

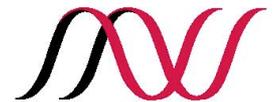
Computational tools for systems analysis

Sushmita Roy

Dept. of Biostatistics and Medical Informatics

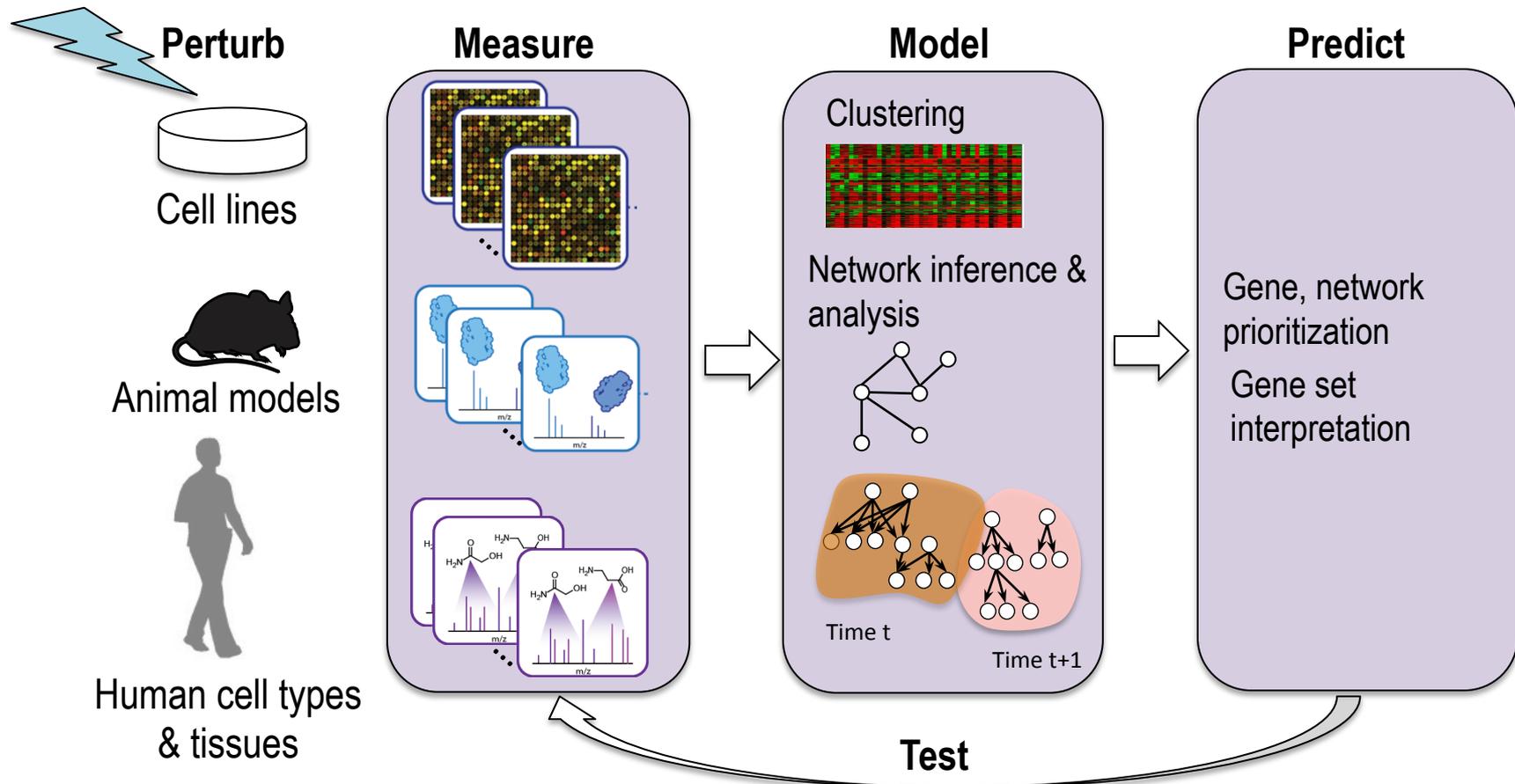
Wisconsin Institute for Discovery

UW Madison



Systems Analysis Goals

To provide an integrated understanding of cellular response to environmental signals



Analysis strategies

- Clustering/Dimensionality reduction approaches
- Network reconstruction approaches
- Network-based data integration
- Network-based interpretation
 - Gene prioritization methods
 - Gene set analysis

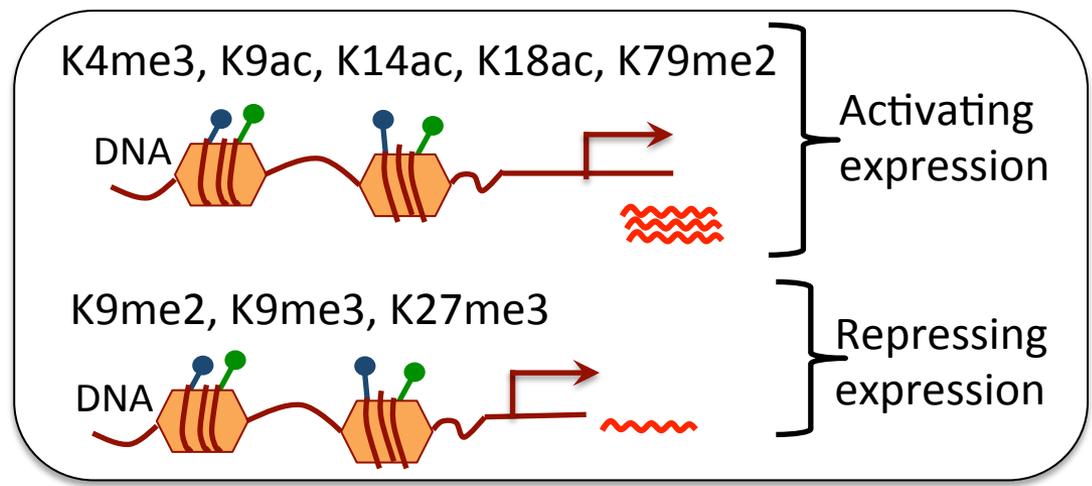
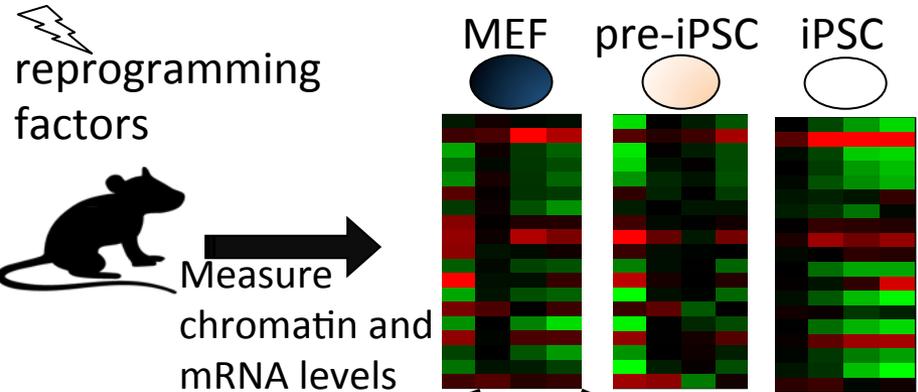
Analysis strategies

- Clustering/Dimensionality reduction approaches
- Network reconstruction approaches
- Network-based data integration
- Network-based interpretation
 - Gene prioritization methods
 - Gene set analysis

