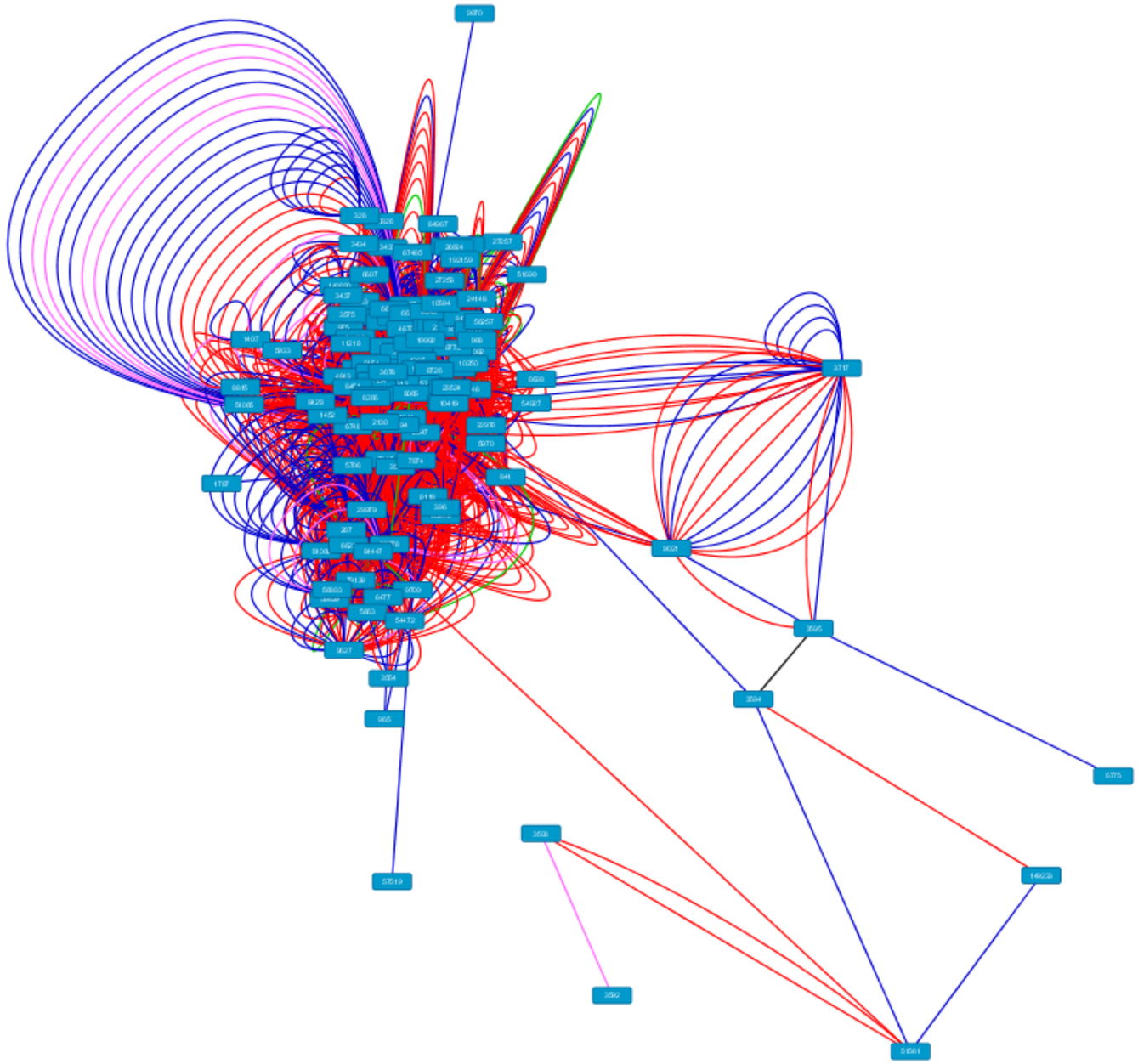
Contrasting Breast Cancer Subtypes by Analyzing Differences in Network Structure

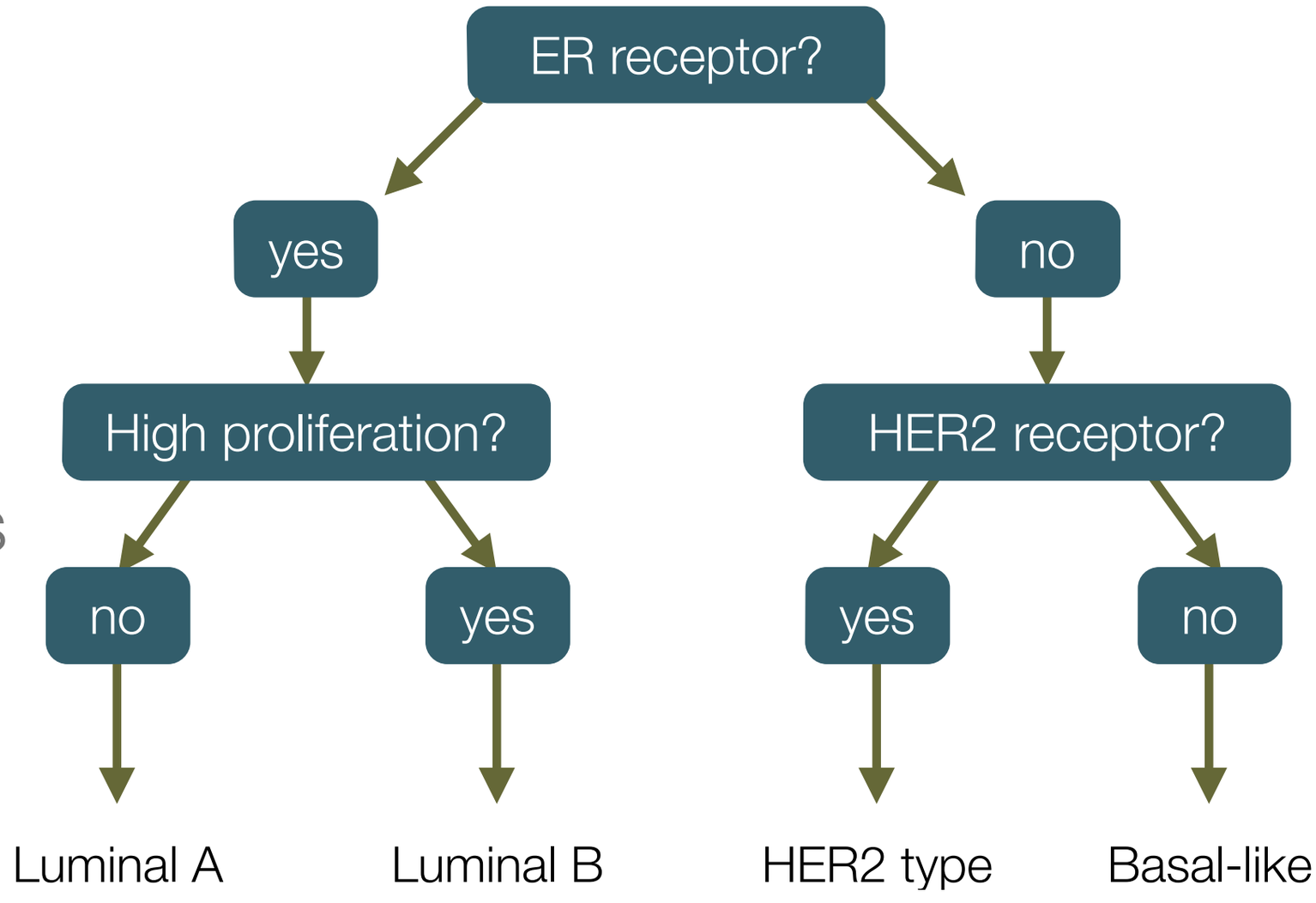
Kamrine Poels[a,d]
Kimberly Glass[b,c]
John Quackenbush[a,d]

[a]Dana Farber Cancer Institute
[b]Brigham and Women's Hospital
[c]Harvard Medical School
[d]Harvard T.H. Chan School of Public Health



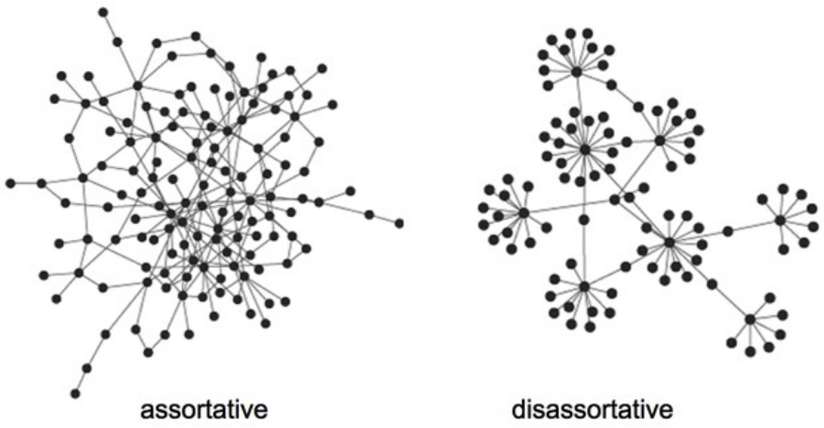
Breast Cancer

- Most common cancer among women worldwide
  - Over 230,000 new cases each year in the U.S.
- Four well-known molecular subtypes
- Classified by the presence of hormone receptors

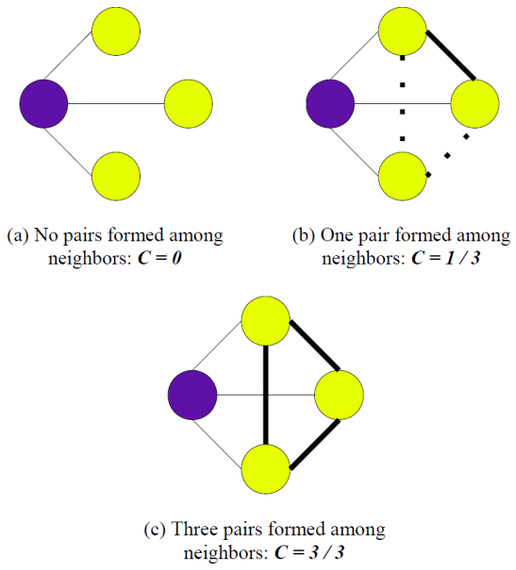


Biological Networks

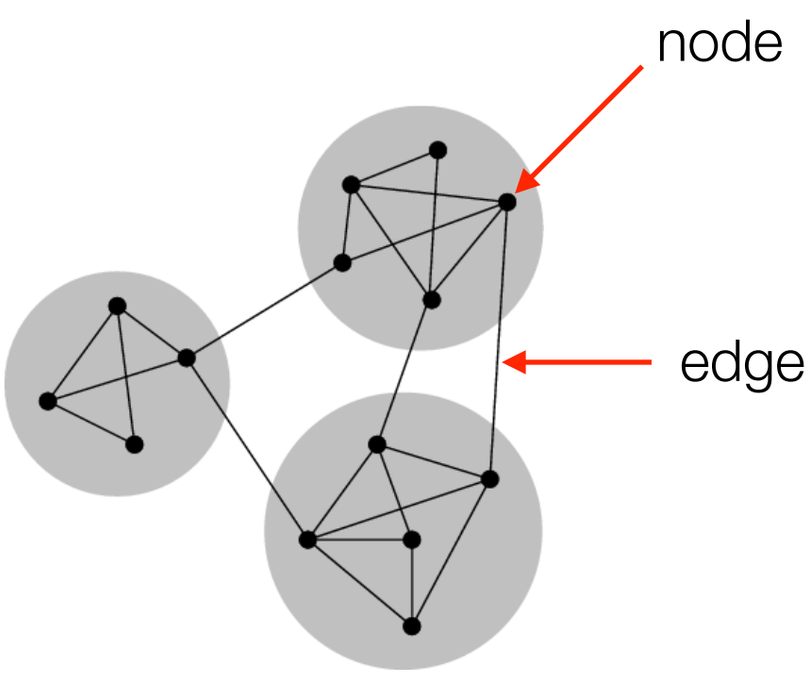
- Graph built of nodes and edges
- Helpful for visualizing or analyzing large sets of data
- Structure of network: assortative, modules, and clustering coefficients



