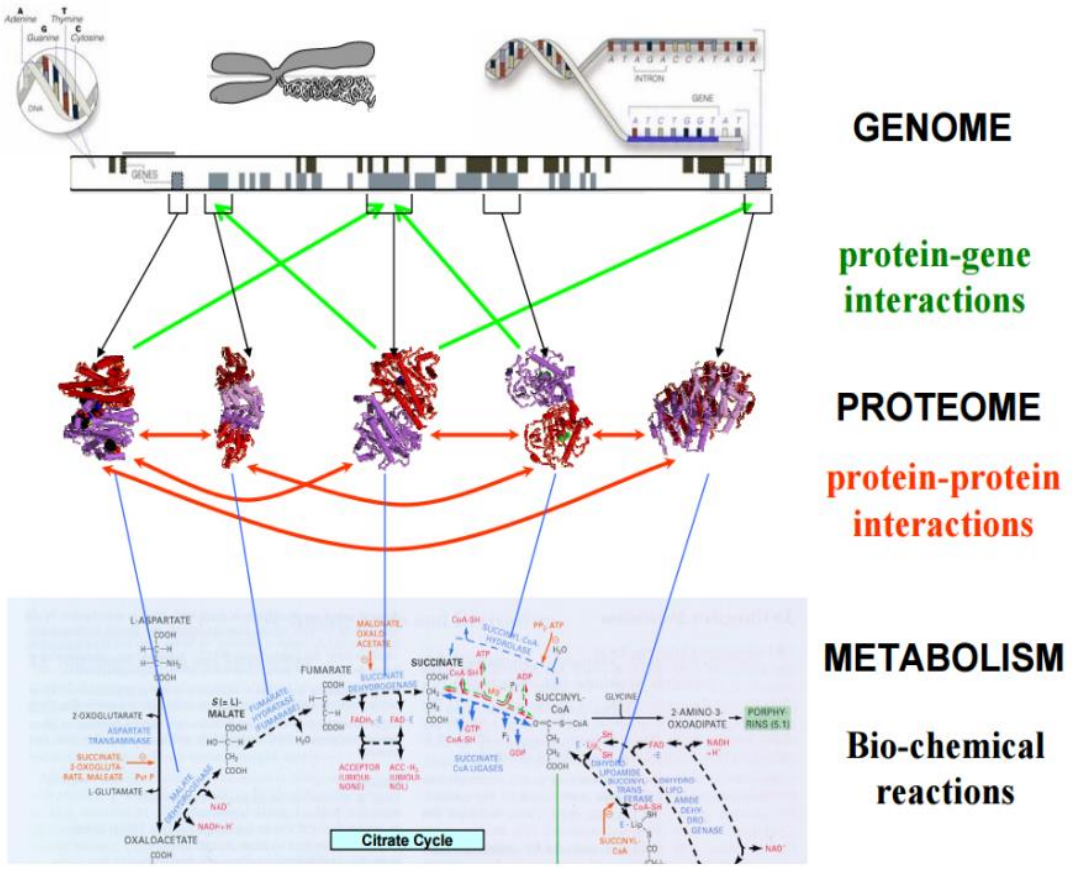
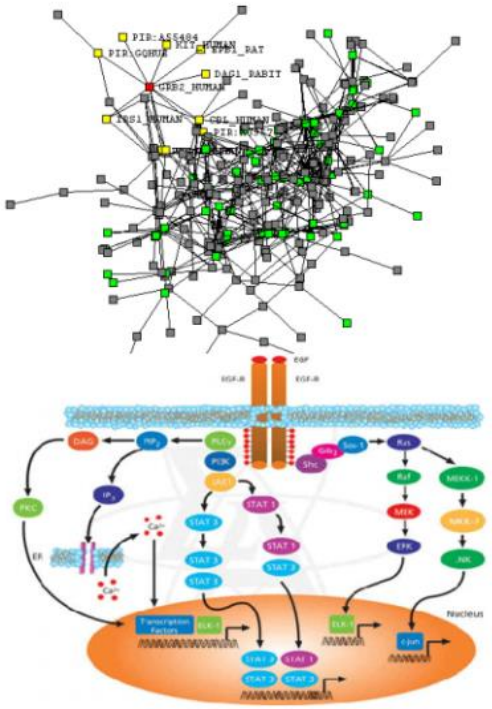
UNIT -4 MODELING THE DYNAMICS OF CELLULAR NETWORKS