Characterizing chromatin state dynamics during cellular reprogramming

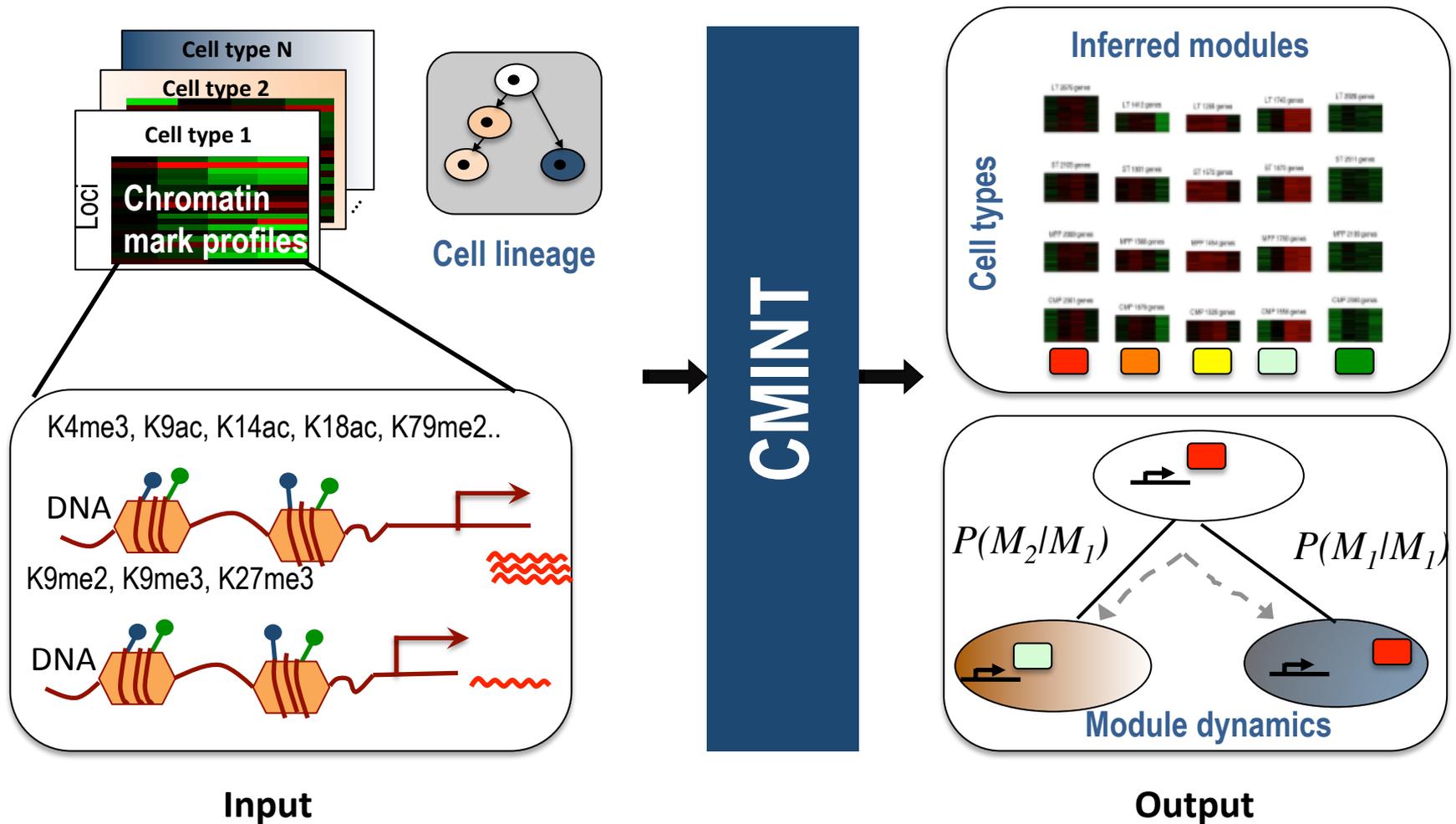


Rupa Sridharan

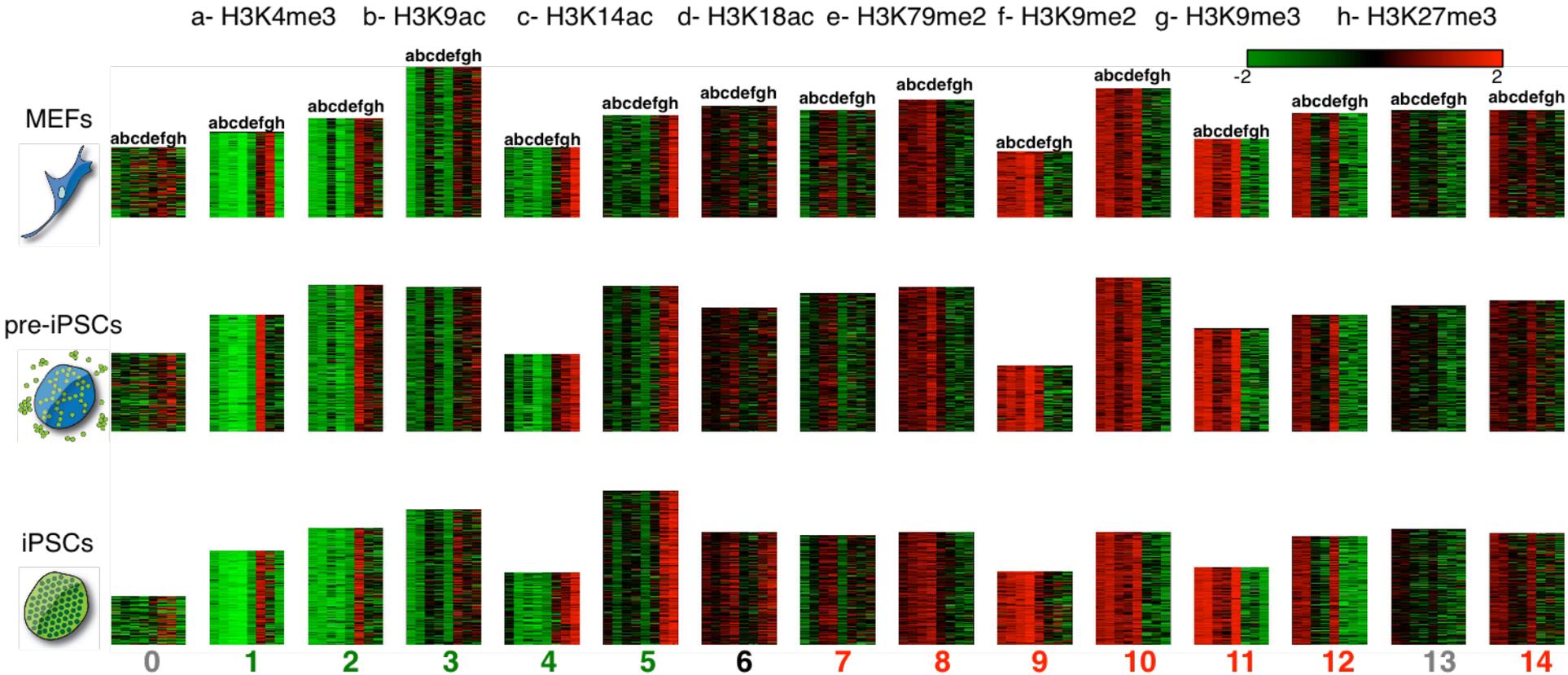


What genes change their chromatin state?

