Principles of gene circuit design

Pablo Meyer (IBM research)
Michael Elowitz (Caltech)
Diego Ferreiro (Universidad de Buenos Aires)
Thierry Mora (Ecole Normale Superieure and CNRS)
Aleksandra Walczak (Ecole Normale Superieure and CNRS)

10, September–15, September, 2017

1 Overview of the Field

Genetic circuits operating within and between cells are the basis for a vast range of cellular and developmental functions in biology. Biologists have now assembled enormous information about individual components, their regulation, and interactions. However, in most cases we still lack understanding of why the resulting circuits have specific architectures or other features. Specifically, what features of these circuits enable them to operate efficiently or robustly in individual cells, and why have some architectures been selected by evolution over others? A lack of understanding of these basic issues currently limits our ability to predictively manipulate cells for biomedical applications.

The goal of this meeting was to bring together researchers from a diverse background, mainly computational and systems biologists including synthetic biologists, seeking to discover and understand the basic design principles of genetic circuits, and to chart a path forward toward a more complete, predictive, and comprehensible understanding of genetic circuitry.

2 Recent Developments and Open Problems

We present here a brief overview of the different approaches used to define and model genetic circuits from different biological systems. Predictive and mechanistic models are powerful tools to understand biological processes at the core of systems biology. Two schools of thought exist when building models of a given process, but both require in first instance a list of components and their interactions. For the computational biologist, this list of components can be defined and assembled from prior knowledge and/or inferred, or reverse engineered, from dedicated experimental data. System biologists build and test experimentally natural motifs previously encountered and modeled.

2.1 Gene Network motifs: Recurrence and inference

A first step towards obtaining the key design principles underlying biological circuits is to assemble the different components that make up the gene regulatory networks and assess what designs are used most often. In a seminal paper Uri Alon's group showed by studying the network of transcriptional interactions in

Escherichia coli that it contains three highly significant motifs: Feed-forward loops (FFL), single input modules (SIM) and dense overlapping regulons (DOR). Each network motif has a specific function in determining gene expression, such as generating temporal expression programs and governing the responses to fluctuating external signals. These different motif structures also allow an interpretation of E.coli's entire transcriptional network and the inference of some of its dynamics. FFLs can act as circuits that reject transient activation signals while responding to persistent ones while accelerating responses, SIMs function in systems that build stoichiometric protein assemblies and temporal expression can be encoded in the promoter expression thresholds, finally DORs mostly interconnect these two motifs but their function is less well defined. Conversely, Network inference algorithms using experimental datasets have been shown to infer with different accuracy diverse types of motifs. It is essential to note that crowd-sourcing efforts such as DREAM (Dialogue for Reverse Engineering and Methods) are essential to test the accuracy of inference methods and their predictions. Common mistakes include the fan-out error, where co-regulated nodes are predicted to be interacting and the fan-in error, that also induces FFL errors, where multiple inputs to a single node are difficult to predict. Mutual information methods are also better at predicting FFLs than Bayesian or linear regression approaches that fare better at predicting cascades.

