Transcriptional regulatory networks: inference and evolution

Sushmita Roy
Goals for today

• Background
  – Components of the regulation machinery
  – Transcriptional gene regulation

• Challenges in regulatory networks
  – Element identification
  – Network identification
    • Extensions to inference
  – Network structure analysis

• Evolution of regulatory networks
  – Comparative functional genomics
Gene Regulation

Collection of biological processes that determine what set of genes get expressed when and where.
What regulates gene expression?

Transcriptional gene regulation

**Input:** Transcription factor level (*trans*)

Transcription factor binding sites (*cis*)

**Output:** mRNA levels

Transcriptional regulatory network connects TFs to target genes
Goals for today

• Background
  – Components of the regulation machinery
  – Transcriptional gene regulation

• Challenges in regulatory networks
  – Element identification
  – Network identification
    • Extensions to inference
  – Network structure analysis

• Evolution of regulatory networks
  – Comparative functional genomics
Challenges in regulatory networks

1. Parts Identification
   - Regulators
   - Regulatory Motifs
   - Target genes

2. Network identification
   - Identifying edges and their function
   - Element activity
   - Structure
   - Function
   - \( X = f(A, B) \)
   - \( Y = g(B) \)

3. Network Structure Analysis
   - Hubs, degree-distributions, Network motifs

- ATTAAT
- CGCTT
- Element activity
Element identification

• Elements
  – Regulators: Transcription factor proteins
  – Targets: Sequence-specific binding sites

• Computational approaches
  – Regulators: Sequence alignment
  – Motifs: De novo motif discovery
  – Targets: Sequence specific motif scanning
Goals for today

• Background
  – Components of the regulation machinery
  – Transcriptional gene regulation

• Challenges in regulatory networks
  – Element identification
  – Network identification
    • Extensions to inference
    – Network structure analysis

• Evolution of regulatory networks
  – Comparative functional genomics
Network identification

Who are the regulators?

How they determine expression levels?

Structure

Function

\( \psi(X_1, X_2) \)

BOOLEAN
LINEAR
DIFF. EQNS
PROBABILISTIC
...
Approaches to Network identification

• Wet-lab approaches
  – ChIPseq/ChIP-chip
  – Genetic perturbations

• Computational approaches
  – What data to learn networks?
    • Motifs, ChIP binding assays, Expression
  – How to learn networks?
    • Supervised network inference
    • Unsupervised network inference
  – How to evaluate network usefulness?
Types of data

• Physical
  – ChIP-chip and ChIP-seq
  – Sequence specific motifs
  – Measure static information

• Functional
  – Gene co-expression
  – Measure dynamic information
Supervised learning of TF-target interactions

Define:

\[ I_{12} = \begin{cases} 
1 & \text{if X and regulates Y} \\
0 & \text{otherwise} 
\end{cases} \]

Given:

XY.features: Attributes of X and Y

We need:

Prob. of regulating: \( P(I_{12}=1|XY.features) \)

Prob. of not regulating: \( P(I_{12}=0|XY.features) \)
Supervised learning of TF-target interactions

Positive examples:

A → B
C → D
E → F

Feature extraction

FEATURE SET

TRAINING CLASSIFIER

Predicted regulatory edges:

E → A
G → L

Training

Testing

Negative examples:

G → H
I → J
K → L

Positive examples
Inferring the regulatory network of the fly

Marbach, Roy et al., 2012
Supervised, integrative approach recovers more ground truth edges

- Integrative networks outperform single-feature networks
Unsupervised network inference

Models differ in the function that maps input system state to output state.