Modeling the dynamics of cellular networks

Systems-level understanding of cellular dynamics is important for identifying biological principles and may serve as a critical foundation for developing therapeutic strategies. To date, numerous developments of therapeutics have been based on identification and comprehensive analysis of cellular dynamics, especially in the involved pathways. In cancer therapy, for instance, many researchers have focused on oncogenic pathways such as the Rb pathway, whose in-depth understanding of the pathway dynamics promises effective therapeutics.



Cellular processes form networks on many levels



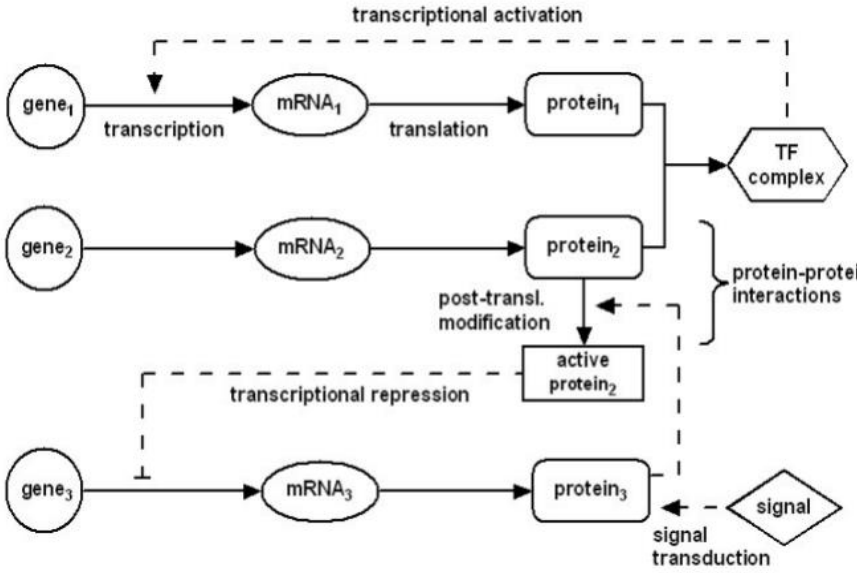
Protein interaction networks

- Nodes: proteins
- Edges: protein-protein interactions (binding)