2.2 Gene Network dynamics: Single cell data are essential to test predictions

Accurate predictions of gene network motif dynamics are limited by the correct inference of the motifs and the parameter values characterizing the dynamics. Once a genetic circuit topology is defined, estimation of the parameters that will help establish the circuit dynamics is a less difficult problem. Nevertheless, while inferring networks, it is necessary to consider that information on gene interactions is not and probably never will be, complete, especially considering all the different kinds of interactions possible. One main caveat to network inference is that it is performed using static datasets extracted from thousands of cells or tissue, but cells are controlled by genetic circuits that show transient, repetitive, pulsatile, and stochastic dynamics even under constant conditions. Since a fundamental problem in biology is to understand how the dynamics of these circuits implement core cellular functions and how different circuit designs are used, it is necessary to complement these studies with dynamical in vivo measurements. Single-cell studies are necessary as many of the dynamical features of gene circuits that are sharp at the level of individual cells get washed out or obscured by population averages. Time-lapse single-cell microscopy combined with fluorescent reporters, single-cell sequencing and other systems biology approaches are beginning to provide a direct view of circuit dynamics in individual cells. We propose that a DREAM competition testing the predictions of models inferred from genome-wide data would help refine gene circuit reconstruction and parameter inference methods as shown previously with in silico data. Indeed, spontaneous dynamic behavior has now been measured in diverse cell types from microbes to mammalian cells. For example, pulsing is generated by genetic circuits that activate and deactivate key regulators and modulate pulse characteristics, such as frequencies and amplitudes. Signal transduction pathways face the challenge of internally representing, or encoding, the identity, amplitude, and timing of many different external signals. Cells address this challenge, by typically encoding stimulus information in the frequency, amplitude, and duration of pulses of pathway activation. Pulsing offers a flexible mode of regulation that can be adapted to many cellular contexts. What are the underlying genetic circuit mechanisms that the cell uses to generate and regulate pulsing? Excitability might seem like an ideal property for any pulse-generating genetic system, but it can also be done with a different, non-excitable circuit architecture or a circuit containing multiple negative-feedback loops. The proposed workshop would also allow the discussion regarding how different circuits with mixed feedback loops exhibit a variety of dynamical behaviors including multi-stability, oscillations, and pulsing, depending on parameter regimes. Indeed several features appear common to many pulsatile genetic circuits. First, negative feedback loops occur in all examples, second, noise helps generate pulses. The discussion can also be extended to the diverse set of theoretical approaches and experimental datasets used to infer the gene regulatory networks.

2.3 Metabolic Network motifs and dynamics

Another example in which the circuitry is important is the regulation of cellular metabolism where cells need to perform and regulate in a confined space a myriad of biochemical reactions. Metabolic activity is controlled by the expression level of the enzymes composing the reactions but is also constrained by the structure of

the circuit underlying the biochemical network. The variability due to fluctuations in enzyme levels is in part smoothened by the architecture of biochemical networks. For example, the arabinose pathway is composed of a FFL that delays the response induced in presence of arabinose and absence of glucose, as the lactose pathway is controlled via a simple regulation in the presence of lactose and absence of glucose. Also, recent advances in bacterial cell biology have definitely put aside the view that the bacterial cytoplasm is a "bag of enzymes". It is now clear that in cells, proteins have to localize at the right time and in the right place in order to perform their functions. Although much is known about the enzymatic cascades that underlie cellular metabolism, such as their genetic identity, their genomic organization into operons and the classical enzymatic rules governing these reactions (see Michaelis-Menten, Beadle and Tatum), little is known about their cellular organization and the genetic circuits underlying such organization. This is in part due to the lack of experimental tools necessary to obtain single-cell metabolic data. This workshop could help discuss the tools necessary to leap forward towards generating the predictable synthetic metabolic circuits that are fundamental to produce new medicines and biofuels. These examples raise also fundamental question: Is all the information necessary for regulating a biological function encoded in the circuit architecture? That is, knowing that the behavior of genetic circuits critically depends not just on their connectivity but also on the dynamics that depend on the parameters of the biochemical reactions. Reconstructing Synthetic Gene Networks Synthetic Biology complements the previous approaches that study natural circuits by designing and testing genetic circuits. Insight gained from analyzing the circuitry of natural processes leads to improved designs of synthetic systems, and the creation of small artificial networks helps to analyze hypotheses on the function of natural ones. Synthetic Biology has evolved from small transcriptional circuits into complex systems that can be applied broadly to all types of biological functions from metabolism to multicellular development. By building and testing these circuits in living cells this approach allows discerning what types of genetic circuit designs are capable of implementing different cellular behaviors, and what trade-offs exist between different designs. Several generations of oscillators and genetic switches have been built in order to work with diverse cellular components and regulatory mechanisms that can also interact with natural gene circuits. Complex metabolic pathways have been engineered to produce useful products, and signaling pathways have been rewired to alter their dynamic behaviors in predictable ways. However, synthetic biology remains extremely primitive owing both to technical challenges and, even more, to fundamental inadequacies in our understanding of biological circuit design. As the successes of synthetic biology become more impressive, the field moves towards building biochemical circuits with more challenging and complex functions. Engineering new circuits without understanding how they work is potentially dangerous and can lead to the accidental creation of unwanted functions, creating potential biohazards. On the fundamental side, as we still have little understanding of how circuit designs can function effectively in cells and tissues and much to learn from natural examples, we are convinced that a workshop bringing together people from different backgrounds will help enhance discussions relative to circuit design. In particular, one of the greatest challenges is to move synthetic biology from circuits operating in individual microorganisms to circuits that function in a truly multicellular fashion, for example, circuits sufficient to implement self-patterning of cells. If successful, we may be able to understand multicellular development from a totally new point of view that could inform tissue engineering and regeneration.

