COLLABORATIVE NSF/BIO-BBSRC. ABI INNOVATION. MULTI-SCALE GENE FUNCTION PREDICTION USING BIOLOGICAL NETWORKS

1. Specific Aims

In recent years, the numerous large-scale sequencing projects combined with fast sequencing techniques have generated enormous amounts of sequence data. This has led to the identification of thousands of previously unseen genes whose function awaits to be characterized. A fundamental goal is therefore to identify the function of uncharacterized genes on a genomic scale. It is difficult to design functional assays for uncharacterized genes so a major current challenge in bioinformatics is to devise algorithmic methods that, given a gene, can predict a hypothesis for its function that can then be validated experimentally.

In this project, we shall focus on three different aspects of gene function: molecular function, cellular role and organismal phenotype. Our aim will be to build a general system that, given a gene, can predict its function at these three different levels. This multi-scale prediction will be carried out exploiting the structure of biological networks.

AIM 1: We plan to develop a computational framework to identify and characterize the cellular role of genes using logic-circuit models and regulatory networks. Specifically we propose to develop a method to analyse logic operations of small regulatory triplets using a two-in-one-out logic gate model. We will use a binarized gene expression data to score how well each triplet matches each of all 16 possible logic gates. A high score implies that the logic operation describes accurately the interactions between elements forming the regulatory triplet. As such a similarity in logic gate matches between various triplets implies a similarity in function between input elements.

AIM 2: We will develop a computational workflow to infer phenotypic function using as input network neighbourhoods and text mining. For this we will use machine learning techniques on a graph model that accounts for latent variables that can explain the association between the data. Here we make the presumption that the relationships existing between the observable variables can be explained by assuming the existence of unobservable (latent) variables and of relations between these latent variables and the observable ones. Note that any matrix can be interpreted as a graph (and vice versa). Therefore decomposing a matrix amounts to inferring a new graph structure where the relation between the latent factors and the measurable data is made explicit. Here the graphs will be constituted by large-scale biological networks.

AIM 3: We plan to integrate our results from **AIM 1,** specifically synthesizing the circuit elements and their domains of influence within a regulatory network, with phenotypic function predictions from **AIM 2** to better demarcate regions of the network associated with distinct phenotypic functions. We we'll also use these results to improve the phenotypic prediction by integrating information from the cellular role inferred in **AIM 1.**

2. Background and Preliminary Results

2.1 General Background

2.2 Background on Networks

2.2.1 Networks Biology: A Growing Field

Biological systems are mediated by interactions between thousands of molecules. Network-based statistical models are particularly useful in unlocking the complex organization of biological systems. In the last decade, biological network analysis has blossomed into a new scientific discipline. Examples are numerous, ranging from protein–protein to genetic interaction networks \cite{17473168}. Usually, networks are depicted as graphs with nodes and edges, where nodes denote biological entities such as proteins or genes, and edges represent interactions between nodes.

Cellular networks are organized in the form of interacting modules, whereby nodes in a module tend to have a larger density of edges connecting them. Biologically, the genes within a module of a genetic regulatory network are co-regulated. Graph models can reveal interesting new features of the analysed biological system \cite{11034217,10521342,10935628,12202830,12399590,16730024}, while network topologies can be used to address fundamental biological questions \cite{18421347,15190252,12134151,17274682,19372386,16311037}.

2.2.2 Preliminary Results on Networks

Gerstein lab key papers:

Gerstein lab has carried out projects in biological networks for over a decade. We have made extensive contributions in the analysis of genomic data, especially with regard to network prediction and analysis\cite{14564010}. We have also integrated regulatory networks with gene expression to uncover different kinds of dynamic sub-networks \cite{15372033}. We developed methods to determine the hierarchical organization of regulatory networks and applied them to analyze the regulatory networks of a variety of species from yeast to human, including networks constructed from ENCODE, modENCODE and MCF7 data \cite{22125477,20439753,22955619,21177976}.