Assortativity



Transitivity



Modularity

PANDA (Passing Attributes between Networks for Data Assimilation)[1]

- Message-passing algorithm
- Main objective:
  - Find concordance between different types of data represented by networks
- Two types of nodes: ***effectors*** and ***affected***
- Three types of edges:
  - Between *effectors*
  - *Effectors* and *affected*
  - Between *affected*

Effectors ⇒ Transcription Factors

Affected ⇒ Genes

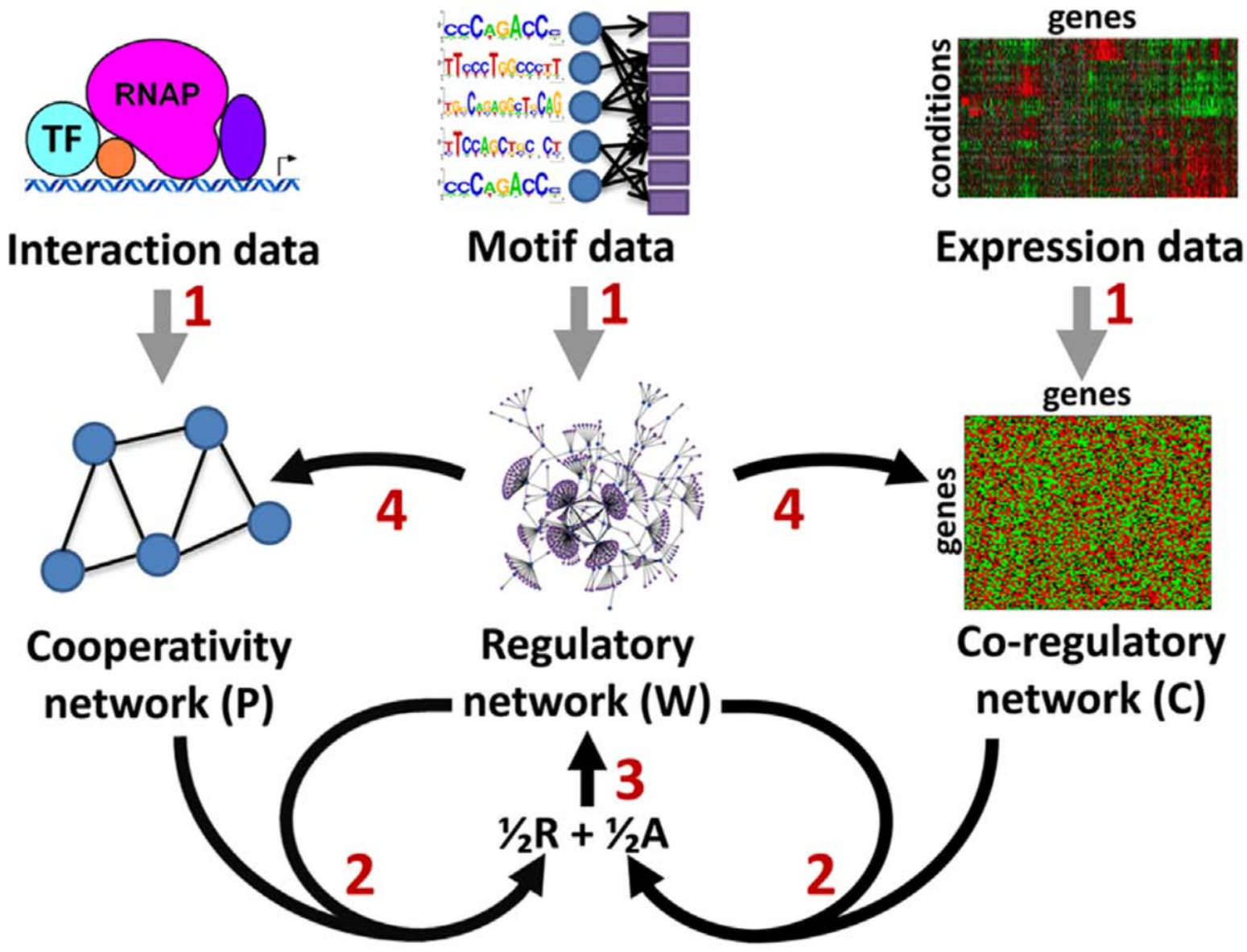
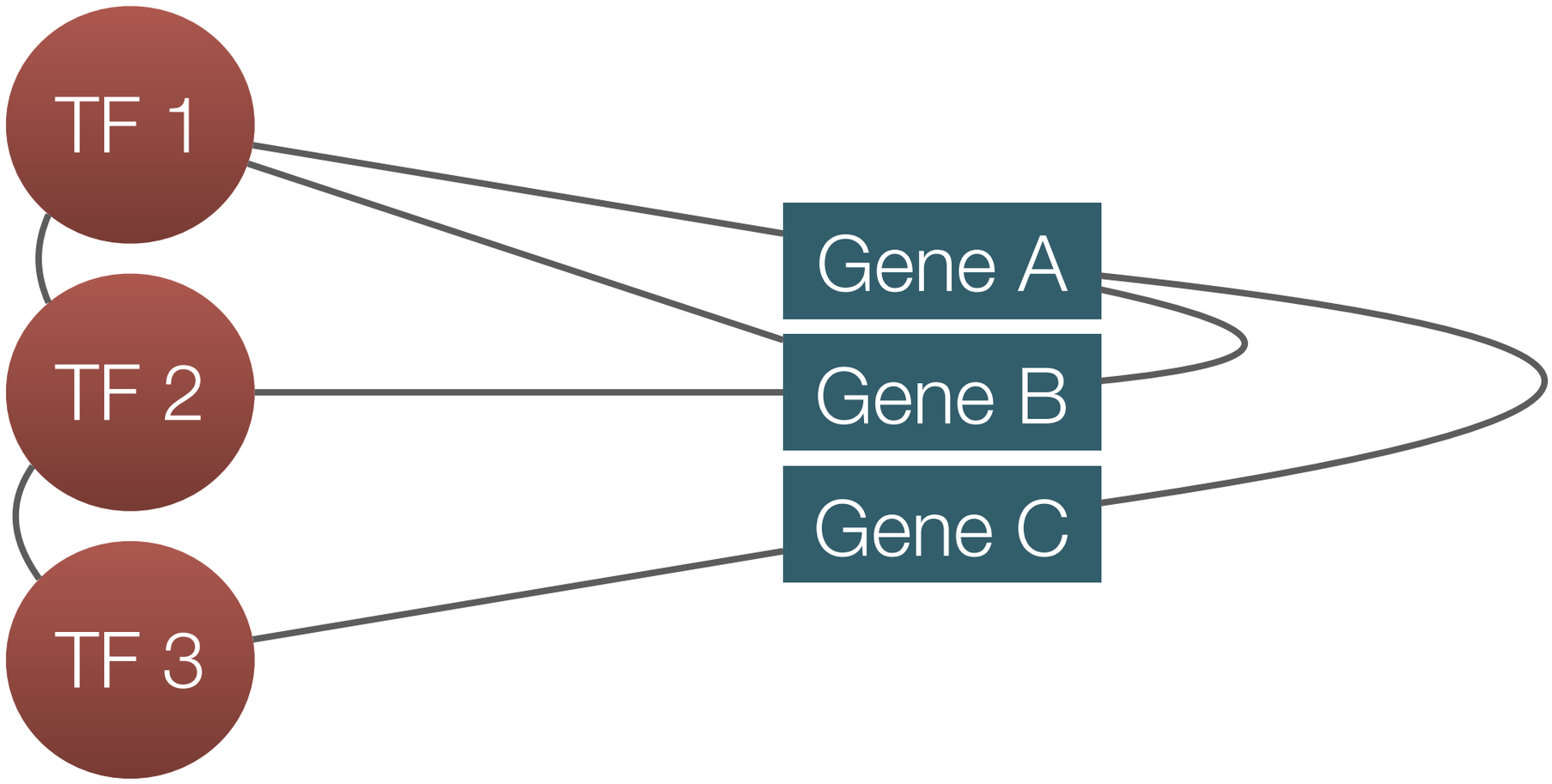


Image taken from *Glass et al* 2014

PANDA's output

3 updated networks

- Cooperativity network (protein-protein)
- Regulatory network (protein-gene)
- Co-regulation network (gene-gene)