Chromatin Module INference on Trees (CMINT)

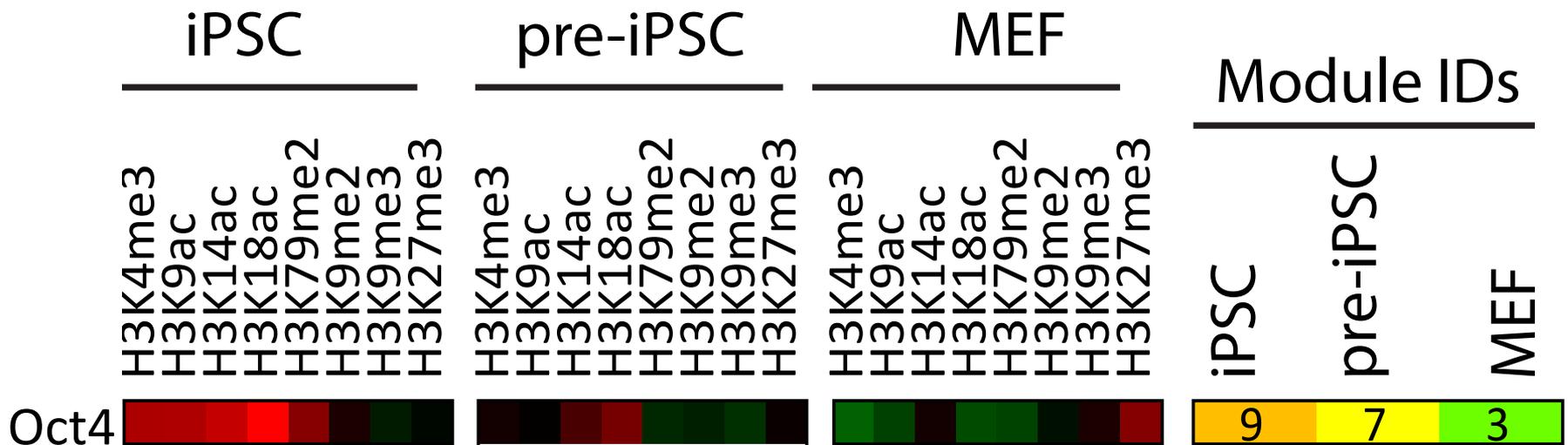


Chromatin state during reprogramming is defined by 15 different patterns



Chromatin state dynamics of a pluripotency regulator

Switched its chromatin state in iPSC cells but not (completely) in pre-iPSCs

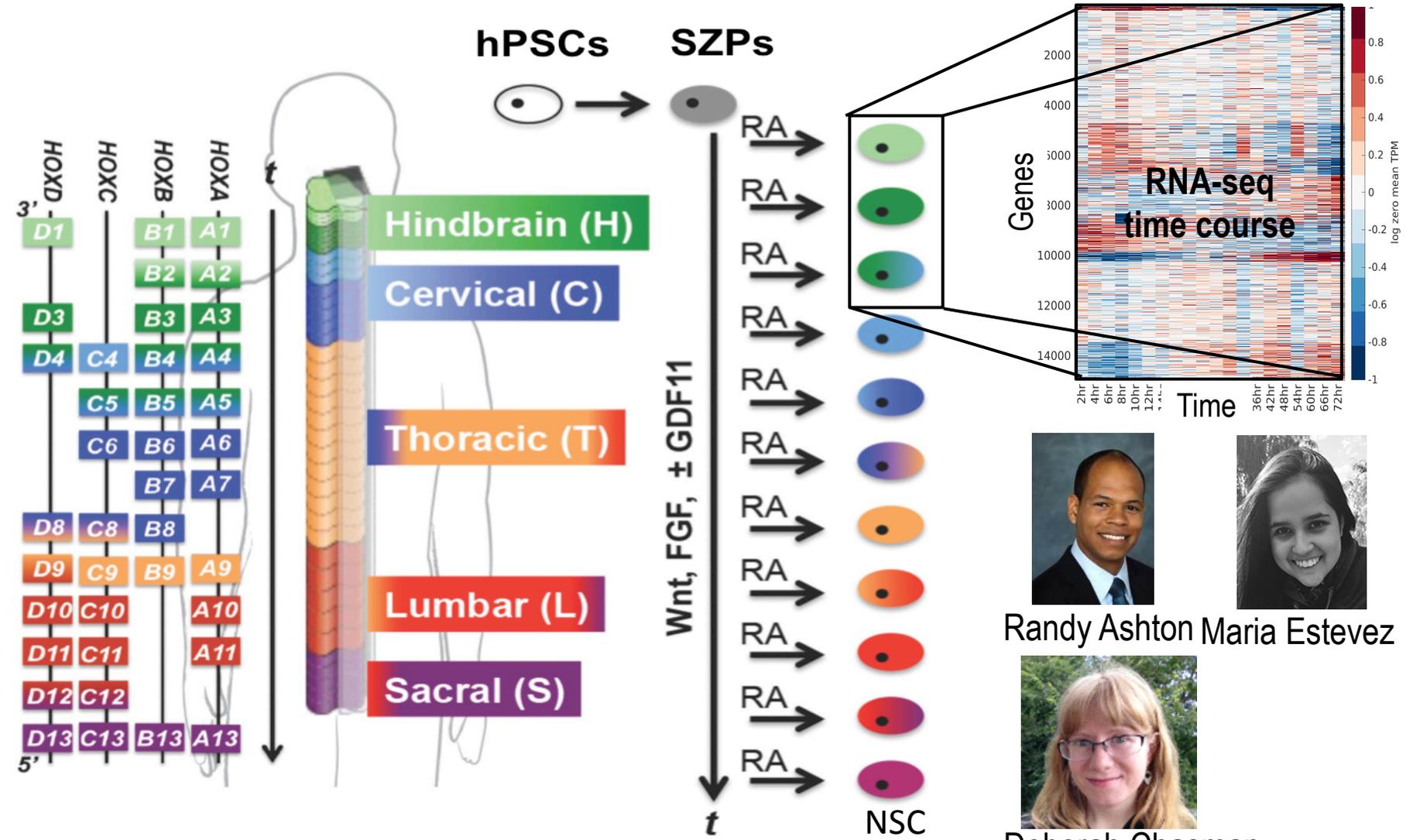


CMINT states can be used to identify novel genes with an Oct4 signature

Analysis strategies

- Clustering/Dimensionality reduction approaches
- **Network reconstruction approaches**
- Network-based data integration
- Network-based interpretation
 - Gene prioritization methods
 - Gene set analysis