Network Construction

We have developed several methods to construct networks based on various genome features \cite{14564010}. We extended this work by combining several heterogeneous biological datasets \cite{12350343,15998909,16413578} and developing new machine learning techniques \cite{19656385} to increase the prediction power. In 2008, our work placed first in the Dialogue for Reverse Engineering Assessments and Methods (DREAM, www.the-dream-project.org) competition for the in silico network prediction challenge. In addition, we have participated in many experimental network determination projects, to refine and keep our methodologies at the cutting edge \cite{16449570,16554755,14704431}.

Recently, we have completed the ambitious goal of constructing draft regulatory networks for humans and model organisms based on the mod/ENCODE datasets \cite{21177976,21430782,22955619}. We have successfully completed this challenge through the development of novel approaches for identifying individual proximal and distal edges, as well as creating new miRNA target prediction algorithms.

Construction of proximal regulatory network. We have developed several computational approaches based on data from cutting-edge next-generation sequencing technologies (such as ChIP-seq) to help construct proximal regulatory networks and identify regulatory targets of transcription factors (TFs). We developed a computational framework, PeakSeq \cite{19122651}, to define the TF binding peaks. Thus we were able to take advantage of the rich ChiP-seq data in constructing a regulatory network. PeakSeq constructs local

thresholds using input signals from genomic DNA without an enrichment process to simulate the null process for the background. The variability in the background signal reflects the accessibility of the DNA given the chromatin state of the genome. PeakSeq then identifies TF binding regions by identifying peaks that are significantly enriched relative to the background signal. PeakSeq is a widely used (the associated paper being cited more than 300 times) and highly versatile tool for identifying TF binding sites from ChIP-seq data.

—LESS

In addition, we have also proposed a probabilistic model, referred to as target identification from profiles (TIP), that identifies a given TF's target genes based on ChIP-seq data \cite{22039215}. Given a ChIP-seq dataset for a particular TF, we start by identifying all genes bound by the respective TF. Next, we characterize the TF binding profile by averaging its binding signal at each position around the transcription start site (TSS) for the related genes. We calculate the binding strength of a given TF for a particular gene using a weighted sum of the TF binding signal from nucleotides surrounding the TSS and estimate the corresponding significance level assuming a normal distribution of regulatory scores. Compared to other peak-based methods, this new approach his more reliable providing highly accurate TF targets. We have successfully used TIP to identify a high confidence set of ERα target genes.

Constructing distal networks and identifying enhancers. We have developed machine-learning methods that integrate ChIP-seq, chromatin, conservation, sequence and gene annotation data to identify gene-distal regulatory regions \cite{20126643}. By correlating the binding signals around DRMs with respect to expression of transcripts, we developed a computational pipeline to identify potential enhancers and the transcripts associated with them (Figure 1A). As published in \cite{22950945}, we have validated some of the results from our pipeline by experiments, which show a fairly high predictive accuracy. The enhancers and their targets form a distal regulatory network (Figure 1C), and when analyzed along with the corresponding proximal regulatory network (Figure 1B), provide a more comprehensive and complete view that incorporates multiple dimensions of transcriptional regulation into the network.