Methods: Running PANDA

Data and Sources

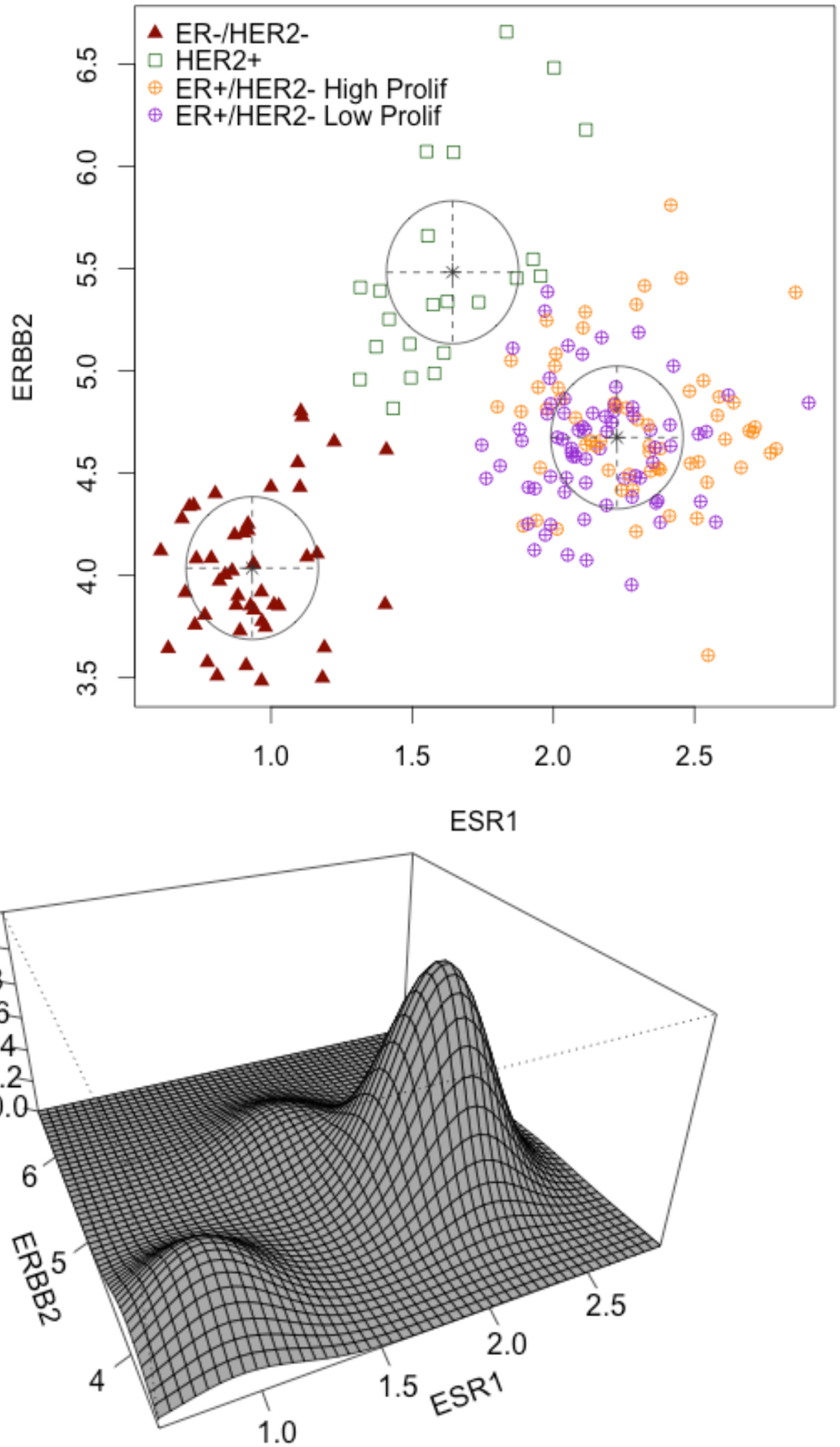
- Protein-interaction network ← Physical TF interactions measured using a mouse-2-hybrid assay[2]

- Regulatory Network ← scanning human promoters for the core vertebrate DNA sequence motifs in JASPAR (where promoter is [-750, +250] around the TSS)

Used for all 4 subtypes

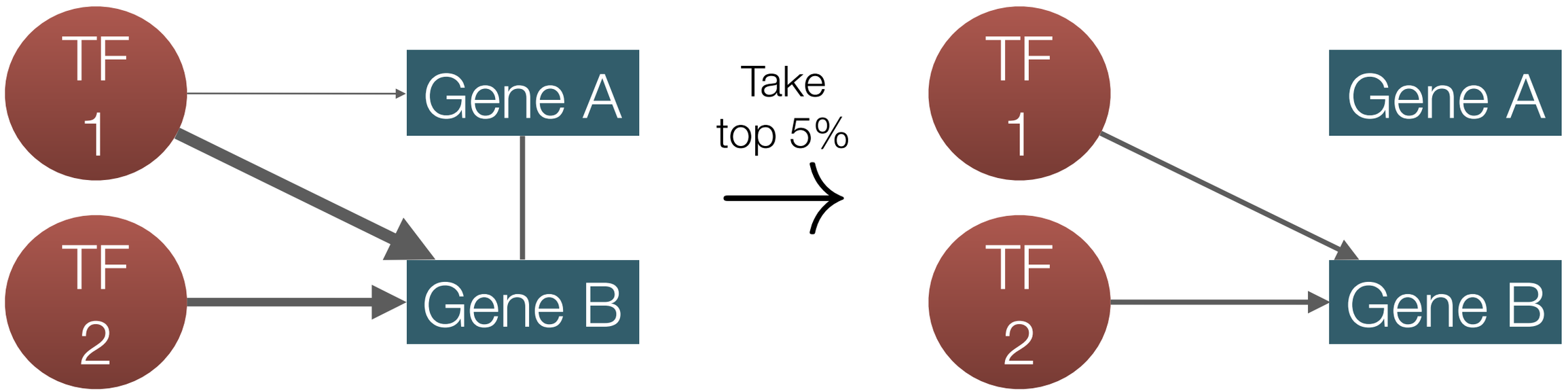
Data and Sources

- Gene expressions from a breast cancer study in 2007 that contained 198 lymph-node negative breast cancers
- Robustly classified samples into molecular subtypes
  - 71 Luminal A
  - 60 Luminal B
  - 45 Basal-like
  - 22 HER2



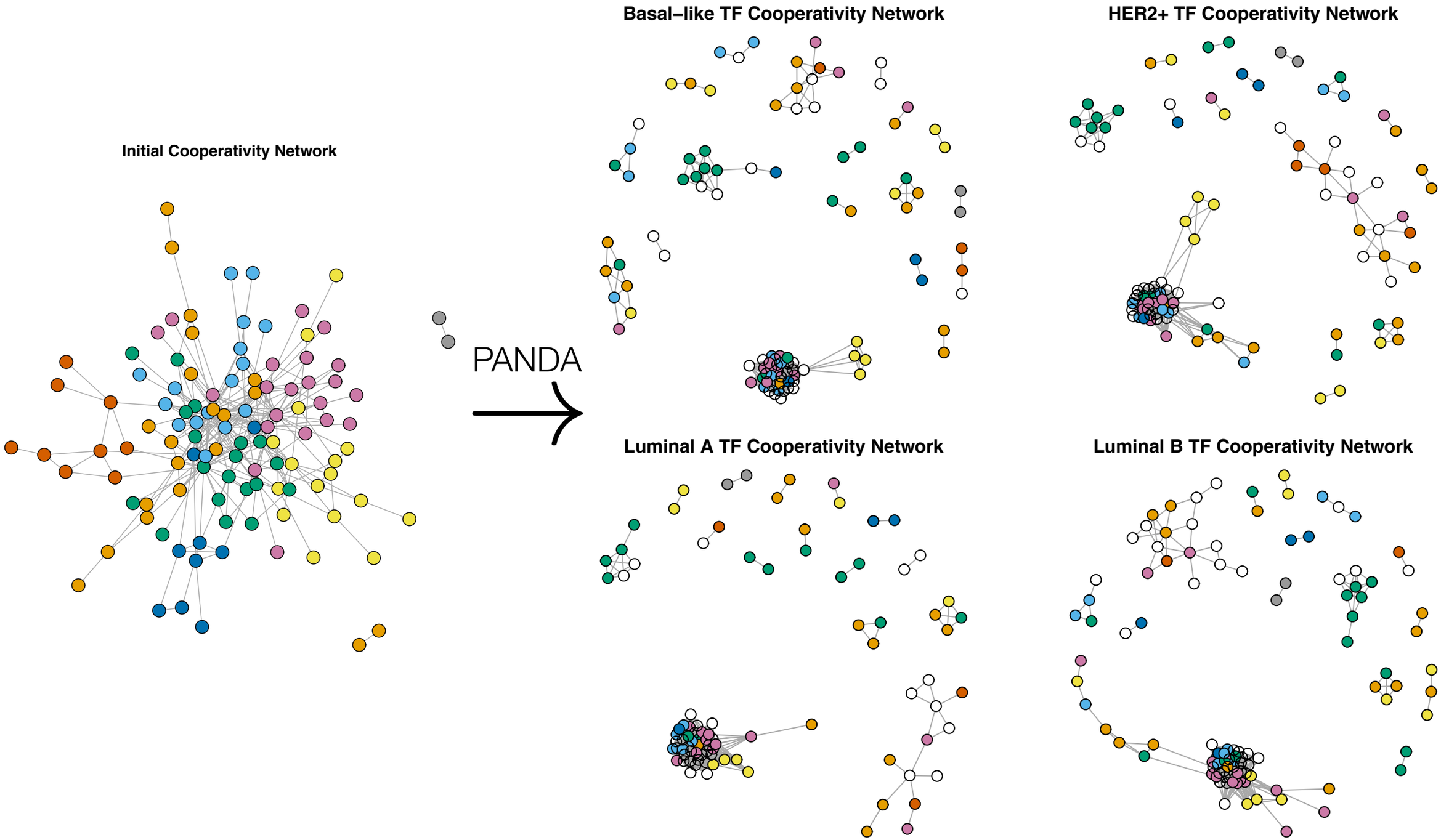
PANDA

- Run PANDA four times, each time with different gene expression
- Returns the probability that an edge exists in terms of standard score units
- We accept top 5% edges
- Networks with binary edges



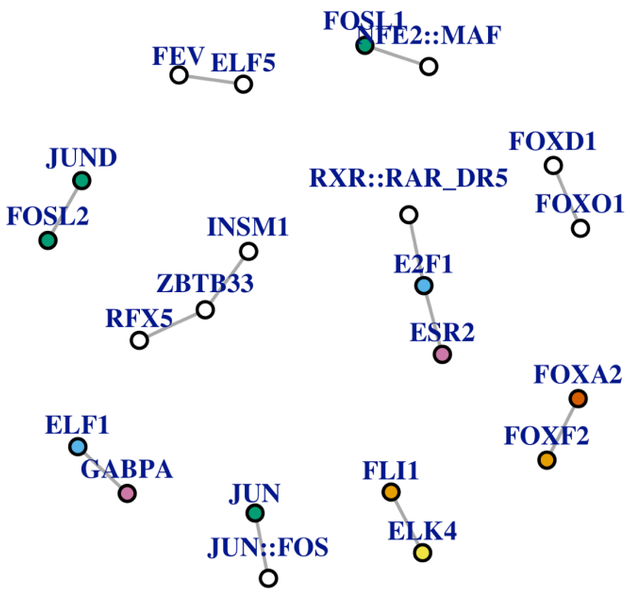
Network Analysis

Cooperativity Networks



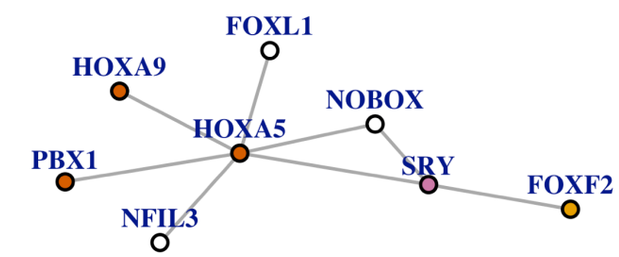
Remove common edges:

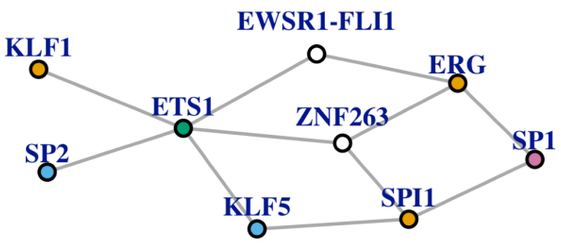
Basal-like TF Cooperativity Network



- Basal-like network is non-assortative
  - *ESR2* interacts with tumor suppressor E2F1
- HER2 network is disassortative with hubs:

HER2+ TF Cooperativity Network

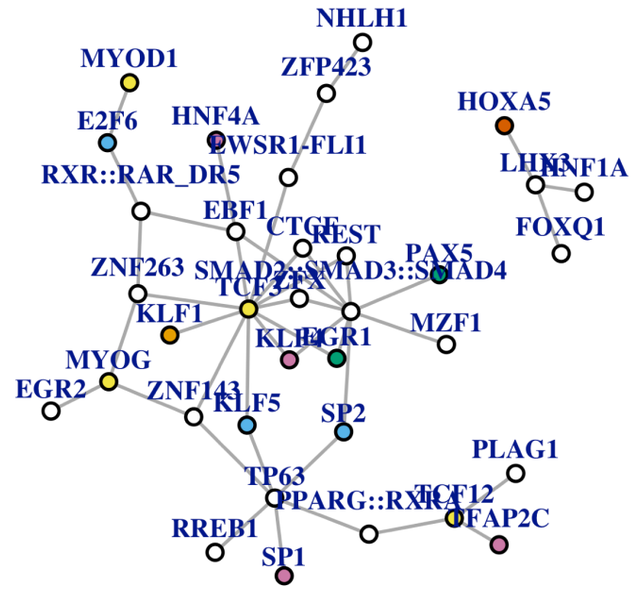




- *ETS1*: high expression indicative of poor prognosis
- *HOXA5:* induces apoptosis

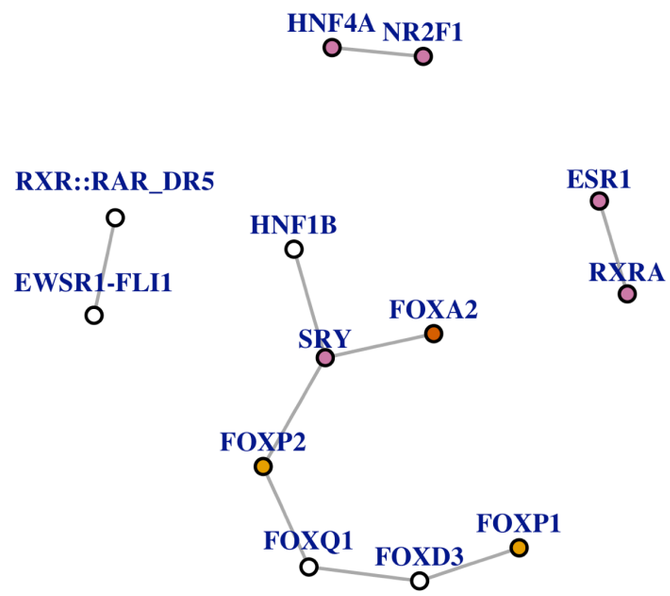
- Luminal A is disassortative
  - *TCF3*, shown to be involved in the regulation of breast cancer.
  - *SMAD* proteins

Luminal A TF Cooperativity Network

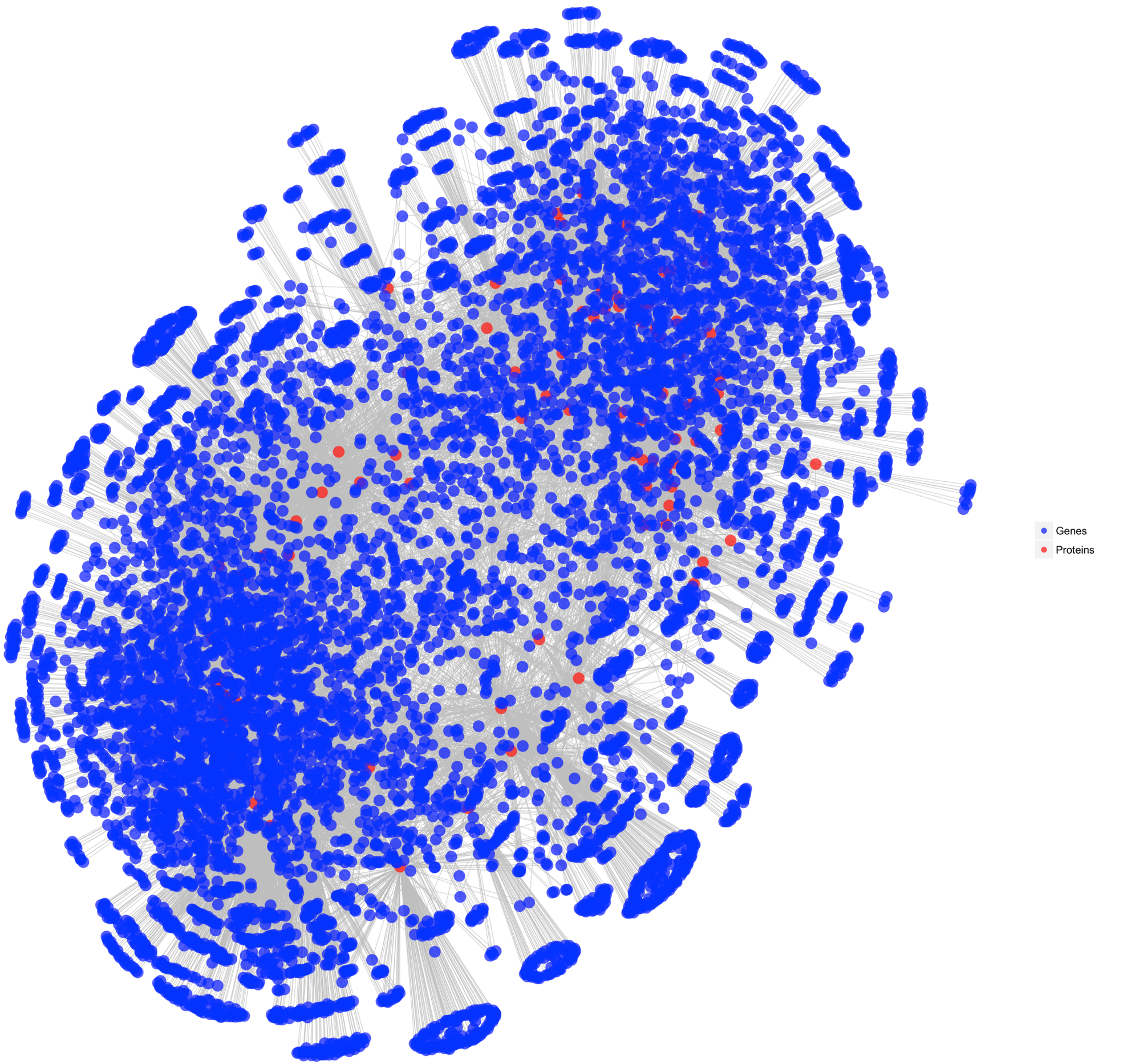


- Luminal B
  - *ESR1* interaction with *RXRA*, a receptor for retinoid acid
  - Study of *RXRA* as potential target therapy[3]

