Methods xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Systems biology approaches to defining transcription regulatory networks in halophilic archaea

Cynthia L. Darnell [a], Amy K. Schmid [a,b,*]

[a] Biology Department, Duke University, Durham, NC 27708, USA
[b] Center for Systems Biology, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form 27 April 2015
Accepted 28 April 2015
Available online xxxx

Keywords:
Archaea
Gene regulatory networks
Systems biology
Halophilic archaea

ABSTRACT

To survive complex and changing environmental conditions, microorganisms use gene regulatory networks (GRNs) composed of interacting regulatory transcription factors (TFs) to control the timing and magnitude of gene expression. Genome-wide datasets; such as transcriptomics and protein–DNA interactions; and experiments such as high throughput growth curves; facilitate the construction of GRNs and provide insight into TF interactions occurring under stress. Systems biology approaches integrate these datasets into models of GRN architecture as well as statistical and/or dynamical models to understand the function of networks occurring in cells. Previously, these types of studies have focused on traditional model organisms (e.g. *Escherichia coli*, yeast). However, recent advances in archaeal genetics and other tools have enabled a systems approach to understanding GRNs in these relatively less studied archaeal model organisms. In this report, we outline a systems biology workflow for generating and integrating data focusing on the TF regulator. We discuss experimental design, outline the process of data collection, and provide the tools required to produce high confidence regulons for the TFs of interest. We provide a case study as an example of this workflow, describing the construction of a GRN centered on multi-TF coordinate control of gene expression governing the oxidative stress response in the hypersaline-adapted archaeon *Halobacterium salinarum*.

1. Introduction

Microorganisms continually face stressful and variable environmental conditions. A central goal of the study of microbial physiology is to understand how organisms maintain homeostasis during fluctuations in environmental conditions. Furthermore, organisms do not experience the environment one stressor at a time, but rather respond to many simultaneous stressors. Integral to this process are gene regulatory networks (GRNs) composed of groups of interacting regulatory transcription factors (TFs) and their target gene promoters. Environmental stimuli are propagated through signal transduction cascades. In response, TFs promote or inhibit RNA polymerase binding to differentially regulate the expression of genes encoding proteins which alter physiology [1]. Transcriptional regulation by interacting TFs is therefore important for signal integration, with the appropriate timing of gene expression enabling adaptation to a variable environment. How do TFs work together to carry out the appropriate response(s)? In most microorganisms, especially in model archaeal and bacterial species that are understudied relative to model organisms such as *Escherichia coli*, the direct effects of TFs on combinatorial gene expression resulting from environmental change are not understood at a genome-wide level. Recent comparative analyses in model systems such as yeast suggest extensive transcription network rewiring even in closely related species [2–4]. Mapping transcription network topology and dynamics across a wide variety of species from the domains of *Bacteria* and *Archaea* is therefore necessary in order to gain a general understanding of how the environment of a microorganism shapes the dynamic interactions between TFs and their target genes.

A systems biology approach to this problem uses iterative experimental and computational methods, with the ultimate goal of quantitative understanding of the physiology and behavior of the cell at multiple levels of information processing. Systems biology is the repeating process of experimental design and model refinement using data types generated from genome-wide experiments (Fig. 1). The integration of genome-wide datasets (e.g. transcriptomics, protein–DNA interactions, proteomics, metabolomics, gene functional annotation) drive the construction of predictive models (statistical and/or detailed dynamical). In turn, hypotheses

* Corresponding author at: Duke University, Department of Biology, 125 Science Dr, Box 90338, Durham, NC 27708, USA. Fax: +1 (919) 660 7293.
E-mail address: amy.schmid@duke.edu (A.K. Schmid).