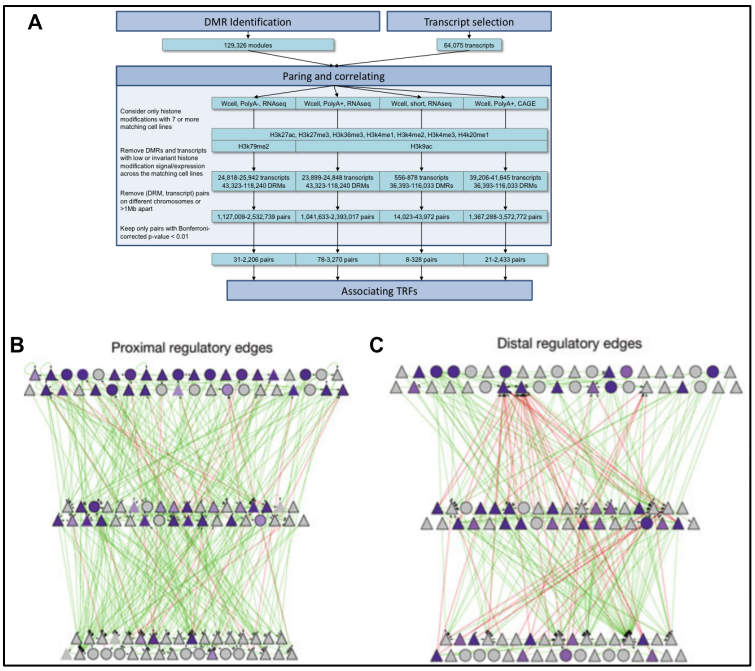


Figure 1 A. Associating transcription factors to target genes \cite{22950945} **B.** Proximal regulatory network. **C.** Distal regulatory network \cite{22955619}.

Network constructions using ENCODE, modENCODE and other system-wide data. Using the machine-learning approaches we have constructed highly integrated regulatory networks for humans and model organisms based on ENCODE \cite{22955619} and modENCODE datasets \cite{21430782}. These integrated networks consist of three major types of regulation: TF-gene, TF-miRNA and miRNA-gene, showing rich statistical patterns. For instance, the human regulatory network uniquely displays distinct preferences for binding at proximal and distal regions. The proximal-distal binding preference is a property of the intergenic space in the human genome, which is much larger relative to the genomes of other model organisms. This difference leads to a larger amount of distal binding. Furthermore, in

the human regulatory network, the more highly connected TFs are more likely to exhibit allele-specific binding and gene expression. More recently, we built a regulatory map for 24 nuclear receptors and 14 breast-cancer-associated TFs that are expressed in the breast cancer cell line MCF-7. The resulting network reveals a highly interconnected regulatory matrix with extensive "crosstalk" between NRs and other breast-cancer-associated TFs. We show that large numbers of factors bind in a coordinated fashion to target regions throughout the genome. The highly occupied targets are associated with active chromatin state and hormone-responsive gene expression.

Network Analysis

Biological networks, normally large in scale, are organized with topological structures in the form of interacting modules. Statistics such as 'eccentricity' and 'betweenness' are helpful to explain the connectivity and behaviour of nodes in a network. We have developed a number of tools to analyze the organization and structure of biological networks including identifying the importance of a node in a single network and identifying the modular structure inherent within several biological networks. For example, by mapping gene-expression data onto the regulatory network of yeast, we identified different sub-networks that are active in different conditions \cite{1537203}. We developed a method to extract metabolic modules from metagenomic data, enabling us to identify pathways that are expressed under different environmental conditions \cite{19164758}. We have also developed a way to identify nearly complete, fully connected modules (cliques) present in network interactions \cite{16455753} and we have been using networks to map various kinds of functional genomics data \cite{22955619}.

Hierarchical Networks

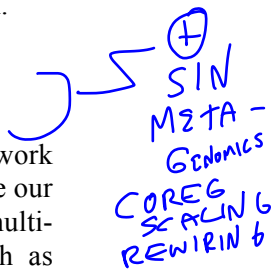
Through our network analysis, we found that gene-regulatory networks are composed of hierarchical structures dominated by downward information flow and that some TFs act as top master regulators to govern the transcription of downstream TFs. We developed methods to determine the hierarchical organization of regulatory networks in a variety of species. In these hierarchical networks, TFs are organized into three levels, whereby TFs at the top tend to act as regulators while TFs at the bottom tend to be targets of regulation.

We also found that regulatory factors are hierarchically organized in all organisms. In a hierarchical organization, the factors at the top of the hierarchy are most influential, as reflected by their highly correlative binding and gene-expression profiles. The factors at the top level are under stronger selective pressure and are more conserved evolutionarily. In comparison, the middle level contains many elements characterized by bottlenecks in information flow and are highly connected by miRNA and distal regulatory elements. These analyses thus highlight the general features of regulatory networks conserved in evolution.

[[CSDS2MG – shall I add here CC's paper on hierarchies or do you want it as an aim?]]