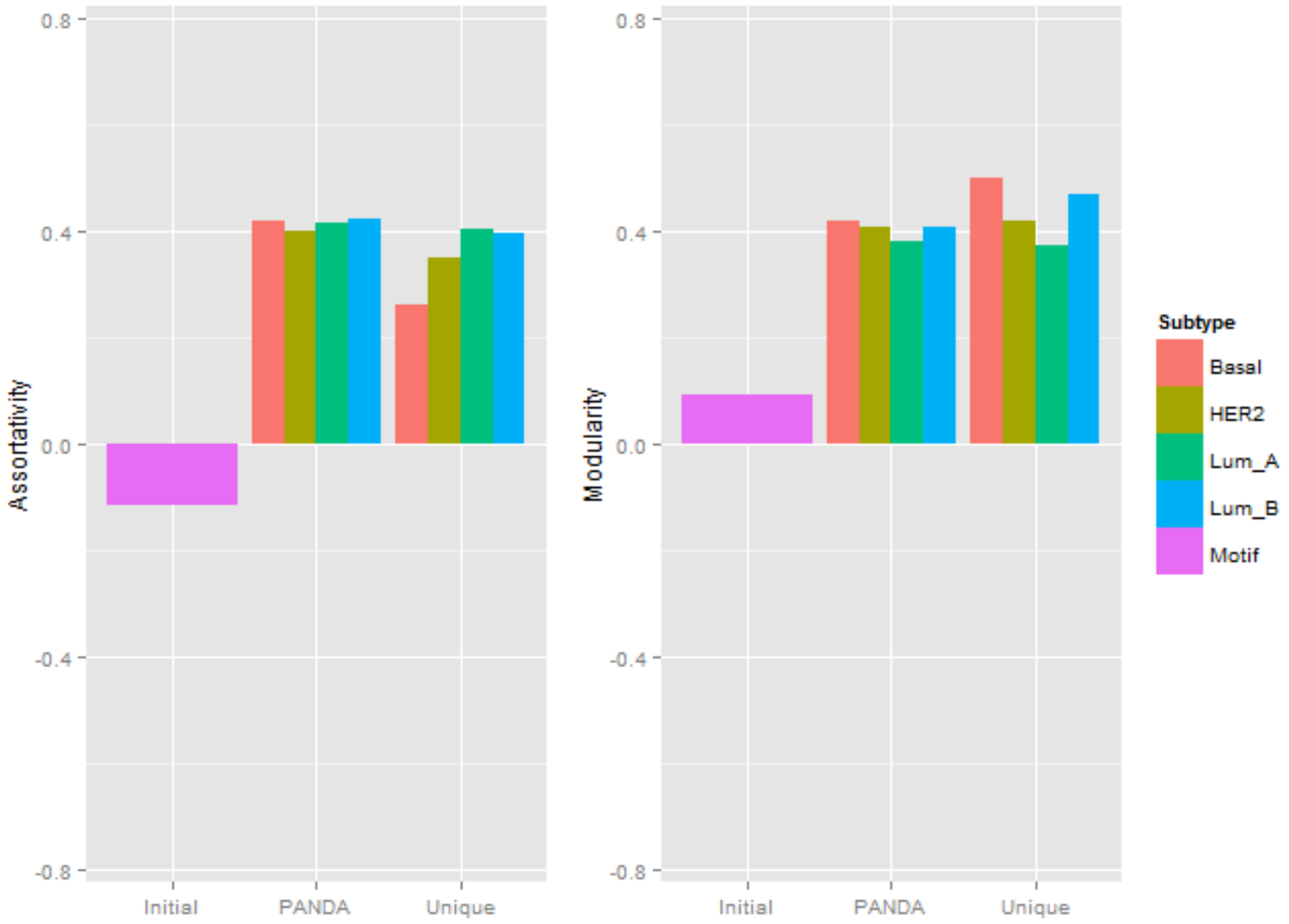
Luminal B TF Cooperativity Network



Regulatory Network

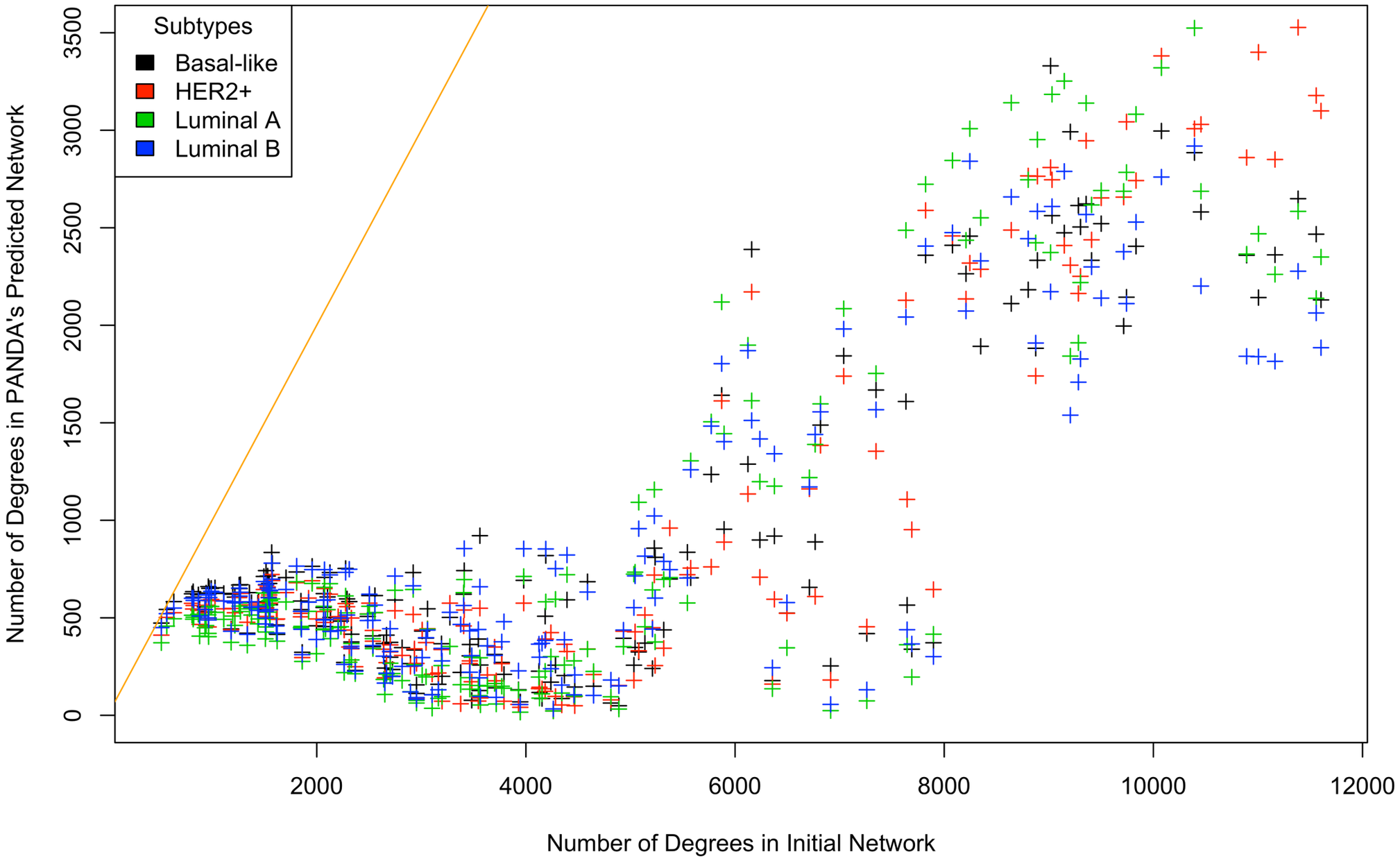


Regulatory Network



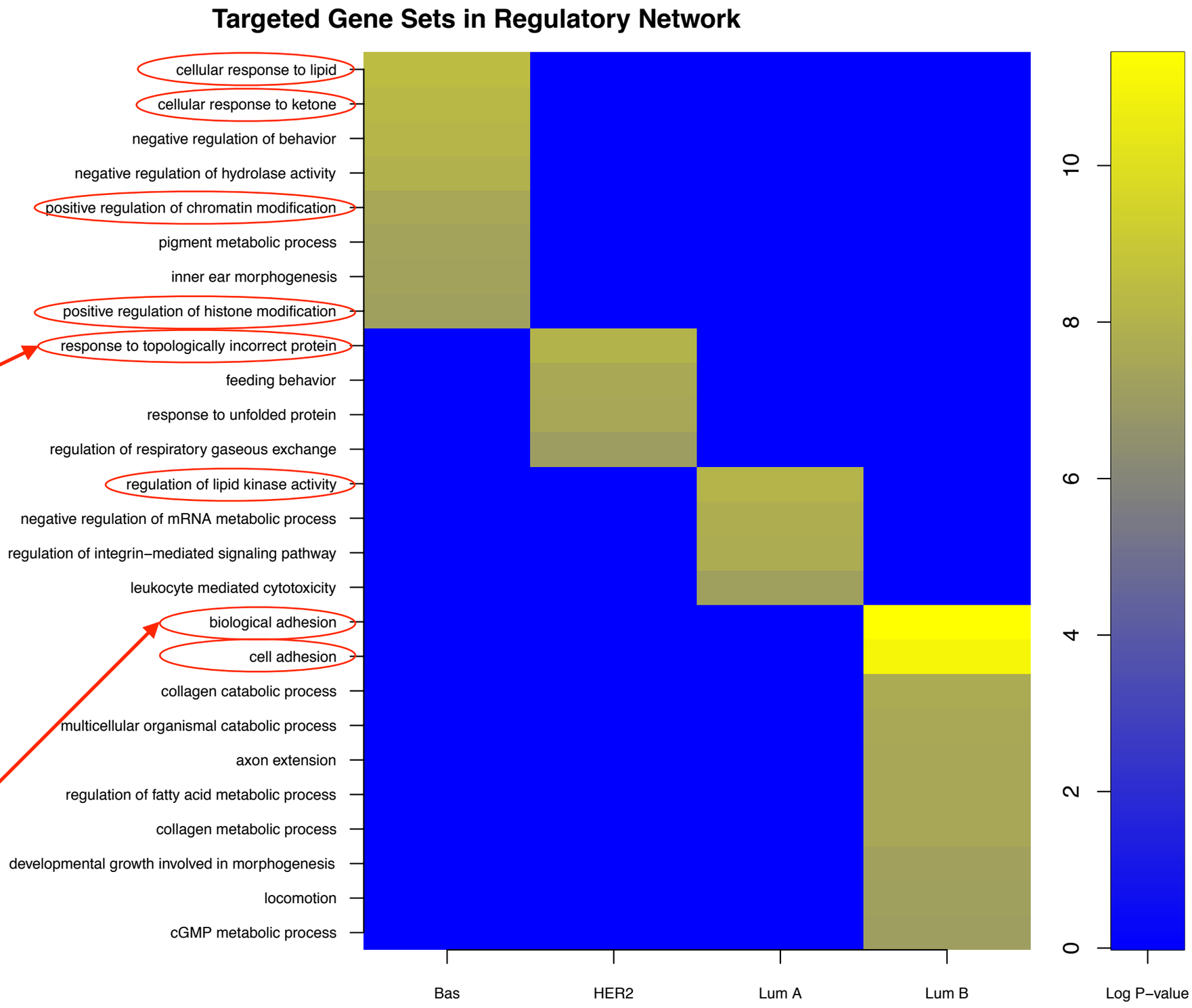
- Disassortative: protein targets many genes and genes are targeted by few proteins
- Assortative: protein targets fewer genes

- PANDA networks are smaller
  - We remove *noise* from motif data that is not corresponding to breast cancer



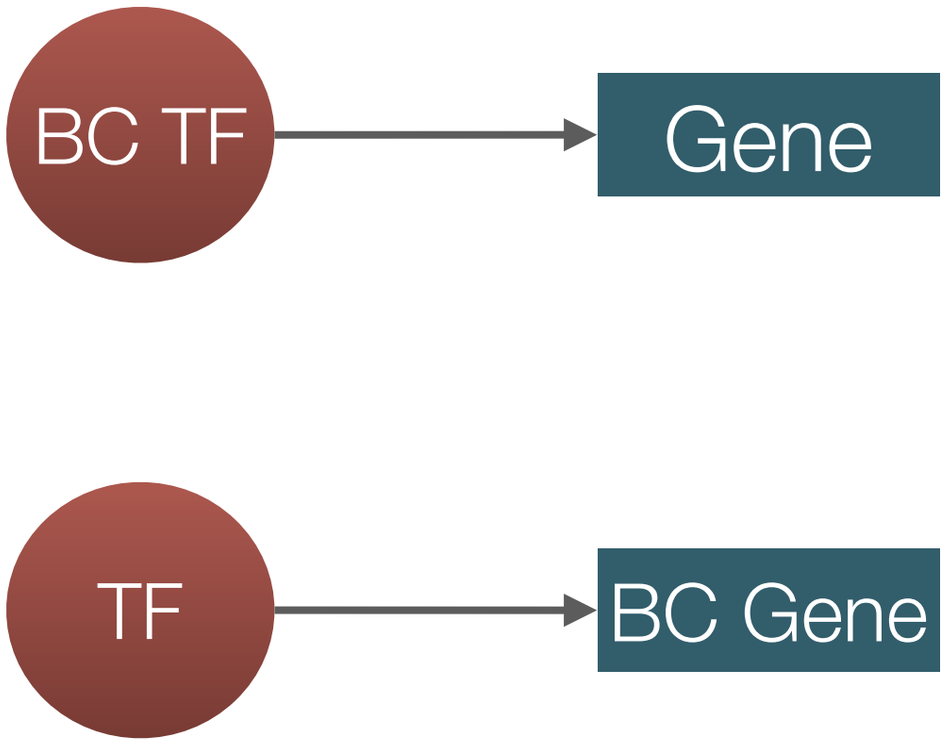
Uniquely targeted genes

- Gene Ontology Analysis
- Activation pathway renders traztuzumab treatment ineffective[4]
- Proliferation