3 Presentation Highlights

The principal topics that rose in this workshop were related to gene circuit design, network reconstructions, chromatin regulation, development networks and regulation of immune cells.

The excitability and bifurcation properties of a developmental circuit in B. subtilis were discussed as well as how oscillations in metabolism allow the coupling between distant bacterial biofilms and emergence of nutrient time-sharing. This topic was extended to a discussion on the difficulty to define design principles for spatio-temporal organisation of cell populations, but the existence of gene circuits that allow dynamical compensation and biphasic control and a antithetic feedback motif leading to perfect adaptation.

The difficulty related to giving an interpretation to large-scale networks across bacteria and multiple pathways in Whole Cell Models brought two approaches to help untangle the biological hairball: 1) Edge simplification of networks through pruning while maintaining its dynamical properties and 2) Reconstruction of a phenomenological model of the networks.

This objective also lead to the discovery of new properties that help explain the redundancy in biological receptors for signaling networks such as the BMP and Notch signaling systems.

Furthermore in-vivo synthetic biology analysis helped perform a quantitative analysis of the Notch transcriptional response. Data driven modeling and imaging single-cell data helped understand the boundaries of transcription in living embryos and the gene regulatory logic underlying the Neural Tube patterning.

Another developmental model helped understand the principles for modeling the B cell differentiation circuitry. Single cell experiments helped define how to build a good immune response through the recognition of self and non-self in the context of the diversity in immune receptor repertoires. Understanding the dynamics of the immune system were helped by the untangling the biological hairball of immune recognition networks and finding the physical limits to concentration sensing in a background of competing ligands.

Recent advances in chromatin biology and in particular Hi-C data helped define the promoter regulatory connections of the pluripotent genome and how chromatin enables fractional gene regulation and epigenetic memory. Chromatin was further shown to be involved in NF-κB signaling to drive diverse viral phenotypes and transcriptional bursting and in the mechanistic basis for the quantitative Epigenetic Memory of vernalization in Arabidopsis Thaliana.

Finally low cost and open source resources for synthetic biology in Latin America were discussed in the context of GOSH, LOOP, as well as other open source, open data initiatives such as TecnoX and DREAM.

4 Scientific Progress Made

Given the satisfaction of the participants, there was a clear need for a workshop to discuss the principles of gene circuit design, a follow up to the 2011 Workshop on Stochasticity in Biochemical Reaction Networks at the Banff International research station. The increasing sophistication and power of synthetic biology to analyze and reconstruct genetic circuits has developed in parallel to the more computational approach of gene regulatory networks inference from large genomic datasets. Revolutionary techniques using single-cell approaches such as single-cell imaging, single-cell metabolomics and single-cell genomics are helping merge these previously separated fields. Through its many applications single cell studies, genomics and synthetic biology are beginning to be used in medical contexts, but there is a growing recognition that circuit level problems are limiting our ability to predictively design therapeutic strategies. As these fields leap forward, many critical questions arise:

Bring together scientists whose research fundamentally depends on understanding different forms of genetic circuitry that ranges from inferring networks using large datasets to signaling networks, metabolic networks, cellular networks and the design of synthetic circuits. What set of rules could exemplify design similarities between genetic circuits implemented in different biological systems and performing different biological functions? What are the key physical principles that limit and allow for the encoding of regulatory function and are they conserved by evolution? What computational/experimental tools are necessary to leap forward towards generating the predictable synthetic metabolic circuits that are fundamental to produce new medicines and biofuels? Discuss a possible DREAM competition testing the predictions of models inferred from genome-wide data with data generated from single cell dynamics. How could we implement a computational platform to build models and test their predictions?

