

Differential Transcription Factor analysis

PPG Meeting
5/1/15

Background

- Transcription factors may behave in different ways in different contexts.
- These changes in “involvement” may not be readily observed using standard differential gene expression analyses.

Goal

- Generate context specific networks and identify structural changes between them.

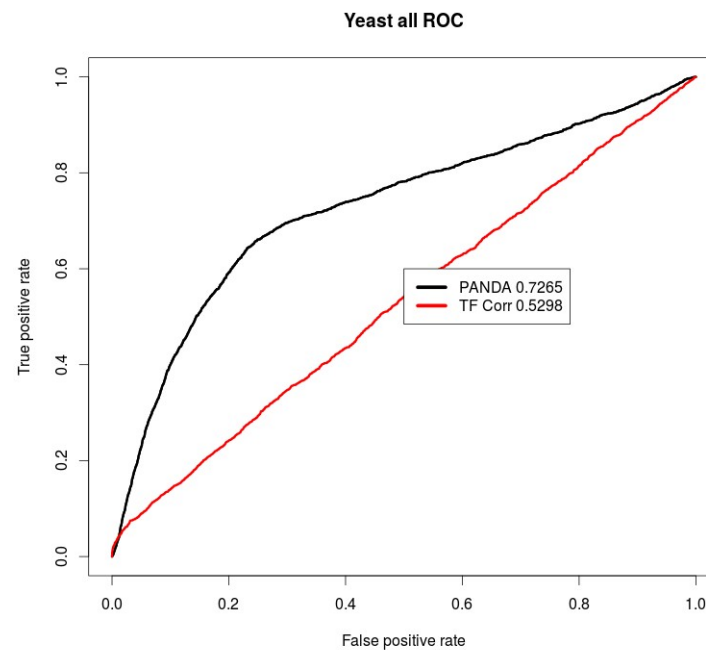
Challenges

- Ultra-high dimensionality for network inference.
- Current network inference methods yield relatively poorly predictive edgeweights at the individual interaction level.
- Comparison of two networks involves the comparison of millions of noisy edges.

Challenges

Prediction of TF-gene regulation.

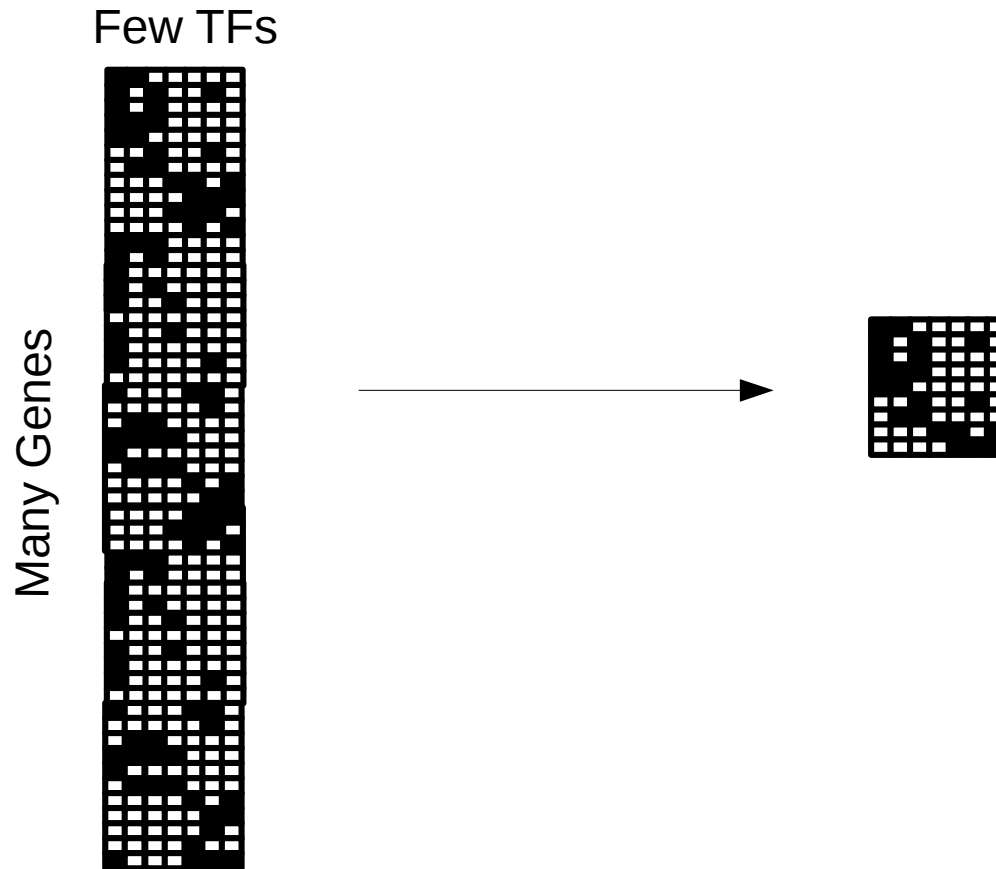
A comparison of gene expression correlation to PANDA algorithm.