Number of genes targeted by BC-TF in unique network

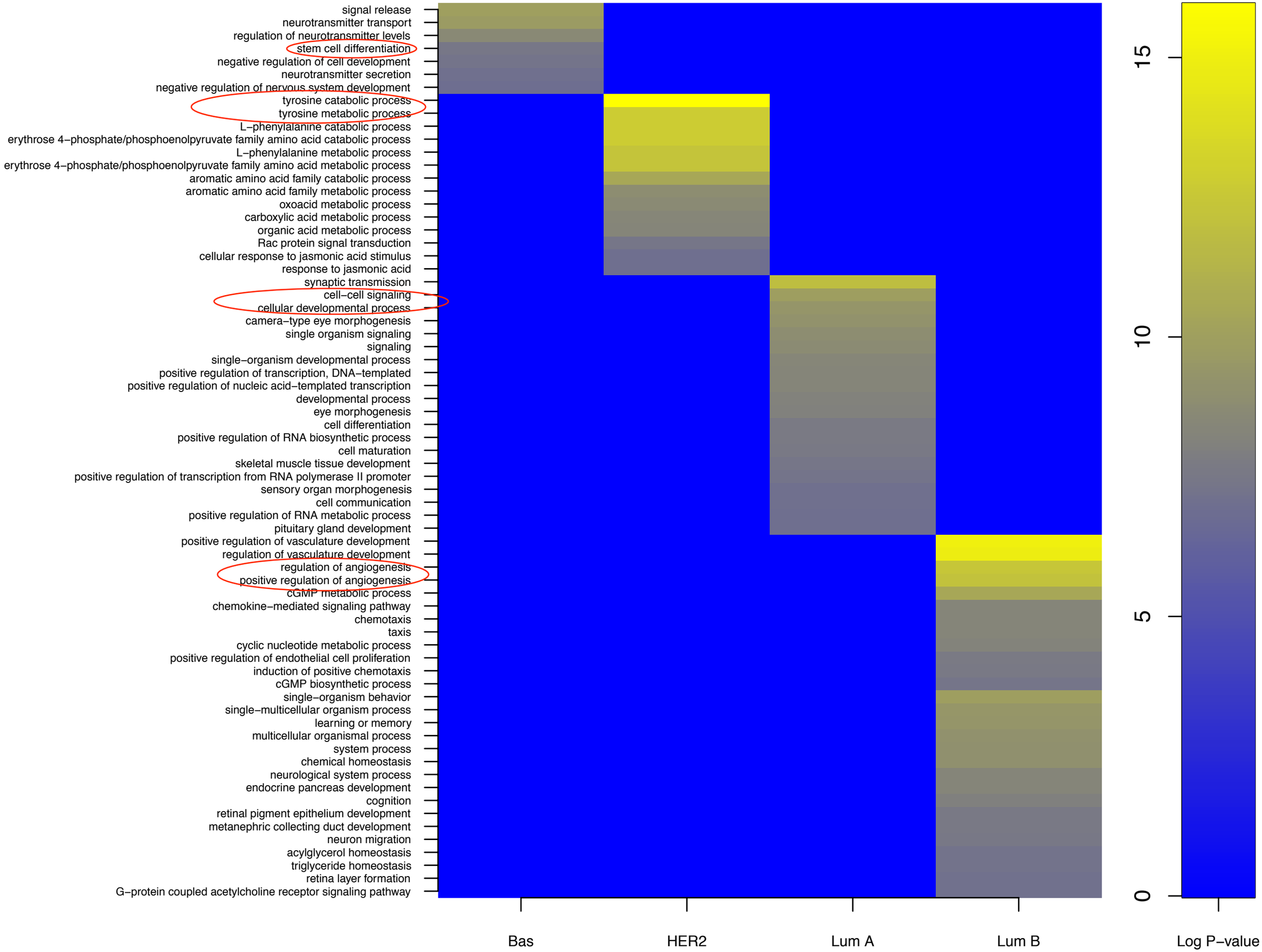
	Basal-like	HER2	Luminal A	Luminal B
AR	35	150	3	30
BRCA1	158	20	23	196
TP53	132	124	9	268
ESR1	438	648	634	454



Number of TF targeting BC-genes in unique network

	Basal-like	HER2	Luminal A	Luminal B
AR	0	1	2	2
ATM	3	0	0	1
BARD1	0	0	3	2
BRIP1	8	27	0	1
DIRAS	8	7	0	0
ERBB2	1	1	4	0
NBN	1	0	0	0
PALB2	0	1	0	10
RAD50	0	6	4	0
RAD51	0	7	3	1
BRCA1	4	1	9	2
BRCA2	0	0	0	3
CDH1	6	6	0	21
STK11	2	0	0	23
TP53	2	0	3	1
ESR1	2	0	13	0

Gene Sets Targeted by AR and ERH1 in Regulatory Network



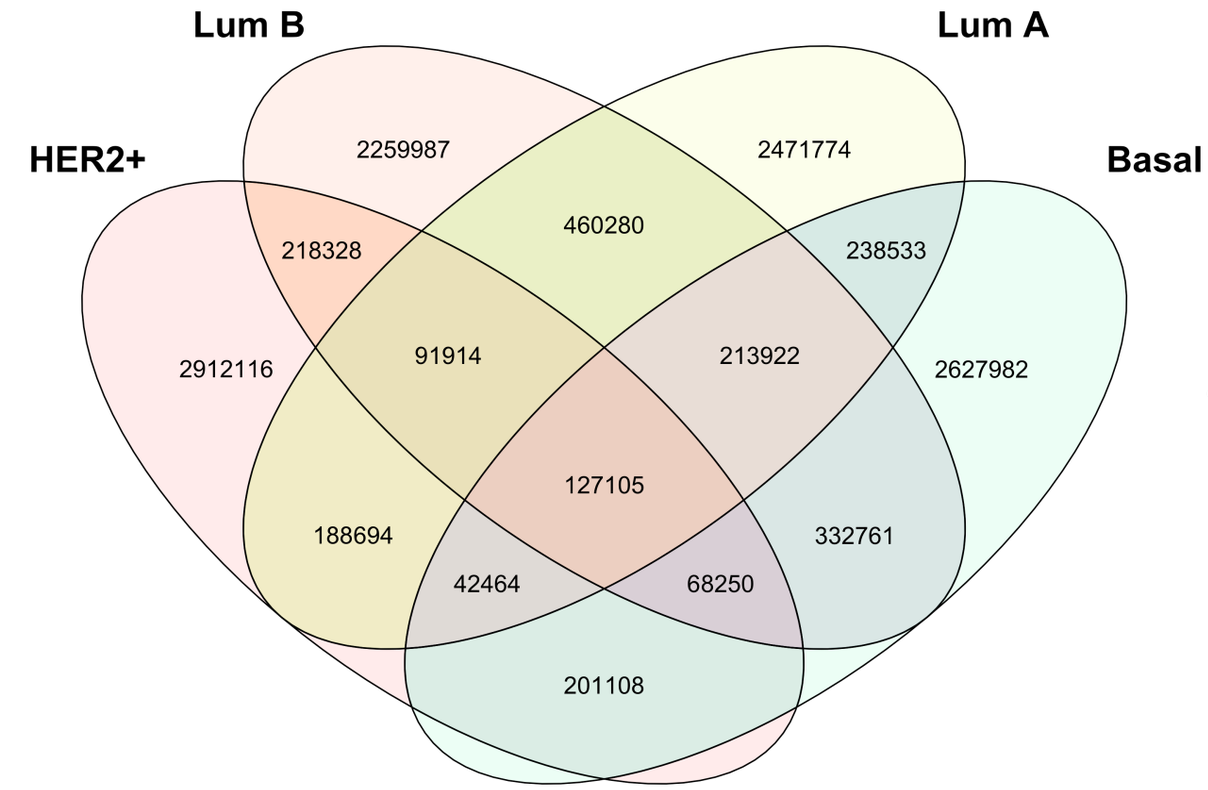
Observations of Regulatory Network

- *AR* uniquely targets 150 genes in the HER2 network
    - Previous study suggested *AR* target therapy in this subtype
- *ESR1* targets genes involved in stem cell differentiation in the Basal-like network
    - Basal-like is estrogen receptor negative, but gene may play important role

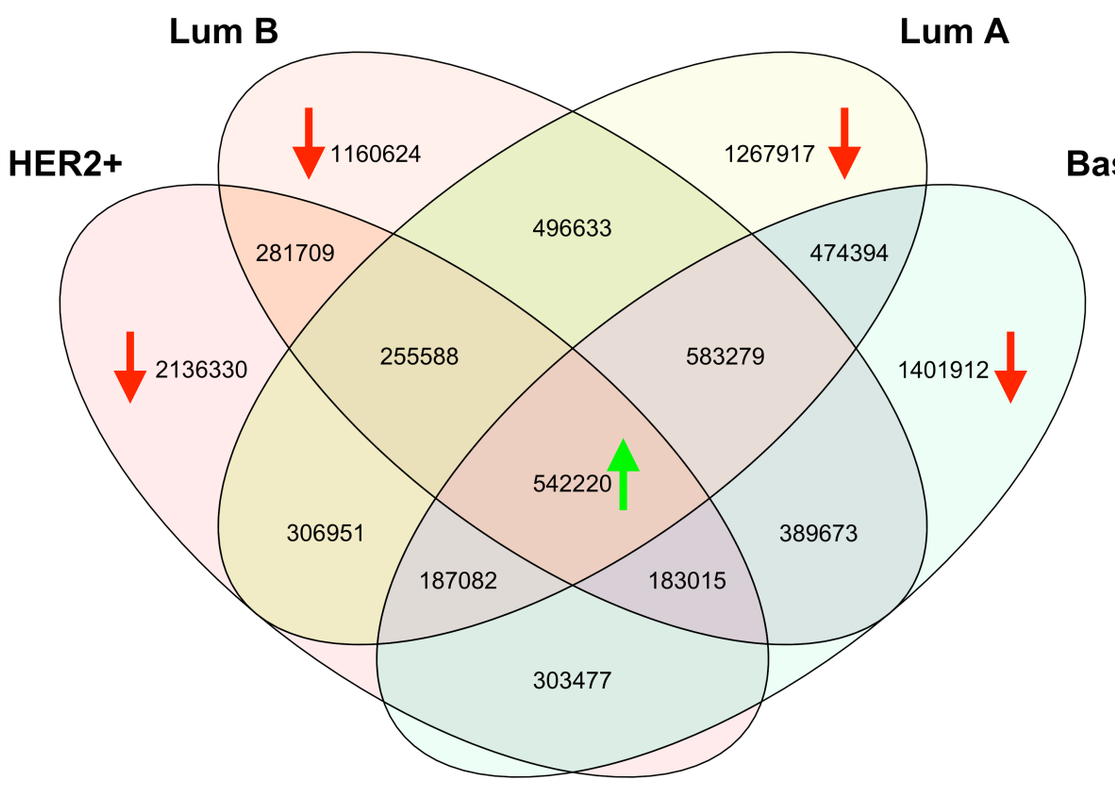
	Basal-like	HER2	Luminal A	Luminal B
AR	35	150	3	30
BRCA1	158	20	23	196
TP53	132	124	9	268
ESR1	438	648	634	454