What types of genetic circuit designs are capable of implementing different cellular behaviors, and what trade-offs exist between different designs?

The conference brought together many of the pioneers and leading experts in these diverse fields for a few days of extensive, interdisciplinary and informal discussion. Our goal was to create a forum where knowledge is shared, hoping that this diverse community will discuss how the lack of understanding of design principles is limiting the advancement of genetic circuit engineering and help define together the agenda for understanding genetic circuit design.

5 Outcome of the Meeting

One of the major positive outcomes of the workshop was successfully bringing together people from extremely different backgrounds. Although many workshops invite people from different communities and create the potential for the exchange of ideas, our workshop was truly unique compared to other meetings we

have attended. Unlike in other events, the participants exchanged ideas and points of view and engaged in discussion both during the sessions and during the plentifully designated discussion time. Due to the variety of backgrounds, not a lot of "basal" knowledge was shared by participants. We strongly encouraged our participants to ask questions, and *every single talk at the workshop elicited several questions and discussions*. This suggestion on one hand resulted in incredibly lively discussions. On the other hand it allowed for a bridging of the gap between the different backgrounds of participants. These prolonged introductions often led to the emergence of interesting questions. Frankly, the level of interactions, exchange of ideas and mixing of fields greatly surpassed the expectations of the organizers.

As discussed above, one of the main goals of this workshop was to encourage collaboration between researchers from diverse fields, who often might not be aware of each others' research. The diverse participant list helped us to achieve this goal. The workshop participants included representatives of diverse fields: chemists, biologists, synthetic biologists, physicists, mathematicians and computer scientists. These researchers also represented a diverse set of locations including US (12), Canada (2), Europe (9), Mexico (13) and South America (4). Most of the participants were young: 14 are pre-tenure faculty members; 7 will have just started their faculty positions this year; and 6 are currently post-doctoral researchers. 8 women attended the workshop: seven invited attendees and one organizer. All in all 7 junior faculty , 2 post-docs and 8 graduate students attended. This is a high ratio for research disciplines which include such male dominated fields as physics, computer science, engineering and theoretical chemistry.

We note that the ability to bring together such a young internationally diverse group of people was helped by the award we received from IBM RESEARCH. This award allowed us to subsidize the rather high travel costs of 3 Faculty.

1. Pablo Meyer, IBM research
2. Michael Elowitz, Caltech
3. Diego Ferreiro, Universidad de Buenos Aires
4. Aleksandra Walczak, Ecole Normale Superieure and CNRS
5. Thierry Mora, Ecole Normale Superieure and CNRS
6. David Sprinzak, Tel Aviv University
7. Mustafa Khammash, ETH Zurich,
8. Jan Skotheim, Stanford
9. Jonathan Karr, Mount Sinai
10. Ilya Nemenman, Emory
11. Maria Rodriguez, IBM Zurich
12. Yana Bromberg, Rutgers
13. Gregoire Altan-Bonnet, Sloan-Kettering
14. Paul Franois, McGill
15. Osbaldo Resendis, INMEGEN
16. Enrique Hernandez, INMEGEN
17. Ignacio Enrique Sanchez, University of Buenos Aires
18. Julio Freyre-Gonzalez, CCG
19. Fernan Federici UCC, Chile
20. Alex de Luna, LANGEBIO
21. Rodrigo Reyes-Lamothe, McGill
22. Tim Rudge UCC, Chile
23. Mayra Furlan, UNAM
24. Ernesto Borrayo Carbajal, U de GUADALAJARA
25. Omer Karin, Weizmann
26. Yaron Antebi Caltech
27. Kathryn Miller-Jensen, Yale University
28. Nathalie Dostatni, University of Paris
29. Lacramioara Bintu Caltech
30. Joe Larkin, UCSD
31. Elisa Franco, University of California Riverside
32. Alcala-Corona, Sergio Antonio, INMEGEN
33. Arzate, Rodrigo,UNAM

34. Avelar-Rivas, J Abraham, CINVESTAV
35. Cruz Maldonado, Carlos Roberto, UNAM
36. Escorcia Rodriguez, Juan Miguel, UNAM
37. Herrera-Valdez, Marco Arieli, UNAM
38. Lpez Castillo, Alfredo Antonio, UNAM
39. Martinez-Sanchez, Mariana, UNAM
40. Briscoe, James, Francis Crick Institute