Conclusion: Using gene expression patterns to discern differential transcription factor involvement may not be a sufficient approach.

Approach

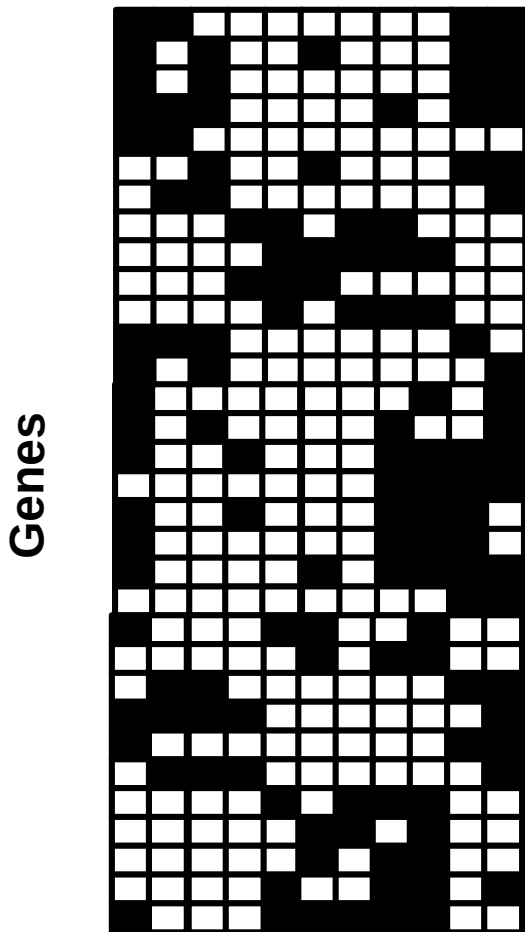
- We can view the problem as a dimension reduction problem.



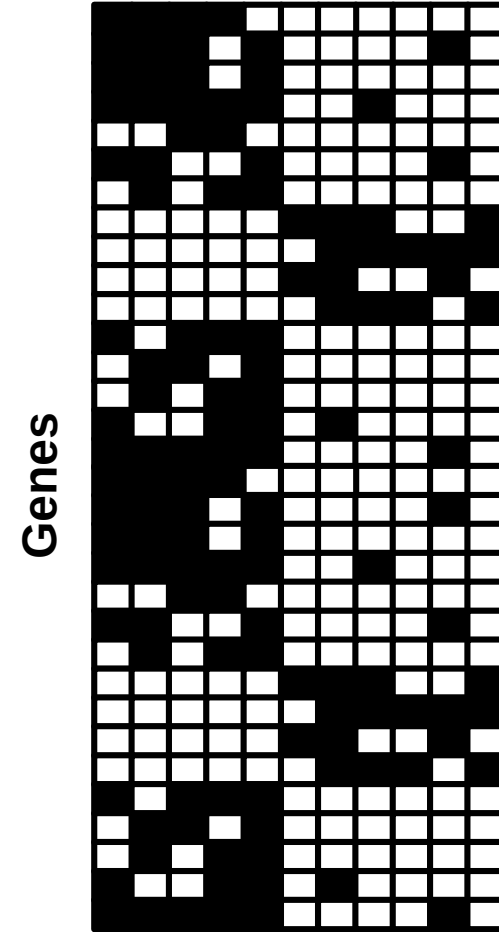
Approach

- Consider two adjacency matrices...