http://dx.doi.org/10.1016/j.ymeth.2015.04.034
1046-2023/

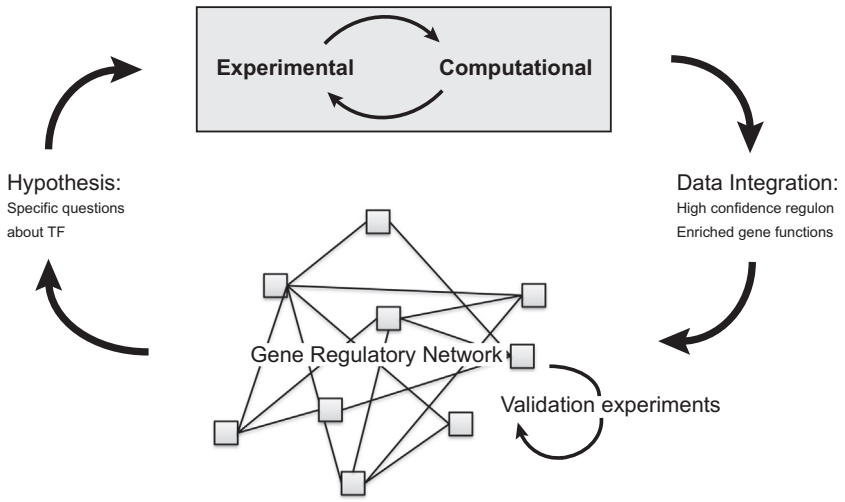

Fig. 1. Overview of the systems biology process and procedures for building gene regulatory networks for multiple TFs of interest. Diverse genome-wide data sources are analyzed and integrated to generate regulatory networks. Focused validation experiments lend additional confidence to network conclusions. Hypotheses are generated from these networks, which are tested in subsequent rounds of genome-wide experimentation.

are generated from predictive models. These hypotheses are tested in future rounds of genome-wide experiments and more detailed, smaller-scale validation experiments.

Thus far, systems biology approaches to characterize GRNs have focused primarily in well-studied model microbial species (e.g. yeast, *E. coli*), with fewer studies mapping GRNs in archaea [5–10]. Previously, progress on archaeal GRNs has been hampered by a lack of tools including genetic manipulation, genome sequences, ability to culture, and large databases of genome-wide datasets. Recently, such tools have become available for several species of archaeal extremophiles, including *Thermococcus kodakaraensis* [11,12], *Pyrococcus furiosus* [13–15], *Sulfolobus acidocaldarius* [16,17], and various species of hypersaline-adapted archaea [18], among others [19]. The stage is set for systems analysis of GRNs across a wide variety of organisms.

In this report, we provide a guide for a suite of experimental and computational genomics methods to characterize GRNs *in vivo* in archaeal model organisms. As an example of this suite of tools, we provide a case study focused on the mapping of the GRN controlling the response to extreme oxidative stress in the hypersaline-adapted model archaeal species, *Halobacterium salinarum*, for which the systems biology approach has already been implemented. However, many of the methods described can be applied to various species for characterizing coordinate regulation and networks, TFs of unknown function, and identification of genes under the control of these TFs (“regulon”). Although each individual method has been described elsewhere, the purpose of this report is to explain details essential to experimental design, interpretation, and systematic deployment of these tools. Together, these methods enable rapid characterization and meaningful biological interpretation of GRNs in understudied organisms.

2. Methods

2.1. Experimental procedures for mapping GRNs

2.1.1. Experimental design

Gathering genome-wide data during an organism’s response to environmental and genetic perturbations is a common and effective method to reveal GRN function and architecture [7–9,20]. Exact matching between the environmental conditions, time points, and genetic backgrounds for preparation of these data types is particularly important. The data types of focus here include genome-wide expression, TF–DNA binding, and growth rates of TF knockouts. Continuity in experimental design facilitates the integration and biological interpretation of these multiple data types. Therefore, aspects of particular importance to experimental design include environmental context (laboratory conditions under which samples are prepared) and the dynamics of a response to environmental perturbation (time points during which measurements are made). These key concepts and other, more detailed aspects of experimental design for systems biology are considered in detail below.

2.1.1.1. Environmental context. Because high throughput datasets such as genome-wide expression reveal the behavior of all genes in response to a given perturbation, any unknown or hidden environmental variables can confound results [21]. It is therefore important to control for as many factors as possible during the preparation of cell cultures for collection of high throughput data. Ideally, cultures would be prepared in a chemostat under steady state conditions during growth in the presence of a limiting nutrient. Any changes in the gene regulatory network following addition of stressors to the vessel can then be confidently attributed to the perturbation (e.g. [22,23]). However, in many archaeal model systems, chemostat studies are not possible because a minimal medium is not available or, as is the case in *H. salinarum*, the organism may use a feast or famine nutritional strategy (i.e. all carbon and energy sources consumed simultaneously [24]). In this case, a turbidostat with controlled temperature, pH, optical density, light, etc., is effective in tracking perturbations to the controlled system [25]. Alternatively, growth in batch culture provides reproducible data; however, sufficient replication (see also Section 2.1.1.3) and age matching of cultures are critical in this case. Regardless of which growth method is chosen, consistency in growth phase at harvest, stress treatment concentrations, time points, and temperatures across genetic backgrounds and datatypes enable reproducibility and comparability between experiments. Other parameters known to be important for the organism of choice should also be carefully controlled. For example, in our experience, the timing of strain recovery from frozen stock impacts *H. salinarum* stress resistance (Darnell, unpublished data). Any inconsistency in these growth parameters can affect the reproducibility of results and interpretation of high throughput data.