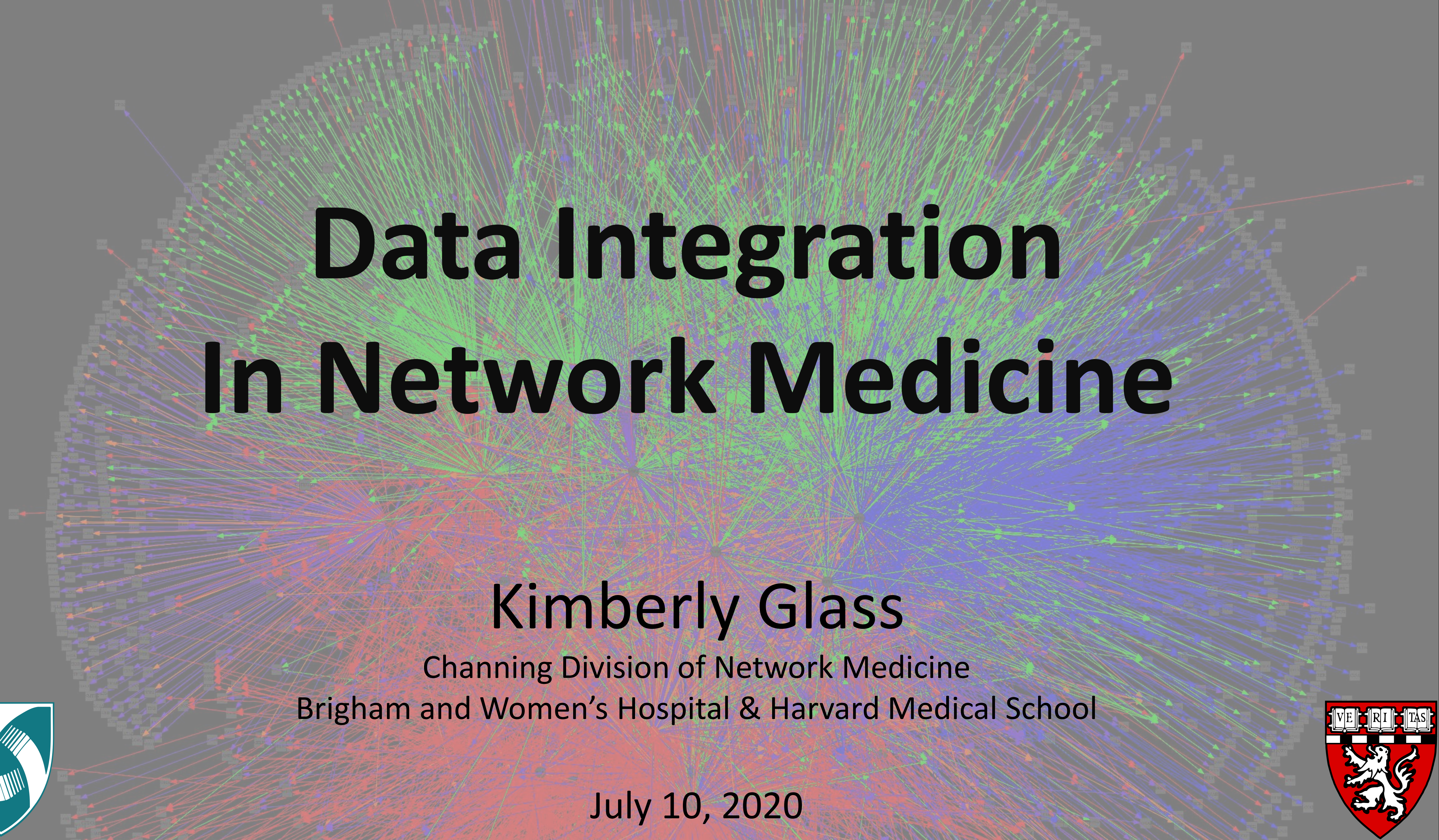


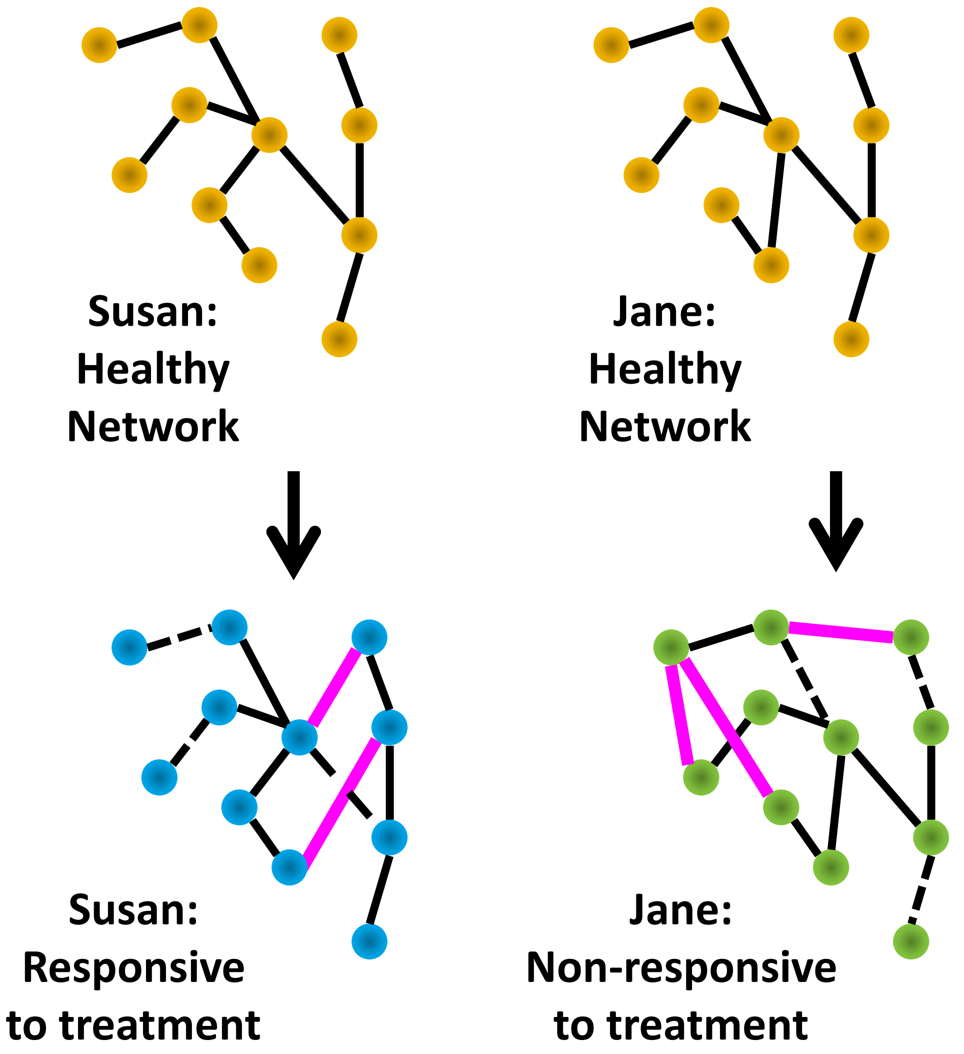
Data Integration In Network Medicine

Kimberly Glass
Channing Division of Network Medicine
Brigham and Women’s Hospital & Harvard Medical School

July 10, 2020

What can networks teach us about medicine?

Susan Jane



Susan: Healthy Network

Jane: Healthy Network

Susan: Responsive to treatment

Jane: Non-responsive to treatment

What Is Network Medicine?

The study of cellular, disease, and social networks which aims to quantify the complex interlinked factors contributing to individual diseases. (Adapted from Barabasi, NEJM 2007; 357:404)

Key components of Network Medicine:

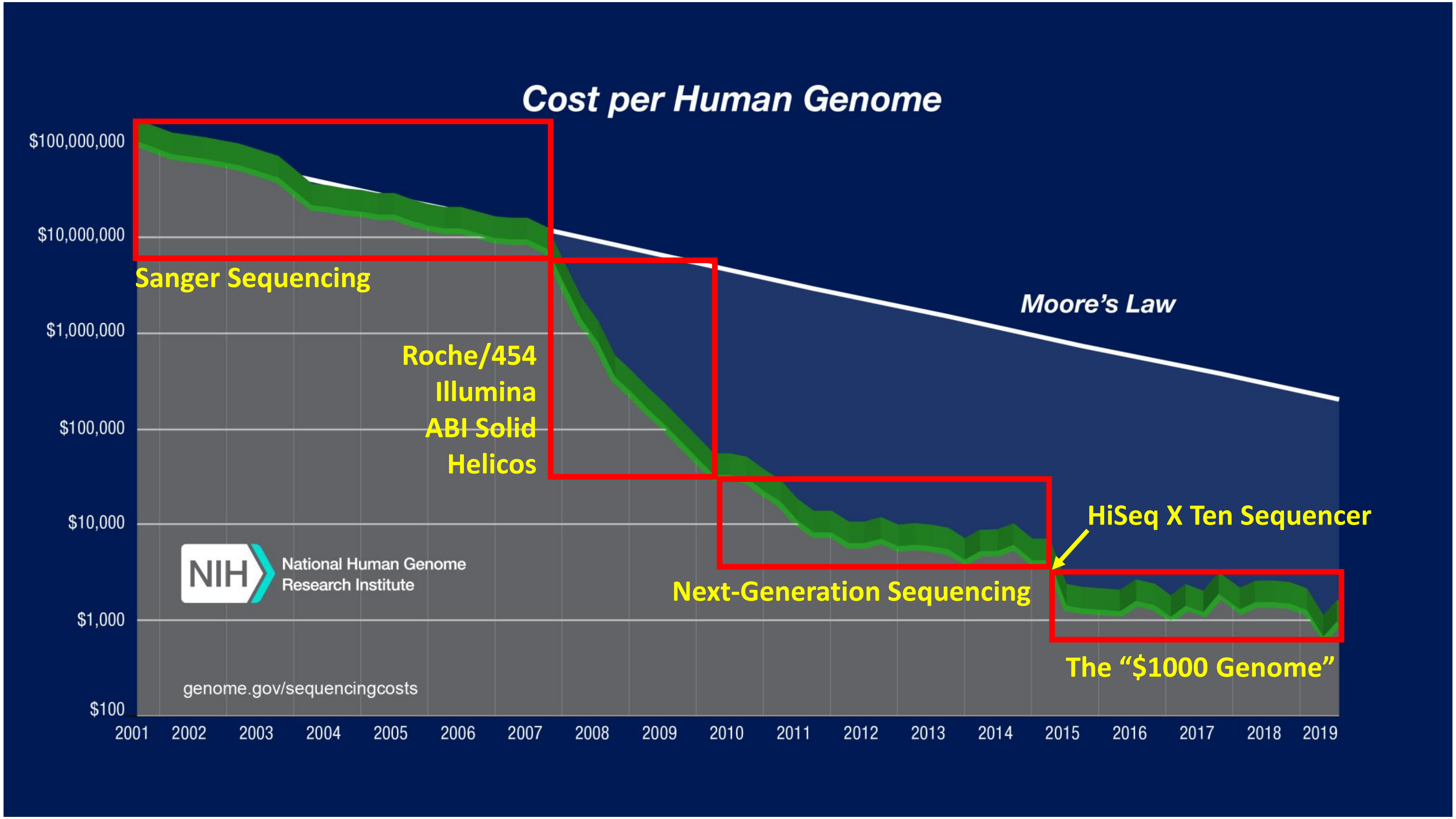
--Holistic rather than reductionist approach

--Emergent properties from entire network

--Employs systems biology methods

--Construction of molecular networks

Falling Costs of Sequencing Data



https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

The Human Genetic Code

ACGTTTAACAGTAACGAAT

ACGTTTAACAGTACCGAAT

ACGTTTAACAGTAACGAAT

ACGTATAACAGTAACGAAT

ACGTTTAACAGTAACGAAT

all humans share
~99.9% of their DNA

The Human Genetic Code

ACGTTTAACAGTAACGAAT

ACGTTTAACAGTAACGAAT

ACGTTTAACAGTAACGAAT

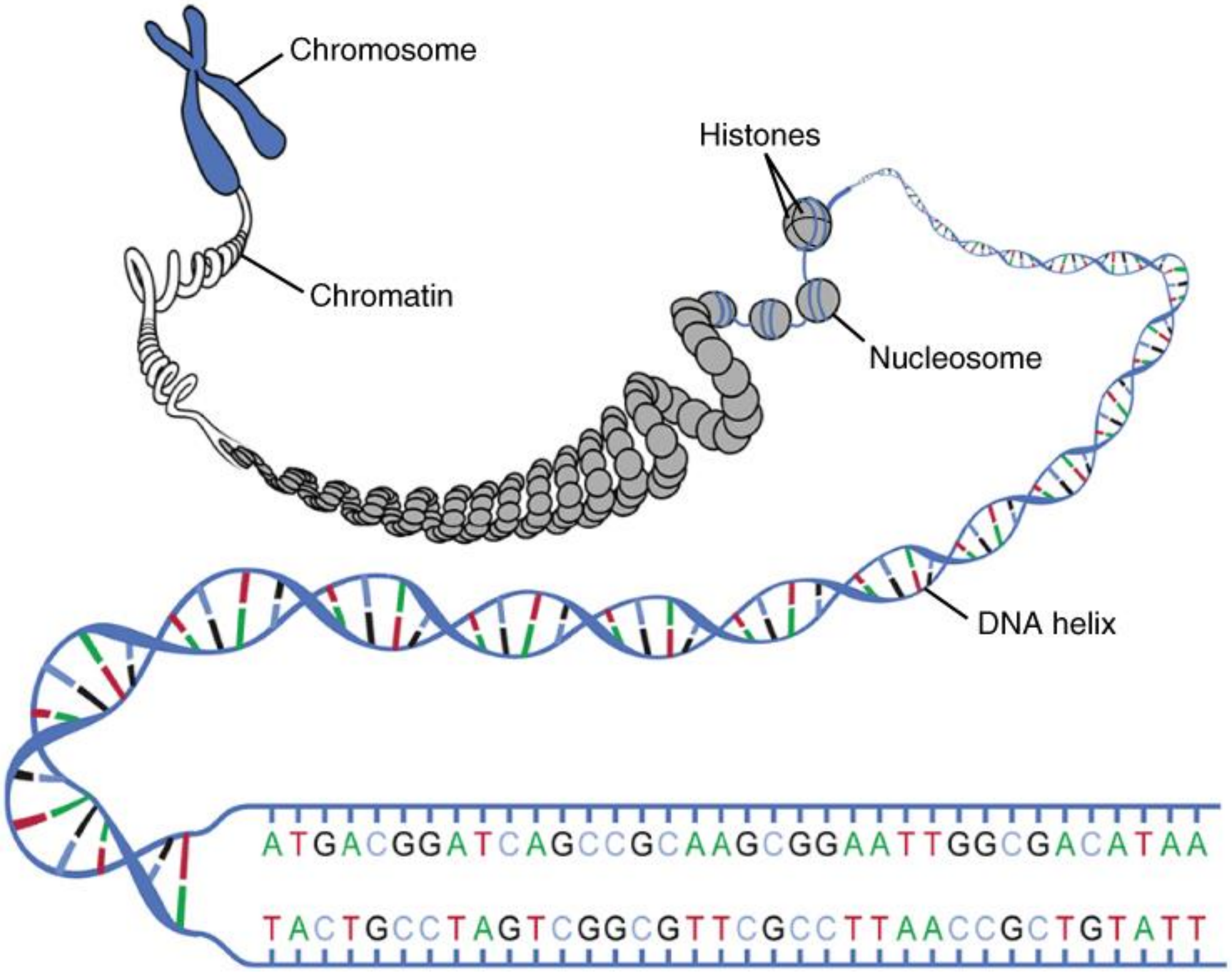
ACGTTTAACAGTAACGAAT

ACGTTTAACAGTAACGAAT

all cells in your body
have identical DNA!

Where do the differences come from?