Integrating networks with other biological data

To further illustrate the value of the network concept, we have also combined network analyses with many other types of biological data. Recently, we used networks to improve our understanding of genomic variants \cite{24092746}. In \cite{23505346}, we built a multi-layered network that incorporated information from heterogeneous data sources such as protein-protein interactions and metabolic, phosphorylation, signaling, genetic, and regulatory networks. In general, population variants are more likely to be deleterious when they occur in genes or in regulatory elements associated with hubs in the multi-layered networks, indicating that a gene's interactions likely influence the selective pressures on acting on it. Connectivity is also related to selective pressure in noncoding regions, as transcription binding motifs with greater connectivity tend to be under stronger evolutionary pressure \cite{24092746}. We built a workflow model to prioritize noncoding mutations in disease variants based on these patterns of negative selection in functional variants. In addition, we showed that proteins



under positive selection are found on the network and on the cellular periphery, an indication of how human variation is arranged with respect to the interactome.

Web Tools for Network Analysis

We have developed numerous network analysis web tools such as TopNet \cite{14724320}, tYNA \cite{17021160}, and PubNet \cite{16168087}. These tools have been widely used by the research community to analyze network topology—i.e., to calculate hubs, "between-ness", shortness of paths and degree of modularity.

2.3 Background for Phenotypic Function Prediction

3. Research Plan and Methods

3.1 AIM 1: Developing a Method to Infer Gene Cellular Role Using Logic Circuit Models and Biological Networks

Our aim is to develop a novel method of inferring a gene cellular role from the analysis of biological networks. More specifically, we will integrate regulatory networks with gene expression data. This will allow us to analyse the interactions between the regulatory factors and target genes using a logic operations based algorithm, Loregic. This study will highlight common behaviour patterns between various RFs as well as groups of genes under similar regulatory constraints. We plan to make available Loregic available as an online tool as well as a stand alone application that can be downloaded and used on various input datasets.

3.1.1 Background on Logic Circuit Models in Biological Networks

Gene expression is a complex process controlled by regulatory factors on multiple dimensions. An increasing number of recent experimental and computational studies suggest that gene transcription is regulated cooperatively by numerous factors (i.e. TFs and miRNAs) \cite{24009496,22955619}. These studies analyse the relationships between the regulatory factors (RFs) from various aspects such as protein-protein interactions, sequence motifs in cis-regulatory modules, co-associations of TFs in binding sites, and co-expressions of TF target genes \cite{14627835,22705667, 21828005}. However, they focus solely on the identification of the wiring relationships between RFs (e.g. co-binding, co-association and co-expression) leaving untouched the cooperative patterns that drive the biological functions behind the wiring diagrams.

At a high level, the gene regulatory network can be regarded as an electronic circuit, with TFs and miRNAs acting as resistors and capacitors. Just as wiring different circuit elements can generate various electrical functions \cite{Rabaey}, connecting various regulatory factors as functional modules will result in different biological functions. Thus, in order to obtain a comprehensive map of gene regulation, it is necessary to go beyond identifying the wiring relationships among individual RFs. Here we propose to study the RFs cooperative patterns, and further regulatory functional modules resulting from those cooperative patterns.

RFs control gene expression in a discrete way, such that in numerous cases gene regulation can be regarded as a logic process where the RFs are the input variables while the target gene is the result \cite{12782112,19180174,14530388,21414487, 22927416,23412653, 21885784}. In this respect, a common regulatory triplet, with two RFs regulating the same gene, can be formally described by a two-in-one-out logic gate.

Moreover, combinatorial logics are much more numerous than the three basic operations: AND, OR, and NOT \cite{14530388}. For example, for any two-in-one-out scenario, there are 16 possible logic gates (including all possible combinations between positive and negative regulators). In order to capture all possible combinatorial cooperations between regulatory factors we need a more comprehensive model. Previous studies took advantage of binarized regulatory data provided by perturbation experiments (i.e. TF knock-outs) and used a Boolean

model to capture this logic processing \cite{Somogyi}. However, previous efforts focused only on a small set of genes, missing the genome-wide identification and characterization of logic operations in gene regulation.

Here we propose a novel approach, Loregic, that will allow a comprehensive analysis of all possible regulatory logic operations from a genome-wide perspective.