Co-regulation Networks

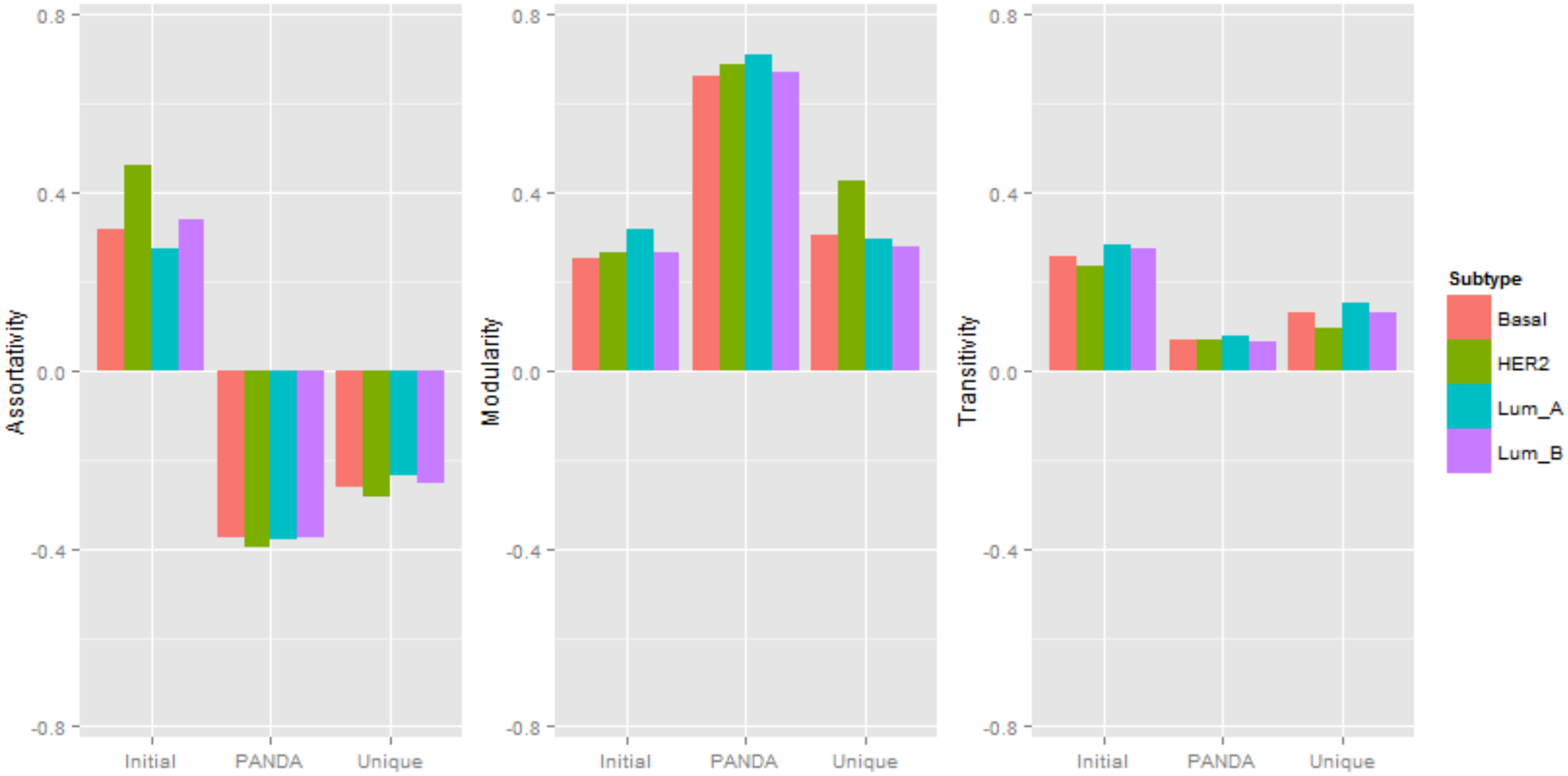
Number of Edges



PANDA



Network Structure

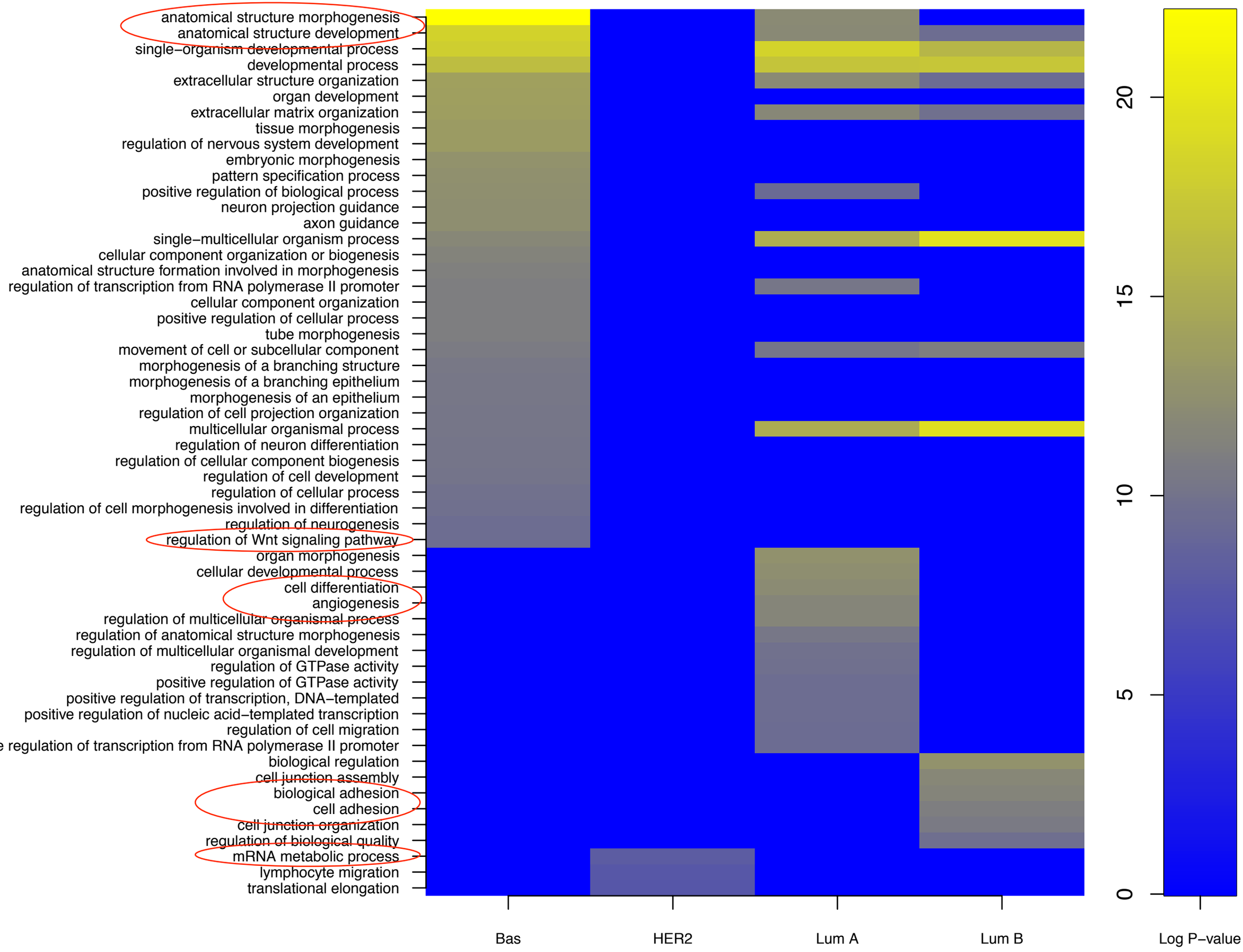


Looking at hubs:

- Within each subnetwork, we identified the genes with highest degrees (~100 genes in each subnetwork) that most probably form hubs

- Analyzed hubs by running Gene Ontology Analysis on genes that are co-regulated with those high-degree genes

Final Network



Conclusions

- Basal-like networks show enriched sets in *Wnt* signaling pathway and stem cell differentiation
- HER2
    - *AR* protein targets many genes involved in metabolism
- Luminal A
    - *SMAD* proteins are of interest
- Luminal B
    - biological and cell adhesion enrichment
    - *RXRA* is a protein of interest

Limitations

- PANDA returns edge weights in z-scores. We took the upper 5% scores and ignored the rest
    - We repeated this with other cut-offs and saw similar results

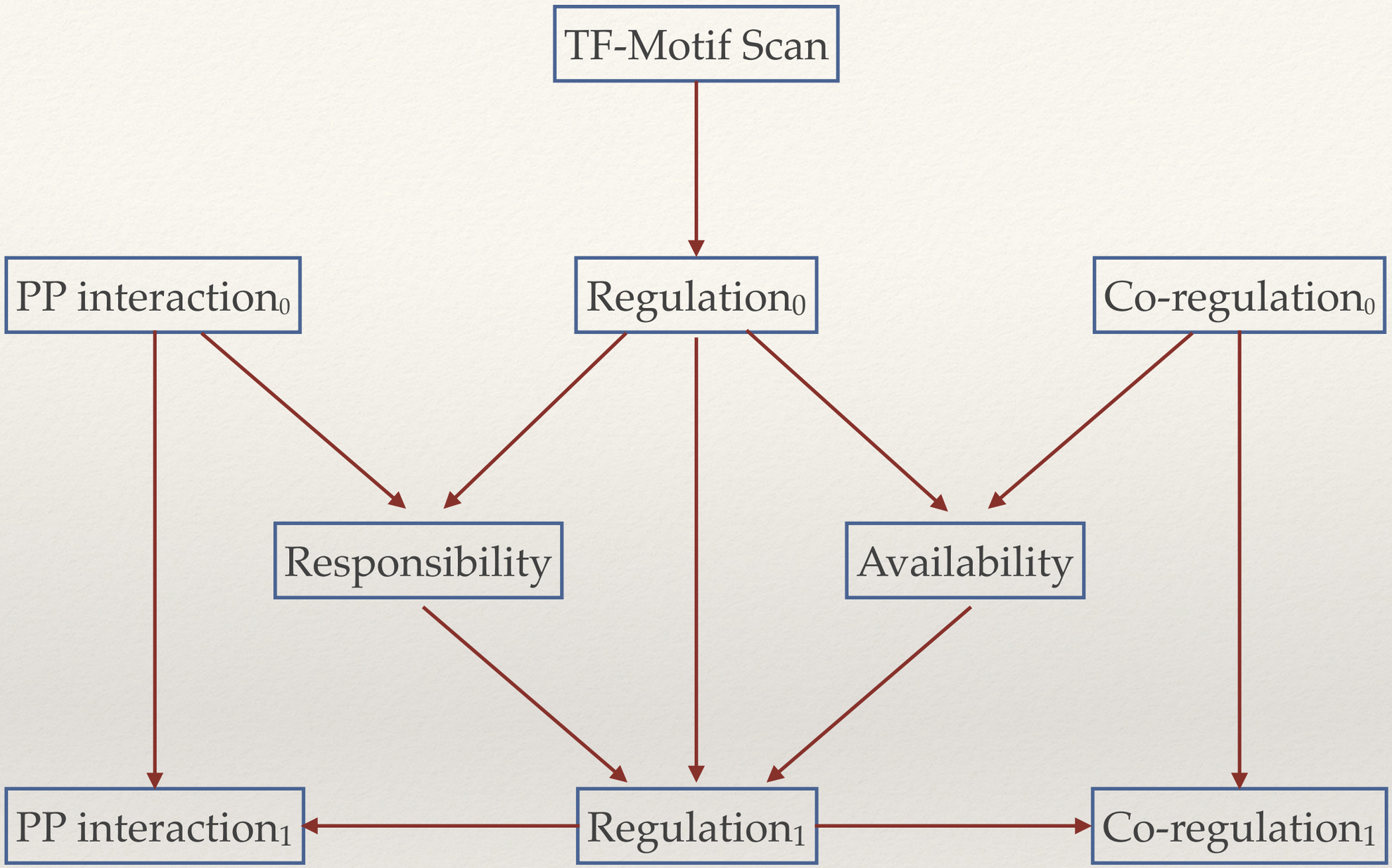
Sources

[1]Glass K, Huttenhower C, Quackenbush J, Yuan G-C (2013) “Passing Messages between Biological Networks to Refine Predicted Interactions”. PLoS ONE 8(5): e64832. doi: 10.1371/journal.pone.0064832

[2]Ravasi, T; Suzuki, H; Cannistraci, et al. (2010). "An atlas of combinatorial transcriptional regulation in mouse and man". Cell 140 (5): 744–52. doi:10.1016/j.cell.2010.01.044