TFs



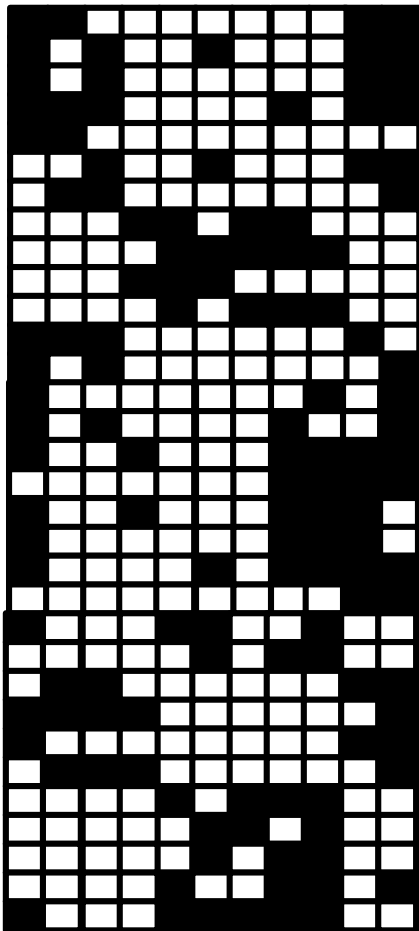
TFs



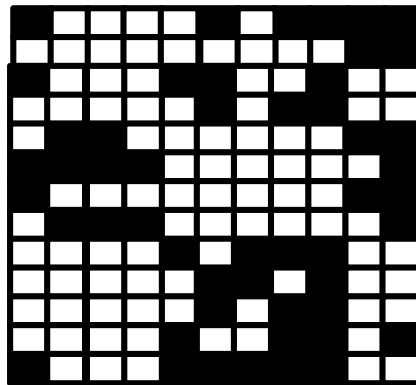
Approach

- Consider two adjacency matrices...

Smoker Control



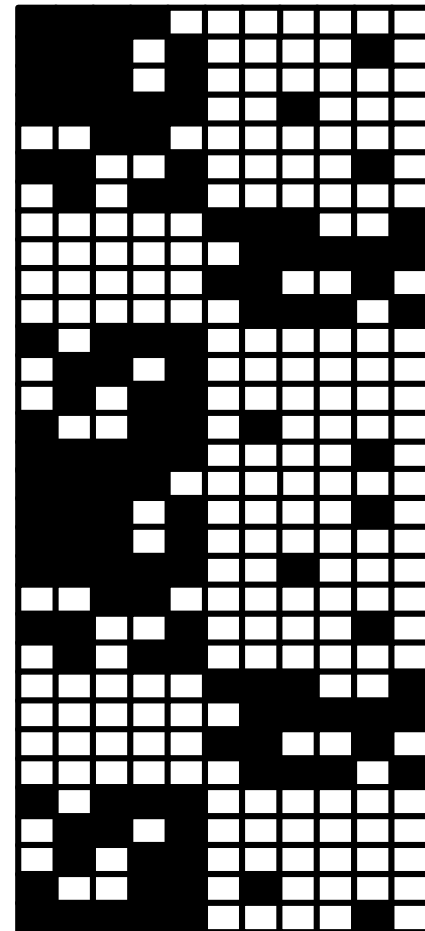
T



\times

\sim

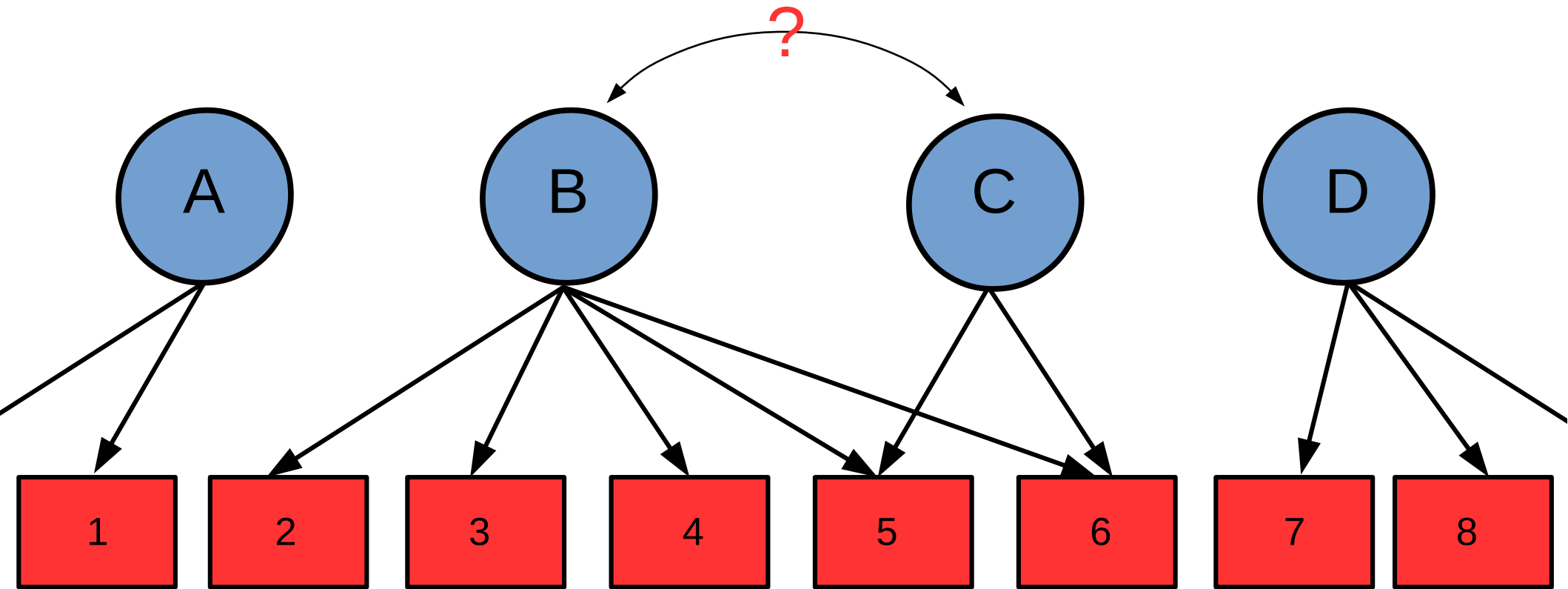
COPD



The Transition Matrix

- Calculated via penalized regression.
- Each column in the TM can be thought of as being the best linear combination of columns in the control AM that “create” the columns in the COPD.

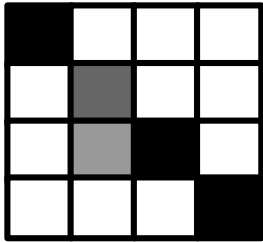
An Example



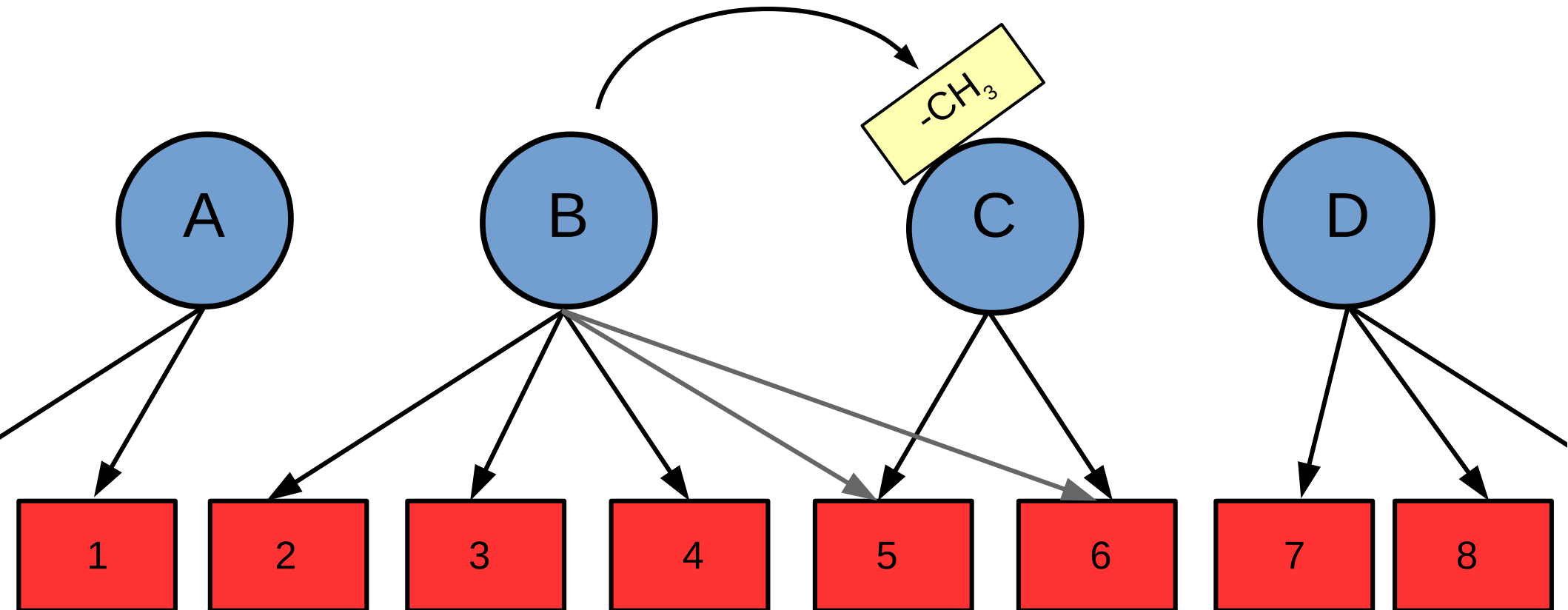
TM =

	A	B	C	D
A				
B				
C				
D				

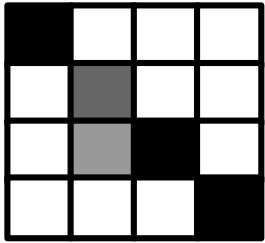
Biological Mechanism?



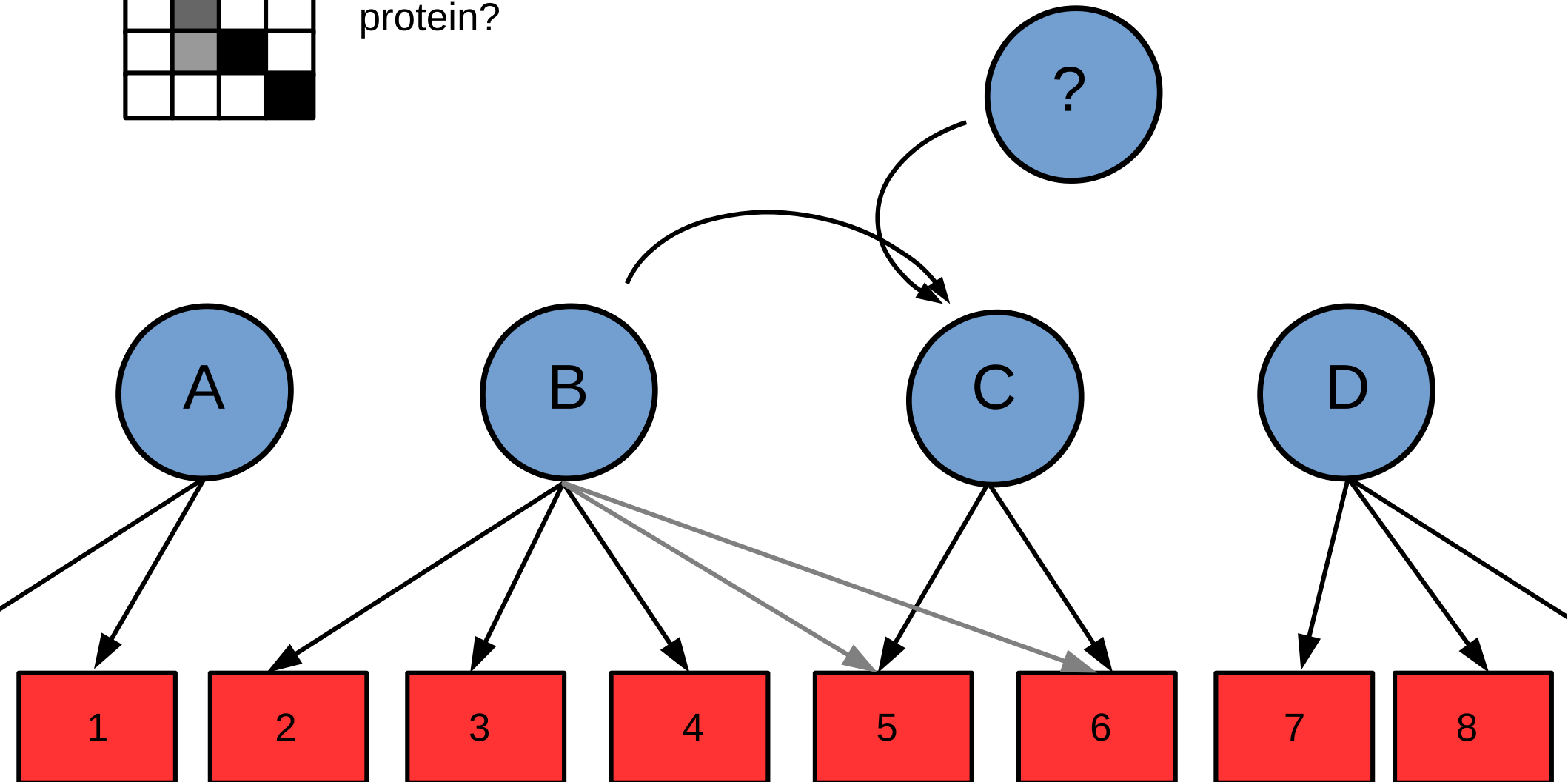
Suggested mechanism #1:
Differential methylation of the gene for TF C?

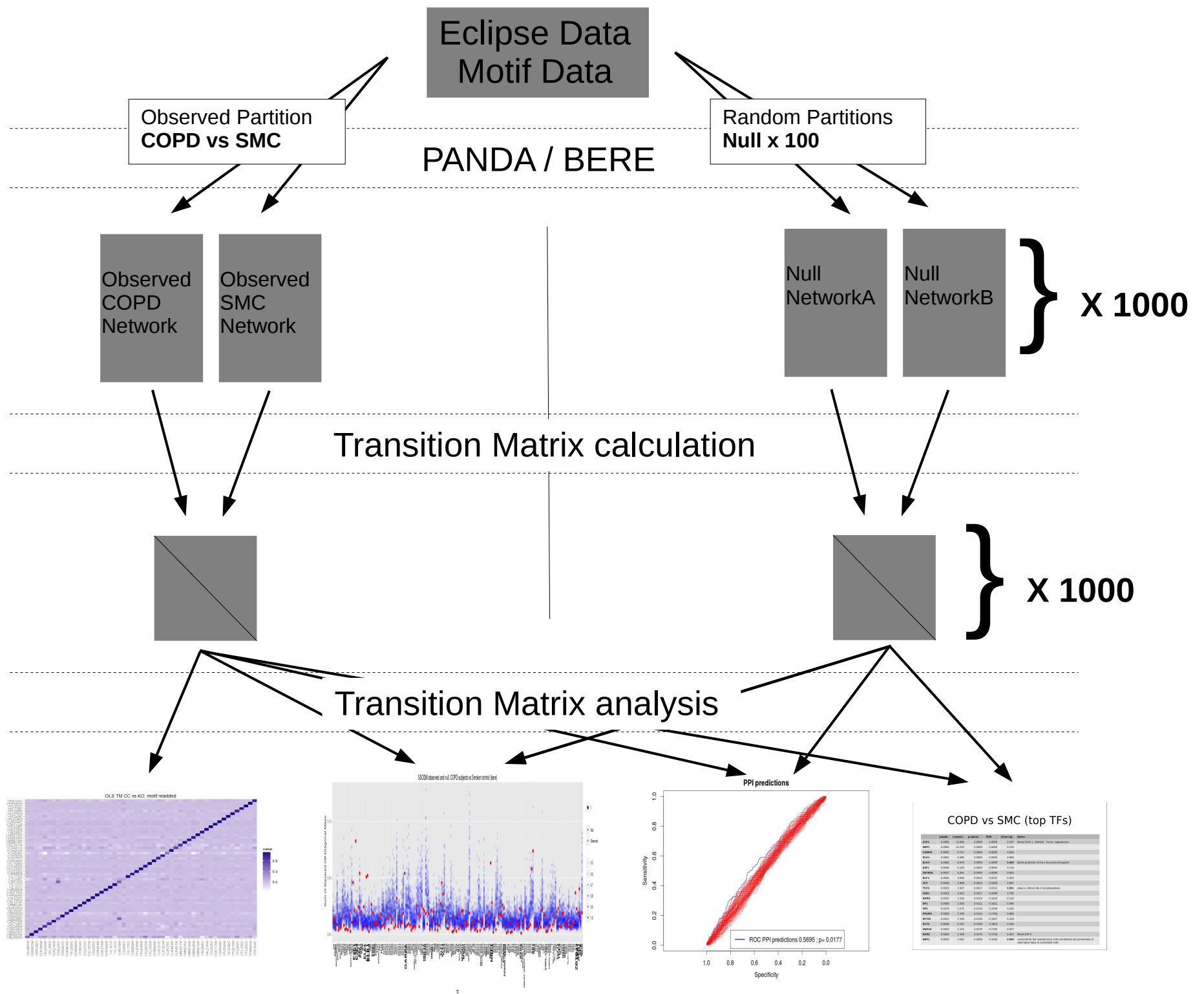


Biological Mechanism?



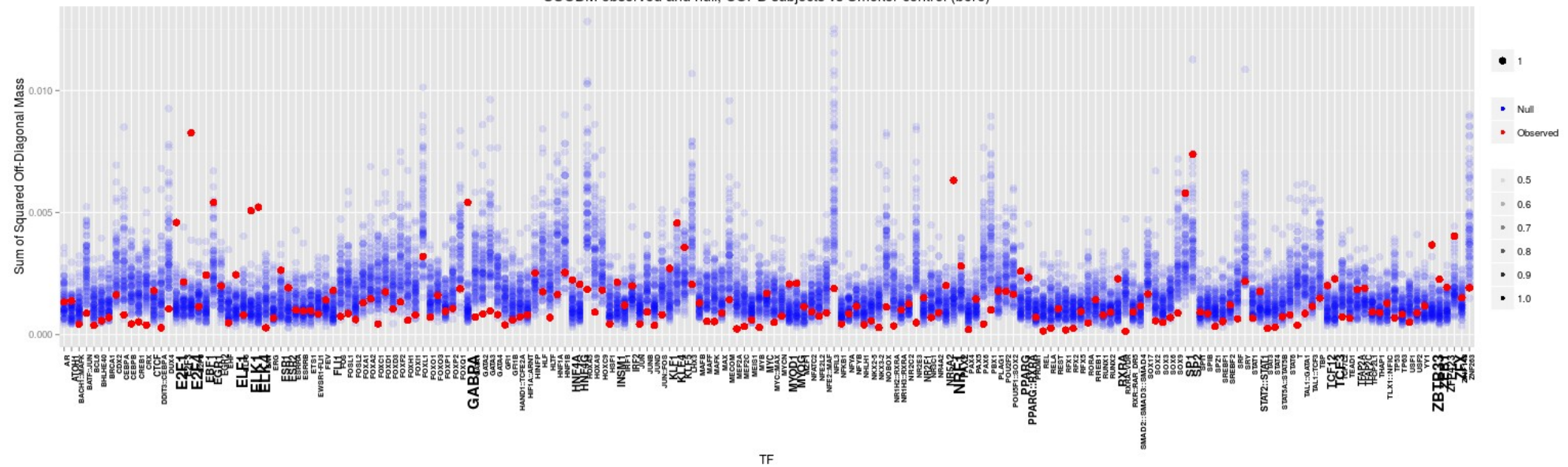
Suggested mechanism #2:
Protein complex of B-C-? with unknown 3rd
protein?



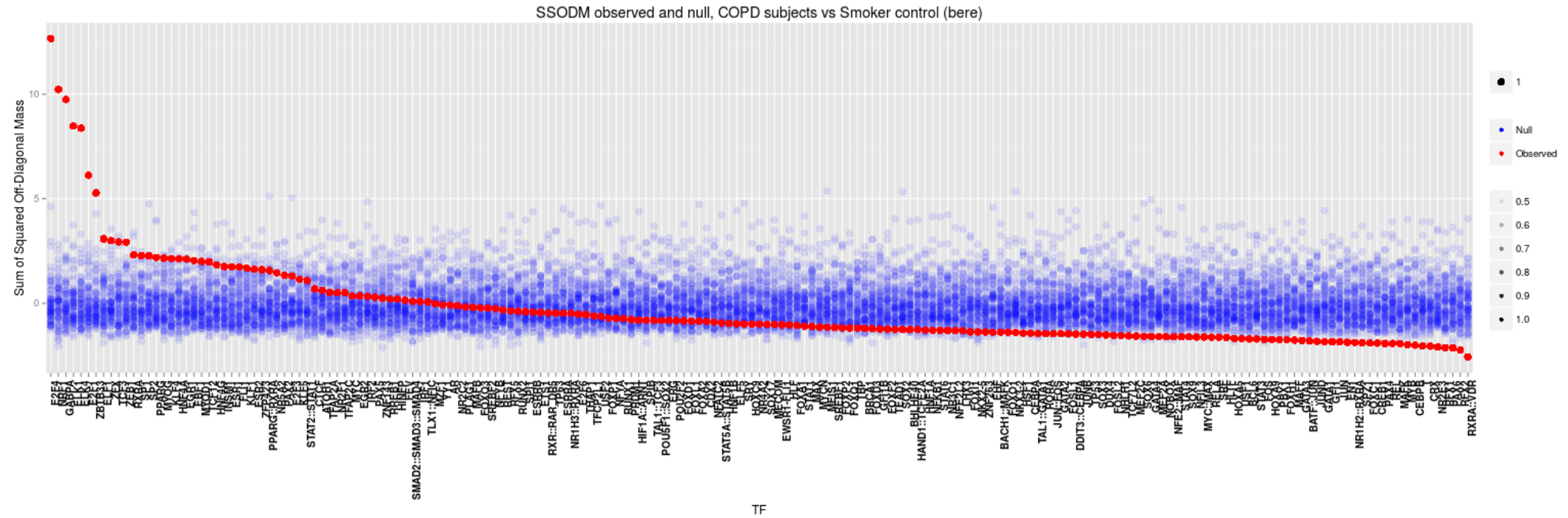


COPD vs SMC

SSODM observed and null, COPD subjects vs Smoker control (bere)



SSODM observed and null, COPD subjects vs Smoker control (bere)



COPD vs SMC (top 20 TFs)

	ssodm	t.values	p.values	FDR	Sig (LIMMA)	Notes
E2F4	0.0083	12.666	0.0000	0.0000	0.337	Binds EGR-1, SMAD3. Tumor suppression.
NRF1	0.0063	10.232	0.0000	0.0000	0.215	acts on nuclear genes encoding respiratory subunits and components of the mitochondrial transcription and replication machinery.
GABPA	0.0054	9.747	0.0000	0.0000	0.816	Related to NRF1, involved in activation of cytochrome oxidase expression and nuclear control of mitochondrial function
ELK1	0.0051	8.480	0.0000	0.0000	0.080	
ELK4	0.0052	8.379	0.0000	0.0000	0.000	Binds promoter of the c-fos proto-oncogene
E2F1	0.0046	6.126	0.0000	0.0000	0.714	E2F family...
ZBTB33	0.0037	5.281	0.0000	0.0000	0.602	shown to interact with HDAC3, Nuclear receptor co-repressor 1
ELF1	0.0025	3.083	0.0010	0.0242	0.301	primarily expressed in lymphoid cells
ZFX	0.0040	2.998	0.0014	0.0285	0.987	
TCF3	0.0023	2.937	0.0017	0.0313	0.001	plays a critical role in lymphopoiesis
ZEB1	0.0023	2.923	0.0017	0.0298	0.785	
RXRA	0.0023	2.316	0.0103	0.1618	0.216	
SP1	0.0058	2.285	0.0111	0.1621	0.385	
SP2	0.0074	2.275	0.0115	0.1546	0.242	
PPARG	0.0026	2.193	0.0142	0.1784	0.866	
MYOG	0.0021	2.156	0.0156	0.1837	0.416	
KLF4	0.0046	2.137	0.0163	0.1814	0.154	
HNF4A	0.0022	2.121	0.0170	0.1782	0.927	
EGR1	0.0054	2.109	0.0175	0.1740	0.497	Binds E2F4
EBF1	0.0024	2.042	0.0206	0.1945	0.000	essential for the maintenance of B cell identity and prevention of alternative fates in committed cells