3.1.2 Development of the Logic Circuit Models Approach - Loregic

Loregic is a logic operation based method that requires two types of input data: a regulatory networks (defined by regulatory factors and their target genes) and a binarized gene expression dataset across multiple samples. The binarized gene expression data (on – 1 and off – 0) is the direct result of the network's regulatory factors activity on the target genes. The inputs can be chosen from different resources to meet user interests. The regulatory network is decomposed into regulatory modules formed by triplets consisting of 2 RFs and a common target gene T. In details, Loregic algorithm comprises of five steps (Figure 2).

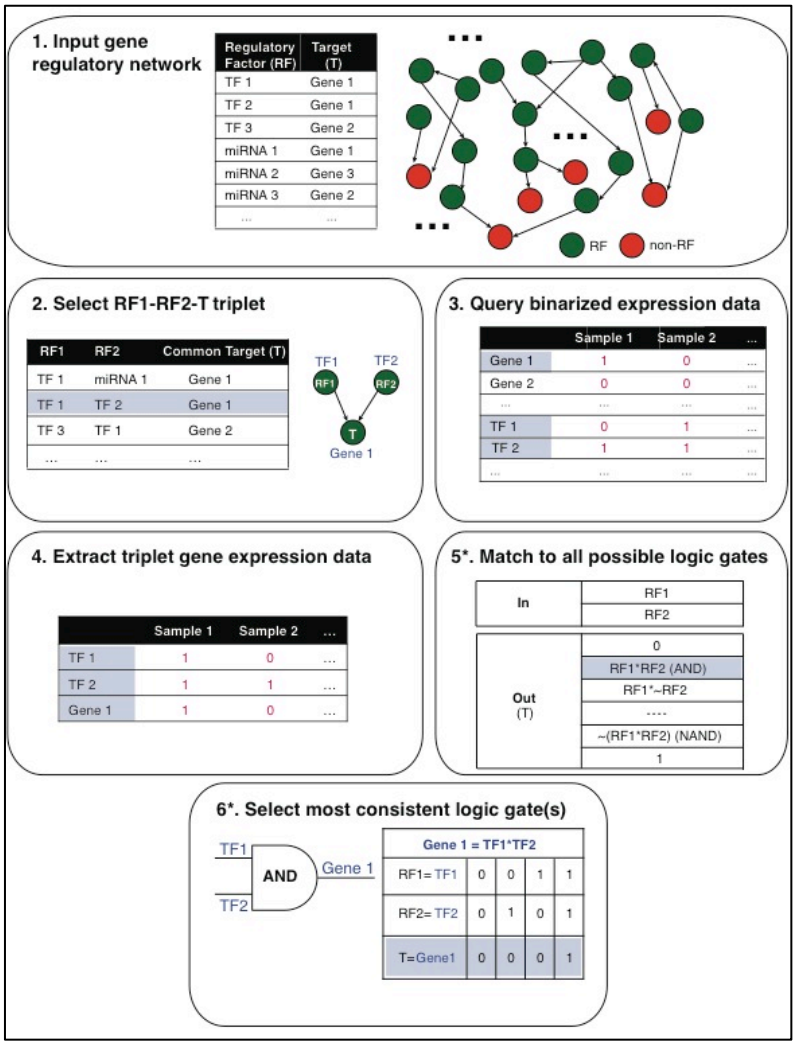


Figure 2 Loregic workflow.

- **Step 1:** Input gene regulatory network consisting of regulatory factors and their target genes;
- **Step 2:** Identify all RF1-RF2-T triplets where RF1 and RF2 co-regulate the target gene T;
- **Step 3:** Given a particular triplet (RF1, RF2 and T) query the binarized gene expression data
- **Step 4:** Match the triplet's gene expressions against all possible two-in-one-out logic gates based on the binary values;
- **Step 5:** Find the consistent logic gate(s) that best matches the expressions and calculate the consistency score. Test the score significance against random effects;
- Repeat **Step 3-5** for all triplets in the regulatory network.

Loregic describes each regulatory module (triplet) using a particular type of logic gate – i.e. the logic gate that matches best the binarized expression data for that triplet across all samples. If such a logic gate is found, we claim that the regulatory triplet is defined by a **consistent logic gate**. Next Loregic calculates the corresponding **consistency score** for the selected gate (Figure 3). Logic gate consistency score is calculated as follows.