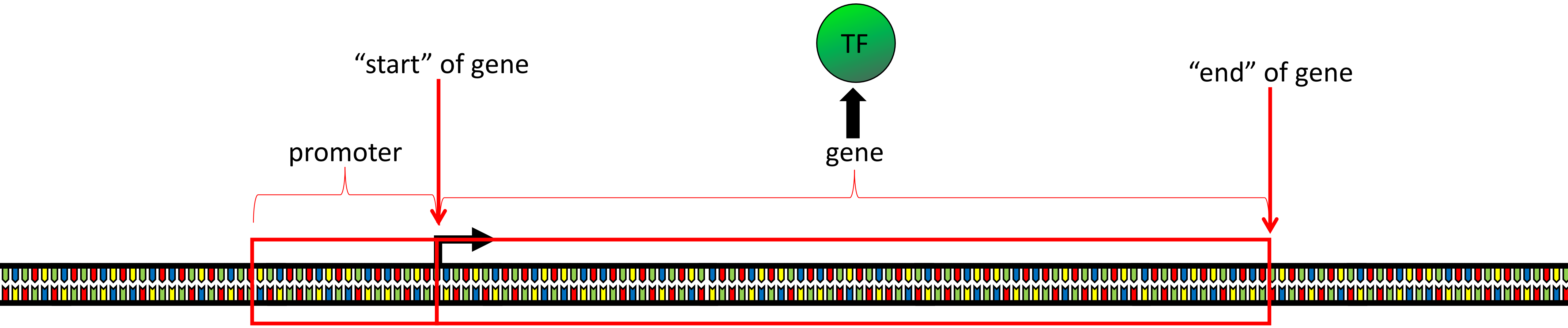
The Physical Structure of DNA



https://cnx.org/contents/9TxHOD3O@4/The-Nucleus-and-DNA-Replicatio

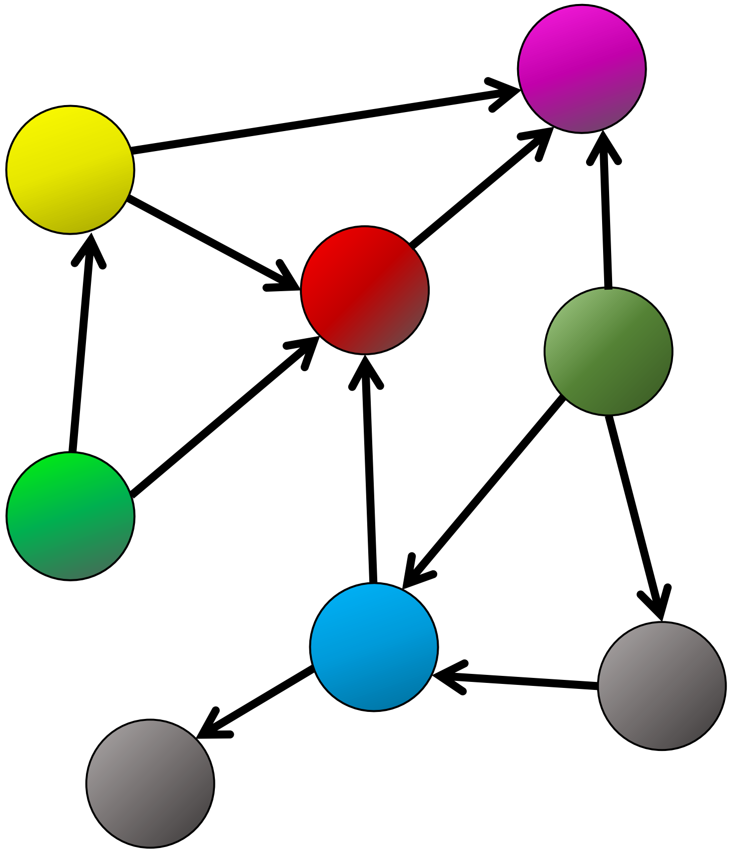
Biological Function is Encoded in Genes

- Human DNA contains approximately 25000 (protein-coding) genes.
- Some genes contain the information for how to make a particular type of protein, called a transcription factor.
- Right before a gene is a special region of DNA, called a promoter, that enables transcription factors to control (regulate) the behavior of a gene.



Why Study Gene Regulation?

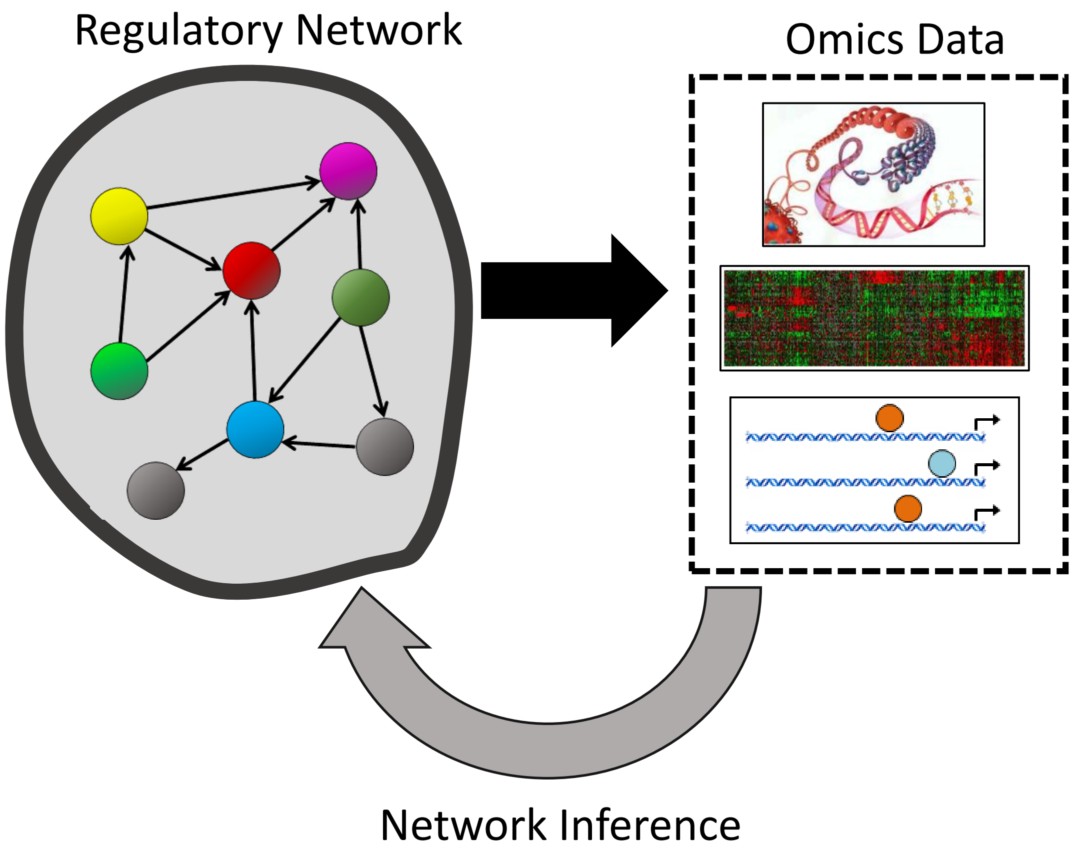
- Biological processes are driven by multiple connections linking together genes, proteins, and other molecules.
- One way to model these processes is through gene regulatory networks.
- By comparing regulatory networks we can gain insights into alterations in biological processes that may underlie differences in various diseases and/or phenotypic states.



Gene Regulatory Network

- *Nodes* are genes.
- *Edges* are relationships between genes.
- Edges are directed to indicate causality.

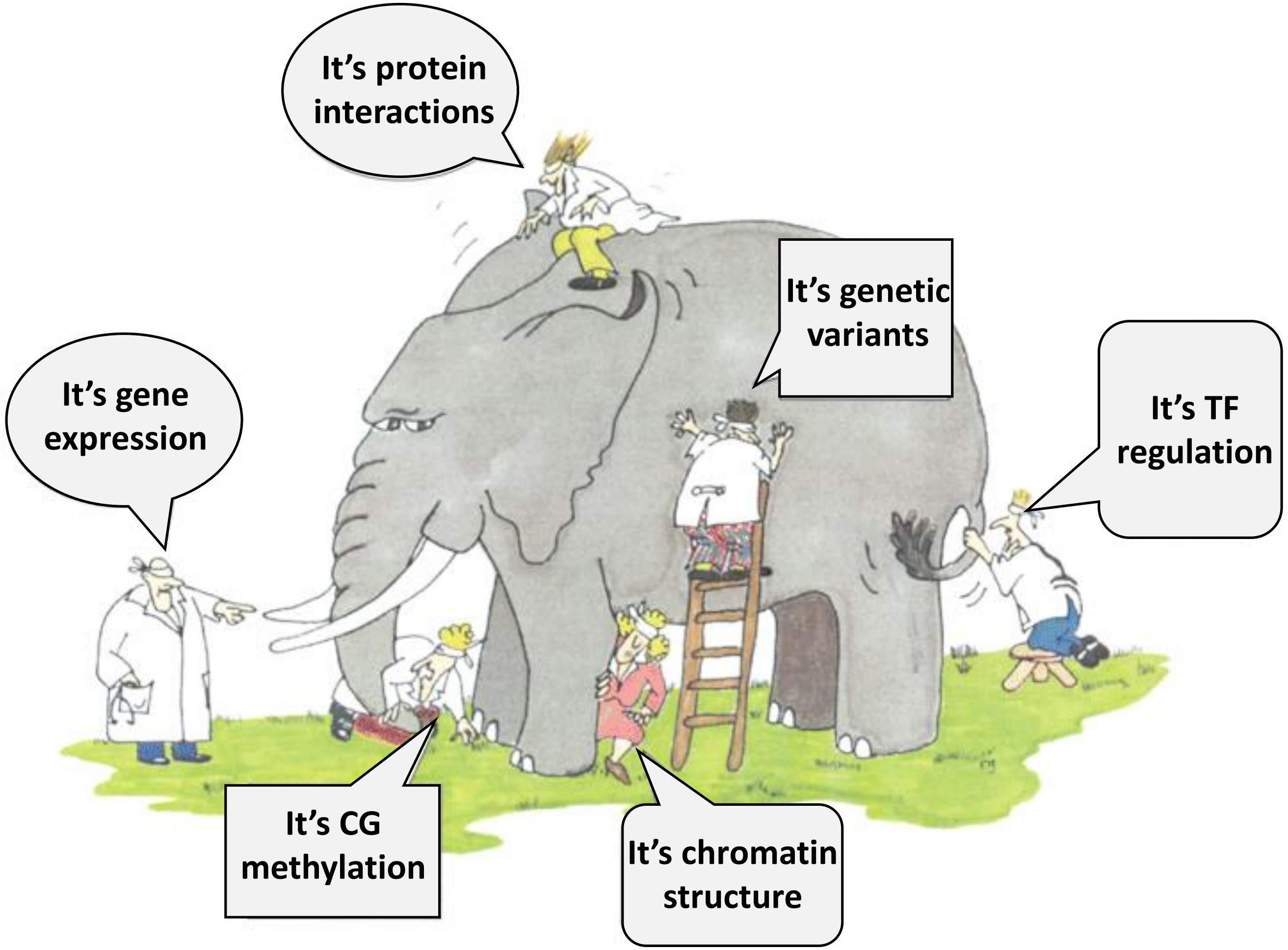
Modeling Gene Regulatory Networks



Regulatory Network

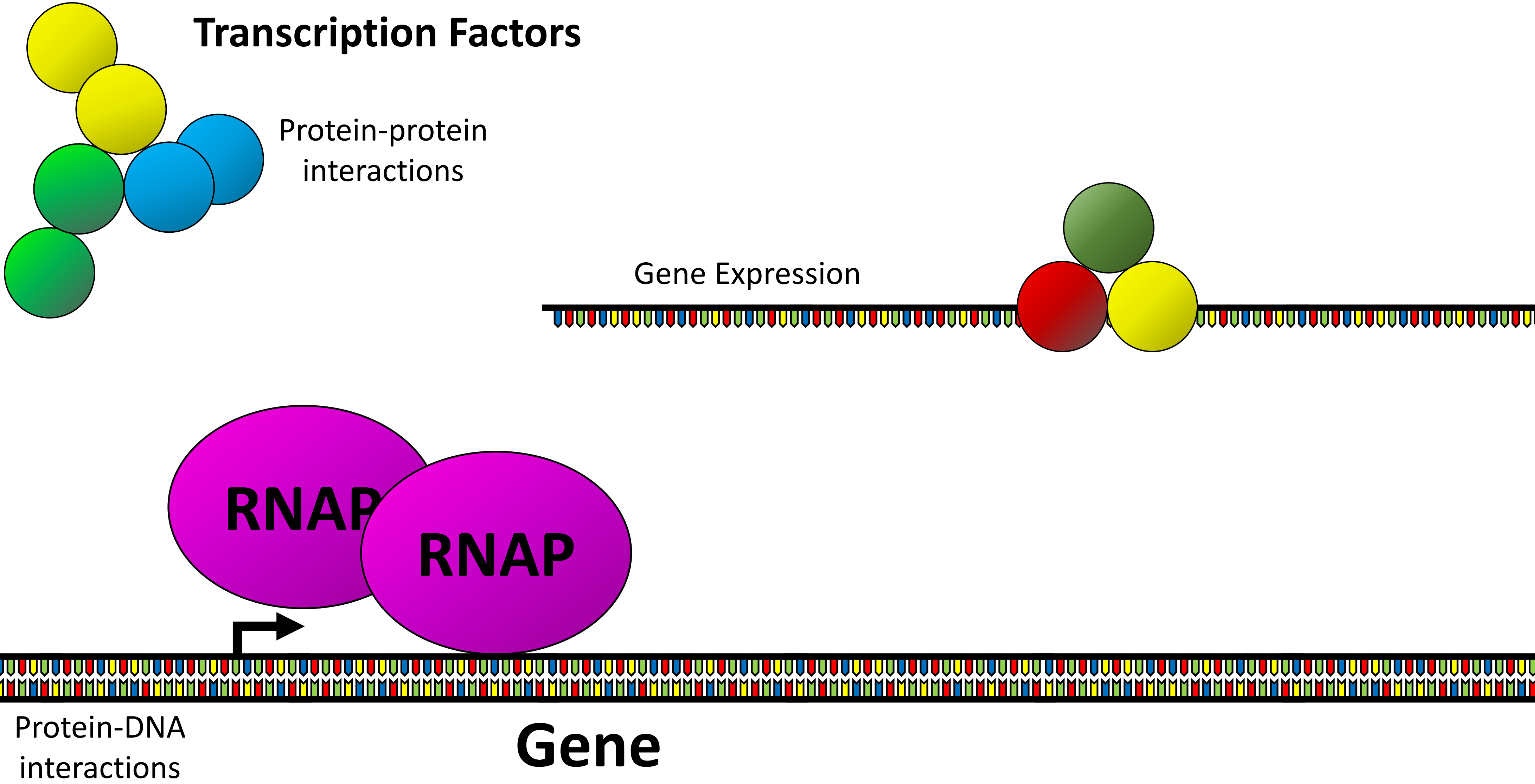
Omics Data

Network Inference



http://expertcytometry.com/the-story-of-the-blind-men-and-the-elephant-or-why-its-better-to-see-your-cells/