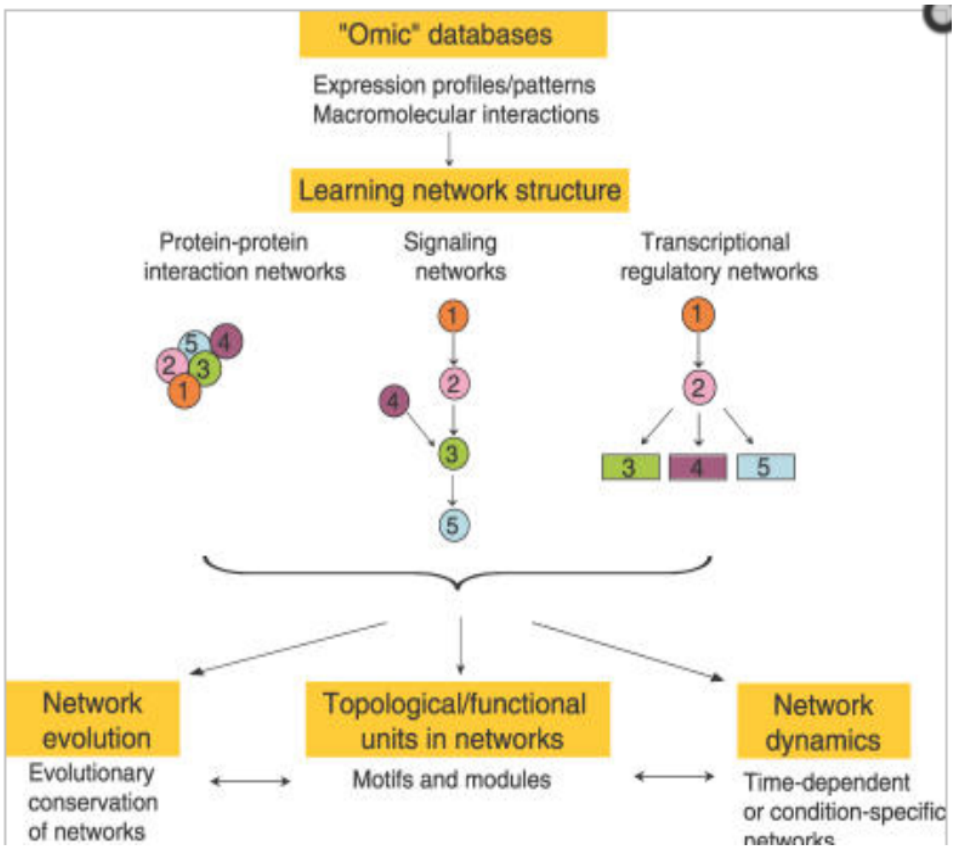
Signal transduction networks

- Nodes: proteins, molecules
- Edges: reactions and processes reflecting information transfer (e.g. ligand/receptor binding, protein conformational changes)

Signaling, gene regulation and protein interactions are intertwined



An Overview of Biological Network Analyses Based on "Omic" Data

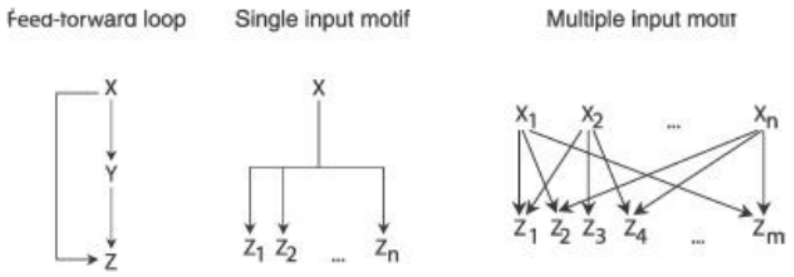


Recent high-throughput technologies have produced massive amounts of gene expression, macromolecular interaction, or other type of "omic" data. Using a computational modeling approach, the architecture of cellular networks can be learned from these "omic" data, and topological or functional units (motifs and modules) can be identified from these networks. Comparisons of cellular networks across different species may reveal how network structures evolve. In particular, the evolutionary conservation of motifs and modules can be an indication of their biological importance. A dynamic view of cellular networks describes active network components and interactions under various conditions and time points. Network motifs and modules can also be time-dependent or condition-specific.

Modularity of Cellular Networks

Unlike random networks, cellular networks contain characteristic topological patterns that enable their functionality. To find the basic building blocks of cellular networks, simple units consisting of a few components were enumerated and some of them were found to be significantly overrepresented . These recurring units were defined as network motifs. For instance, transcriptional network motifs include feed-forward loops, single-input motifs, and multi-input motifs. A feed-forward loop describes a situation in which a transcription factor (TF) regulates a

second TF, and these two TFs jointly regulate a common target gene. A single-input motif contains one TF which regulates a set of target genes, such as subunits of a protein complex. A multi-input motif consists of multiple TFs that regulate a set of target genes, providing the possibility of combinatorial controls. These motifs are found in multiple organisms such as bacteria, yeast, and human. This structural conservation suggests functional importance of network motifs for transcriptional regulation.