- Create the truth table. A logic gate with two inputs (RF1, RF2) and one output (T) can be determined by a combination of four (RF1, RF2, T) binary vectors, ***v1***=(RF1=0, RF2=0, T), ***v2***=(RF1=0, RF2=1, T), ***v3***=(RF1=1, RF2=0, T), and ***v4***=(RF1=1, RF2=1, T) with specific values (0 or 1) for T.
- Given a RF1-RF2-T triplet, match output T (0 or 1) for each of four input combinations of RF1 and RF2, and find the logic gate(s) that describes best the truth tables.
- Calculate the consistency score: For any triplet with ***m*** binary vectors and any gate ***g*** the gate **consistency score** of the triplet is, $S(g)=(n_1+n_2+n_3+n_4)/m$, where n_i as number of vectors matching $v_i(g)$ with i=1,2,3,4.

Also in order to verify the authenticity of the the consistent logic gate given a triplet of (RF1, RF2, T), we calculate its significances over the 16 logic gates' scores. We suppose that it matches the k^{th} logic gate, G_k. We replace the target gene, T by a randomly selected gene N times ($N=1000$), and calculate its significance score, as $p(Gk)=(Gk)/N$. A high significance score implies that random effects may cause the matched logic gate. We suggested to select the consistent logic gates within top 2% of consistency and significance scores.

In the case where no there is no consistent logic gate found, we claim that the triplet is inconsistent with all logic gates. The negative result indicates that the cooperativity of the two RFs cannot be described by a standard logic operation.

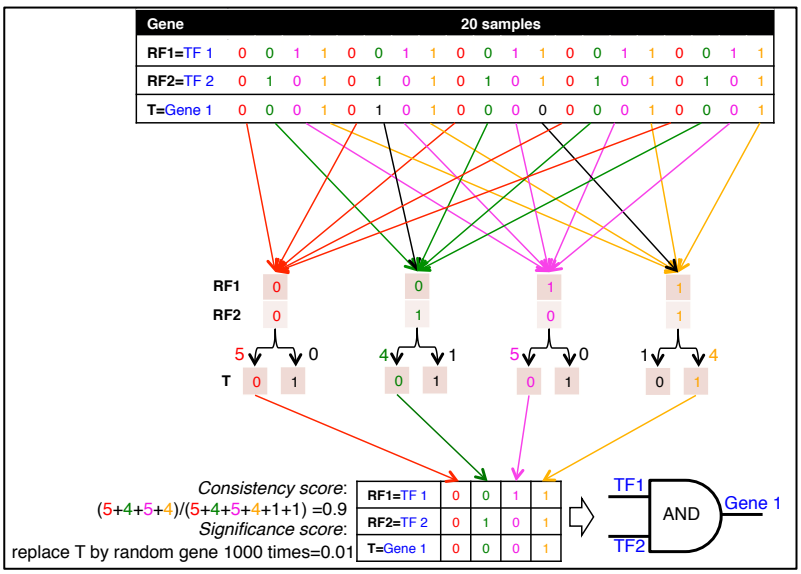


Figure 3 Consistency score evaluation

All the triplets that can be described by logic gates can be further mapped onto other regulatory networks. As such the logic gates information brings a new dimension to the interaction between regulatory elements and targets. We tested and validated the enhancement in the information obtained through Loregic on the study of yeast regulatory network. We studied the TF promoter sequence motifs and used predicted TF logics to infer potential indirect bindings.

3.1.4 Preliminary Results on Logic Circuit Models

Cooperative behaviour between yeast TFs during cell cycle

We have used logic circuit models to study the cooperation between yeast TF during cell cycle. We identified ~39k TF-TF-target triplets from 176 different TFs using TF-target assignments in \cite{15343339,19690563}. We used Loregic to characterize the TF-TF-target logics during yeast cell cycle across 59 time points. We found 4126 TF-TF-target triplets with consistent logic gates (Figure XXX). Among those, we found that "T=RF1*RF2" (AND), "T=~RF1*RF2", and "T=RF1*~RF2" logic gates, have more triplets matched than all the others. The AND triplets mean that both TFs have to be present to activate the expression of their target gene.