The views cited above (and the ones we did not have room to quote) clearly show that the most important outcome of the workshop was for people to meet, exchange ideas and start new collaborations, often with people whose research they were previously unfamiliar with. For example, Gregoire Altan-Bonnet, Thierry Mora and Aleksandra Walczak started a collaboration on immune system diversity regulation. Maria Rodriguez and Aleksandra Walczak started a collaboration on B-cell development. James Briscoe and Nathalie Dostanti started a collaboration on embryonic regulation of precision in gene expression. Michael Elowitz, Martin Howard and Lacra Bintu are writing a review on Chromatin regulation. Ilya Nemenman started a collaboration with Thierry Mora on Network regulation and Michael Elowitz on modelisation of signaling pathways. Yana Bromberg and Mayra Furlan started a collaboration on analyzing Hi-C data. Pablo Meyer and Rodrigo Reyes started a collaboration on bacterial intra-cellular imaging.

A large number of participants noted the importance of fostering a collaborative environment, the time and space to exchange ideas, and meeting other young researchers. Practically all participants emphasized the non-standard nature of this meeting (the large allocation of time to questions and discussions) as opposed to the many talks + posters format present elsewhere.

6 Schedule

Monday, Sep 11
Bintu Lacramioara: Chromatin enables fractional gene regulation and epigenetic memory

Rodrigo Reyes: Speed dating while in a stable relationship: the dynamics of replisome subunits during active DNA replication

Jonathan Karr: Principle for modeling multiple pathways towards Whole Cell Models

Kathryn Miller-Jensen: Exploring how NF-?B-chromatin interactions drive diverse viral phenotypes and transcriptional bursting

Paul Francois: Untangling the biological hairball of immune recognition networks

Jan Skotheim: How Cell Growth Drives Proliferation: On the scaling (or not) of biosynthesis with cell size
Mayra Furlan: Promoter regulatory connections of the pluripotent genome

Yana Bromberg: Predicting modifiable protein residues for effective analysis of exonic variation

Mustafa Khammash: Antithetic Feedback motif leading to perfect adaptation

Tuesday, Sep 12

David Sprinzak: Quantitative analysis of Notch transcriptional response using in-vivo synthetic biology

Nathalie Dostatni: Imaging transcription in living embryos : how data driven modeling could help understand patterning

Julio Augusto Freyre-Gonzalez: Towards a large-scale comparative systems biology across bacteria

Aleksandra Walczak: Diversity in immune receptor repertoires.

Enrique Hernndez-Lemus: Loss of inter-chromosomal regulation in Breast Cancer

Omer Karin: Dynamical compensation and biphasic control in tissue circuits

Yaron Antebi: Signal perception in the BMP signaling system

Mara Rodrguez Martnez: Circuit modelling principles of B cell differentiation

Martin Howard: Mechanistic Basis of Quantitative Epigenetic Memory

Thursday, Sep 14

Tim Rudge: Design principles for spatio-temporal organisation of cell populations

Marco Arieli Herrera-Valdez: Transcription regulation and cellular differentiation in B subtilis: Excitability and bifurcation structures

Alexander De Luna: Genome-wide mechanisms of longevity by dietary restriction in the budding yeast

Diego Ferreiro: Nothing in Biology makes sense except in the light of BS

James Briscoe: Gene Regulatory Logic of Neural Tube Patterning

Grgoire Altan-Bonnet: How to build a good immune response

Ignacio Enrique Snchez: Do genome and proteome dynamics pose universal constraints to gene circuit design?

Joe Larkin: Coupling between distant biofilms and emergence of nutrient time-sharing

Ilya Nemenman: Phenomenological model of (biological) networks

Friday, Sep 15

Osbaldo Resendis: Personalized medicine and the quantitative design of cancer metabolism

Thierry Mora: Physical limit to concentration sensing in a background of competing ligands

Fernan Federici: Low cost and open source resources for synthetic biology in Latin America

Pablo Meyer: Origin of fractional control in regulated cell death

Michael Elowitz: Multiplexed Messaging