Network transitions during NSC differentiation



Randy Ashton



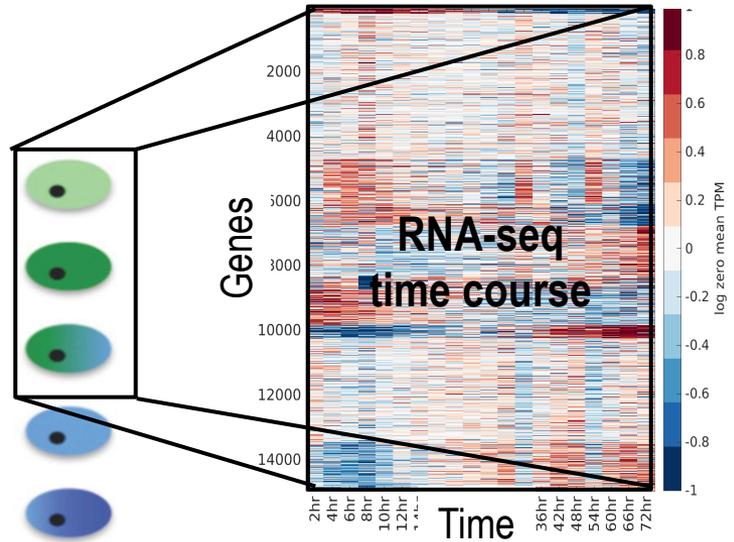
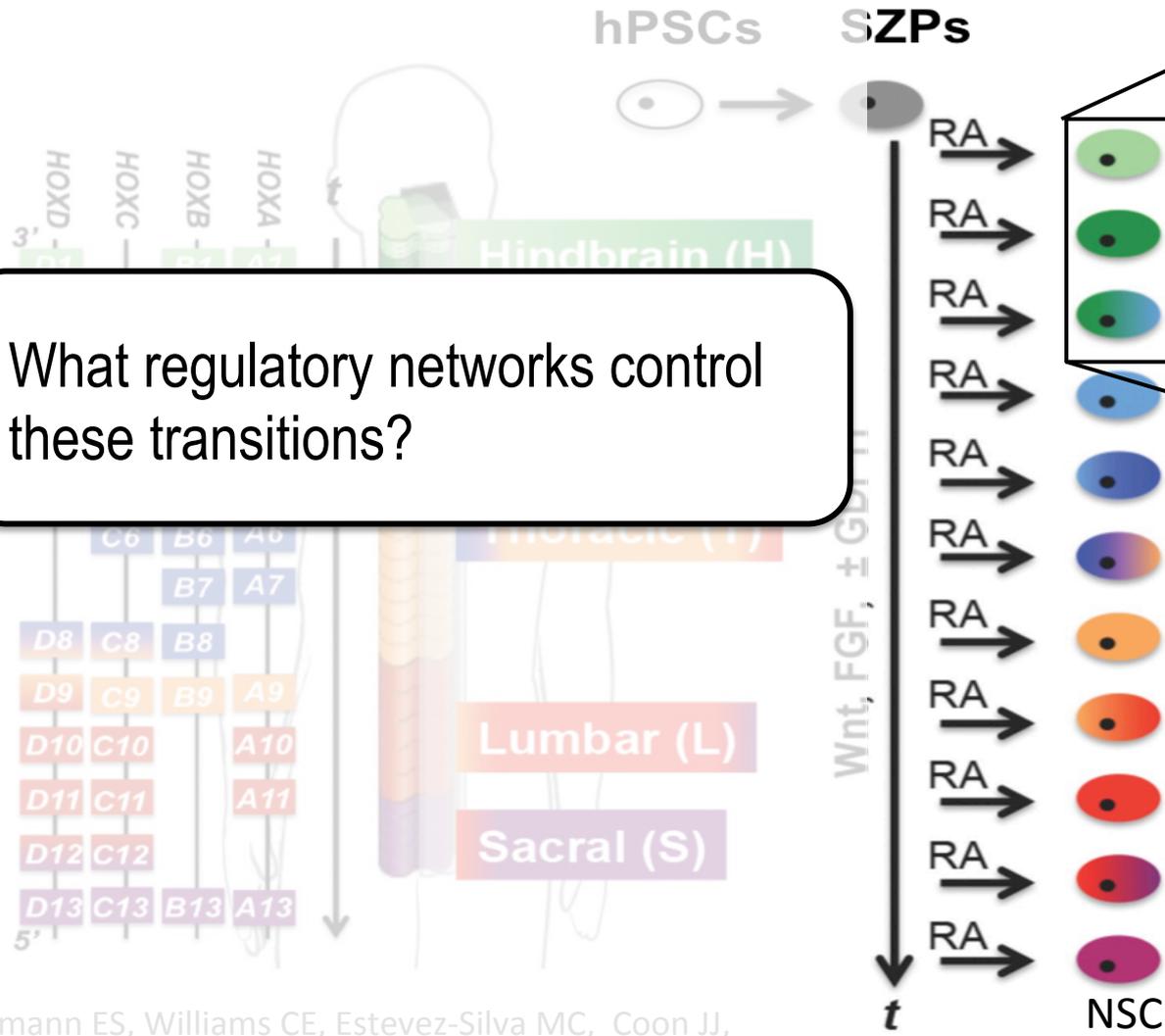
Maria Estevez



Deborah Chasman

Network transitions during NSC differentiation

What regulatory networks control these transitions?



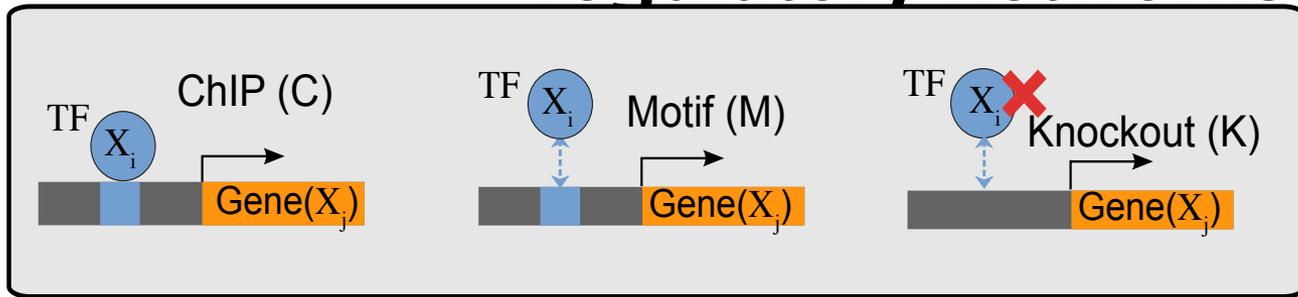
Randy Ashton Maria Estevez



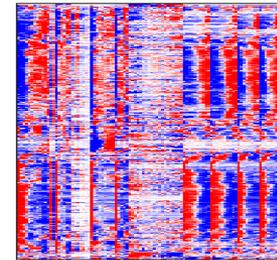
Deborah Chasman

Lippmann ES, Williams CE, Estevez-Silva MC, Coon JJ, and Ashton RS. *Stem Cell Reports* 4 (2015)

MERLIN-P: Integrative inference of gene regulatory networks



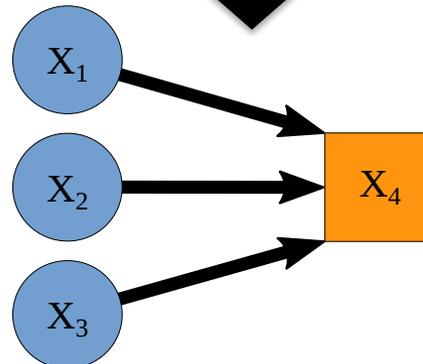
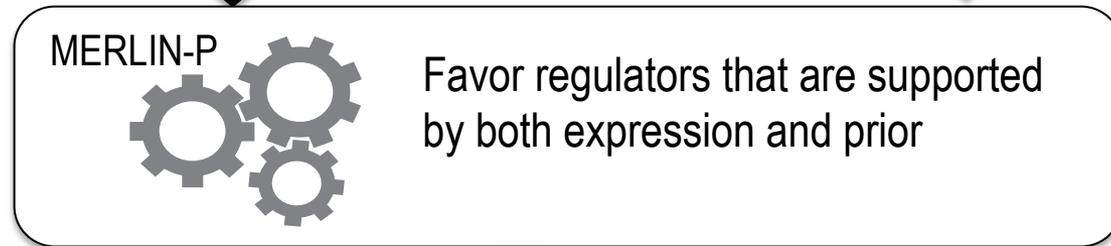
Samples



