (dashed line) signal is transduced by X. Y at the same moment is discarded. Thus, the feedback implements a memory. And that is how dGRNs transduce signal after vanishing of signal input.

Because of these features study of dGRNs has several difficulties. As we now understand in study of developmnent we always deal with dynamic biological systems. That is why researcher have to monitor multiple time points. Moreover he has to understand the start and the end point of studied process.

Next we discuss how solve such difficulties and try to review common tech-

nologies which is involved in GRN building.

Experiment

In order to build GRN related to any development process we need to answer the following questions:

1) What TFs or signal molecules are participate in the process?
2) When do genes of those factors express? (Temporal expression)
3) Where do that happens? (Spatial expression)
4) And how they interact with each other?

We can distinguish two main approaches in building GRN. First is exprimental approch. In this way we predominantly operate experimental data (microarrays, qPCR, hybridization *in situ*, morphology etc). Second is computation approach where modeling of GRN is based on some descriptive and mathematical models using already generated data. In this review we will concentrate on first approach. In subsequent literature reviews we will cover second approach.

Regulatory genes identification

The first step in building GRNs is regulatory genes identification. If complete genome of the organism is available, than the best way is genome-wide survey with identification of all predicted regultory genes. RGs predominantly are very conserved. Therefore, transcription factors and signal molecules encoded by the regulatory genome can be identified by sequence-based homology searching using BLAST. Also DNA-binding motifs can help identify potencial TF. The next step after identification of the majority of regulatory genes is functional annotation of every gene. Most likely, most of them will be identified. If not, the best way is to use mutant resources. They include information about chemically induced mutants, ectopic expression, and gene silencing. Use external data and papers in order to reveal problem genes.

Phylogenetic conservation.

Researchers have identified potential cis-regulatory sequences in many organisms by searching for evolutionary con- servation

Temporal expression

After regulatory genes identification we have to characterise temporal properties of gene expression during studied process. And at this stage, we encounter several problems. Develppment is a longitudional event which takes lots of time. Also it is highly dynamic serial event. Thus, researcher have to chracterise this great multitude of regulatory genes (RGs) in different stages or conditions. It is not a trivial task. For this often use several methods:

1) Microarrays

Microarray technology has revolutionized our ability to monitor the output of transcriptional regulation at the level of the whole genome. These expression analyses have been further refined by our ability to monitor expression, at high resolution, within individual cell types, tissues, and organs; at different

Temporal restriction
a. Direct transcriptional activation requires coincident expression of driver and target or driver expression earlier than targets **b.** Direct activation by ligand-binding in the recipient cells requires that the downstream genes not be expressed prior to the ligand
Spatial restriction
a. Direct transcriptional activation possible when driver and target are co-expressed in same domain **b.** Indirect interaction through signal(s) is implied if effect is observed in the domain in which the perturbed gene is not expressed **c.** Direct transcriptional repression requires that the repressor and its targets not overlap in their expression domain, or in the same domain at successive developmental stages
Parsimonious topology
a. Direct target of a transcription factor is more strongly affected than indirect targets **b.** A given interaction is presumed indirect if identified direct linkages suffice to explain the data
Special linkages
Double negative control logic is required when interference with a canonical repressor causes decrease in expression of target genes

Table 1: Rules for Building Gene Regulatory Networks Based on Perturbation Data

developmental time points; in different mutant backgrounds; and in response to hormones and various stress conditions.

cell sorting
(GFP)-marked

2) RNA-sq

3) qPCR
4) NanoString

When temporal expression of particular genes is identified, the next step is spatial expression revealing.

Spatial expression

The most common and straight-forward way is hybridization *in situ*. This allows you reveal expression pattern of known regulatory gene. But unfortunatelly this stage is very time-consuming. Additional combination with image analyzing techniaques allows one to achieve high resolution. Mapping regulatory gene expression to various territorial domains is essential for subsequent analysis of the GRN.

Linkages

protein-protein

However, the most commonly used methods for protein in- teraction are yeast-two-hybrid assays and affin- ity purification of tagged protein complexes followed by mass spectrometry (AP-MS). Tandem affinity purification (TAP) of tagged proteins coupled to mass spectrometry has been used to study the cell cycle interactome in Arabidopsis cell culture and protein kinase interactions in rice (Rohila et al. 2006, Van Leene et al. 2007). Protein microarrays can also identify protein-protein interactions in a high- throughput manner. Popescu et al. (2007) used a high-density protein array to detect interac- tions between Arabidopsis calmodulins and other plant proteins (Popescu et al. 2007). In addition, new techniques such as fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) are being used in plants to visualize protein-protein in- teractions in vivo (Xu et al. 2007, Zelazny et al. 2007).

dna-protein

yeast-one-hybrid system

this method should greatly facilitate the identification of protein-DNA interactions in plants. Chromatin immunoprecipitation (ChIP) coupled with quantitative real-time PCR is another technique that plant researchers are increasingly employing. This technique allows one to isolate proteins bound to target chromatin

ChIP has been used to validate putative downstream targets of DELLA proteins, transcriptional regulators that control many aspects of signaling by the plant hormone gibberellin (Zentella et al. 2007). The technique has also been used to validate a physical interaction

epigenetic modification

The evidence suggests that a complex network of epigenetic mecha- nisms acts to regulate gene expression

Metabolite profiling. Obtaining data from the transcriptome, its epigenetic modifications, and the proteome provides the pieces that can then be assembled into a model of tran- scriptional regulatory networks. However, re- searchers are assembling one other large data source that, in part, can represent the terminal output of transcriptional regulation. The iden- tification of small-molecule metabolites can provide a direct link to a cell's physiology and, in some cases, signaling networks. These metabolites are important in the differentiation of tissues and organs and in the responses of plants to the environment. In some cases, metabolites act in signaling cas- cades that influence transcriptional networks.

Modeling

1) Boolean network
2) Bayesian networks

Visualisation

Conslusion

Therefore research of such a complex regulatory systems can shed more light on developmenatal processes and help in the curing of molecular genetic diseases associated with the violation of the networks.

Also such GRNs play a crucial role in evolution. They can buffer gene expression variation and stabilizes phenotypes while adaptation permanently changes them.

An outstanding challenge for both systems biology and evolutionary biology is understanding the molecular mechanisms that allow development to buffer phenotypes while retaining flexibility

Transcription factors, the chro- matin landscape, and signaling pathways act together to regu- late gene expression during cell development

Bibliography

[1] U. Alon. Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.

[2] D. H. Erwin and E. H. Davidson. The evolution of hierarchical gene regulatory networks Douglas. *Nature Reviews Genetics*, 10(2):141–148, 2009.