Network Motifs Found in *E. coli* Transcriptional Regulatory Networks

(Left) Feed-forward loop: TF X regulates TF Y, and both X and Y jointly regulate gene Z.

(Middle) Single-input motif: TF X regulates genes Z_1, Z_2... and Z_n.

(Right) Multi-input motif: a set of TFs X_1, X_2... and X_n regulate a set of target genes Z_1, Z_2... and Z_m.

The components of cellular networks, including proteins, DNA, and other molecules, act in concert to carry out biological processes. These functionally related components often interact with one another, forming modules in cellular networks. While motifs represent recurrent topological patterns, modules are bigger building units that exhibit a certain functional autonomy. Modules may contain motifs as their structural components. Modules may maintain certain properties such as robustness to environmental perturbations and evolutionary conservations.

Modularity exists in a variety of biological contexts, including protein complexes, metabolic pathways, signaling pathways, and transcriptional programs. For transcriptional programs, modules are defined as sets of genes controlled by the same set of TFs under certain conditions. Gene expression experiments often do not reveal direct regulations. However, if we assume that the expression profiles of regulators provide information about their activities, expression data

contains information about regulatory relationships between regulators and their target genes. Bayesian networks, directed probabilistic graphical models , were applied to obtain a modular map of *Saccharomyces cerevisiae* transcriptional regulatory networks based on multiple microarray datasets. Protein–DNA binding data provides direct physical evidence of regulatory interactions. Therefore, combining genome-wide protein–DNA binding data with gene expression data improves the detection of transcriptional modules over using either data source alone. While each module has a distinct combination of regulators, modules that share regulators can be grouped together

Cellular Networks as a Dynamic System

A living cell is a dynamic system, where gene activities and interactions exhibit temporal profiles and spatial compartmentalizations. Interactions presented in a static network may not necessarily occur simultaneously. A typical example is Cdc28p, a cyclin-dependent kinase with a constant expression profile, which interacts with a variety of cyclins at different phases of the cell cycle. Dynamic descriptions of networks are necessary for an accurate understanding of cellular events. By integrating yeast PPI networks with gene expression data, In a study that described dynamic protein complex formation during cell cycles, it was found that constitutively expressed and cell cycle–regulated proteins together form protein complexes at particular time points during the cell cycle. This suggested a general mechanism of "just-in-time-assembly," where only some subunits of protein complexes are regulated during cell cycle progression and the synthesis of these subunits control the timing of complex assembly. "Just-in-time-assembly" may be a more efficient way of regulation compared with "just-in-time-synthesis," in which case all subunits of protein complexes are regulated and synthesized at the same time during the cell cycle.

Computational Methods in Network Modeling Using "Omic" Data

Clustering

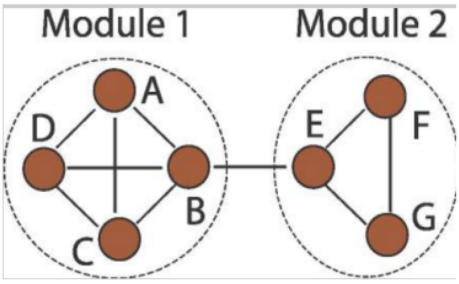
Clustering methods are widely used to find modules in transcriptional regulation. An expression profile dataset can be represented as a two-dimensional matrix where rows index genes and columns index experimental conditions. Clustering methods partition genes into groups such that

genes in each group show similar expression across conditions or through a time series. Since regulation by common TFs may only occur under certain conditions, bi-clustering methods have been developed to identify genes that express similarly under a subset of conditions. It should be noted that genes with similar expression may not all be co-regulated, and that clustering does not necessarily identify the corresponding regulators. Therefore, genes clustered together may not fully represent modules in transcriptional regulatory networks.