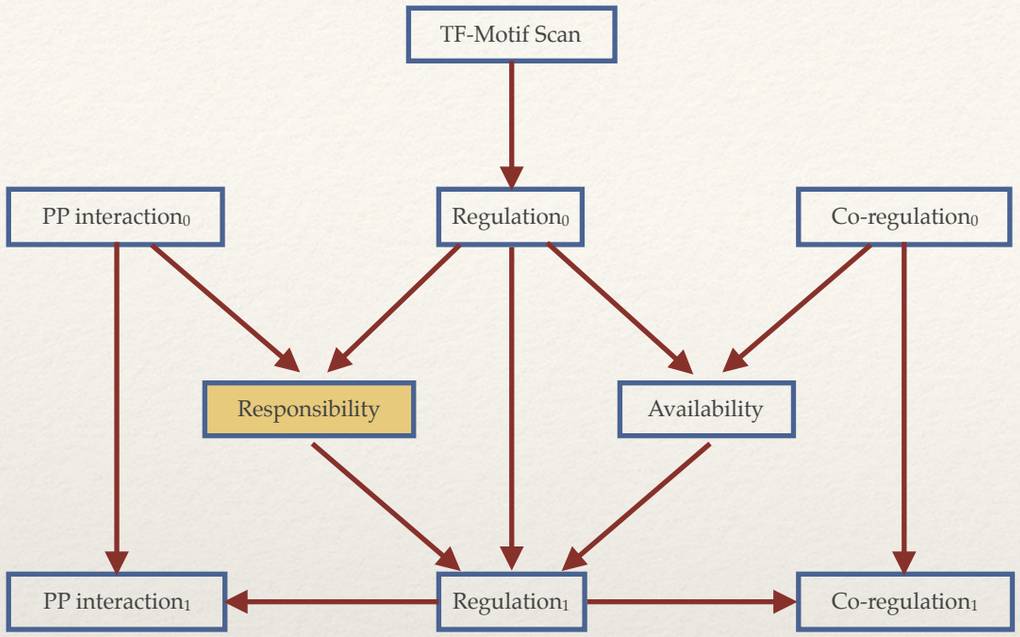
[3]Crowe D.L, Chandraratna R.A. A retinoid X receptor (RXR)-selective retinoid reveals that RXR-alpha is potentially a therapeutic target in breast cancer cell lines, and that it potentiates antiproliferative and apoptotic responses to peroxisome proliferator-activated receptor ligands. Breast Cancer Res. 2004;6(5):R546–R555. doi: 10.1186/bcr913

[4] Kumandan, Sreekanth et al. Activation of the unfolded protein response bypasses trastuzumab-mediated inhibition of the PI-3K pathway. Cancer Letters , Volume 329 , Issue 2 , 236 - 242



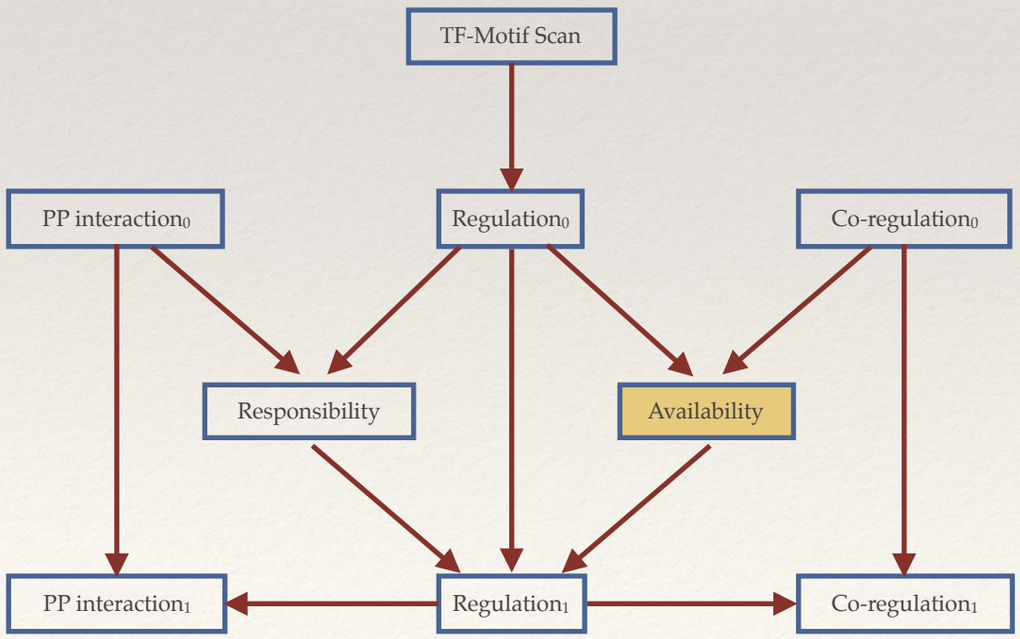
PANDA estimates the probability that an edge exists in a network and returns that estimate in terms of Z-score units

- *Responsibility(R_{ij}):* information flowing from TF *i* to gene *j*



$$R_{ij}^{(t)} = \frac{\sum_m P_{im}^{(t)} W_{mj}^{(t)}}{\sqrt{\sum_m \left(P_{im}^{(t)}\right)^2 + \sum_m W_{mj}^{(t)^2} - \left|\sum_m P_{im}^{(t)} W_{mj}^{(t)}\right|}}$$

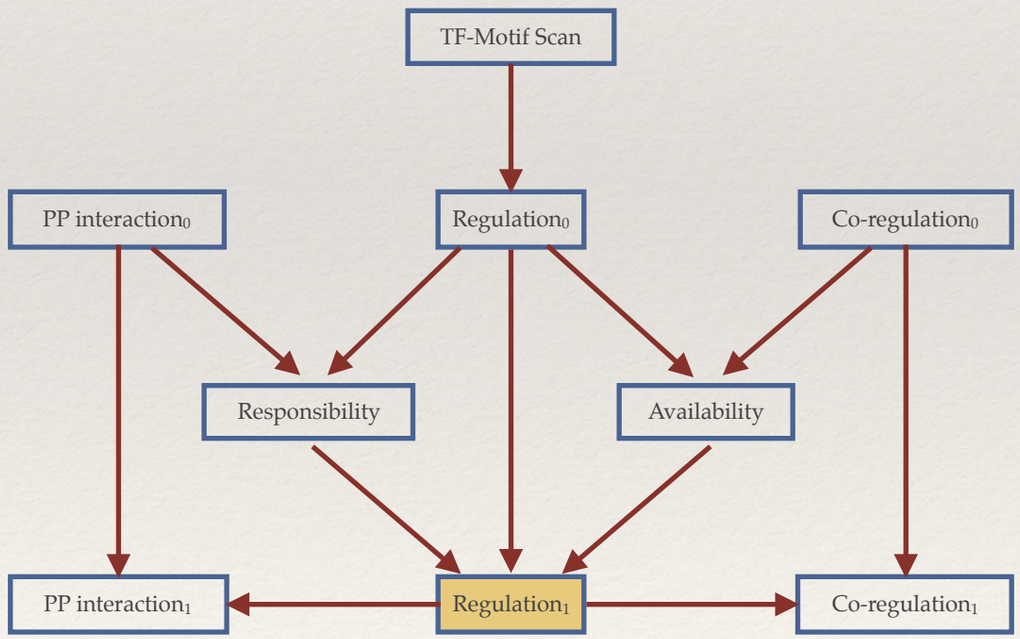
- *Availability (A_{ij}):* information flowing from gene *j* to TF *i*



$$A_{ij}^{(t)} = \frac{\sum_k W_{ik}^{(t)} C_{kj}^{(t)}}{\sqrt{\sum_k \left(W_{ik}^{(t)}\right)^2 + \sum_k \left(C_{kj}^{(t)}\right)^2 - \left|\sum_k W_{ik}^{(t)} C_{kj}^{(t)}\right|}}$$

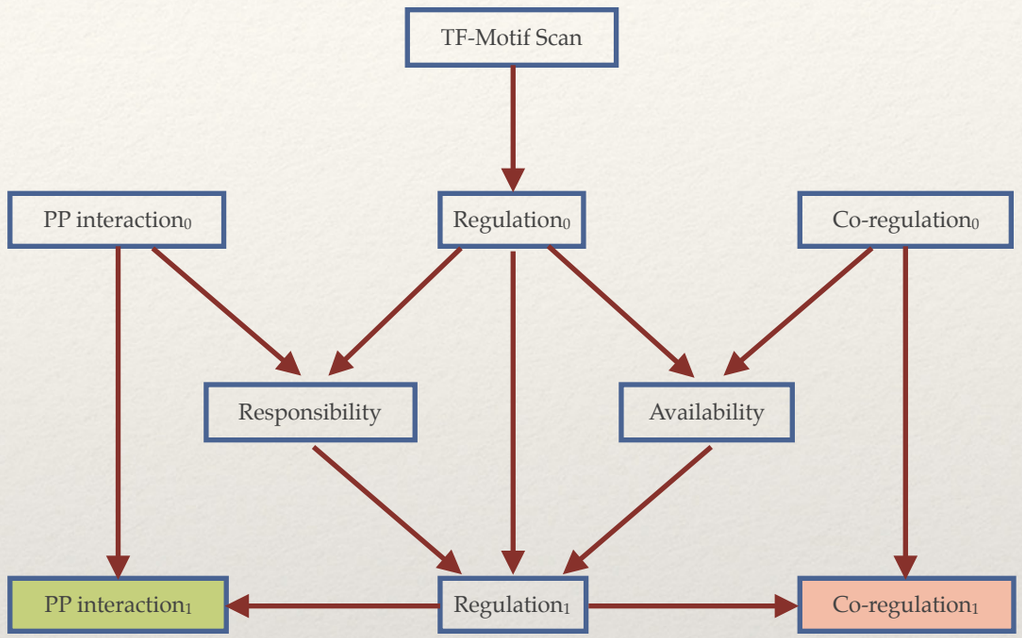
- Since regulation requires both that TF is responsible for the regulation of a certain gene and that gene to be available for regulation by that TF:

$$\tilde{W}_{ij}^{(t)} = \frac{A_{ij}^{(t)} + R_{ij}^{(t)}}{2} \qquad W_{ij}^{(t+1)} = (1-\alpha)\, W_{ij}^{(t)} + \alpha \tilde{W}_{ij}^{(t)}$$



- Cooperation between TF's *i* and *m*:

$$\tilde{P}_{im}^{(t)} = \frac{\sum_j W_{ij}^{(t)} W_{mj}^{(t)}}{\sqrt{\sum_j \left(W_{ij}^{(t)}\right)^2 + \sum_j \left(W_{mj}^{(t)}\right)^2 - \left|\sum_j W_{ij}^{(t)} W_{mj}^{(t)}\right|}}$$



- Co-regulated genes *j* and *k*:

$$\tilde{C}_{kj}^{(t)} = \frac{\sum_i W_{ik}^{(t)} W_{ij}^{(t)}}{\sqrt{\sum_i \left(W_{ik}^{(t)}\right)^2 + \sum_i \left(W_{ij}^{(t)}\right)^2 - \left|\sum_i W_{ik}^{(t)} W_{ij}^{(t)}\right|}}$$

- Update matrices with *update parameter*:

$$P_{im}^{(t+1)} = (1-\alpha)\, P_{im}^{(t)} + \alpha \tilde{P}_{im}^{(t)} \qquad C_{jk}^{(t+1)} = (1-\alpha)\, C_{jk}^{(t)} + \alpha \tilde{C}_{jk}^{(t)}$$

Initial Network