We tested our TF cooperativity results analysing the fold changes in the target gene expression as consequence of deleting one of the regulatory triplet TF.

The yeast TF knockout experiments gave us fold changes in gene expression as a result of deleting a single TF \cite{17417638,20385592}. If a target gene is regulated by two cooperative TFs in an "AND" relationship, deletion of either TF may corrupt the cooperativity and that impacts gene expression. For example, for the triplets with high significant scores at "AND" gate, we found that deleting either of their TFs gave rise to considerably down-regulated target genes, i.e., negative expression fold changes (t-test p-value =0.068). For non-cooperative TFs such as "T=RF1" or "T=RF2" gates, i.e., one of TFs (dominate TF) fully determines target gene expression, we found that target genes are more af- fected (down-regulated) by the removal of the dominant TFs rather than by deleting the other TFs (t-test p-value < 0.05 for T=RF1, <0.005 for T=RF2).

3.2 AIM 2: Inferring Phenotypic Functions Through Network Mining (Using Graph Models and Latent Variables)

[[AP- to add]]

3.3 AIM 3: Optimizing Phenotypic Function Prediction Using Logic Circuit Models on Biological Networks

[[CSDS, AP- to add]]

4. Tools and Sustainability

The sustainability of the developed resources is very much contingent on the hardware and servers on which they are stored and run. The existing infrastructure at Yale and Royal Holloway has served investigators well, but we aim to improve the current setup by making it reliable and robust for supporting all the proposed tools, as well as more accessible to the scientific community. To this end, we intend to make use of new technologies such as web services and cloud computing.

Specifically, we intend to use Amazon Web Services (AWS) for distributing most of developed tools, and intend to make use of the Amazon Elastic Compute Cloud EC2 (processing) and S3 (storage). AWS EC2 enables flexible, resizable online resources, and would serve as a sensible means for distributing the contents and services our tools, as it provides high performance computing, processing resources which adjust to user demand, reliability, and greater security.

Here, we summarize the various distinct components the function predictions tools, along with the means by which we intend to disseminate each:

i) **Source code**, as used in constructing the various software components will be made available from open access repositories, such as sourceforge, github, and/or google code.

ii) **Web-services** (LoREgic, AP-tools, etc.): each of the different servers would be encapsulated and made available as a virtual machines (see below for a description and the advantages of virtual machines), which may be downloaded from our servers, and then stored locally by the user, or uploaded by the user to AWS EC2.

iii) **Databases**: the regulatory network datasets will be bundled up into a single virtual machine, and distributed in a similar manner. We intend to move all the working datasets to AWS S3 for storage, and would periodically (about once a month) make backups of this data locally.

[[AP- to add]]

5. Broader Impacts

5.1 Integration of Research into Education

We propose to integrate the above described research activities into graduate and undergraduate education.

Mark Gerstein is the Co-Director of the Computational Biology and Bioinformatics (CBB) PhD program (cbb.yale.edu) at Yale University, and he has been designing and teaching graduate courses in bioinformatics, genomics, and data mining for over ten years. These activities could easily be translated into class projects, which may help recruit undergraduates into Yale labs. In addition, we focus on students of underrepresented groups through a Yale program called "Science, Technology and Research Scholars" or STARS (science.yalecollege.yale.edu/stars-home), which includes Computer Science, Bioinformatics, and Genomics components.

All the tools developed for gene function prediction will be integrated into Computational Biology and Bioinformatics 752 (*Bioinformatics: Practical Application of Simulation and Data Mining*), a course directed by Dr Gerstein, and taught to undergraduates and graduate students. The course is an introduction to the computational approaches used for addressing questions in genomics and structural biology. The function component of the course can be substantially improved by introducing the students to innovative tools to predict gene function using a variety of data. This resource represents the integration of many facets of bioinformatics, including functional data, biological network analysis, programming, as well as sets of algorithms applied to address questions about gene function discovery. It will also be integrated into final year projects, and as part of these projects, students will develop online libraries for gene function.