Basic Mechanics of Gene Regulation



Transcription Factors

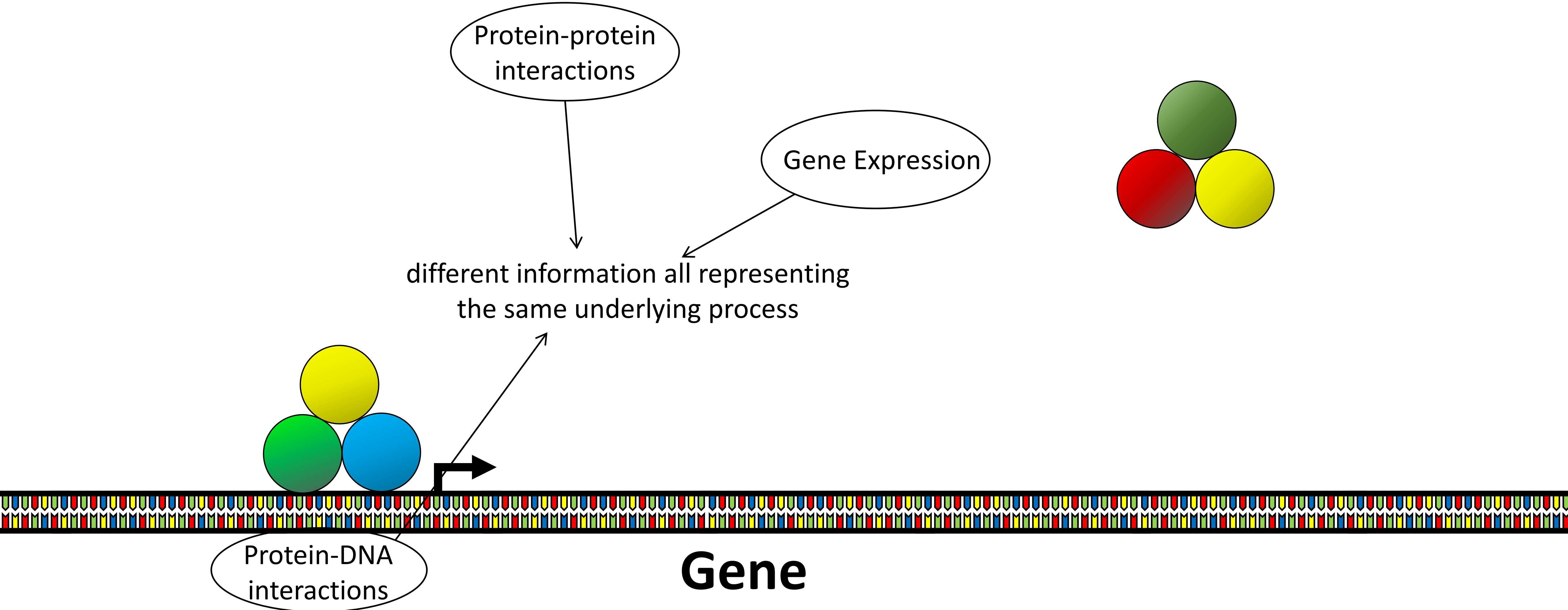
Protein-protein interactions

Gene Expression

Protein-DNA interactions

Gene

Basic Mechanics of Gene Regulation



Protein-protein interactions

Gene Expression

different information all representing the same underlying process

Protein-DNA interactions

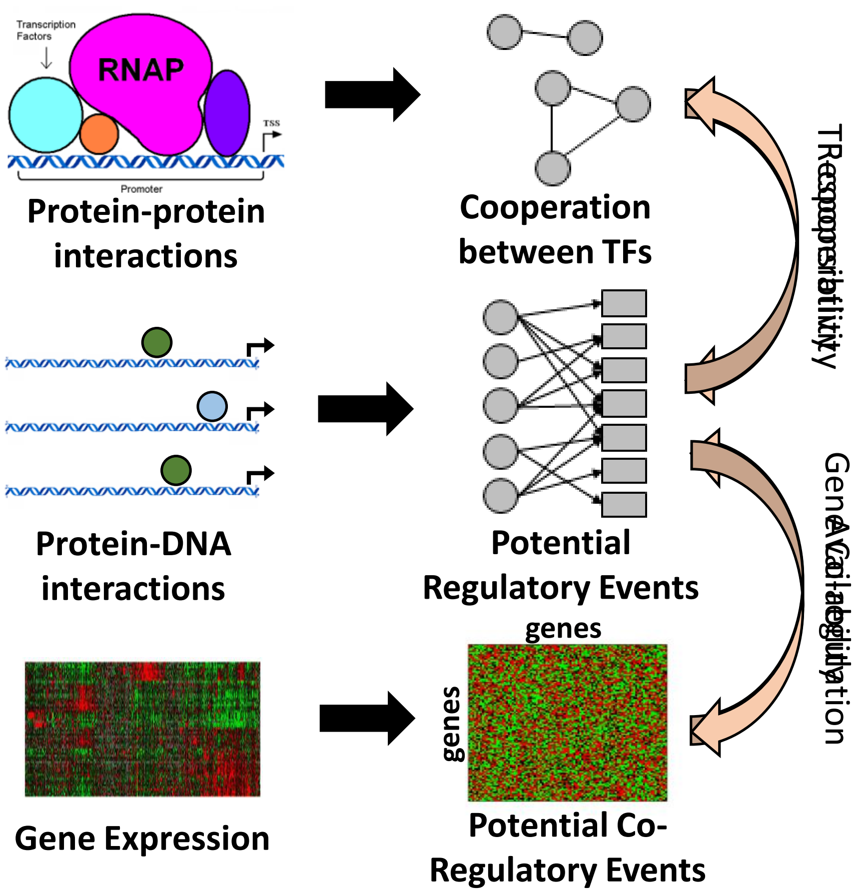
Gene

Message-Passing Networks: PANDA

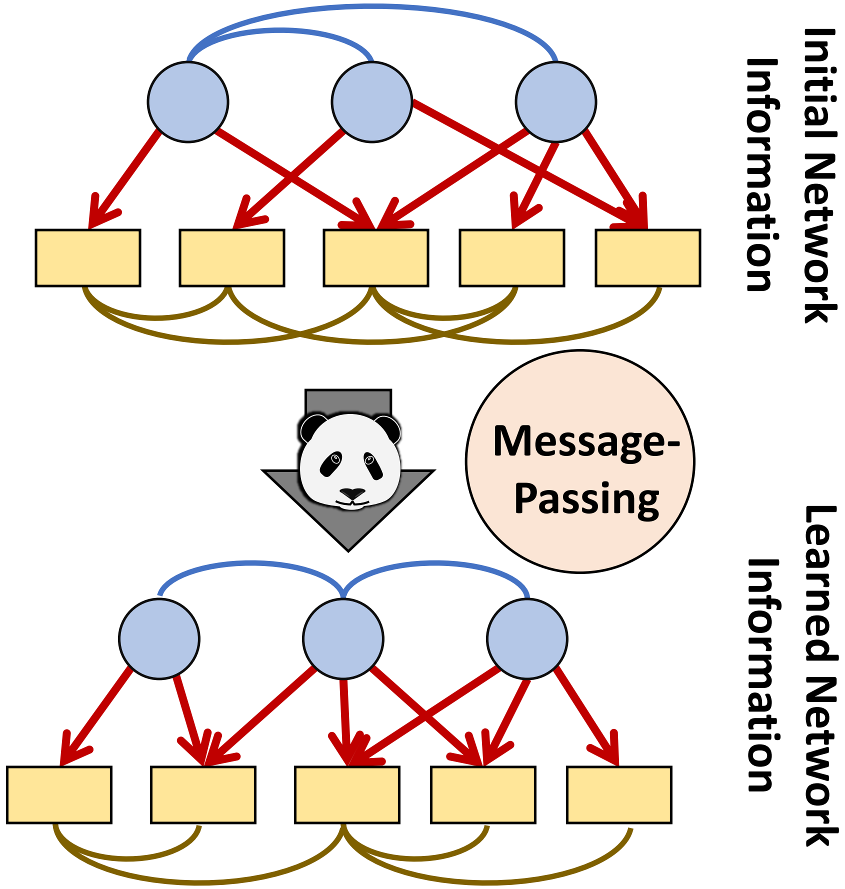
(Passing Attributes between Networks for Data Assimilation)

Genomic Data

Network Representation

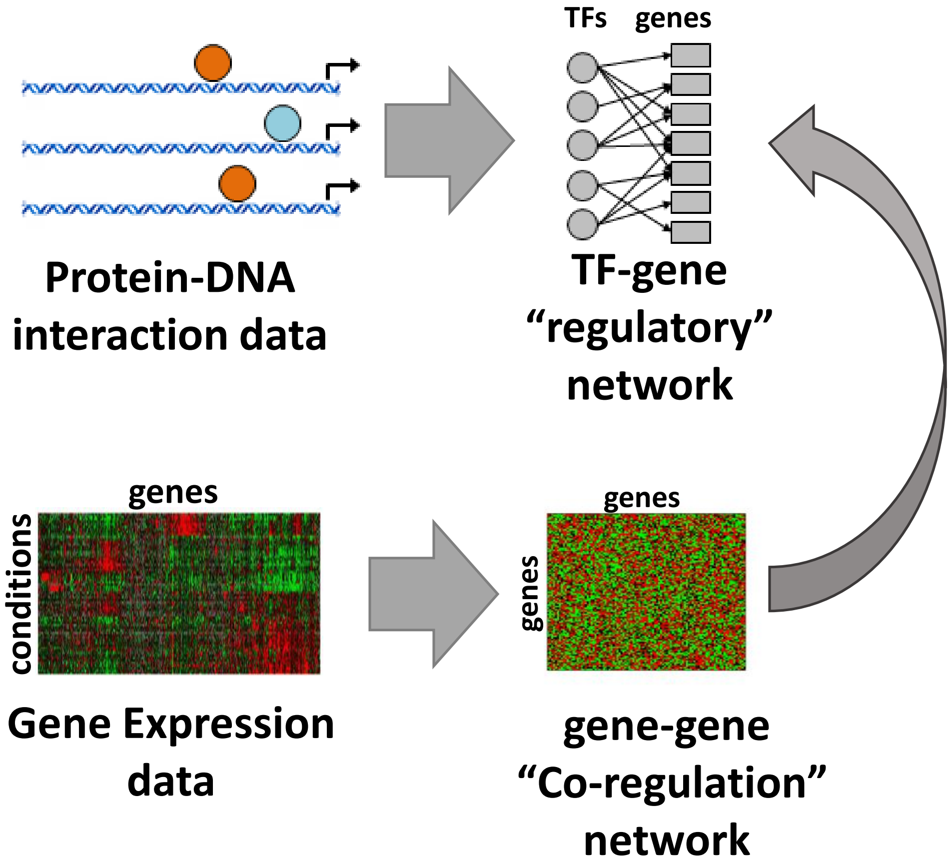


Use message-passing to find a consensus among the data



Glass et. al. "Passing Messages Between Biological Networks to Refine Predicted Interactions." PLoS One. 2013 May 31;8(5):e64832.

Estimating Availability by comparing gene expression data and regulatory information



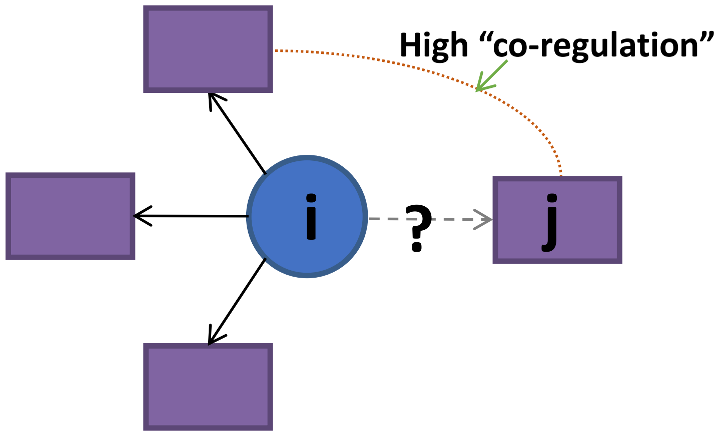
Protein-DNA interaction data

TF-gene "regulatory" network

Gene Expression data

gene-gene "Co-regulation" network

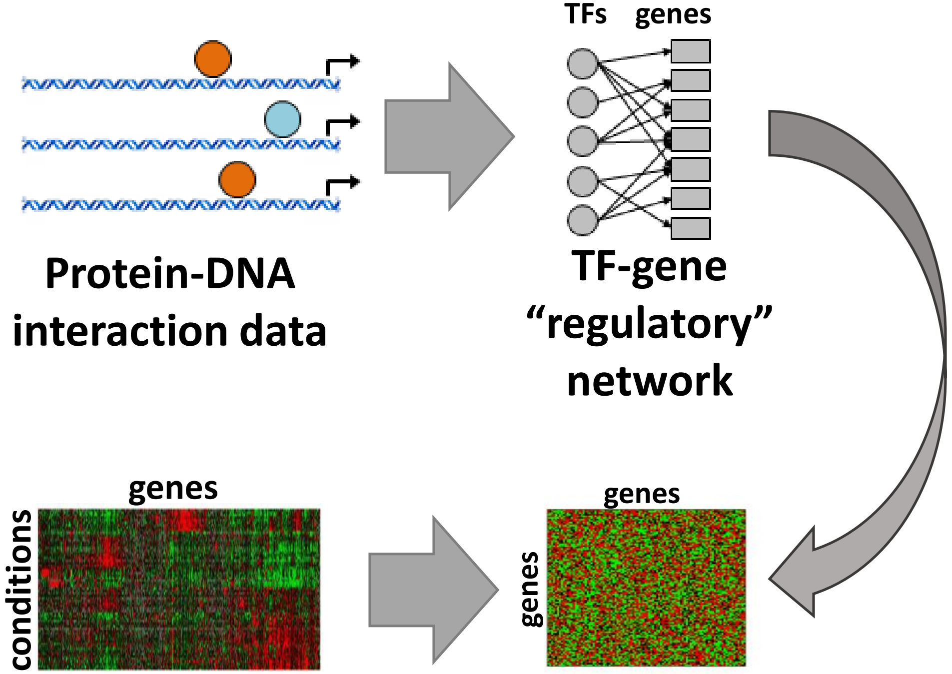
Passing message from the co-regulation to the regulatory network



Gene *j* may be regulated by TF *i* if it is co-regulated with the other targets of TF *i*.

Integrating expression data with regulation data lets us estimate the **availability**.

Estimating Availability by comparing gene expression data and regulatory information



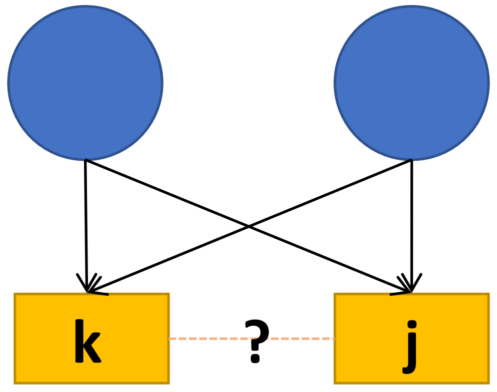
Protein-DNA interaction data

TF-gene “regulatory” network

Gene Expression data

gene-gene “Co-regulation” network

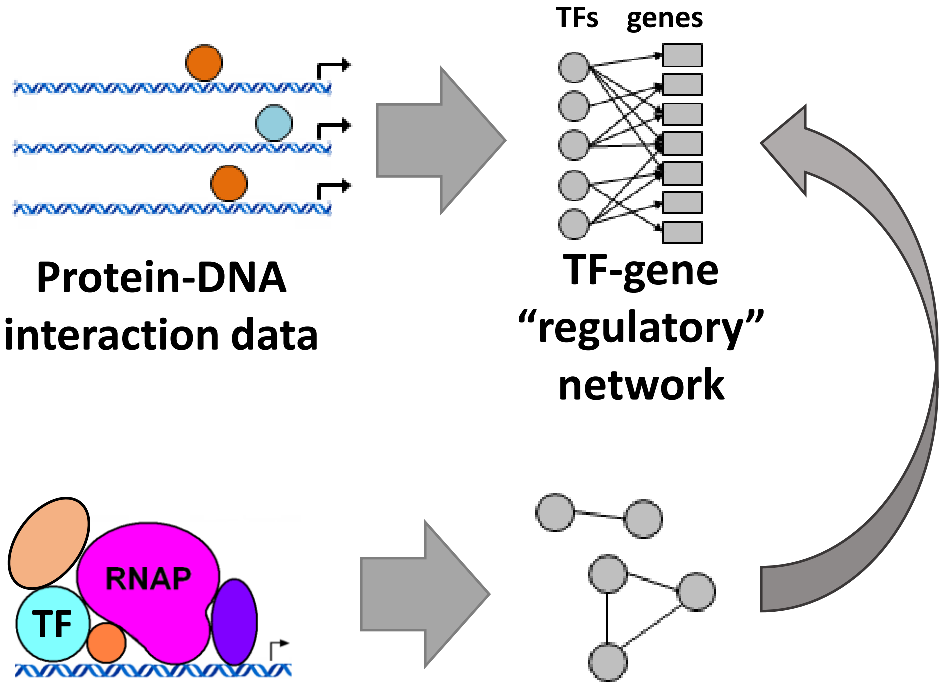
Passing message from the regulatory to the co-regulation network



Genes *j* and k may be co-regulated if they are both targeted by the same TFs.