Traditional clustering methods, such as K-means, require a predefined and fixed number of gene clusters, which may be hard to assign in practice and greatly influence the results. They also do not model temporal dependence between expression profiles. To address these issues, Schliep and coworkers and Beal and Krishnamurthy applied Hidden Markov Models to cluster gene expression time course data. Specifically, both of them used Hidden Markov Models to model temporal dependence of gene expression, instead of treating different time points independently. While Schliep and coworkers proposed a heuristic approach to determine the number of clusters, Beal and Krishnamurthy used a nonparametric prior distribution on mixture weights, such that the genes can be clustered without a predefined number of clusters.

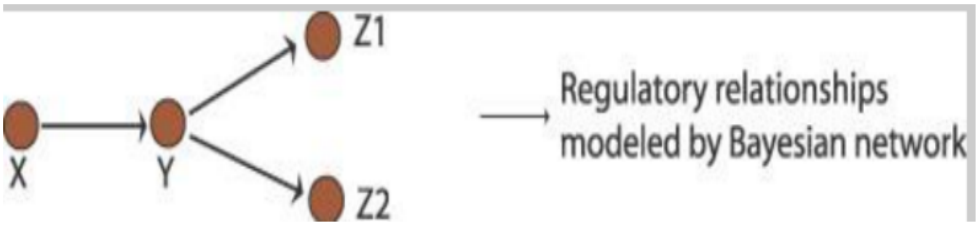
(b) Topology-based analysis

Interaction networks are often visualized as graphs where nodes represent genes/proteins and edges represent interactions. Modular structures can be inferred based on topological features of the networks. For example, densely connected subgraphs can be exhaustively identified in PPI networks. These suggest the existence of multi-protein complexes. Also, modules can be identified using topological distances in the networks. More specifically, the distance between two nodes is defined as the length of the shortest path(s) between them. A matrix of distances between all pair-wise combinations of nodes can be used for clustering. The underlying assumption is that proteins in a module have similar distances to proteins outside of the given module.



c) Probabilistic graphical models

Nodes of probabilistic graphical models represent variables, and edges represent independency relations among the variables . According to the directionality of edges, graphical models can be classified into two major categories: Bayesian networks and Markov random fields. A Bayesian network is a directed acyclic graphical model: if there is an edge from node *X*pointing to another node *Y,* then values of variable *Y* depend directly on values of *X* and *X* is called a *parent* of *Y*. Coupled with intervention data, Bayesian networks can be used to learn causal relationships, and are thus suitable to model transcriptional regulatory networks or signaling pathways. In contrast, the edges in Markov random fields are undirected, which makes them suitable to model PPI networks or other networks of symmetric interactions



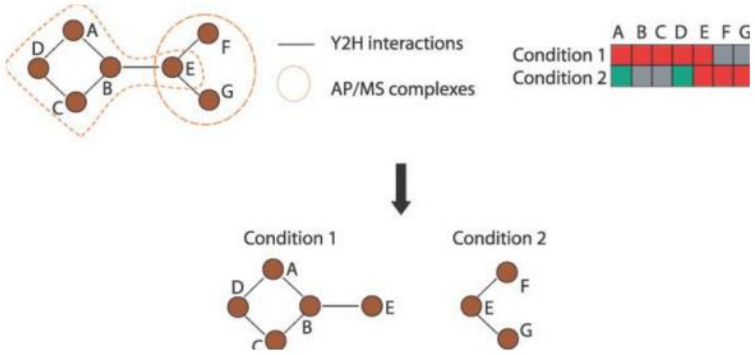
To use graphical models, we need to systematically learn the structures of networks based on biological data and to estimate the parameters of these networks. The learned graphical models reveal how proteins and genes interact, which can be applied to answer different biological queries as an inference problem. For example, when the activities of a protein are suppressed, cells may respond by changing the expression levels of other genes. Such responses can be predicted based on a learned regulatory network