5.2 Workshops and Webinars: From Our Computers to Everyone's

Alberto Paccanaro is Associate Professor in Computational Biology at Centre for Systems and Synthetic Biology, Department of Computer Science, Royal Holloway, University of London. [[AP – to add]]

5.3 Advances in Function Prediction

[[CSDS – to add]]

6. Project Management Plan

The research will be conducted by graduate students and early career personnel under the supervision of Dr Mark Gerstein at Yale University, and Dr Alberto Paccanaro at Royal Holloway University of London.

In leading this collaborative project, we will draw on considerable experience we have had with other integrative consortium projects. In particular, Dr Gerstein has been an integral part of the ENCODE Project as well as the modENCODE Project since its inception. Within these he has had a number of leadership roles, as he has co-directed the Networks/Elements Group. He has co-led high profile papers focusing on networks and was the leader of the numerous collaborative papers. [[AP – to add]]

Integrating the network science expertise of Dr Gerstein with the knowledge domain of protein science and also software development expertise of Dr Paccanaro, will bring a fresh new perspective to protein function prediction.

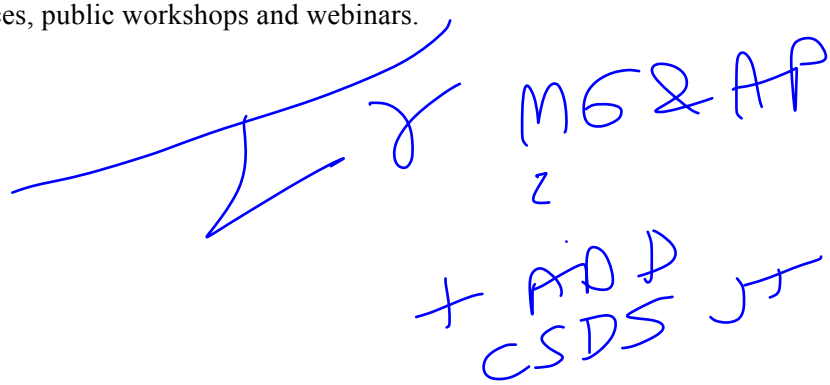
Dr Gerstein will be responsible for the coordination, designing and development of tools associated with **AIM 1**. Dr Paccanaro will be involved in the design and development of phenotype prediction tools associated with **AIM 2**. While these two aims are lead mostly by each lab independently, both groups will collaborate towards their completion. As such,

Paccanaro group will help with model development and implementation for **AIM 1**, while Gerstein group will help with assessment of data quality, standardization and biological interpretation of **AIM 2** results. The two groups will work closely together to facilitate the implementation of **AIM 3**.

The overall progress of the project is summarized in yearly milestones. As such, during the first year of the project, the Gerstein lab work will be devoted to the construction of a biological network library and the development of logic circuit models for network analysis. The successful development of computational models will be assessed by a pilot study on yeast regulatory network. The second year will be focused on developing of a robust and friendly interface for LoREgic, deploy it on the host website and also make it available through various open access repositories. The final year will be used to implement **AIM 3**. [[AP – to add milestones]]

The two groups will also coordinate the analysis and writing of collaborative papers. To achieve this, we plan to implement regular conference calls between the two groups, but also open them to the larger networks and protein function community.

We will also work closely with outside investigators from UK and US to identify additional regulatory networks datasets for integrative analysis, and coordinate the sharing of information with the larger biological research community. On a regular basis, the project results will be disseminated to a broad audience (from senior researchers to middle and high school teachers) through conferences, public workshops and webinars.



CheckList:
Important Proposal Preparation Information: FastLane will check for required sections of the full proposal, in accordance with Grant Proposal Guide (GPG) instructions described in Chapter II.C.2. The GPG requires submission of:

- Project Summary;
- Project Description;
- References Cited;
- Biographical Sketch(es);
- Budget;
- Budget Justification;
- Current and Pending Support;
- Facilities, Equipment & Other Resources;
- Data Management Plan; and
- Postdoctoral Mentoring Plan, if applicable.

If a required section is missing, FastLane will not accept the proposal.