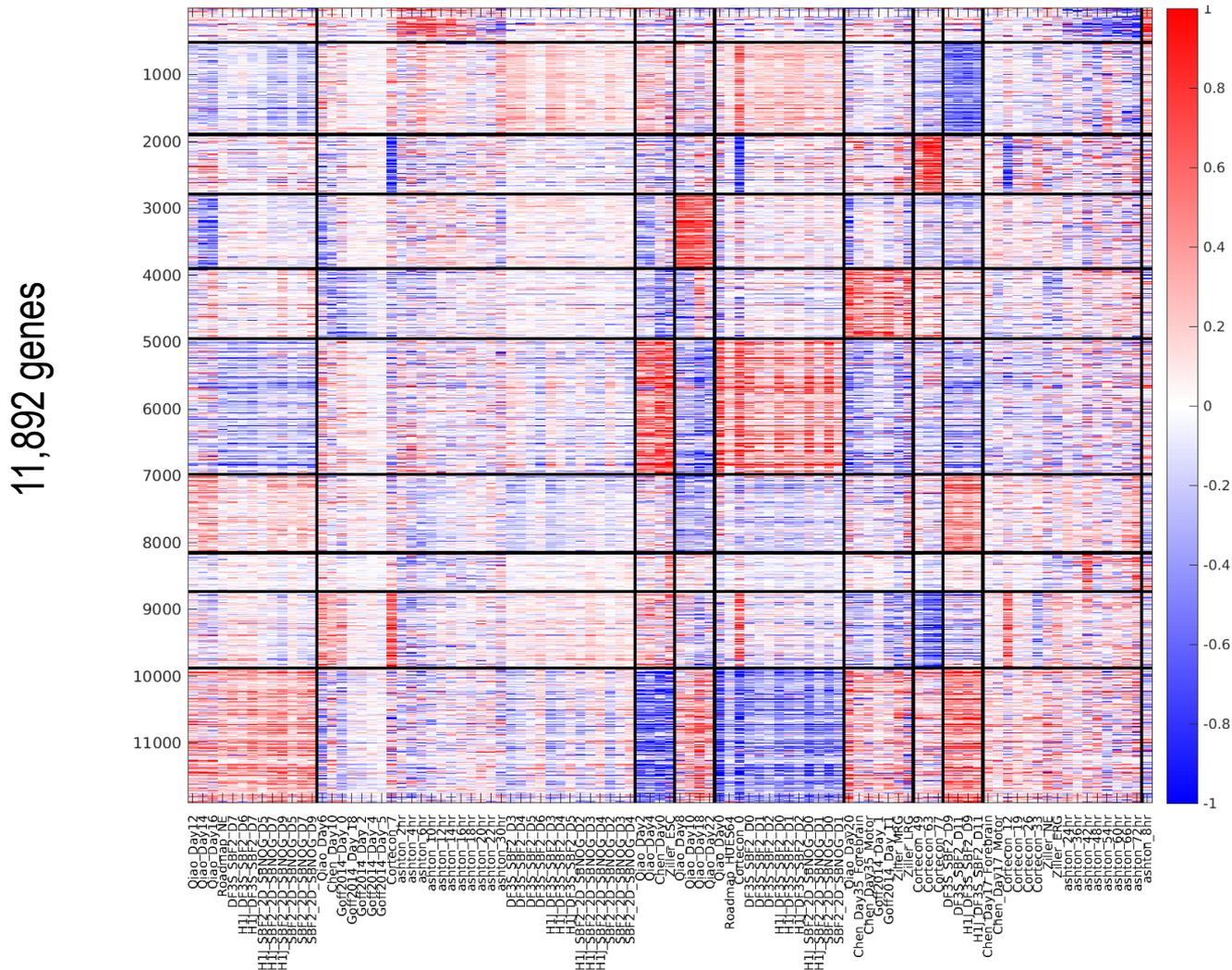
Genes

Different types of prior knowledge

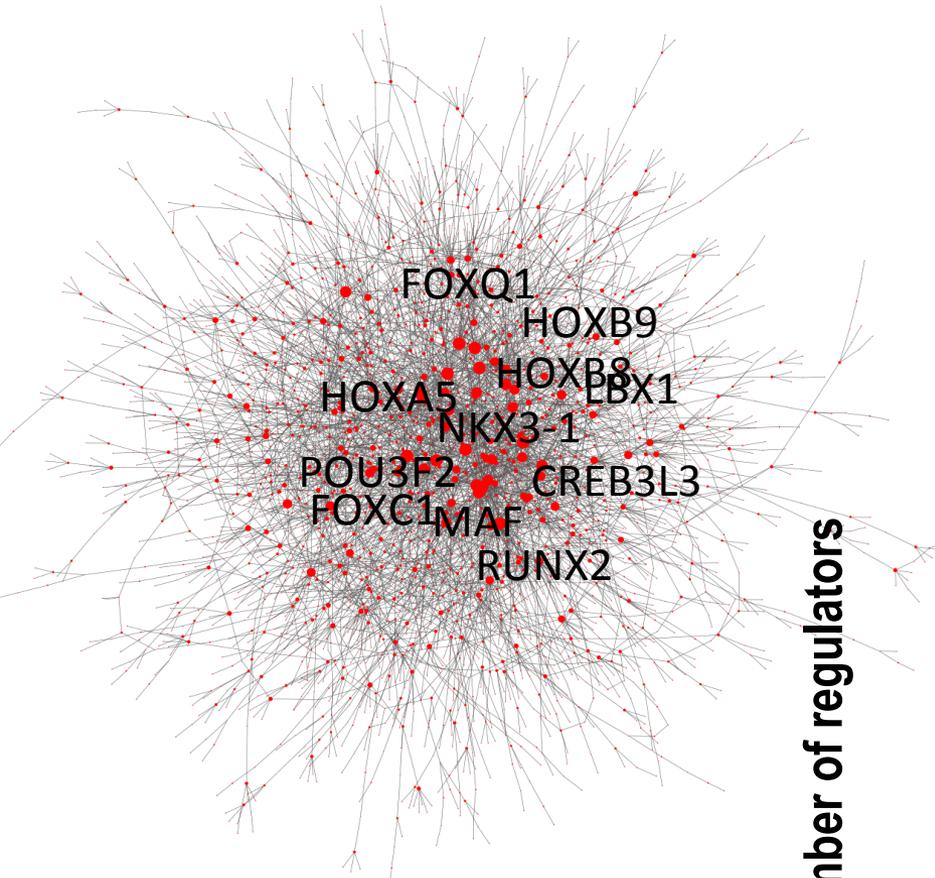
Gene expression data



Publicly available and new RNA-seq data used to construct network

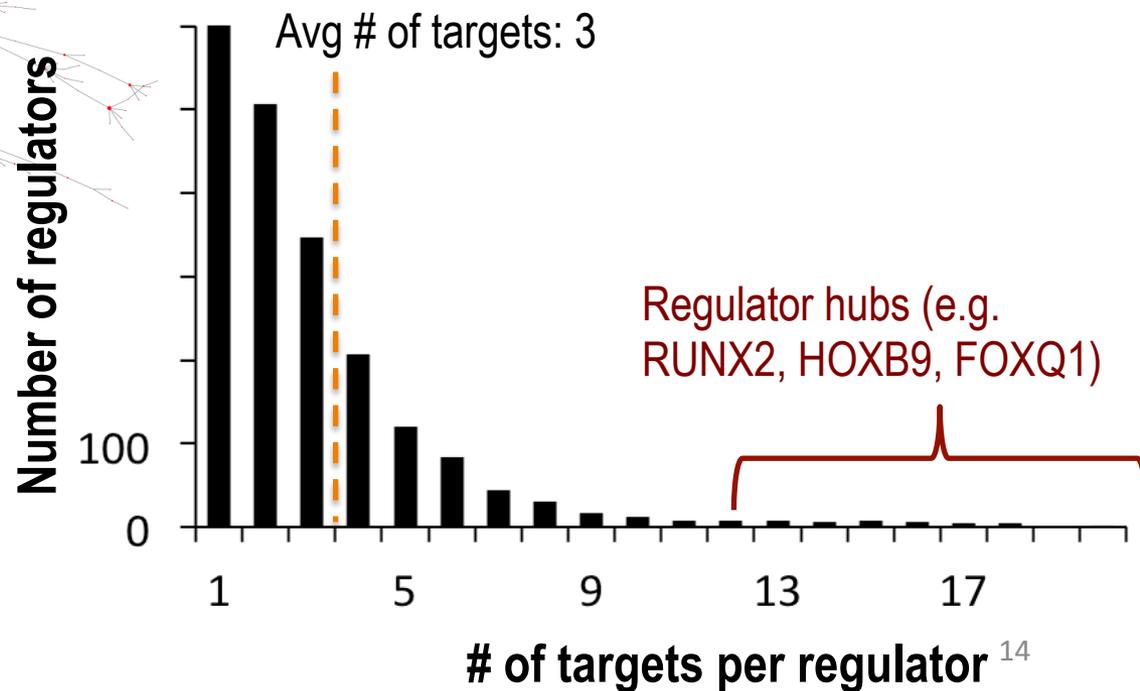


Overview of inferred NSC regulatory network

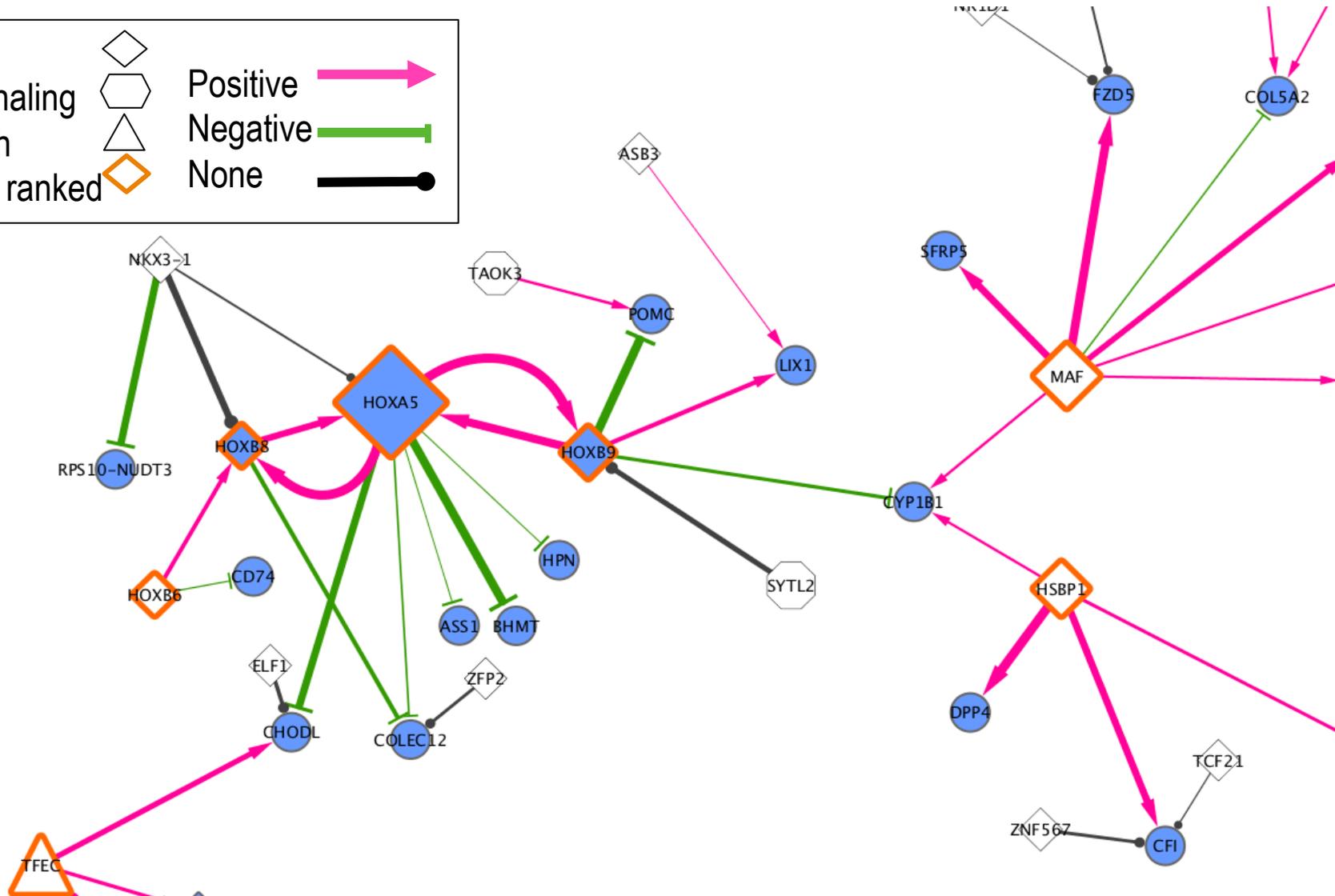
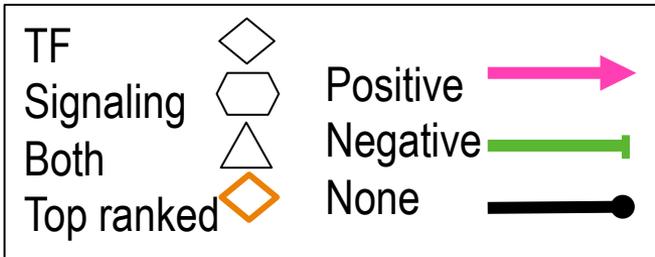


Network stats (FDR<0.01)

Network property	Value
Regulators	2,081
Targets	4,072