Integrating expression data with regulation data lets us estimate the **availability**.

Estimating Responsibility by comparing PPI data and regulatory information



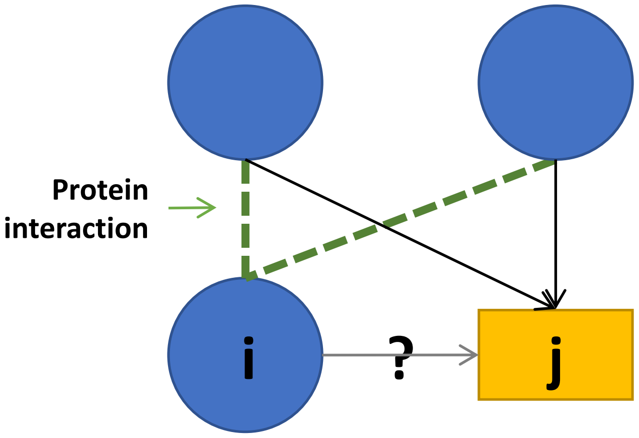
Protein-DNA interaction data

TF-gene "regulatory" network

Protein-Protein interaction data

Protein-protein interaction network

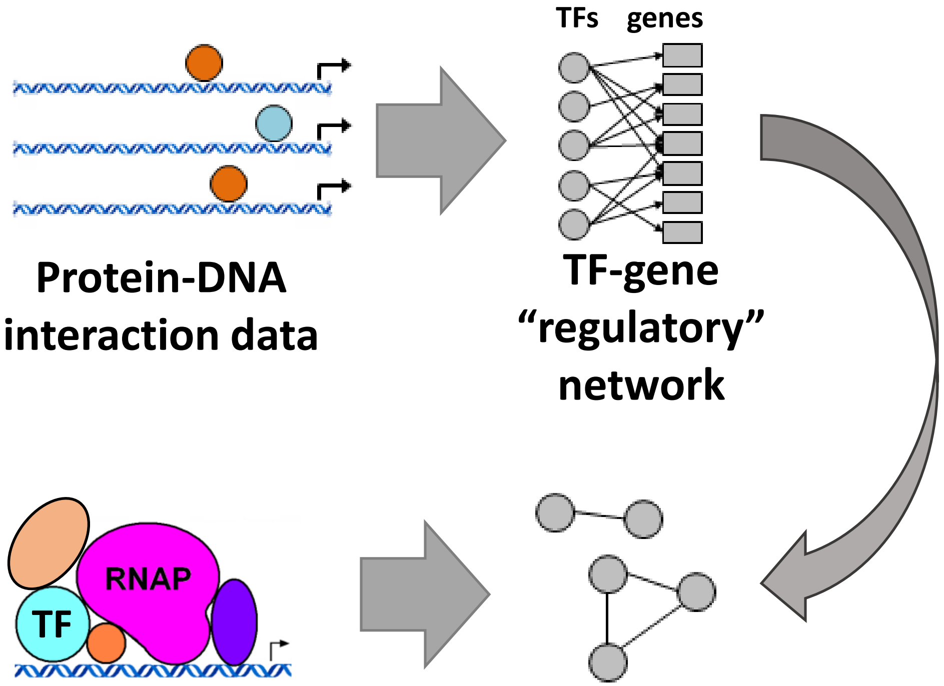
Passing message from the PPI to the regulatory network



TF *i* may regulate gene *j* if TF *i* interacts with many of the TFs targeting gene *j*.

Integrating PPI data with regulation data lets us estimate the **responsibility**.

Estimating Responsibility by comparing PPI data and regulatory information



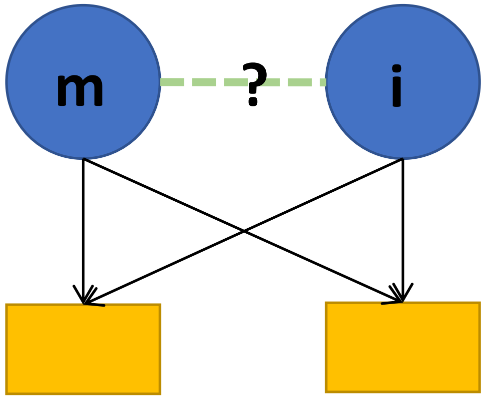
Protein-DNA interaction data

TF-gene “regulatory” network

Protein-Protein interaction data

Protein-protein interaction network

Passing message from the regulatory to the PPI network

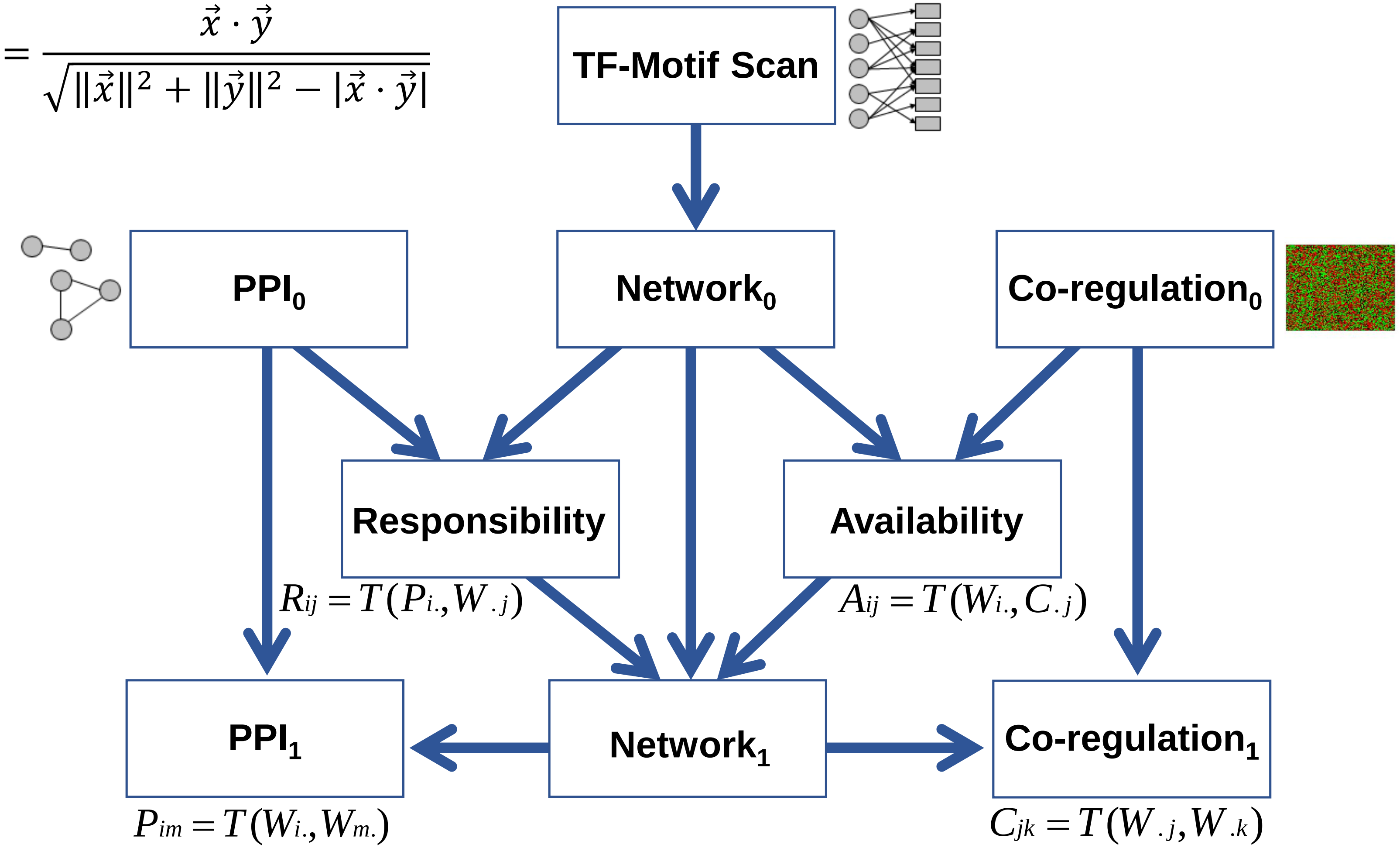


TFs *i* and *m* may interact if they both target many of the same genes

Integrating PPI data with regulation data lets us estimate the **responsibility**.

Passing Messages Between Networks

$$T(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{\sqrt{\|\vec{x}\|^2 + \|\vec{y}\|^2 - |\vec{x}\cdot\vec{y}|}}$$

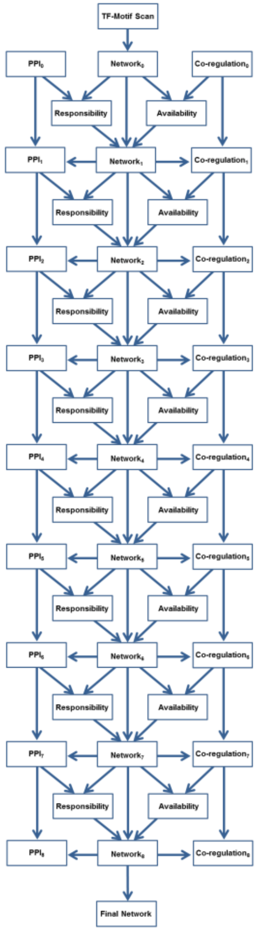


$$R_{ij} = T(P_{i\cdot}, W_{\cdot j})$$

$$A_{ij} = T(W_{i\cdot}, C_{\cdot j})$$

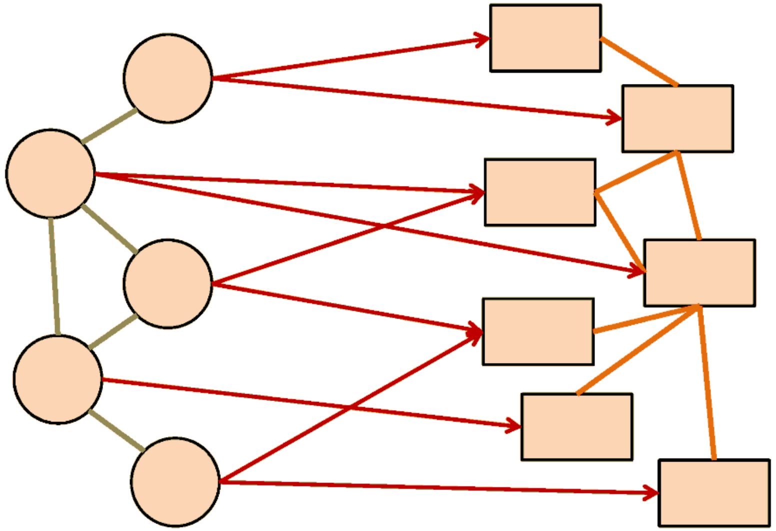
$$P_{im} = T(W_{i\cdot}, W_{m\cdot})$$

$$C_{jk} = T(W_{\cdot j}, W_{\cdot k})$$



Glass et. al. “Passing Messages Between Biological Networks to Refine Predicted Interactions.” *PLoS One.* 2013 May 31;8(5):e64832.
Code and related material: https://sites.google.com/a/channing.harvard.edu/kimberlyglass/tools/panda

Conceptual Framework of PANDA



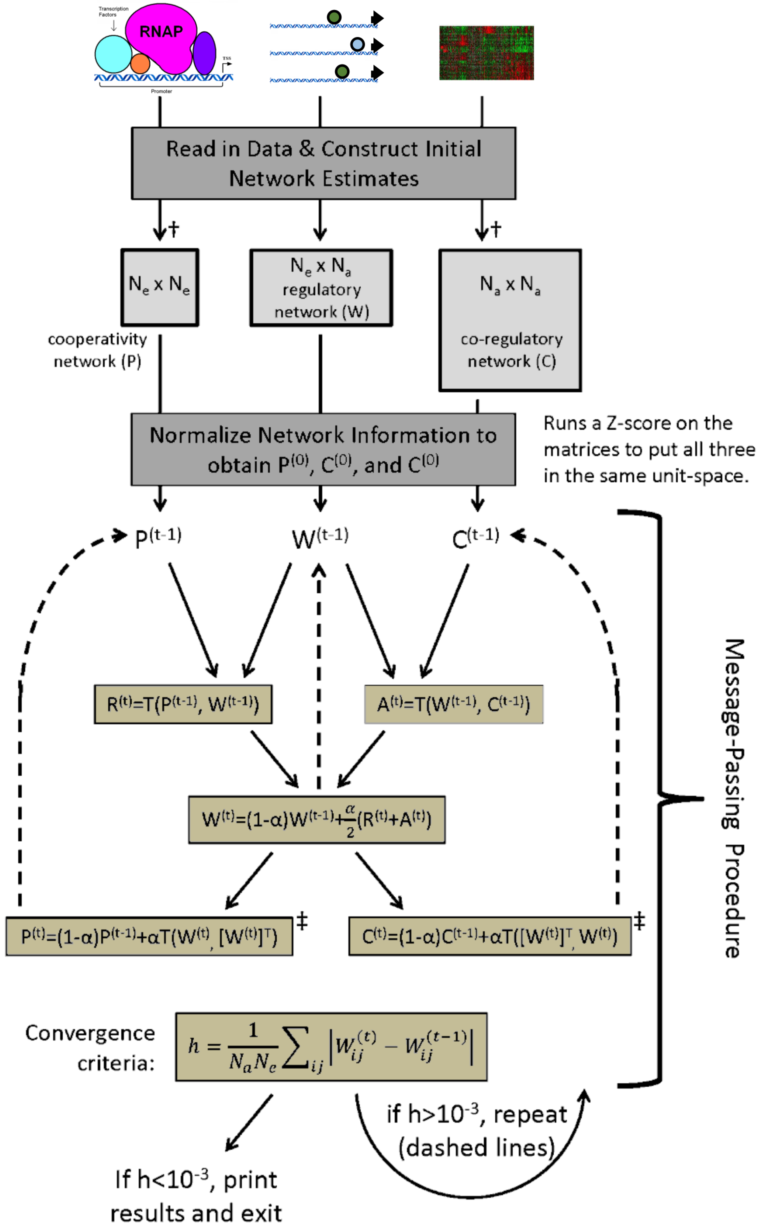
PANDA considers:

Two types of nodes

Three types of edges

In practice we have used PANDA to construct regulatory networks.

In principle it can be used to integrate any network data that follows this framework.



Read in Data & Construct Initial Network Estimates

†

†

N_e x N_e

cooperativity network (P)

N_e x N_a regulatory network (W)

N_a x N_a

co-regulatory network (C)

Normalize Network Information to obtain $P^{(0)}$, $C^{(0)}$, and $C^{(0)}$

Runs a Z-score on the matrices to put all three in the same unit-space.