Booleans Networks

Differential equations

\[
\frac{dX_3(t)}{dt} = \kappa g(X_1(t), X_2(t)) - rX_3(t)
\]

Probabilistic graphical models

\[
P(X_3|X_1, X_2) = N(X_1a + X_2b, \sigma)
\]
Probabilistic graphical models (PGMs)

- A marriage between graph and probability theory
  - Handle noise and uncertainty
  - Nodes: Random variables
  - Edges: statistical dependency among random variables
- Model the joint probability distribution
  - Parameters: mathematical description of relations
- Enable incorporation of prior knowledge
Graphical models for unsupervised network inference

• Bayesian networks
• Dependency networks

Random variables encode expression levels

Goal: learn the structure and function of these networks
Some notation

• Random variables
  \[ \mathbf{X} = X_1, \cdots, X_N \]

• Joint assignment
  \[ \mathbf{x}_d = x_{1d}, \cdots, x_{Nd} \]

• Dataset
  \[ D = \{ \mathbf{x}_1, \cdots, \mathbf{x}_d \} \]

• Joint probability distribution
  \[ P(\mathbf{X} = \mathbf{x}_d) \]
Bayesian networks: estimate a set of conditional probability distributions

$P(Y_i|\text{Pa}(X_1, \cdots, X_p))$

Function: Conditional probability distribution (CPD)

JPD: product of conditionals per variable
The learning problems

• Parameter learning on known structure
  – Estimate $\theta_i$ of the conditionals

• Structure learning
  – Find the statistical dependency structure
  – Subsumes parameter learning
Parameter learning

Maximum likelihood parameter estimation

\[ \hat{\theta} = \arg \max_\theta P(D|\theta, G) \]

Data likelihood

\[ P(D|\theta, G) = \prod_{d=1}^{\mid D \mid} P(X = x_d | \theta, G) \]
Structure learning

- Maximum likelihood framework

\[ \hat{\mathcal{G}} = \arg \max_{\mathcal{G}} \max_{\theta} P(D|\theta, \mathcal{G}) \]
Structure learning using score-based search

\[ \text{Score}(\mathcal{G}) = P(D|\mathcal{G}, \theta) \]

\[ \hat{\mathcal{G}} = \arg \max_{\mathcal{G}} \max_{\theta} P(X|\theta, \mathcal{G}) \]

Best graph

Maximum likelihood
Learning network structure is computationally expensive

- For $N$ variables there are $2^{\binom{N}{2}}$ possible networks:
- Set of possible networks grows super exponentially

<table>
<thead>
<tr>
<th>$N$</th>
<th>Number of networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>1024</td>
</tr>
<tr>
<td>6</td>
<td>32768</td>
</tr>
</tbody>
</table>

Need approximate methods to search the space of networks
Approximation strategies

• Search the parent set independently
• Restrict the size of the parent set
• Assume linear relationships
Dependency networks: a set of regression problems

\[ Y_i = X_{1:}^{p} \]

Regularization term

\[ b_i^* = \arg \min_{b_i} ||Y_i - X_i \ast b_i|| + f(b_i, \lambda) \]

1 \leq i \leq m

Number of genes
Regularized linear regression

• Lasso: sparsity

\[ b_i^* = \arg \min_{b_i} ||Y_i - X_i \ast b_i|| + \lambda |b_i| \]

• Ridge regression: smoothness

\[ b_i^* = \arg \min_{b_i} ||Y_i - X_i \ast b_i|| + \lambda ||b_i|| \]

• Elastic net: sparsity + smoothness
Goals for today

• Background
  – Components of the regulation machinery
  – Transcriptional gene regulation

• Challenges in regulatory networks
  – Element identification
  – Network identification

  • Extensions to inference
    – Network structure analysis

• Evolution of regulatory networks
  – Comparative functional genomics
Extensions to vanilla network inference approaches

• Making methods more scalable
• Imposing biological constraints
• Integrating other types of data
Concept: Expression modules

\[
\min_{C} \sum_{k=1}^{\mid C \mid} \sum_{i,j} d(X_i, X_j)
\]
Learning regulatory programs of modules instead of genes

- Bayesian Network formalism
- No cyclic dependencies
- Target genes share CPDs
- Modules are re-visited