Edges	6,448



Network predicts regulatory connections among HOX genes



Analysis strategies

- Clustering/Dimensionality reduction approaches
- Network reconstruction approaches
- Network-based data integration
- Network-based interpretation
 - Gene prioritization methods
 - Gene set analysis

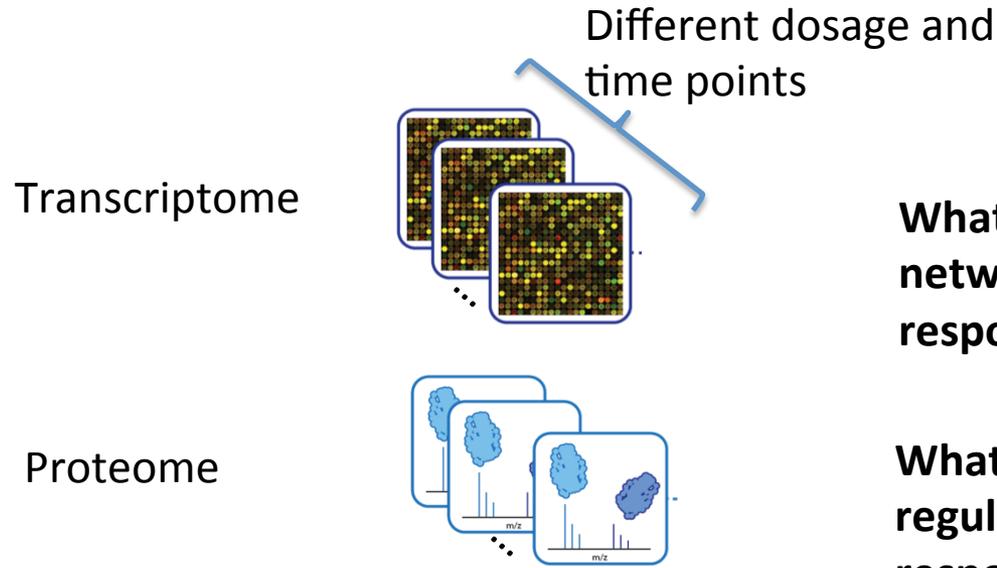
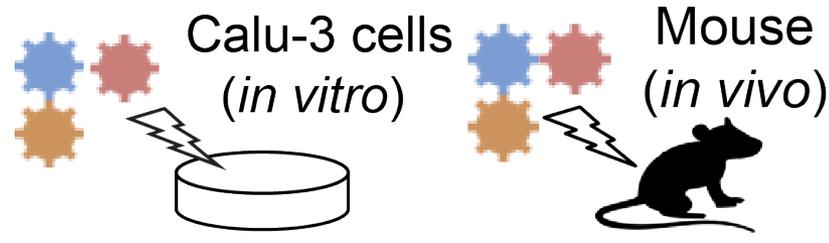
Systems biology studies to understand host response Influenza virus infections