$P^{(t-1)}$

$W^{(t-1)}$

$C^{(t-1)}$

$R^{(t)}=T(P^{(t-1)}, W^{(t-1)})$

$A^{(t)}=T(W^{(t-1)}, C^{(t-1)})$

$W^{(t)}=(1-\alpha)W^{(t-1)}+\frac{\alpha}{2}(R^{(t)}+A^{(t)})$

$P^{(t)}=(1-\alpha)P^{(t-1)}+\alpha T(W^{(t)}, [W^{(t)}]^T)$ ‡

$C^{(t)}=(1-\alpha)C^{(t-1)}+\alpha T([W^{(t)}]^T, W^{(t)})$ ‡

Message-Passing Procedure

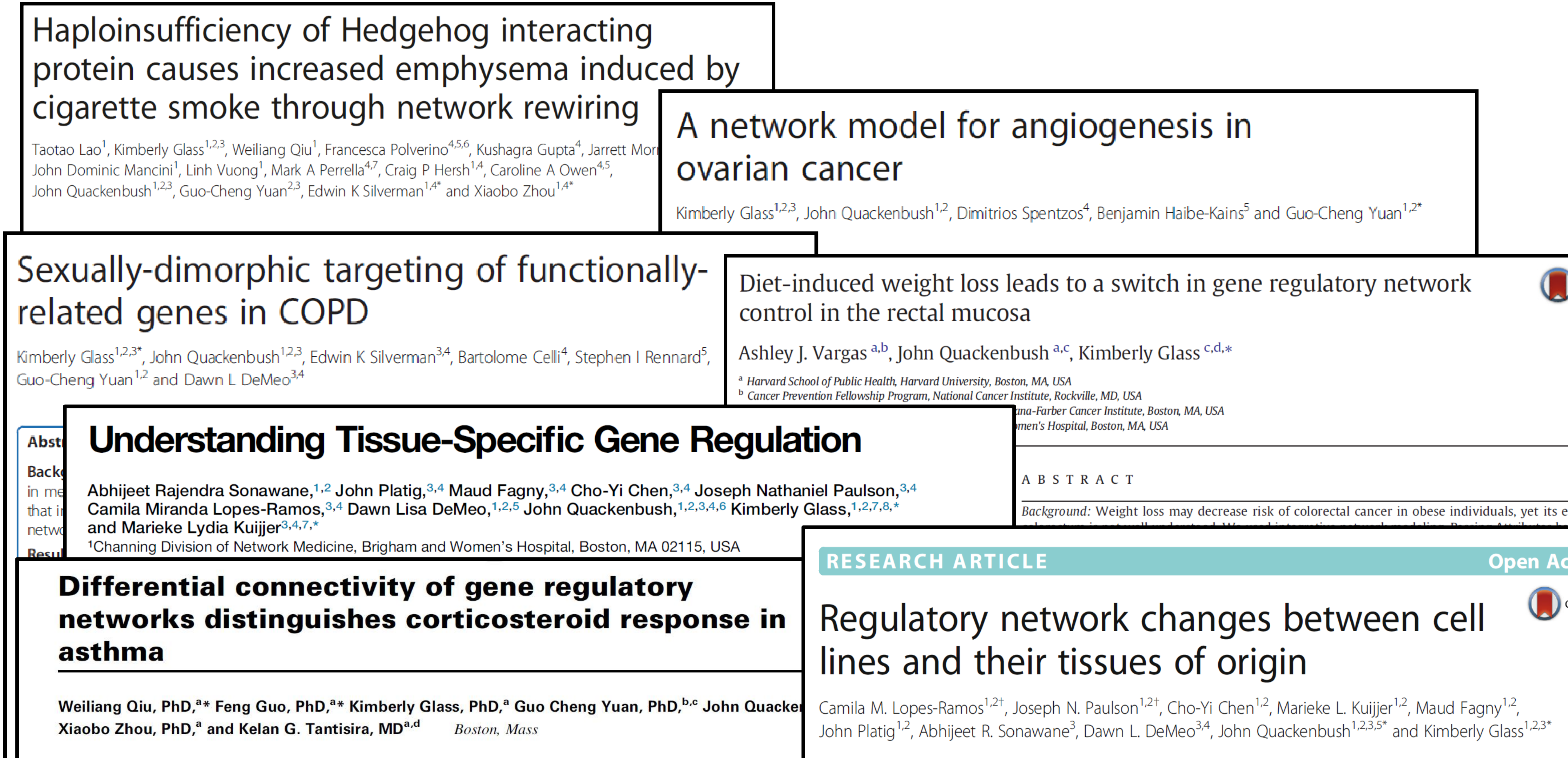
Convergence criteria:

$$h = \frac{1}{N_a N_e} \sum_{ij} \left| W_{ij}^{(t)} - W_{ij}^{(t-1)} \right|$$

If $h<10^{-3}$, print results and exit

if $h>10^{-3}$, repeat (dashed lines)

PANDA Applications



Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring

Taotao Lao[1], Kimberly Glass[1,2,3], Weiliang Qiu[1], Francesca Polverino[4,5,6], Kushagra Gupta[4], Jarrett Mor
John Dominic Mancini[1], Linh Vuong[1], Mark A Perrella[4,7], Craig P Hersh[1,4], Caroline A Owen[4,5],
John Quackenbush[1,2,3], Guo-Cheng Yuan[2,3], Edwin K Silverman[1,4*] and Xiaobo Zhou[1,4*]

A network model for angiogenesis in ovarian cancer

Kimberly Glass[1,2,3], John Quackenbush[1,2], Dimitrios Spentzos[4], Benjamin Haibe-Kains[5] and Guo-Cheng Yuan[1,2*]

Sexually-dimorphic targeting of functionally-related genes in COPD

Kimberly Glass[1,2,3*], John Quackenbush[1,2,3], Edwin K Silverman[3,4], Bartolome Celli[4], Stephen I Rennard[5],
Guo-Cheng Yuan[1,2] and Dawn L DeMeo[3,4]

Abst

Back
in me
that i
netw

Resu

Diet-induced weight loss leads to a switch in gene regulatory network control in the rectal mucosa

Ashley J. Vargas[a,b], John Quackenbush[a,c], Kimberly Glass[c,d,*]

[a] *Harvard School of Public Health, Harvard University, Boston, MA, USA*
[b] *Cancer Prevention Fellowship Program, National Cancer Institute, Rockville, MD, USA*
ana-Farber Cancer Institute, Boston, MA, USA
men's Hospital, Boston, MA, USA

A B S T R A C T

Background: Weight loss may decrease risk of colorectal cancer in obese individuals, yet its e

---

Understanding Tissue-Specific Gene Regulation

Abhijeet Rajendra Sonawane,[1,2] John Platig,[3,4] Maud Fagny,[3,4] Cho-Yi Chen,[3,4] Joseph Nathaniel Paulson,[3,4]
Camila Miranda Lopes-Ramos,[3,4] Dawn Lisa DeMeo,[1,2,5] John Quackenbush,[1,2,3,4,6] Kimberly Glass,[1,2,7,8,*]
and Marieke Lydia Kuijjer[3,4,7,*]
[1]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma

Weiliang Qiu, PhD,[a]* Feng Guo, PhD,[a]* Kimberly Glass, PhD,[a] Guo Cheng Yuan, PhD,[b,c] John Quacker
Xiaobo Zhou, PhD,[a] and Kelan G. Tantisira, MD[a,d] *Boston, Mass*

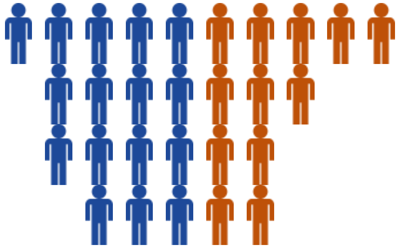
RESEARCH ARTICLE Open Ac

Regulatory network changes between cell lines and their tissues of origin

Camila M. Lopes-Ramos[1,2†], Joseph N. Paulson[1,2†], Cho-Yi Chen[1,2], Marieke L. Kuijjer[1,2], Maud Fagny[1,2],
John Platig[1,2], Abhijeet R. Sonawane[3], Dawn L. DeMeo[3,4], John Quackenbush[1,2,3,5*] and Kimberly Glass[1,2,3*]

Reconstructing Tissue-Specific Networks

GTEx



Multi-tissue transcriptomes
(53 sites, 9590 samples from 550 individuals)

Preprocessing

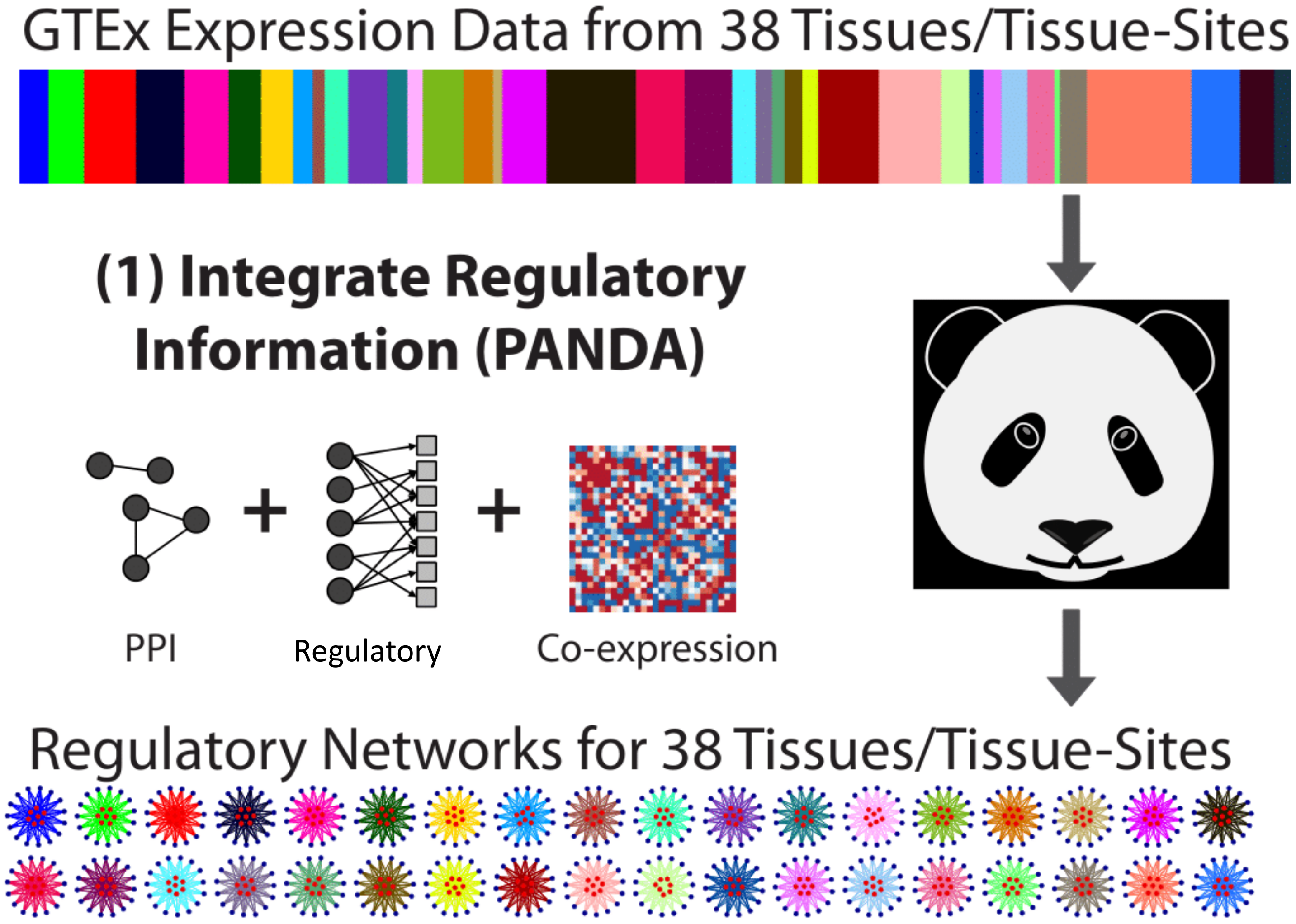
Data QC, merging & normalization
(38 primary sites, 9435 samples from 549 individuals)

Network Analysis

- Identify tissue-specific edges
- Characterize tissue-specific regulation of biological processes

Expression Analysis

- Identify tissue-specific genes and transcription factors
- Investigate their regulatory context



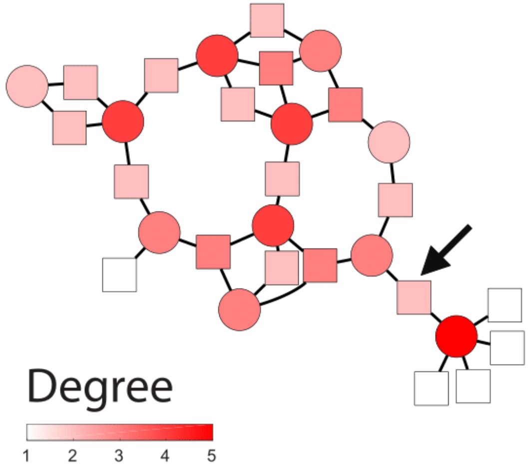
GTEx Expression Data from 38 Tissues/Tissue-Sites

(1) Integrate Regulatory Information (PANDA)

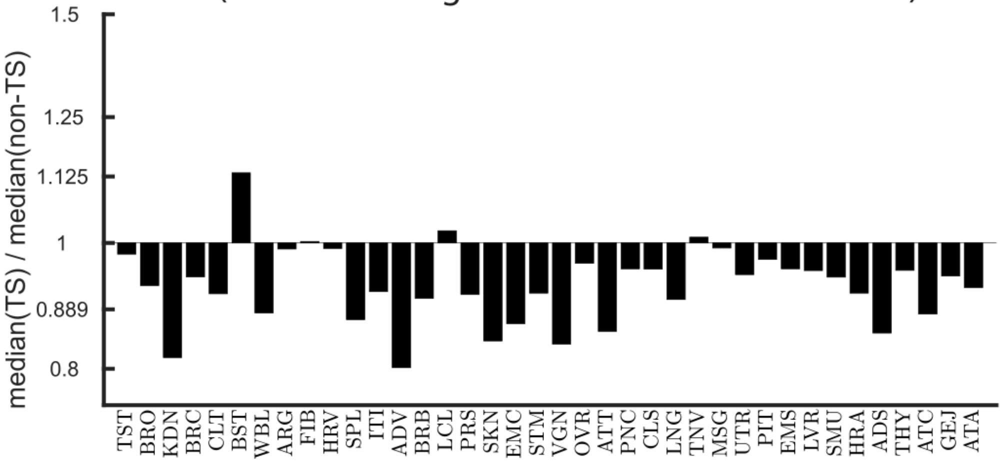
Regulatory Networks for 38 Tissues/Tissue-Sites

Sonawane et. al. “Understanding Tissue-Specific Gene Regulation” Cell Reports. 2017.

Degree of Tissue-Specific Genes



Network Centralities



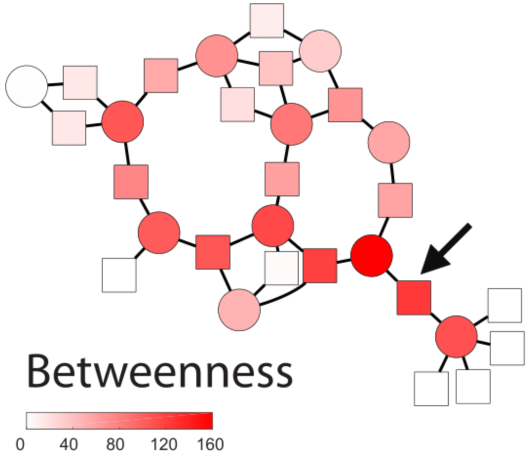
Ratio of Median Centrality

(TS vs. non-TS genes in Each Tissue-Network)

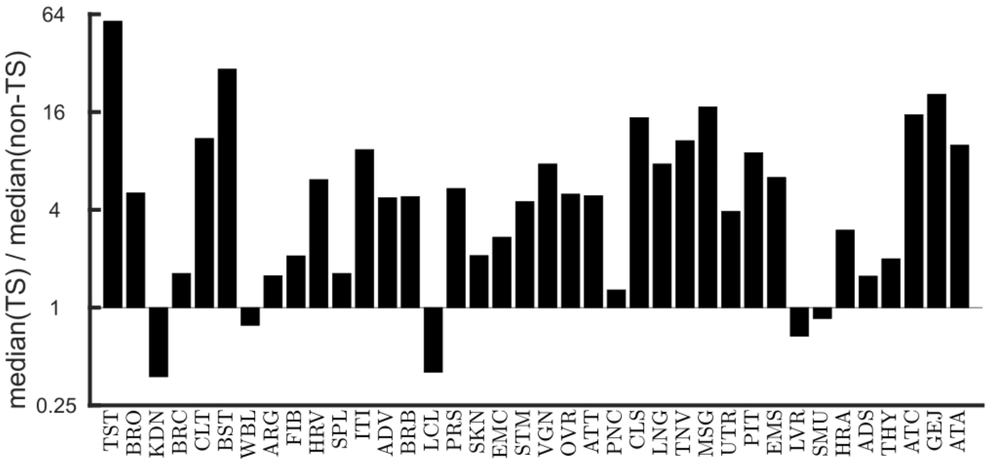
- Calculated the degree of genes in each of the 38 PANDA-predicted network.
- Compared the degree of tissue-specific genes to the degree of non-tissue-specific genes in each of these networks.
- Tissue-specific genes are depleted for regulatory edges.