But every gene has the same set of parameters

Lee et al 2009, Segal et al 03
Combine per-gene and per-module network inference methods

How to impose module constraints?
Keep regulators same but params different
Group lasso for module constrained per gene model

\[
\begin{align*}
\mathbf{Y}^{d \times 1} & = \mathbf{X}^{p \times d} \mathbf{B}^{p \times k} \\
\mathbf{B}^{*} & = \arg \min_{\mathbf{B}} \| \mathbf{Y} - \mathbf{X}\mathbf{B} \|^2_2 + \lambda \sum_{i=1}^{p} \| \mathbf{B}_{i,:} \|^2_2 
\end{align*}
\]
Example coefficient matrix

A regulator is selected for all or no genes
Integrating data as structure priors
Revisiting Structure learning

- Bayesian framework
- $\mathcal{G}$ is an unknown random variable
- Optimize posterior distribution of graph given data

\[
P(\mathcal{G}|D) = P(D|\mathcal{G})P(\mathcal{G})
\]

\[
P(\mathcal{G}|D) \propto P(\mathcal{G}) \int P(D, \theta|\mathcal{G}) d\theta
\]

\[
P(\mathcal{G}|D) = P(D|\mathcal{G}, \theta_{MAP})P(\mathcal{G})
\]
A structure prior to integrate data

• Let $P(G)$ distributes independently over edges

$$P(G) = \left[ \prod_{X_i \rightarrow X_j} P(X_i \rightarrow X_j) \right] \left[ \prod_{X_i \not\rightarrow X_j} (1 - P(X_i \rightarrow X_j)) \right]$$

• Define prior probability of edge presence/absence

$$P(X_i \rightarrow X_j) = \frac{1}{1 + \exp\left(-\left(\beta_1 + \beta_2 w_{ij}\right)\right)}$$

Graph structure complexity Prior strength Edge prior strength
Behavior of graph structure prior

\[ P(X_i \rightarrow X_j) = \frac{1}{1 + \exp(-\beta_1 - \beta_2 w_{ij})} \]
Effect of prior on graph structure

| β₁  | 0.5 | -2  | -4  | -4  | -4  |
| β₂  | 0.4 | 0.4 | 0.4 | 2   | 4   |
| TFs | 199 | 141 | 92  | 96  | 108 |
| Known Edges | 3% | 3.2% | 3.25% | 6.4% | 18.5% |
| Score | -6890 | -8319.53 | -9216 | -9187 | -9055 |
Goals for today

• Background
  – Components of the regulation machinery
  – Transcriptional gene regulation

• Challenges in regulatory networks
  – Element identification
  – Network identification
    • Extensions to inference
      – Network structure analysis

• Evolution of regulatory networks
  – Comparative functional genomics
Hierarchical nature

- Regulators are hierarchically organized with different roles per level
  - Top: Master regulators influence many genes
  - Middle: Bottle necks directly targeting most genes
  - Bottom: Essential regulators

Yu & Gerstein 2006, Jothi et al. 2009
Modularity of regulatory networks

- Modular: Graph with densely connected subgraphs
- Genes in modules involved in similar functions and co-regulated
- Modules can be identified using graph partitioning algorithms
  - Markov Clustering Algorithm
  - Girvan-Newman Algorithm

Structural network motifs

Auto-regulation

Multi-component

Feed-forward loop

Single Input

Multi Input

Regulatory Chain

Feed-forward loops involved in speeding up in response of target gene

Lee et al. 2002, Mangan & Alon, 2003
Network motifs often have specific functions

![Diagram](attachment://network_motifs.png)

Simple regulation and autoregulation.