Yoshi Kawaoka



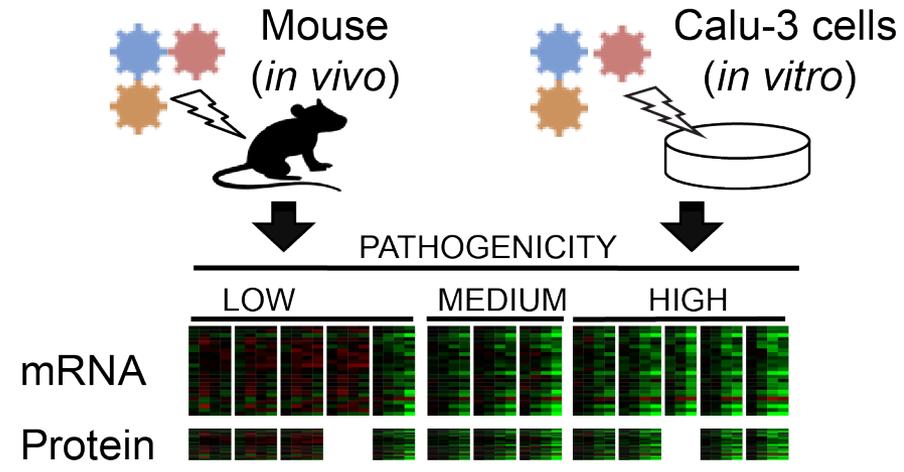
Amie Einfeld



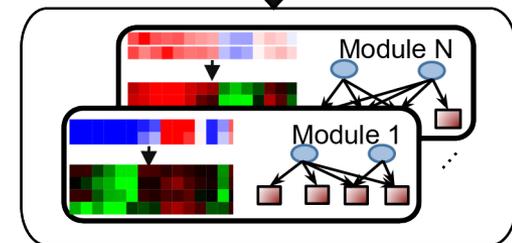
What are the regulatory networks driving host response?

What are key regulators of host response?

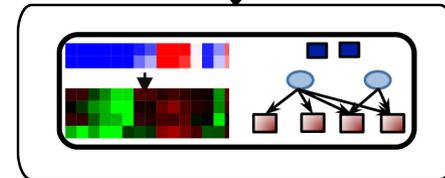
Integrative reconstruction of the host response regulatory network



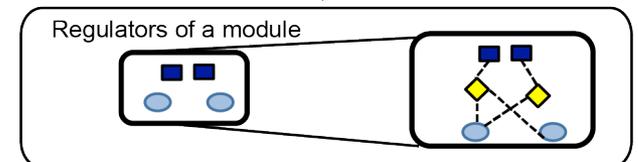
Step 1. Learn host response regulatory module networks from mRNA using MERLIN



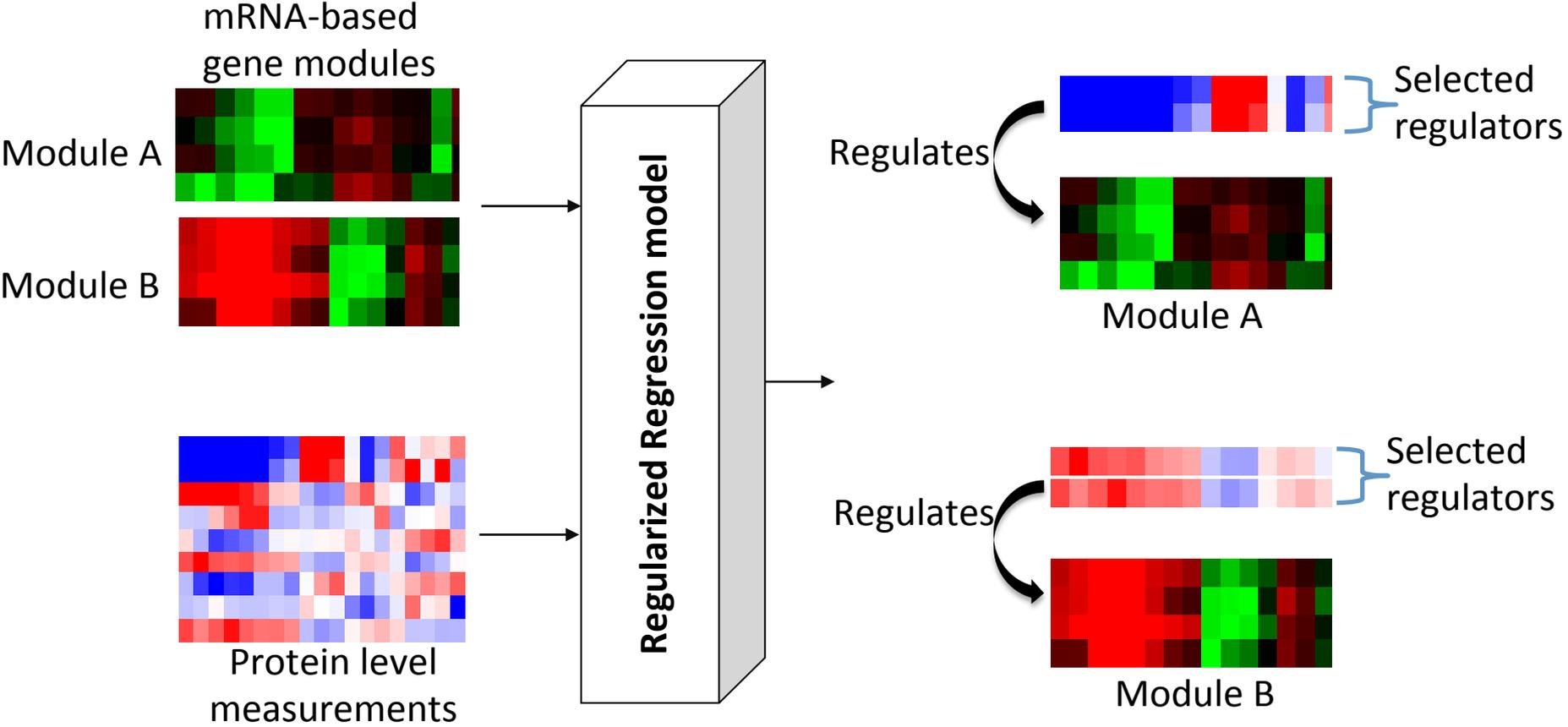
Step 2. Identify protein regulators for each module



Step 3. Predict physical regulatory subnetworks connecting mRNA- and protein-based regulators