While the task of learning Bayesian networks has been well-addressed, learning Markov random fields is still in its early stage. If we use graphical models to model large-scale biological

networks containing structural loops such as PPI networks, the inference problem is not trivial. Monte Carlo methods or approximate inference methods can be used to solve such problems

(d) Integration of various data sources

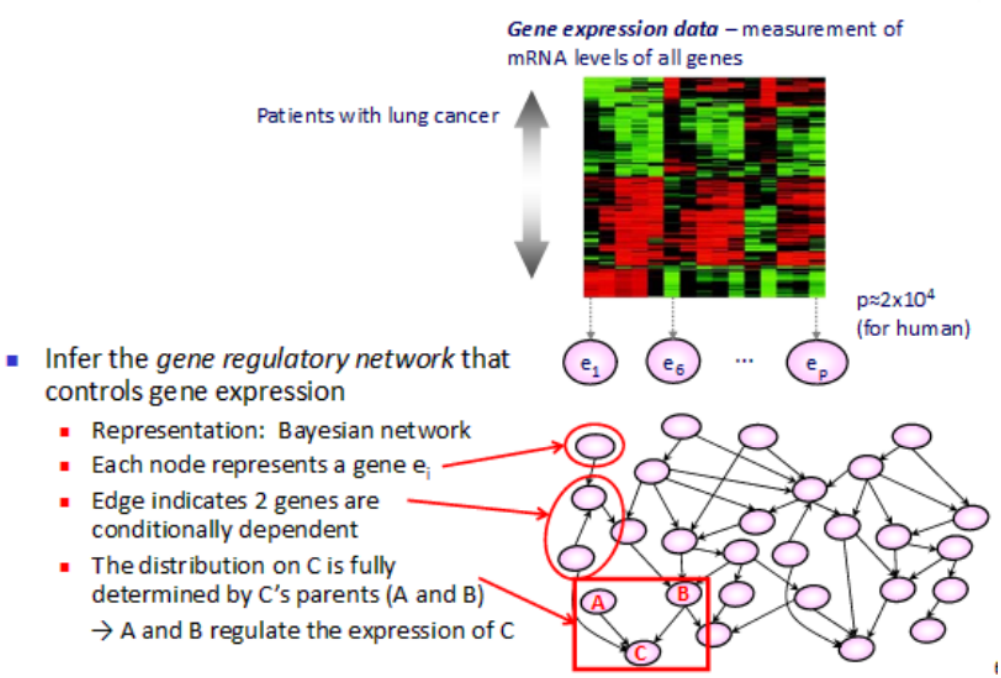
Individual high-throughput biological datasets are usually both incomprehensive and error-prone. Therefore, data integration becomes indispensable in order to model cellular networks accurately and to make functional inferences. For example, both yeast two-hybrid and affinity-purification/mass-spectrometry experiments have been applied to the mapping of PPI networks. Overlapping the two data sources enables the identification of high-confidence interactions. In addition, yeast two-hybrid detects binary relationships while affinity-purification/mass-spectrometry detects proteins as members of a complex. Integrating these two types of data helps to model the actual topology of protein complexes. Furthermore, if temporal, spatial, or conditional expression data are available, it may be possible to provide a dynamic view of protein complexes under physiological conditions



Why network?

- DNA,RNA,protein,and other biological don't Operate alone.
- Instead,they operate as part of complex pathways Or networks.
- Understanding networks can lead to betterunderstanding Of gene regulatory mechanism,disease process,Evolutionary process, etc.

Inferring biological network:



Transcriptome analysis of regulatory networks.

Biology Knowledgebase Implementation Plan describes scientific objectives the Knowledgebase will support in microbial, plant, and metacommunity research areas.

Once RNA-Seq data (short sequences) are collected from a particular growth state for a specific species (step 1),

they will be mapped back to their associated genome sequence (step 2)

and the reads/bp (reads per base pair) will be calculated as a measure of each gene's or operon's expression level (step 3).

The reads/bp will be displayed in conjunction with the genome sequence (step 4) using the latest version of Artemis, which already has this capability.

Rules will be generated to define operons (step 5) based solely on these data.

The output of this analysis will be a list of operons and their expression level for each growth state of every species analyzed.

Using OrthoMCL to help define orthologous genes , orthologous operons will be identified in related genomes (step 9) and used to identify as many orthologous promoters as possible (step 10).

Next, the transcription factor binding sites (TFBS) for these promoters will be predicted using two separate techniques. One will involve multiple sequence alignment of the orthologous promoters in an attempt to define the TFBS (step 11) based on their conservation.

This technique depends upon the number of sequenced, related genomes and the total genetic distance between all the organisms in each alignment.

The average nucleotide identity (ANI) thus will be used to estimate if there will be sufficient sequence divergence in an alignment.

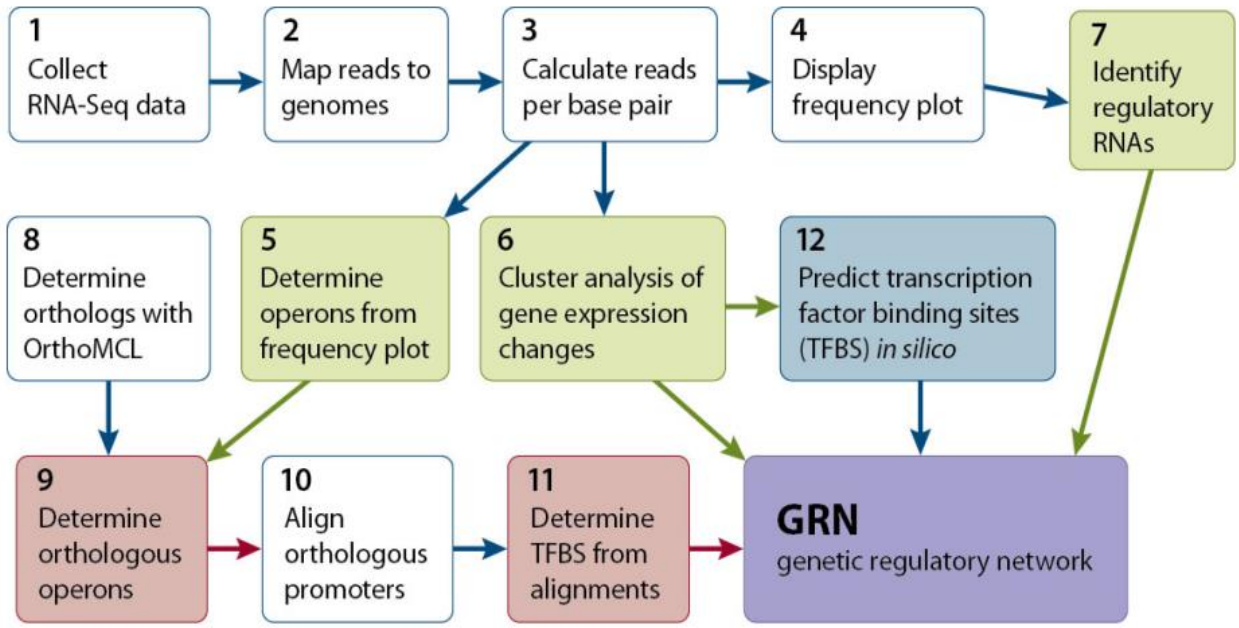
If the orthologous operons can be identified in more distant relatives, attempts will be made to expand the alignments.

The second technique will use more traditional TFBS prediction algorithms. Results from both techniques will be compared for consistency.

Next, cluster analysis will be performed on the differences in gene (operon) expression identified in the RNA-Seq data (step 6).

Finally, small regulatory RNAs will be identified from the frequency plot (step 7).

In this workflow, white boxes represent current capabilities and procedures.



Intracellular analysis