**Positive autoregulation (PAR)**: X activates its own promoter.

**Negative autoregulation (NAR)**: X represses its own promoter.

**Simple regulation**: X regulates Y, which in turn regulates Z, with no self-regulation.

**Positive autoregulation**: X activates its own promoter, leading to an increase in expression.

**Negative autoregulation**: X represses its own promoter, leading to a decrease in expression.

**Response acceleration (or speed-up) by NAR**: has been accompanied by an overshoot or damped oscillations. The dynamics of NAR show a rapid initial rise followed by a slower decrease towards the steady state.

**Flow of protein production**: The rate of production is determined by the concentration of X. High concentrations lead to a rapid increase in protein level, whereas low concentrations cause a slower increase.

**Positive autoregulation** demonstrates delay after stimulus (S→X→Y→Z) and no delay after its removal, which shows signal integration and memory storage in simply regulated genes. These systems are often used in natural contexts, such as the arabinose system in E. coli, which uses a synthetic repressor, TetR, that is designed to repress its own expression in a manner similar to how naturally occurring repressors work.

**Negative autoregulation** in the SOS DNA-repair system of E. coli also shows characteristic behavior. This system uses a negative feedback mechanism to regulate the expression of genes involved in DNA repair.

**Delay (represented by purple squares)**: Z concentration relative to the steady state Z = due to an inherent source of noise: the production rates of proteins that are not actively degraded, as is the case for the degradation rate, the shorter the response time. For example, in the repression threshold of X to its own promoter, the production rate of X increases when a transcription factor enhances its own expression, whereas low concentrations cause a decrease. NAR can reduce these variations experimentally by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured.

**Experimentally**: The response acceleration (speed-up) by NAR has been demonstrated by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured. The transcription level shows a rapid initial rise followed by a slower decrease towards the steady state.

**Delay (represented by red circles)**: shows a delay after addition of the input signal (cAMP), and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.

**Experimentally**: The coherent type-1 feedforward loop (C1-FFL) and its dynamics were discussed in this Review. It shows delay after stimulus (S→X→Y→Z) and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.

**Experimentally**: The coherent type-1 feedforward loop (C1-FFL) and its dynamics were discussed in this Review. It shows delay after stimulus (S→X→Y→Z) and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.

**Flow of protein production**: The rate of production is determined by the concentration of X. High concentrations lead to a rapid increase in protein level, whereas low concentrations cause a slower increase.

**Response acceleration (or speed-up) by NAR**: has been accompanied by an overshoot or damped oscillations. The dynamics of NAR show a rapid initial rise followed by a slower decrease towards the steady state.

**Negative autoregulation** in the SOS DNA-repair system of E. coli also shows characteristic behavior. This system uses a negative feedback mechanism to regulate the expression of genes involved in DNA repair.

**Delay (represented by purple squares)**: Z concentration relative to the steady state Z = due to an inherent source of noise: the production rates of proteins that are not actively degraded, as is the case for the degradation rate, the shorter the response time. For example, in the repression threshold of X to its own promoter, the production rate of X increases when a transcription factor enhances its own expression, whereas low concentrations cause a decrease. NAR can reduce these variations experimentally by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured.

**Response acceleration (or speed-up) by NAR**: has been demonstrated by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured. The transcription level shows a rapid initial rise followed by a slower decrease towards the steady state.

**Delay (represented by red circles)**: shows a delay after addition of the input signal (cAMP), and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.

**Experimentally**: The coherent type-1 feedforward loop (C1-FFL) and its dynamics were discussed in this Review. It shows delay after stimulus (S→X→Y→Z) and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.

**Flow of protein production**: The rate of production is determined by the concentration of X. High concentrations lead to a rapid increase in protein level, whereas low concentrations cause a slower increase.

**Response acceleration (or speed-up) by NAR**: has been accompanied by an overshoot or damped oscillations. The dynamics of NAR show a rapid initial rise followed by a slower decrease towards the steady state.

**Negative autoregulation** in the SOS DNA-repair system of E. coli also shows characteristic behavior. This system uses a negative feedback mechanism to regulate the expression of genes involved in DNA repair.

**Delay (represented by purple squares)**: Z concentration relative to the steady state Z = due to an inherent source of noise: the production rates of proteins that are not actively degraded, as is the case for the degradation rate, the shorter the response time. For example, in the repression threshold of X to its own promoter, the production rate of X increases when a transcription factor enhances its own expression, whereas low concentrations cause a decrease. NAR can reduce these variations experimentally by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured.

**Response acceleration (or speed-up) by NAR**: has been demonstrated by Besckei and Serrano using a fluorescent-reporter assay, which allows the transcription level to be measured. The transcription level shows a rapid initial rise followed by a slower decrease towards the steady state.

**Delay (represented by red circles)**: shows a delay after addition of the input signal (cAMP), and no delay after its removal, which demonstrates signal integration and memory storage in simply regulated genes.
Goals for today

• **Background**
  – Transcriptional gene regulation
  – *cis* and *trans* elements

• **Challenges in regulatory networks**
  – Element identification
  – Network identification
    • Extensions to inference
  – Structural properties of networks

• **Evolution of regulatory networks**
  – Comparative functional genomics
Why understand evolution of regulatory networks

Importance in evolution of complex body plan:

"Although a variety of ways of thinking about evolution have been proposed, the evolution of the body plan is fundamentally a system-level problem to which GRN structure/function provides the most compelling direct access" Peter & Davidson, 2011
Factors affecting regulatory network evolution

**trans**

- environmental signals
  - sensors
  - signal transduction

**cis**

- chromatin modifiers
  - feedback

**trans-factors**

1. Transcription factors
2. Chromatin modelers
3. Signaling proteins
4. Environment

**cis-factors**

1. Binding sites
2. Nucleosomes
3. Histone marks

Key questions

• How conserved are regulatory networks?
  – Elements
  – Connections
• How are different conservation/divergence scenarios implemented?
• What is the ancestral state?
• Do regulatory differences explain functional innovation?
Comparative genomics approaches to understanding regulation evolution

Phylogenetic relationships to compare sequence differences, and relate to phenotypic traits.
Scenarios of conservation & divergence

Network of Organism 1

Network of Organism 2

Conserved

Target is conserved but program is not

Program is conserved but targets are not

Hao Li and Johnson, 2010
How do regulatory networks rewire?

We found surprisingly little divergence of the molecular function of paralogs, but substantial divergence in gene regulation, as reflected in the cis-regulatory elements, the transcription factors (TFs) bound to the genes' promoters, and the gene regulon to which they belong. This is consistent with the idea that regulatory divergence occurs at an elevated rate compared to divergence of the coding sequence of paralogs.

Two likely scenarios can underlie diverged expression of paralogs. Regulatory subfunctionalization occurs when multiple, distinct regulatory elements controlling expression of the single ancestral gene are 'split' between its two descendants, such that each paralog retains only some of the regulatory inputs. In contrast, regulatory neo-functionalization occurs when one paralog evolves a new control not used by the ancestral gene. A recent study proposed neo-functionalization as the dominant mode of diverged expression of *S. cerevisiae* paralogs, since often the expression pattern of only one paralog was distinct from that of the single preduplication ortholog in *C. albicans*.

ED of paralogs may have an important adaptive role. First, regulatory divergence can effectively lead to sub-functionalization of paralogs, even if paralogs have the same biochemical function (but serve different roles because of expression differences). This may have occurred in the post-WGD regulatory divergence of glycolysis/gluconeogenesis enzymes, such as the hexose kinases, Hxk2 and Hxk1, that share hexokinase activity but display distinct expression patterns. Notably, sub-functionalization may be partial and remnants of the shared (joint) ancestral control may still allow paralogs to 'backup' each other.

Finally, constraints on expression levels could in turn influence the evolution of gene copy number, since low-ED genes also exhibit few duplication and loss events, whereas high-ED genes have substantial variation in copy number between species.

Flexibility in regulatory mechanisms can drive expression divergence. Genetic changes in both cis and trans elements can contribute to ED. A genetic change can affect expression in cis, either directly by altering regulatory sequences controlling gene expression, or indirectly by modifying the activity of the gene's product and consequently affecting expression through feedback. Polymorphisms in cis appear to contribute most to ED in phylogenetically close species.

Regulatory evolution in Ascomycota fungi

Alternative modes for the regulatory evolution of transcriptional modules. Each panel shows a distinct scenario of the inferred evolution of an ancestral program (at tree root) into programs observed in extant species (leaves).

(a) Conservation of both cis-elements (boxes) and TFs (ovoids).

(b) Elaboration of a program by the emergence of a new cis-element in addition to the ancestral one.

(c) Divergence of a program by the loss of cis-element.

(d–f) Divergence of a program through cis-redundancy or trans-redundancy. (d) In cis, a switch from one cis-element to another through an intermediate that has both sites, bound by distinct factors. (e) In trans, two paralogous TFs bind the same site following gene duplication. After subfunctionalization, only one of the factors controls the module. (f) In trans, switching of one TF for another that does not share ancestry, but can bind a similar site.

Wolbach, Thompson et al., Current Opinions in Genetics & Dev. 2009
But, we know only a handful of examples from the pre-mRNA era.
Functional genomics approaches to understanding regulation evolution
Systematic approaches to compare regulatory networks

• One species at a time
  – Infer a regulatory network per species
  – Compare networks across species

• Learn multiple networks simultaneously
  – Use phylogenetic relationships to constrain the network structure
Learning networks one species at a time

Genes

Genome-wide mRNA profiles

Species X

Species Y

Species Z

Module Identification

Per-module network identification
Comparing networks across species

Easy case: One to one orthologs:

Not so easy cases: One to many orthologs:
Defining an edge match

\[ E_{AB}^Y : \{(i, j) \in \{A_{y1}, A_{y2}\} \times \{B_{y1}, B_{y2}\}\} \]

\[ A_X \rightarrow B_X \text{ is conserved in } Y \text{ if } E_{Y}^{AB} \neq \emptyset \]
Using yeast Ascomycetes to understand regulatory evolution

**Respiro-fermentative**: use fermentation (ethanol production) when grown on glucose

**Respirative**: use respiration when grown on glucose

300 million years of evolution
Experiments for capturing functional response

Dawn Thompson

Heat shock

Heat shock (severe stress)

Jenna Pfifner
Ilan Wapinski

Topologically networks look similar, but have very few common edges

<table>
<thead>
<tr>
<th></th>
<th>Scer</th>
<th>Cgla</th>
<th>Scas</th>
<th>Kwal</th>
<th>Klac</th>
<th>Calb</th>
<th>Sjap</th>
<th>Spom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scer</td>
<td>1.00</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Cgla</td>
<td>0.04</td>
<td>1.00</td>
<td>0.03</td>
<td>0.09</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Scas</td>
<td>0.02</td>
<td>0.03</td>
<td>1.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Kwal</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>1.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Klac</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>1.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Calb</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>1.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Sjap</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>1.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Spom</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Conservation

| 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |

Pairwise network similarity

C. glabrata network

S. cerevisiae network
Take-away messages

• Transcriptional regulatory networks determine context specific gene expression
  – Important in development and disease
• Most of the regulatory network is not known
• Machine learning approaches to network inference
  – Supervised
  – Unsupervised
• Extensions to existing inference algorithms
  – Incorporate biological intuition
  – Integrate different types of datasets
• Evolution of regulatory networks
  – Major player for diversifying phenotypic diversity of organism
  – Comparative functional genomics brings new opportunities
    • Need phylogenetically-aware network analysis algorithms
For further reading, discussions, chats

sroy@biostat.wisc.edu