Betweenness of Tissue-Specific Genes

Network Centralities



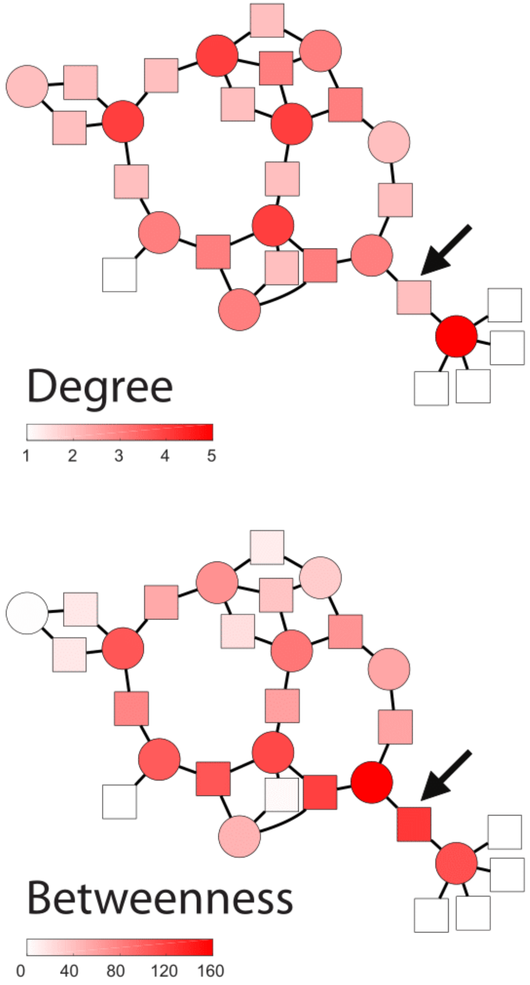
Ratio of Median Centrality



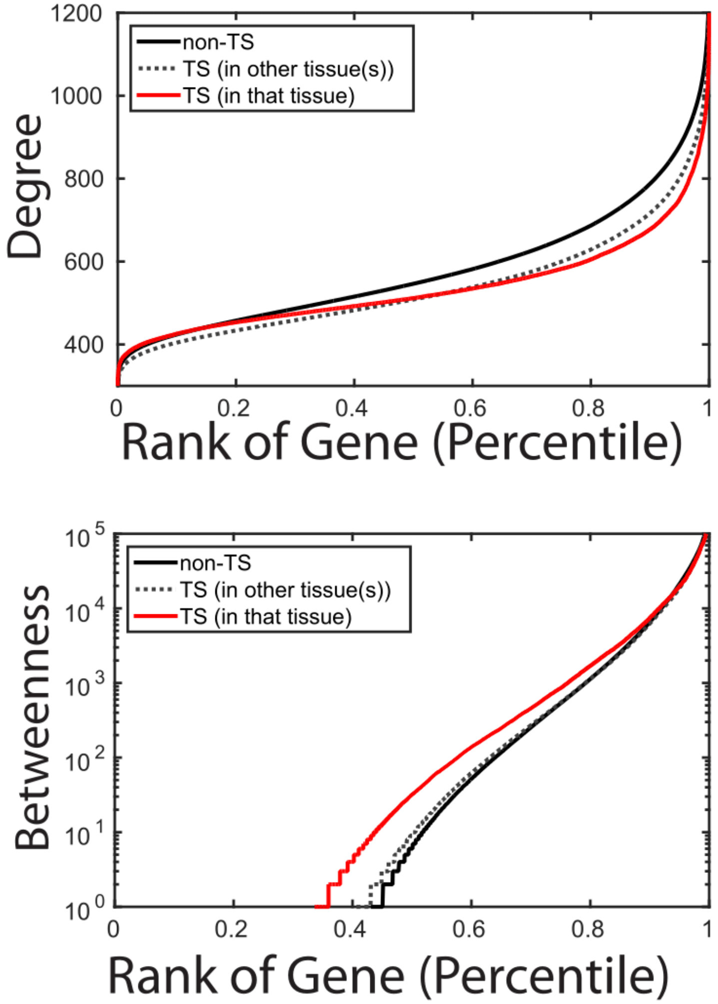
- Calculated the <u>betweenness</u> of genes in each of the 38 PANDA-predicted network.
- Compared the betweenness of tissue-specific genes to non-tissue-specific genes in each of these networks.
- Tissue-specific genes are <u>enriched</u> for regulatory paths.

Centrality of Tissue-Specific Genes

Network Centralities

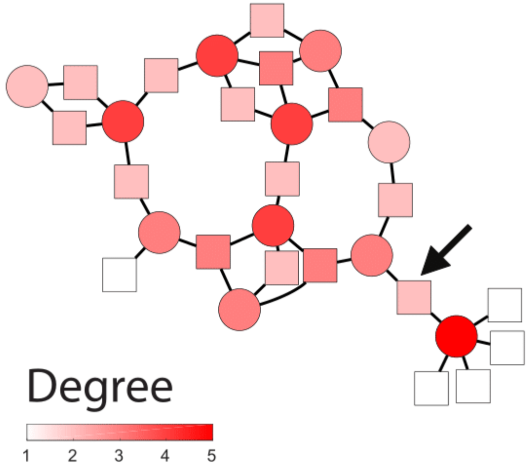


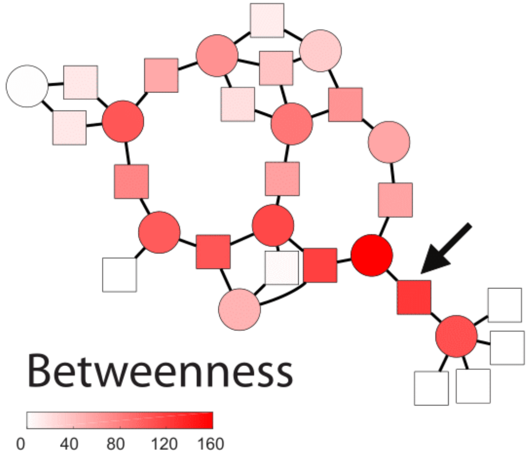
Distribution of Centrality Values
(Across All Tissue-Networks)



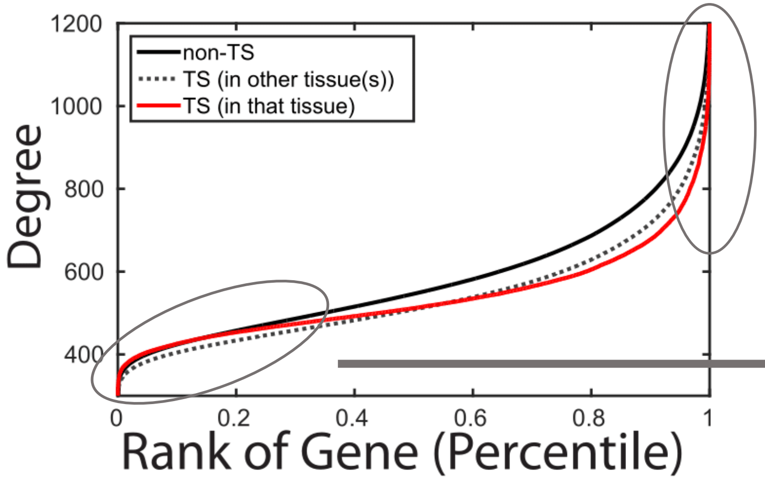
Centrality of Tissue-Specific Genes

Network Centralities



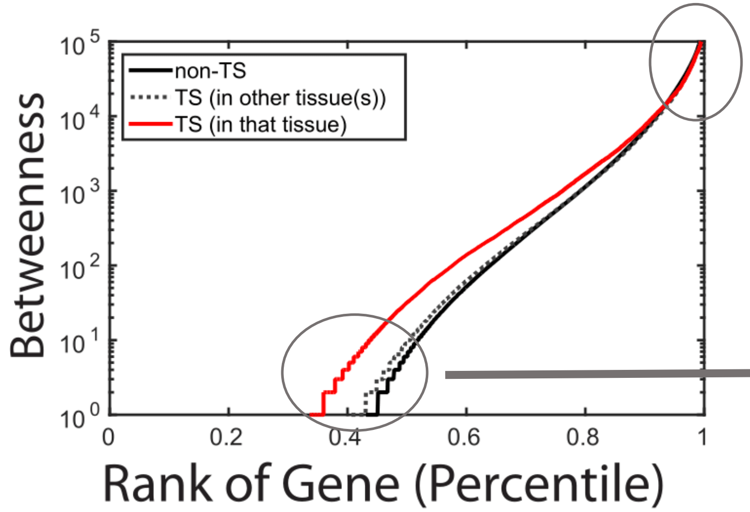


Distribution of Centrality Values
(Across All Tissue-Networks)



TS-genes are depleted among network hubs

Low-degree genes are just as likely to be TS as non-TS



High-betweeny genes are just as likely to be TS as non-TS

Non-TS genes are more likely to have no paths (betweenness=0)

What does this mean?

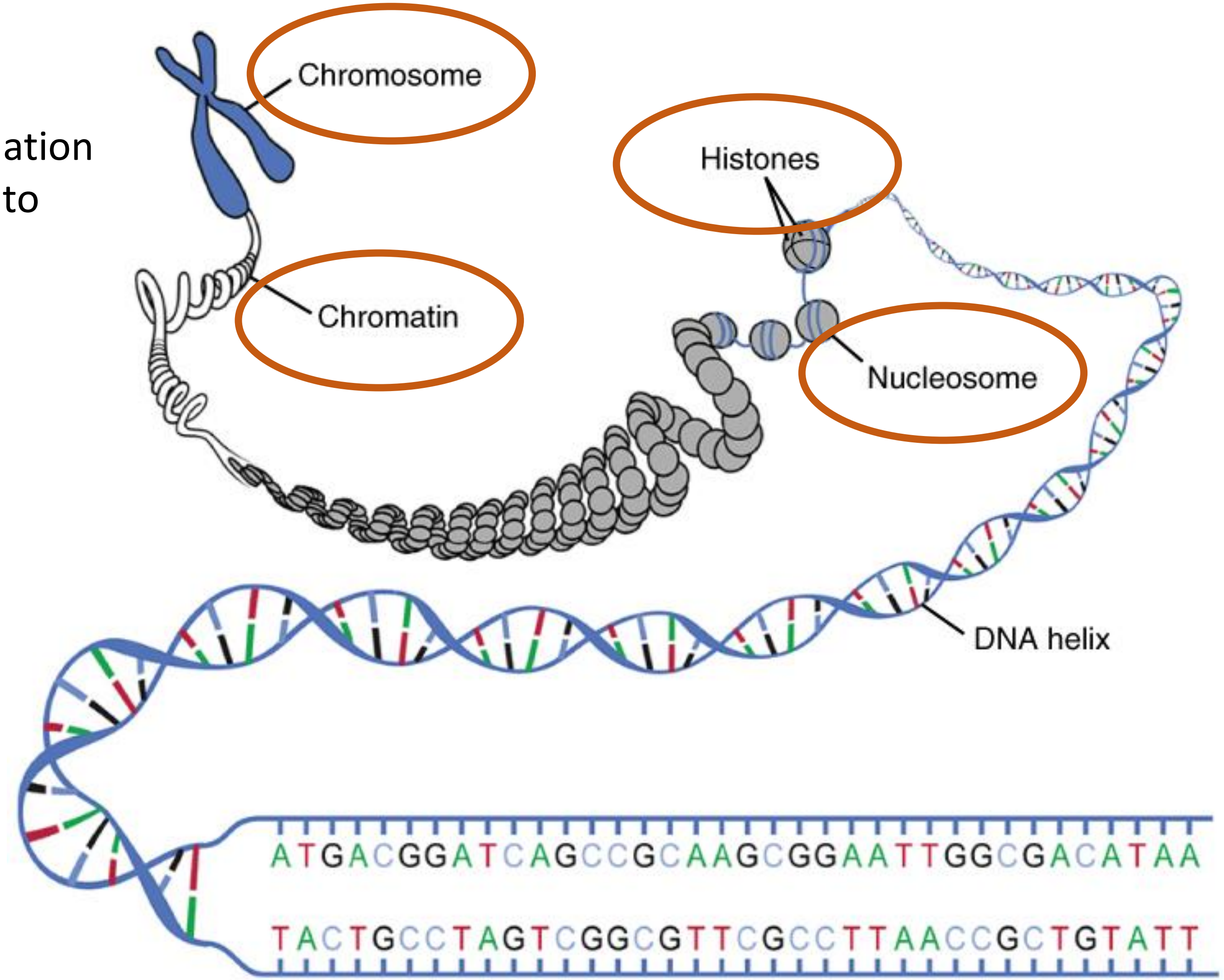
- Tissue-specific regulation does not occur in dense portions of the regulatory network, or by the formation of tissue-specific hubs.
- Tissue-specific genes are central to the regulatory network on an <u>intermediate</u> scale due to the influence of tissue-specific regulatory paths.
- Taken together these results suggest that tissue-specific edges build on an existing regulatory network structure in a way that allows tissue-specific regulatory access to certain genes.

Sonawane et. al. “Understanding Tissue-Specific Gene Regulation” Cell Reports. 2017.

The Physical Structure of DNA

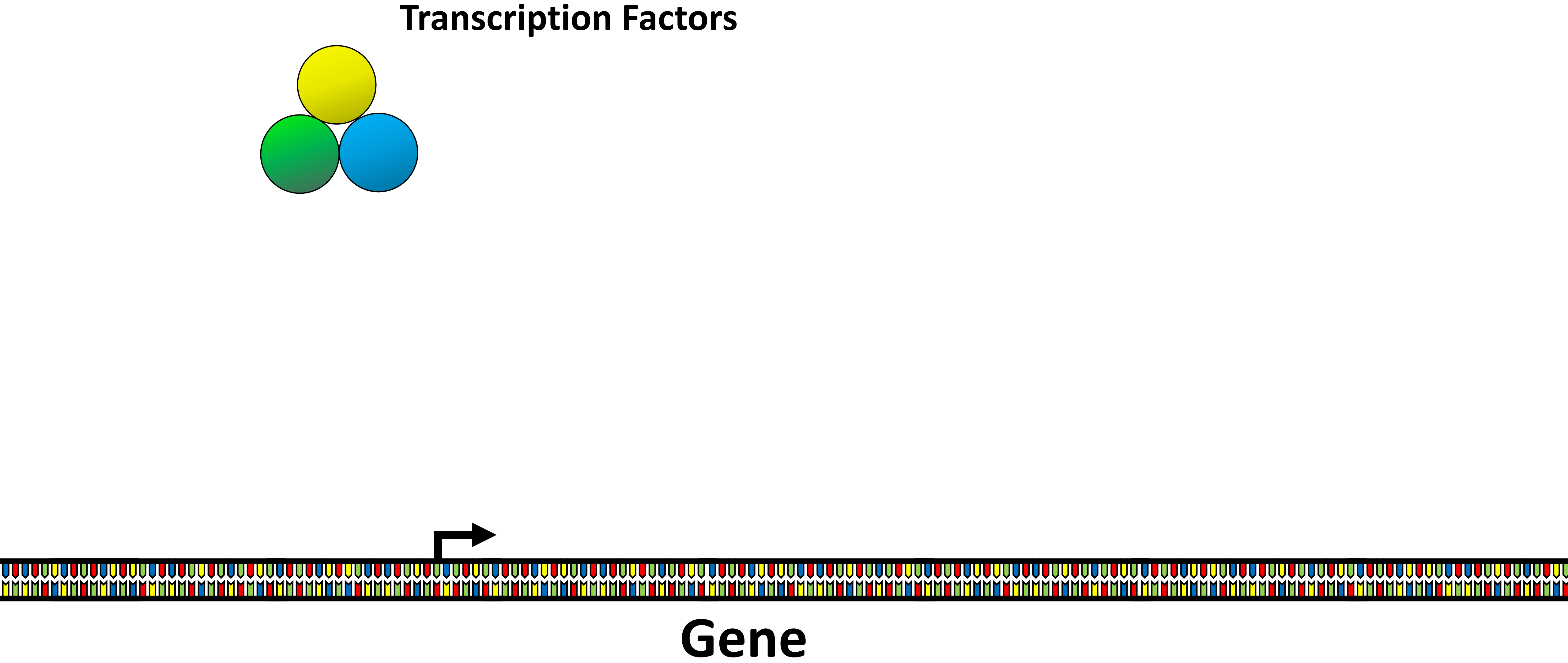
Epigenetics:

Elements of genetic information that don’t involve changes to the DNA sequence

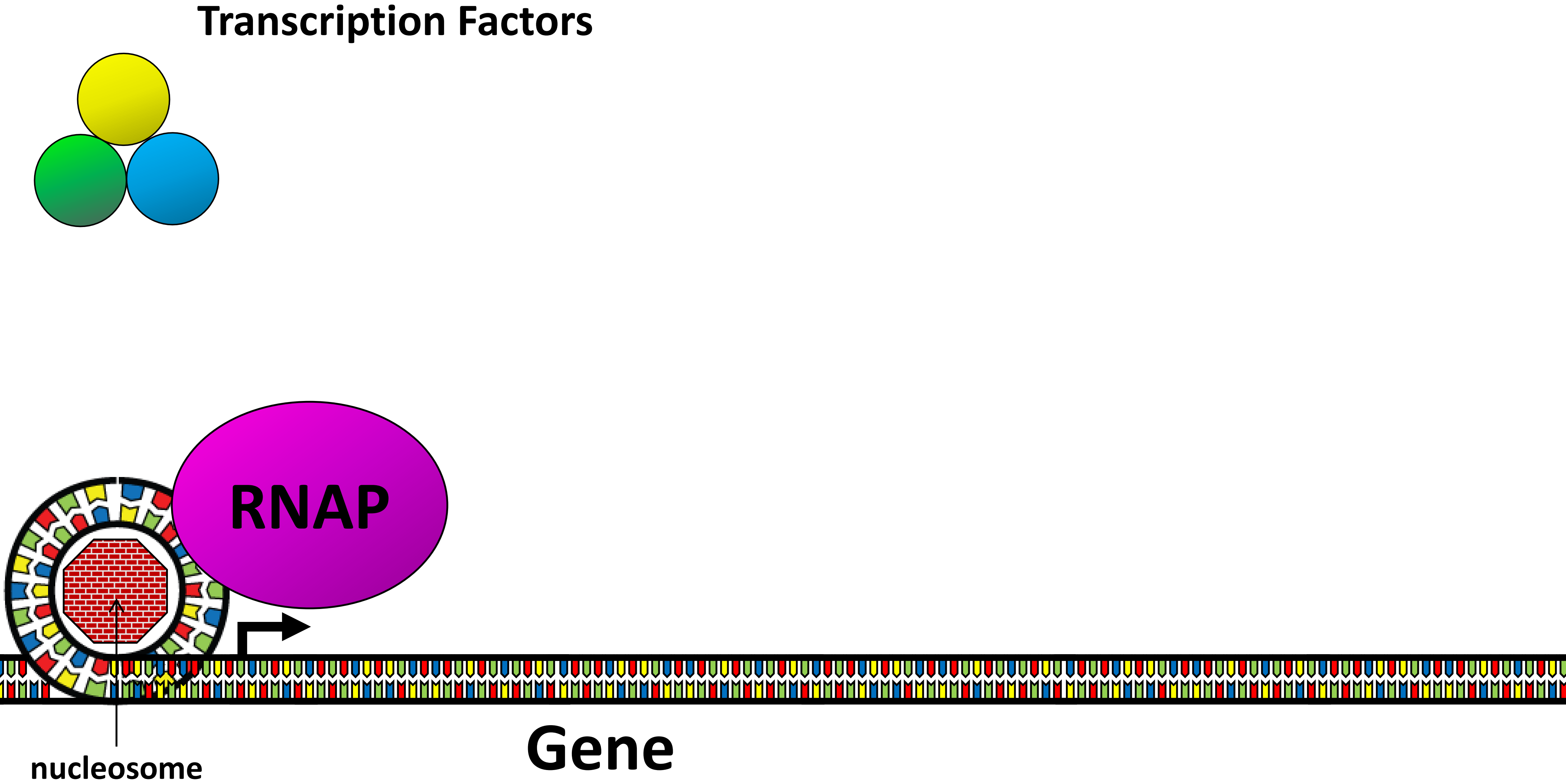


https://cnx.org/contents/9TxHOD3O@4/The-Nucleus-and-DNA-Replicatio

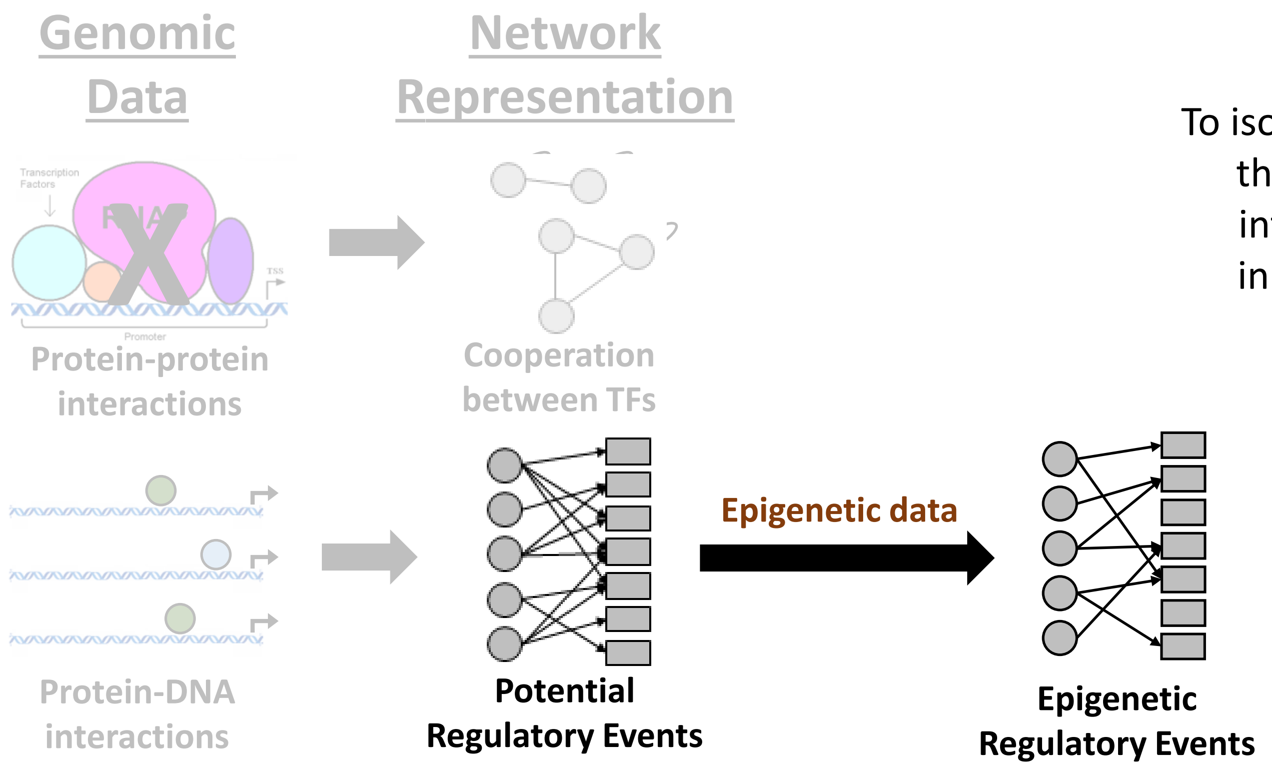
Impact of DNA Structure on Gene Regulation



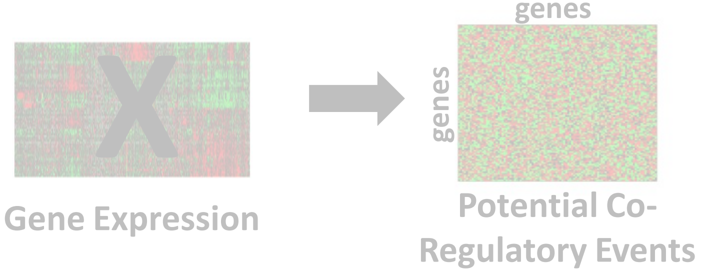
Impact of DNA Structure on Gene Regulation

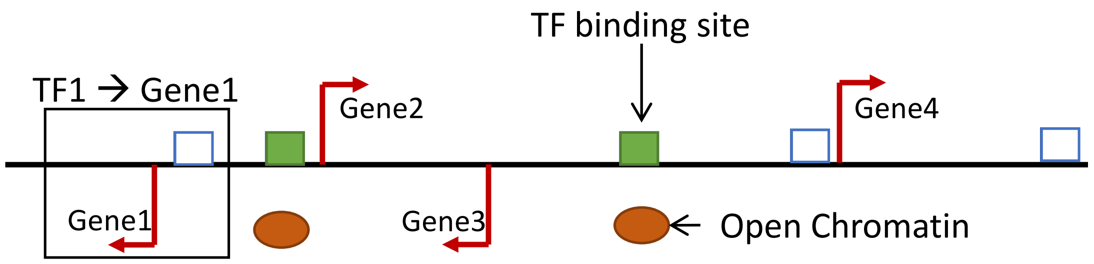


Incorporating Epigenetic Data



To isolate the role of epigenetic data, the co-expression and protein-interaction information can be initialized to an identity matrix

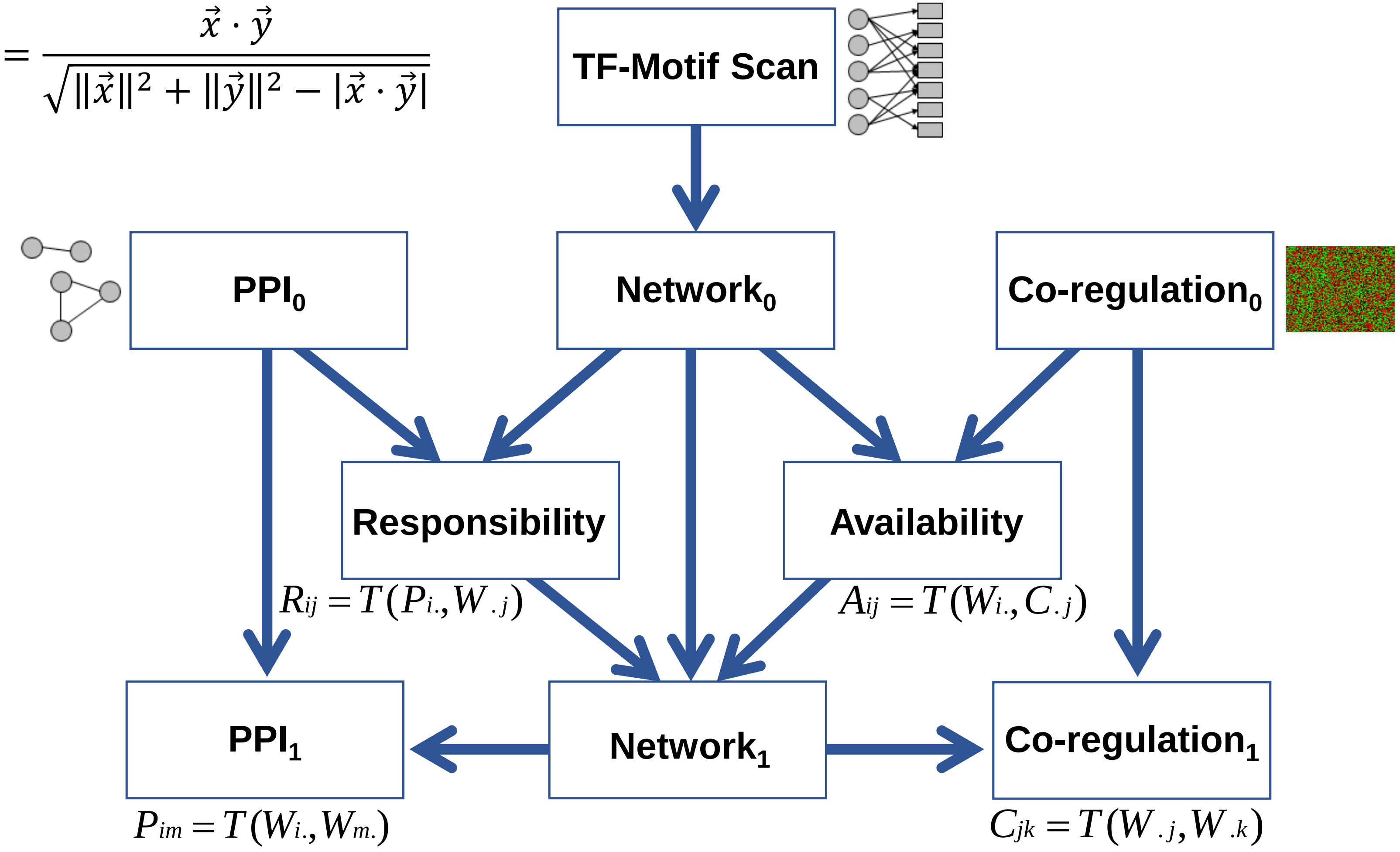




Incorporating Epigenetics with SPIDER

(Seeding PANDA Interactions to Derive Epigenetic Regulation)

$$T(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{\sqrt{\|\vec{x}\|^2 + \|\vec{y}\|^2 - |\vec{x}\cdot\vec{y}|}}$$

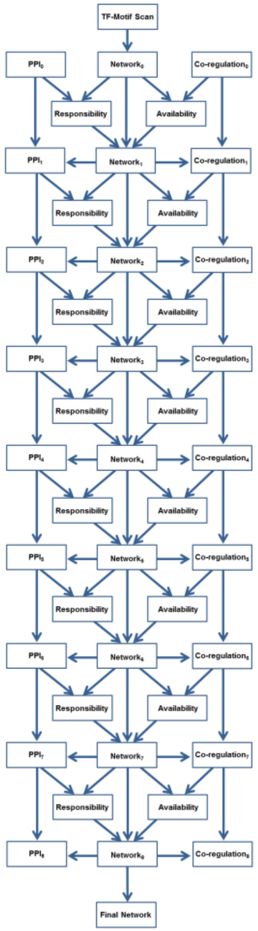


$R_{ij} = T(P_{i\cdot}, W_{\cdot j})$

$A_{ij} = T(W_{i\cdot}, C_{\cdot j})$

$P_{im} = T(W_{i\cdot}, W_{m\cdot})$

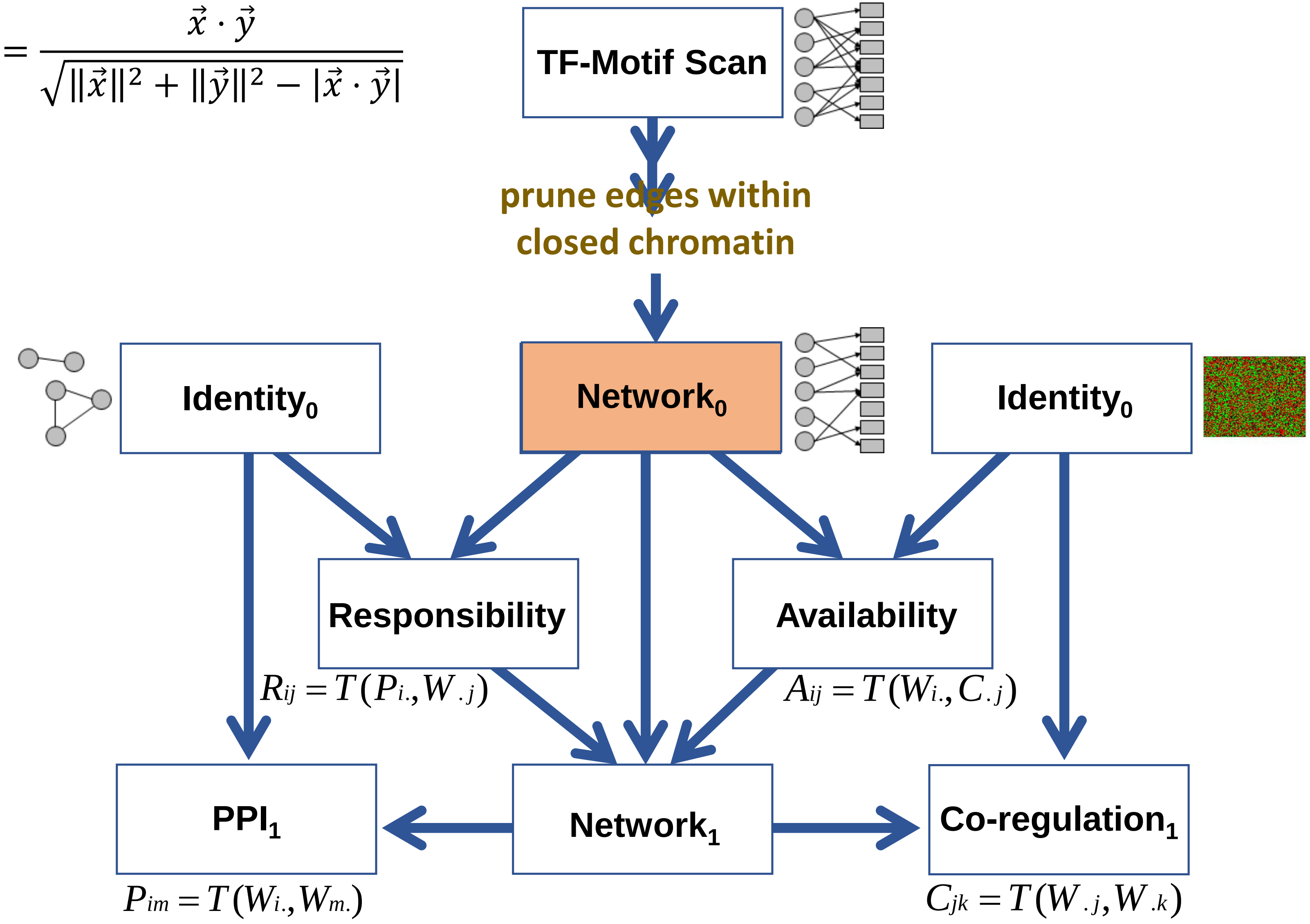
$C_{jk} = T(W_{\cdot j}, W_{\cdot k})$



Incorporating Epigenetics with SPIDER

(Seeding PANDA Interactions to Derive Epigenetic Regulation)

$$T(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{\sqrt{\|\vec{x}\|^2+\|\vec{y}\|^2-|\vec{x}\cdot\vec{y}|}}$$



TF-Motif Scan

prune edges within closed chromatin

Identity$_0$

Network$_0$

Identity$_0$

Responsibility

Availability

$R_{ij} = T(P_{i\cdot}, W_{\cdot j})$

$A_{ij} = T(W_{i\cdot}, C_{\cdot j})$

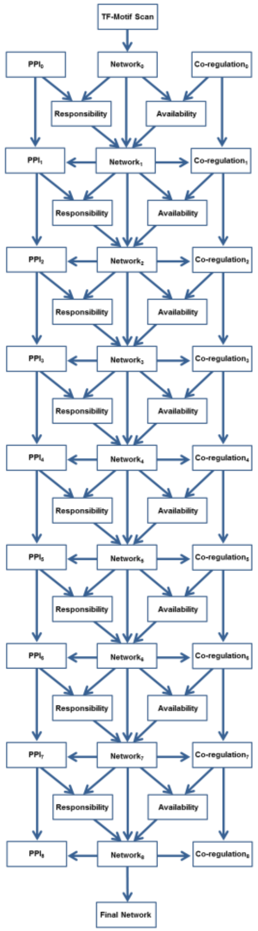
PPI$_1$

Network$_1$

Co-regulation$_1$

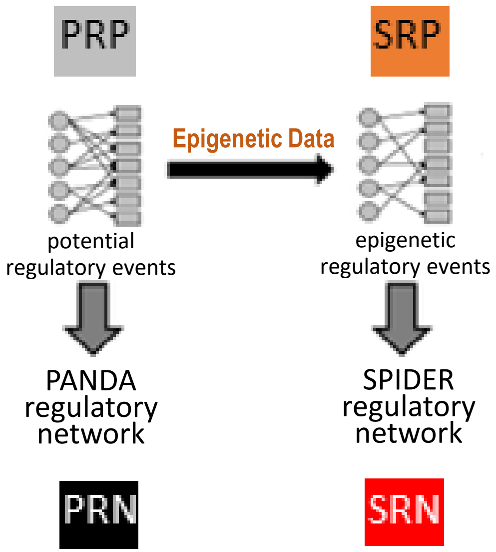
$P_{im} = T(W_{i\cdot}, W_{m\cdot})$

$C_{jk} = T(W_{\cdot j}, W_{\cdot k})$

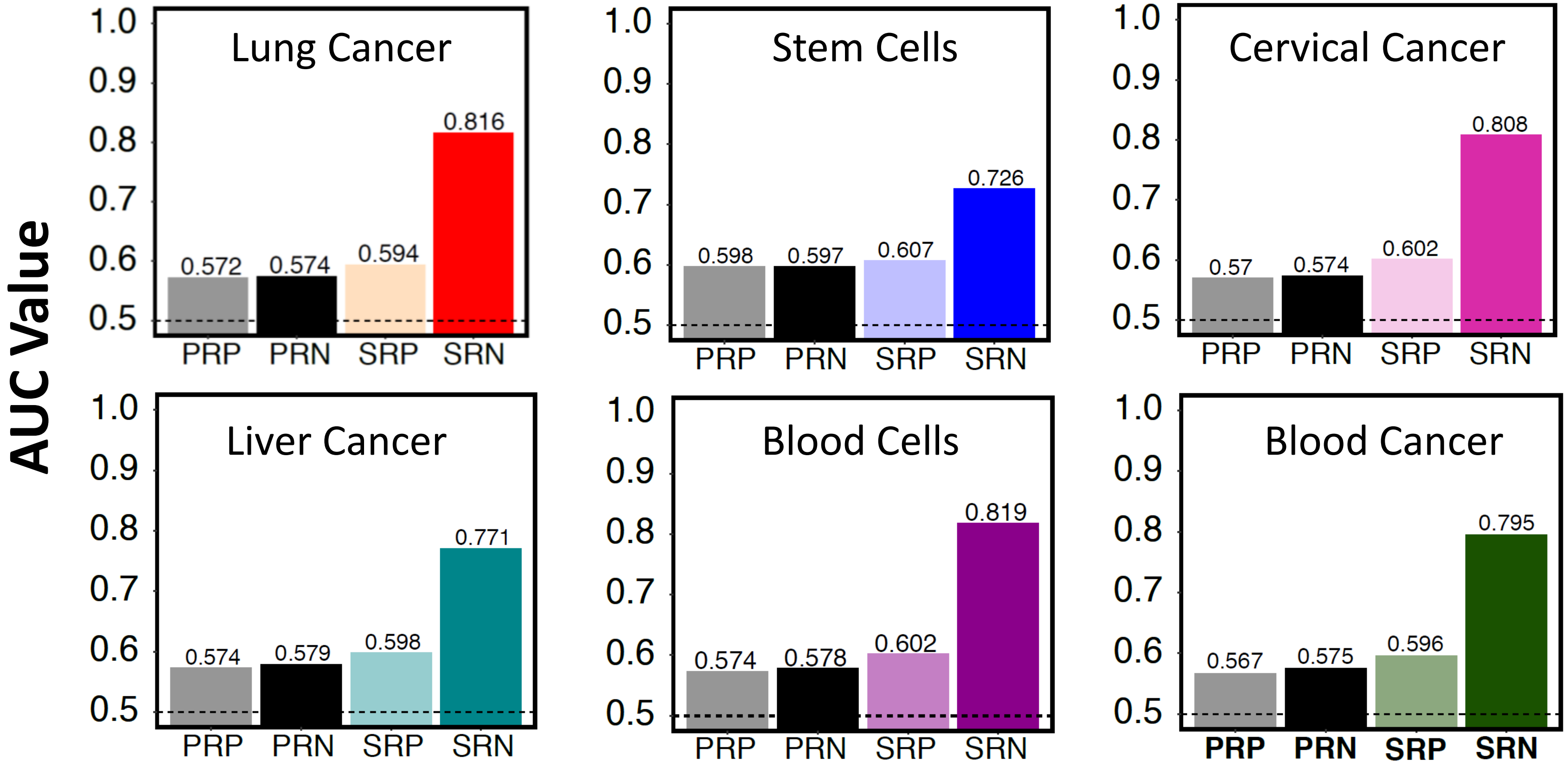


SPIDER predicts accurate networks

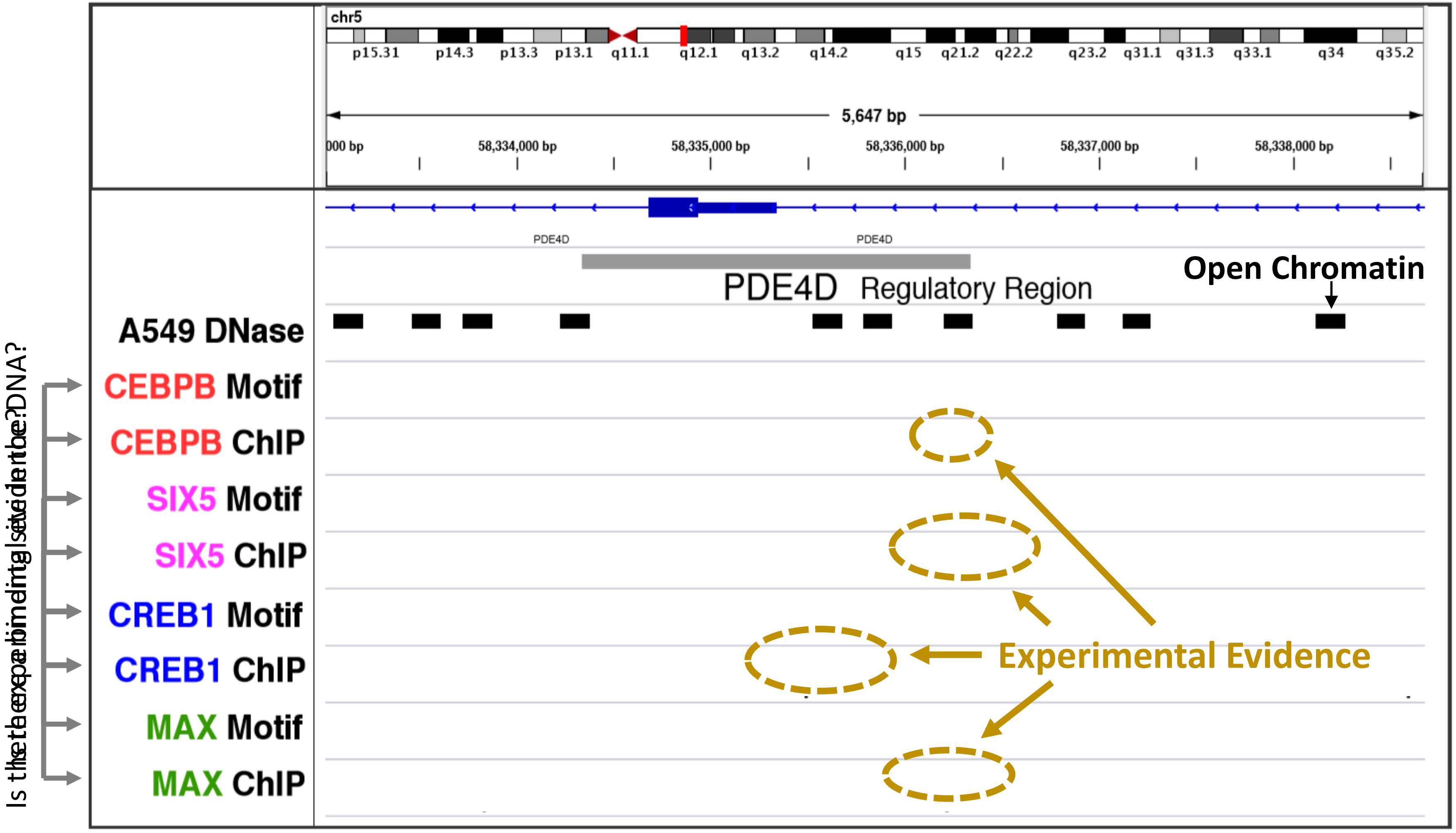
LEGEND



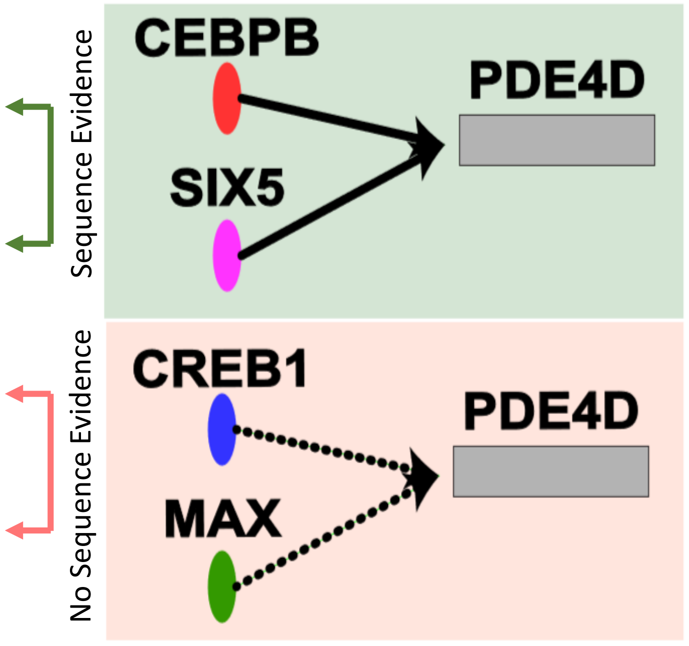
PRP = PANDA regulatory prior
PRN = PANDA regulatory network
SRP = SPIDER regulatory prior
SRN = SPIDER regulatory network



SPIDER recovers “missing links”

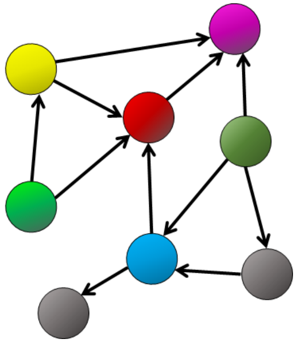
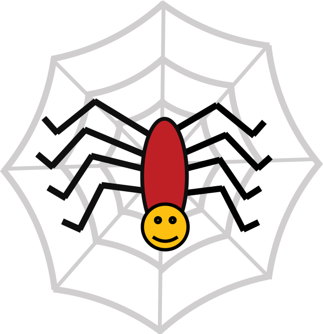
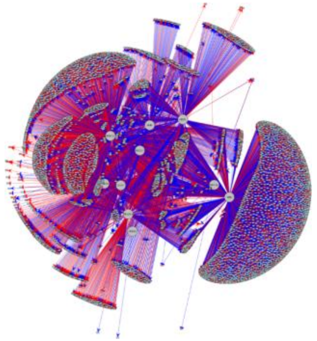


Top-weight edges predicted by SPIDER (Lung Cancer):



Summary

- Gene regulatory networks provide a framework for modeling and analyzing *biological processes*
- Effectively *integrating multi-Omics* data is important in developing a holistic model of biological systems
- Analyzing gene regulatory networks can lead to *novel insights* into how biological systems function.


Acknowledgements

PANDA Development
Curtis Huttenhower
Guo-Cheng Yuan
John Quackenbush

SPIDER Development
Abhijeet Sonawane
Dawn DeMeo

Channing Fellows
Alex Song
Margherita De Marzio
Arda Halu
Enrico Maiorino
Annika Rohl
Seung Han Baek

GTEx Network Analysis
Marieke Kuijjer
Abhijeet Sonawane
John Platig
Maud Fagny
Cho-Yi Chen
Joe Paulson
Camila Lopes-Ramos

funding: NIH K25HL133599, Geoffrey Beene Foundation
contact: rekrg@channing.harvard.edu
website: https://sites.google.com/a/channing.harvard.edu/kimberlyglass/

Why is this important for disease?

- Identified genes that:
  - → Have a significant cis-eQTL in a tissue.
  - → Have a significant cis-eQTL with a GWAS SNP in a tissue.
- Determined the median rank of those genes (compared to all other genes) based on their:
  - → Degree
  - → Betweenness
- Genes associated with GWAS SNPs, although depleted for tissue-specific genes, also are “bottlenecks” (comparatively lower degree, higher betweenness)