Module-level regression models to integrate mRNA and protein levels



Human Module 1549

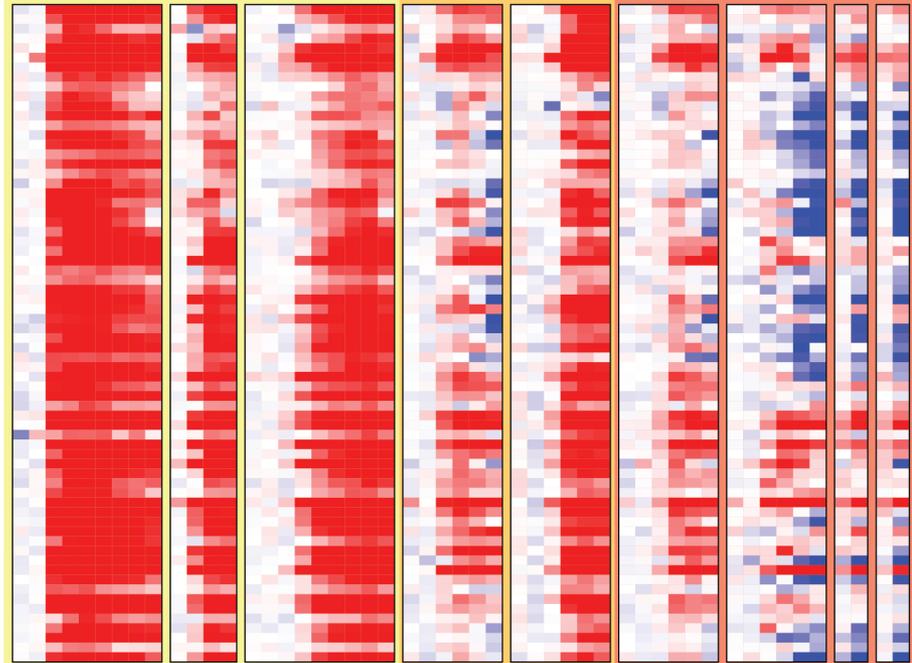
Calu-3 ISGs



Low

Medium

High



**Expression-based
Regulators
(MERLIN)**

★ FGFR3
IRF7
LMO2
MLKL
MOK
NMI
NUPR1
SP110
STAT1
TRIM21

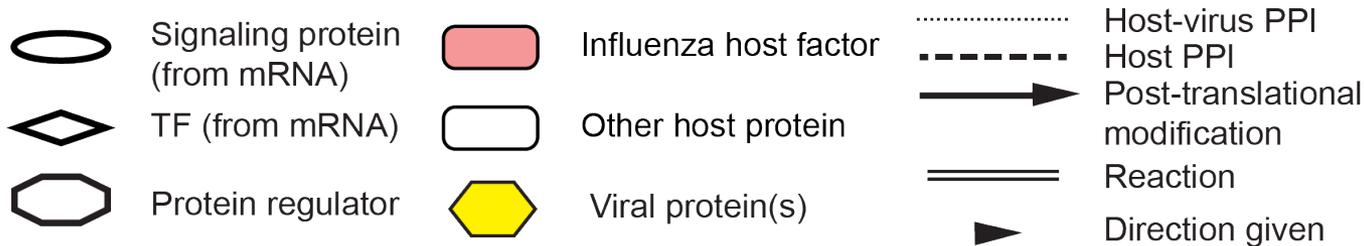
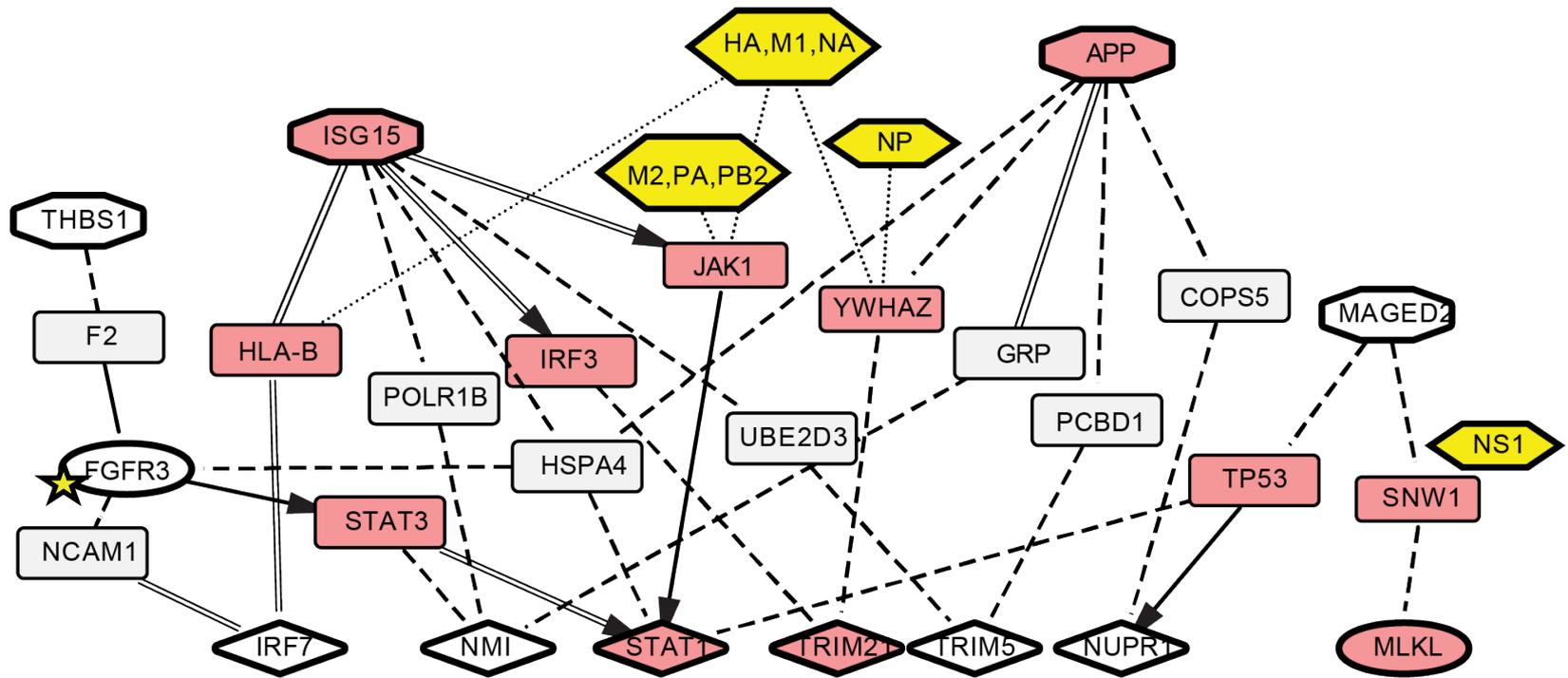
**Type I Interferon signaling,
production and response**

**Protein-based
Regulators
(Sparse regression)**

TRIM5
ISG15p
COLGALT1p
LAMB1p
MAGED2p
APPp
KLHL33p
THBS1p

**Interferon signaling, apoptosis,
membrane, extracellular matrix**

Physical regulatory program connects known immune response regulators to novel genes



Concluding thoughts

- Clustering/dimensionality reduction
 - General patterns of expression/activity
 - Useful for exploratory analysis and interpretation of large datasets
 - Co-clustering could be used to characterize unknown genes and samples
 - Example tools: **CMINT, PCA, Kmeans clustering**
- Network reconstruction
 - Infer the connectivity among the genes and regulators
 - Needed for prioritization of important nodes and interpretation
 - Example tools: **MERLIN, MERLIN-P, GENIE3, Inferelator**
- Network-based integration and interpretation
 - Integrating gene hits from different assays, e.g. differentially expressed genes with genes from screening assays
 - Gene prioritization
 - Example tools: **Subnetwork identification, OmicsIntegrator**

Summary

MERLIN

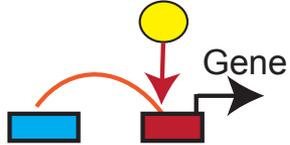
CMINT

RIPPLE

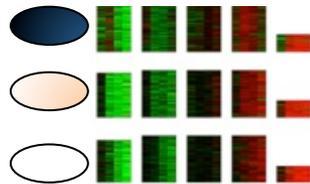
SIGNATURE
SUBGRAPHS

NETWORK INFO
FLOW

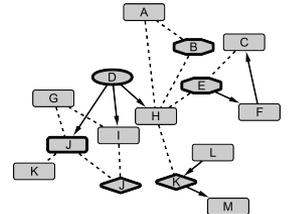
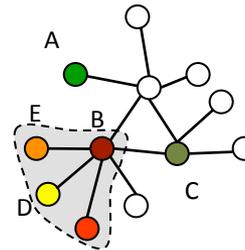
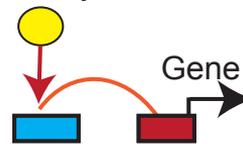
TF binds
proximally



Modules per cell type



TF binds
distally



Mapping regulatory connections

Interpreting regulatory connections



WISCONSIN
INSTITUTE FOR DISCOVERY
AT THE UNIVERSITY OF WISCONSIN-MADISON

HT CENTER FOR
HIGH THROUGHPUT
COMPUTING



NIH U19AI106772 (Kawaoka), US EPA 83573701 (Murphy)

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON