

Coexpression and Collaborative Gene Networks

Hairong Wei

Michigan Technological University

References

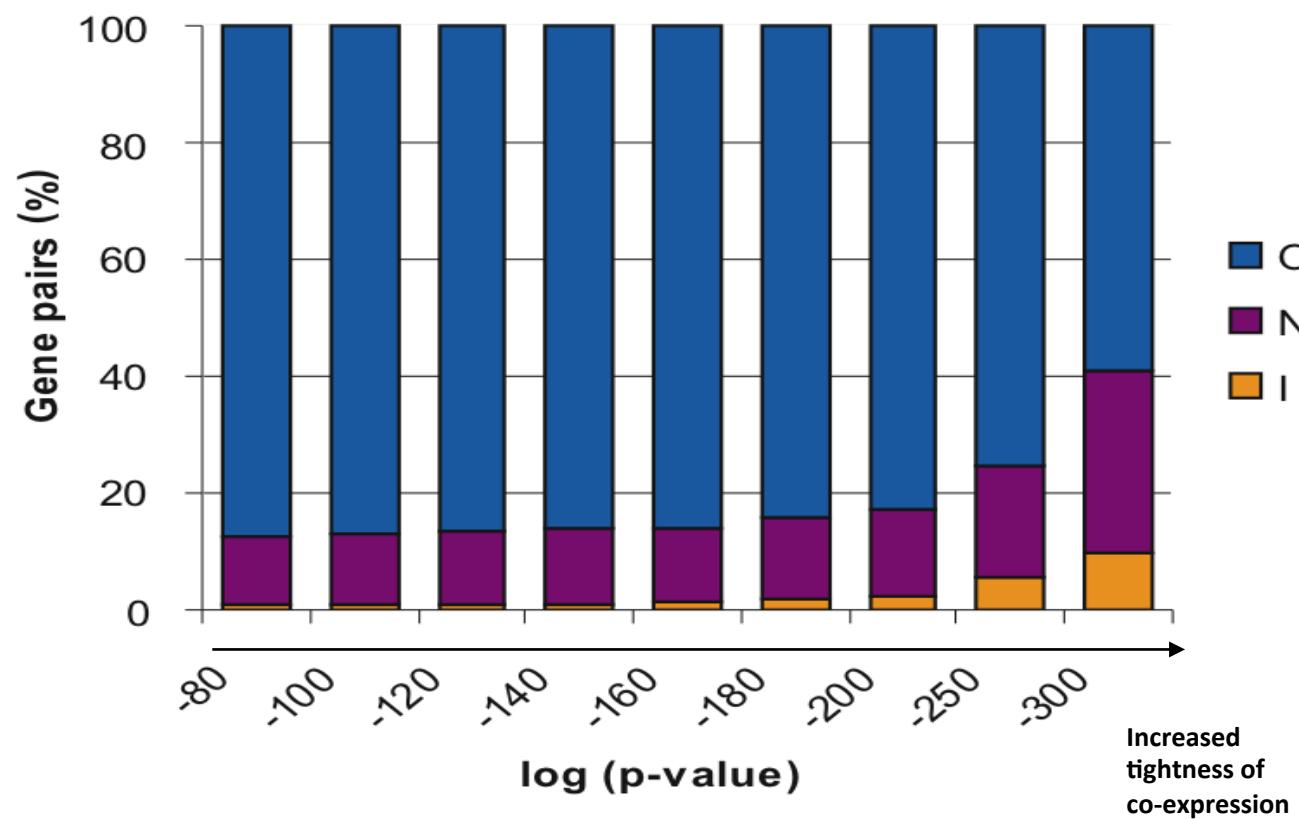
- 1) Persson, S., **H. Wei**, J. Milne, G. Page, C. Somerville. 2005. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc Natl Acad Sci USA*, 102: 8633-8638. (Faculty 1000 evaluation)
- 2) **Wei, H.** S. Persson, T. Mehta, V. Srinivasasainagendra, L. Chen, G. Page, C. Somerville, A. Loraine. 2006. Transcriptional coordination of the metabolic network in *Arabidopsis thaliana*. *Plant Physiology*. 142(2):762-74.
- 3) Nie, J., R. Stewart, F. Ruan, J. Thomson, H. Zhang, X. Cui and **H. Wei**. 2011. TF-Cluster: a pipeline for identifying functionally coordinated transcription factors via network decomposition of the shared coexpression connectivity matrix (SCCM). *BMC Systems Biology*, 5:53.
- 4) Ji, X. S. Chen, J. Li, W. Deng, Z. Wei and **H. Wei**. 2017. SSGA and MSGA: two seed-growing algorithms for constructing collaborative subnetworks. *Scientific Reports*. (in press)

What is coexpression network ?

A **gene co-expression network (GCN)** is an undirected graph, where each node corresponds to a gene, and a pair of nodes is connected with an edge if there is a significant co-expression relationship between them.

Are Genes involved in the same biological pathway coexpressed?

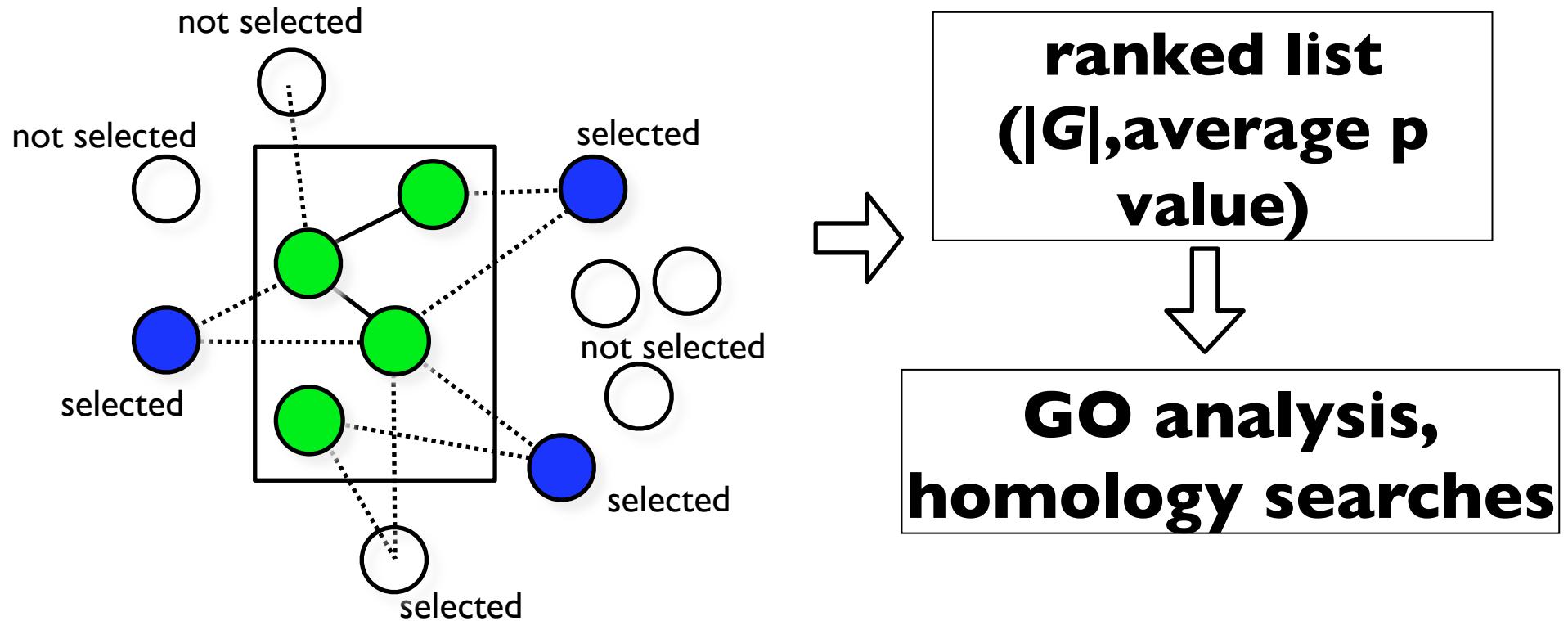
Hypothesis: The genes having functional association, e.g. in the same pathway, are more tightly co-expressed.



216 biological pathways that contain 1330 genes in Arabidopsis

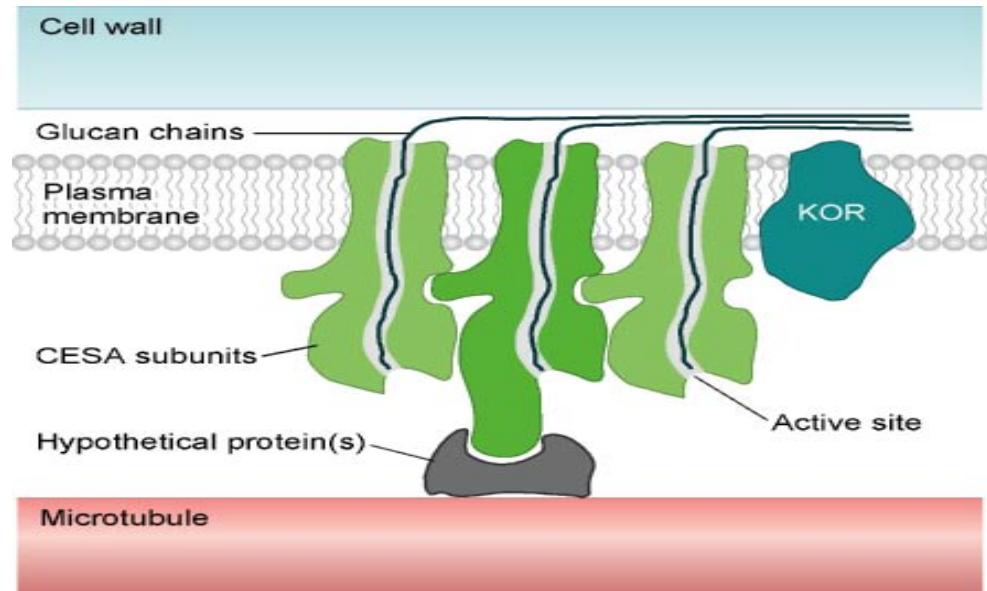
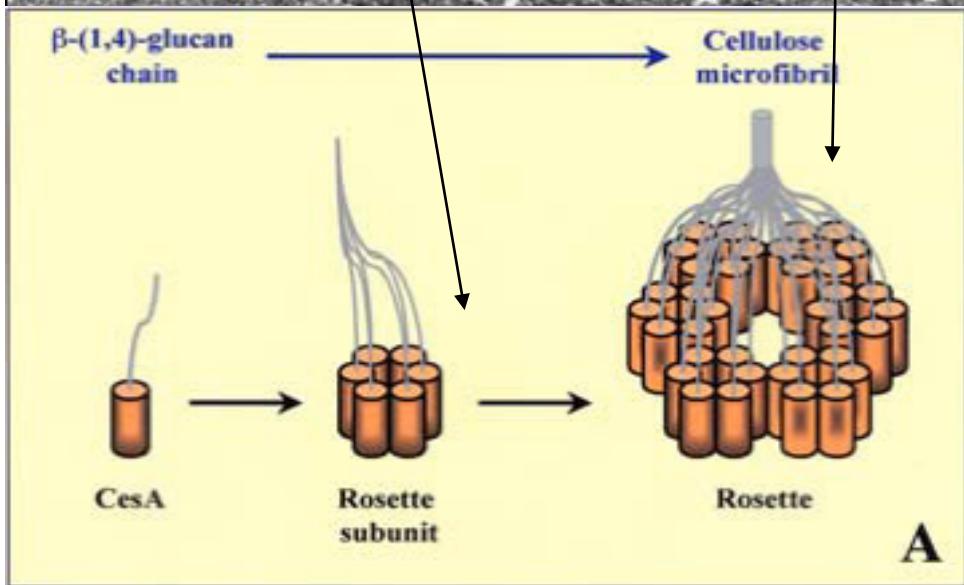
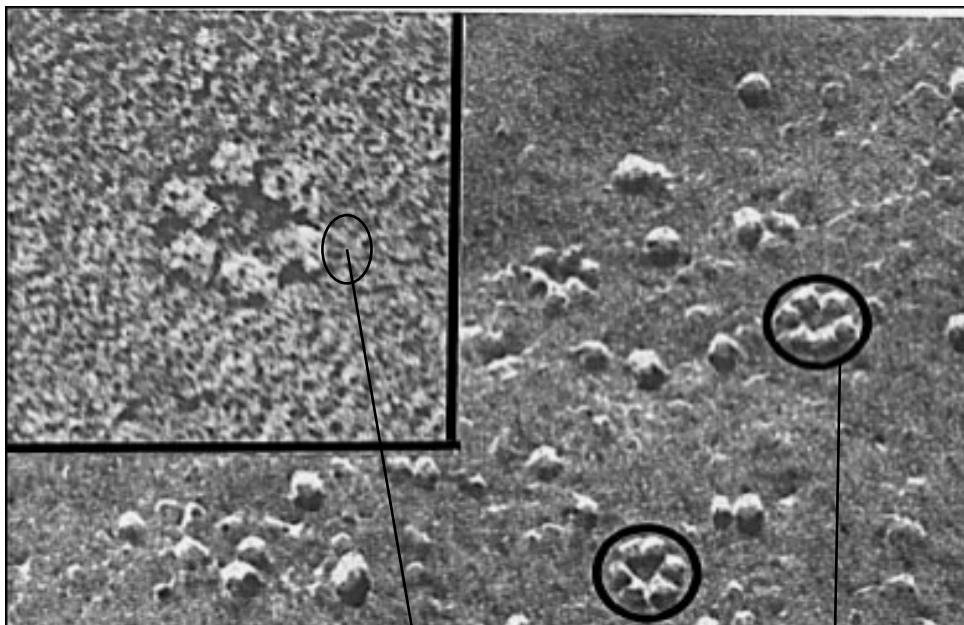
O-outside
N-different pathway
I-same pathway

Intersection of Co-Expression (ICE)



1. Define co-expression network.
2. Select a group of “bait” genes. (e.g., a pathway, a complex). Green nodes
3. Perform genome-wide co-expression analysis (correlation or regression).
3. Select genes outside the group of baits that are linked with >1 group members G.
4. Sort them by $|G|$, then average p value (or r^2)

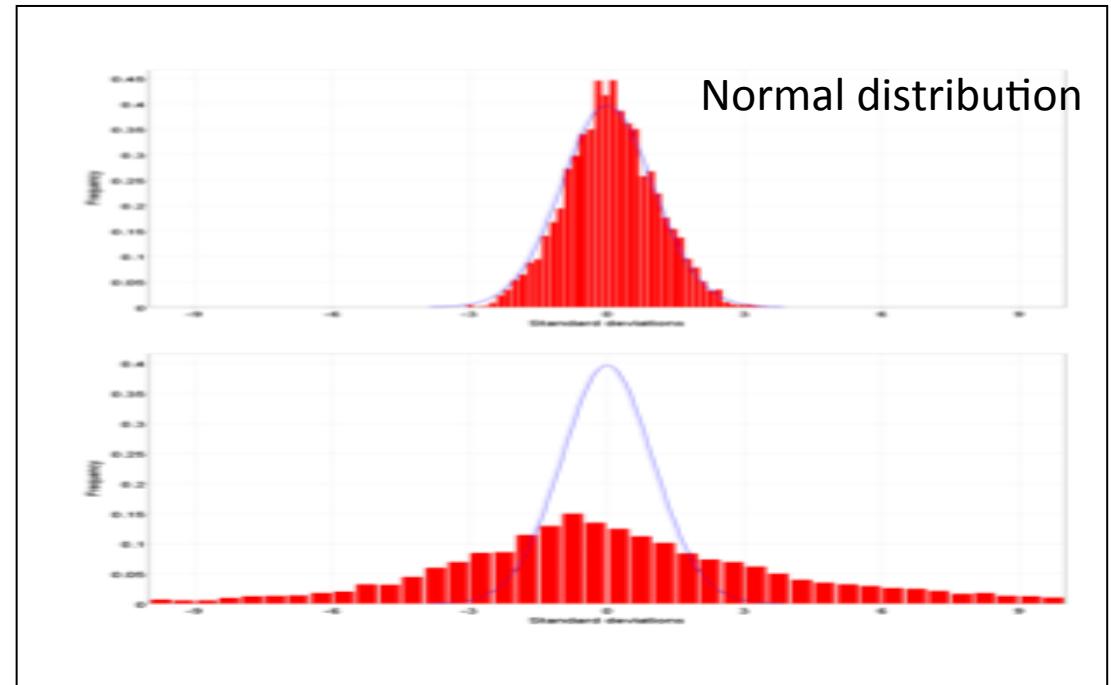
ICE Implementation 1: finding functionally associated genes



A very challenging problem:

What genes in addition to
CEAS genes (4, 7 and 8) are required
for secondary cell wall synthesis?

Primary: CESA1,3,6 Secondary: CESA4, 7, 8



For n genes and m data sets,
deleted residual

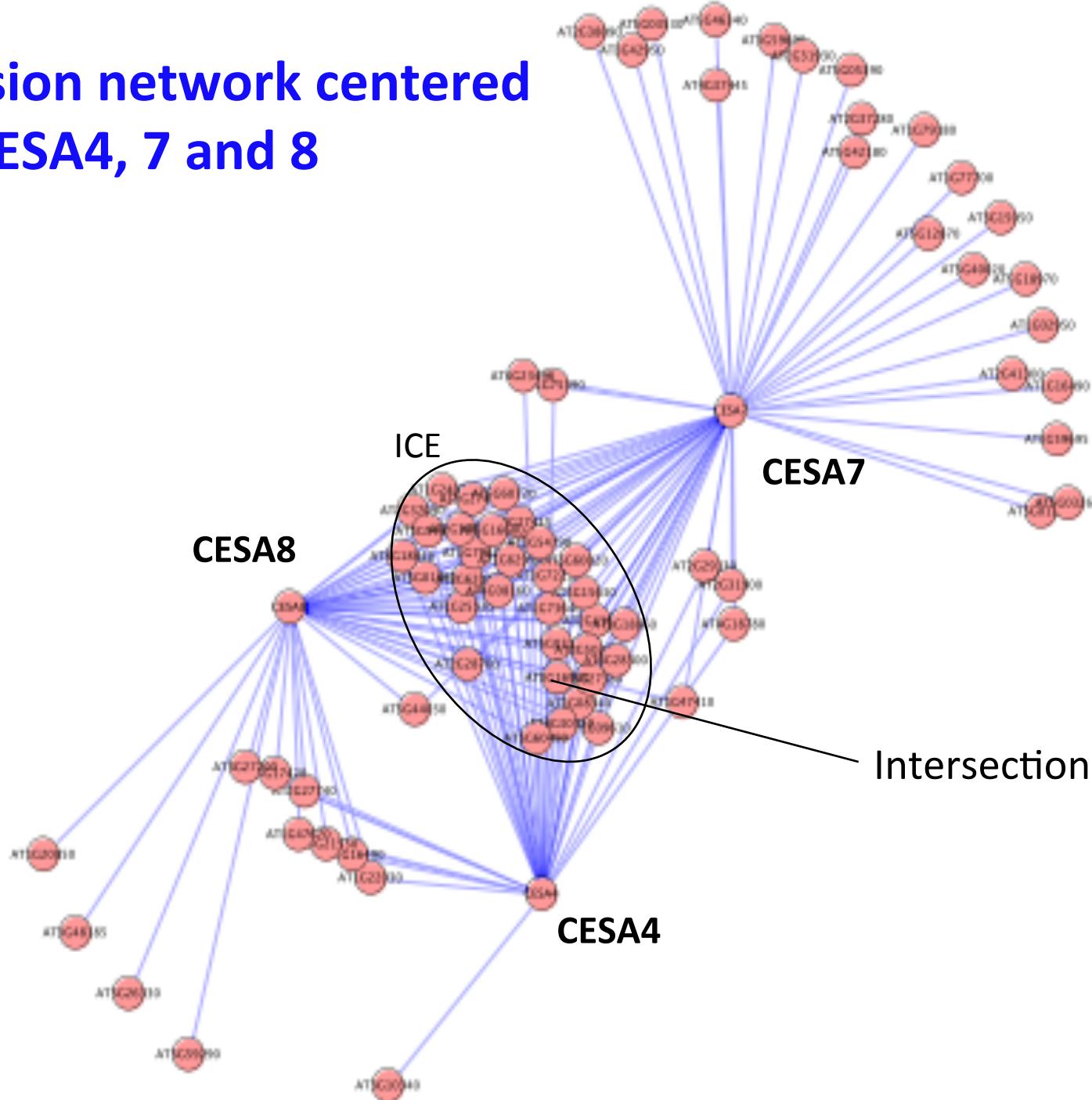
$$d_{ij} = X_{ij} - \bar{X}_{i-j},$$

$$d_i^* = d_i / s(d_i)$$

obeys t distribution with $m-2$ df

(Persson and Wei et al, 2005, PNAS, Vol102:8633)

A coexpression network centered around AtCESA4, 7 and 8

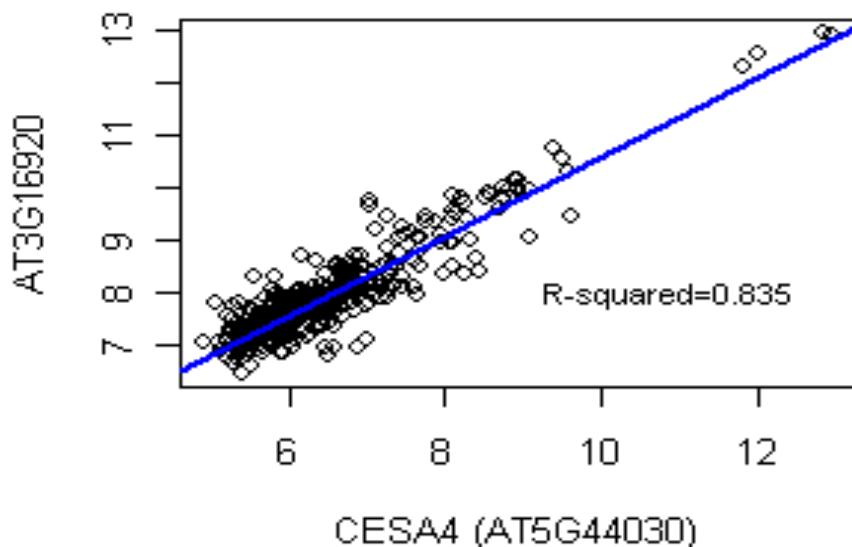
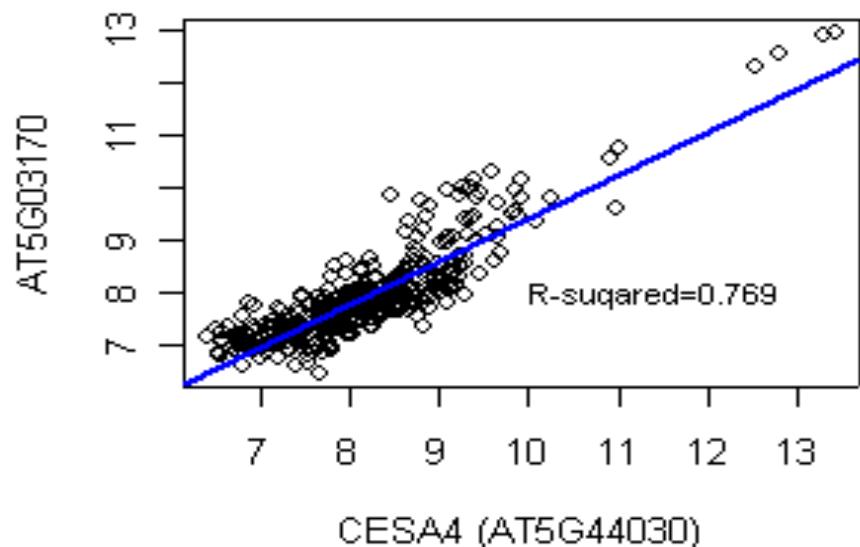
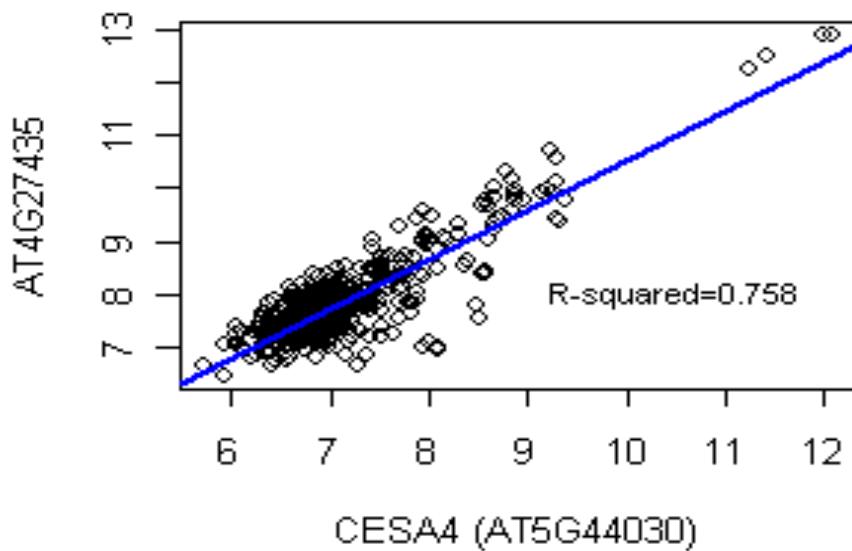
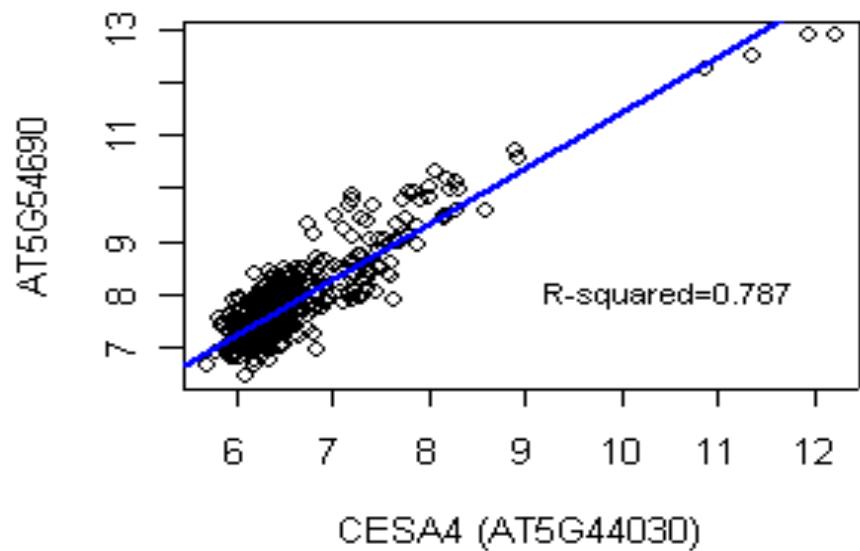


Intersection

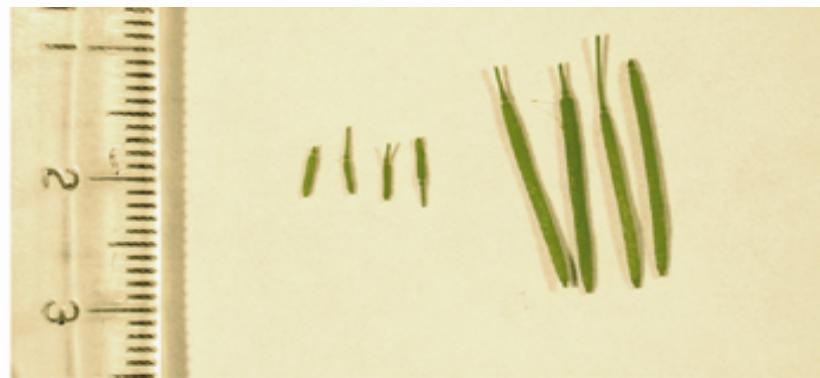
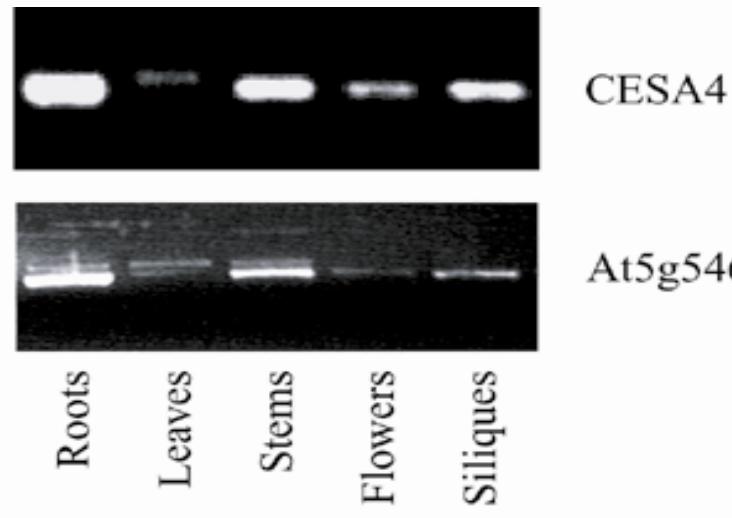
Genes selected from ICE

AGI	Gene title	Score*	P-value
AT5G44030	cellulose synthase, catalytic subunit (IRX5/CESA4)	4	4.1E-35
AT4G18780	cellulose synthase, catalytic subunit (IRX1/CESA8)	5	7.1E-33
AT5G54690*	glycosyl transferase family 8 protein	5	4.1E-35
AT5G17420	cellulose synthase, catalytic subunit (IRX3/CESA7)	5	3.7E-30
AT3G16920*	CTL1-like	6	9.3E-33
AT2G38080	laccase, putative / diphenol oxidase, putative	6	1.4E-34
AT5G15630	COBRA cell expansion protein COBL4	7	1.6E-32
AT5G03170*	fasciclin-like arabinogalactan-protein (FLA11)	8	6.7E-32
AT2G37090	glycosyl transferase family 43 protein	10	2.3E-29
AT3G18660	glycogenin glucosyltransferase (glycogenin)-related	29	4.9E-23
AT4G27435*	expressed protein	30	2.6E-24
AT5G60720	expressed protein	37	1.2E-20
AT3G62020	germin-like protein (GLP10)	42	9.2E-21
AT5G60020	laccase, putative / diphenol oxidase, putative	58	1.1E-17
AT4G28500	no apical meristem (NAM) family protein	62	9.7E-24
AT5G60490	fasciclin-like arabinogalactan-protein (FLA12)	64	7.9E-21

These genes highlighted in blue are four selected for experiment validation

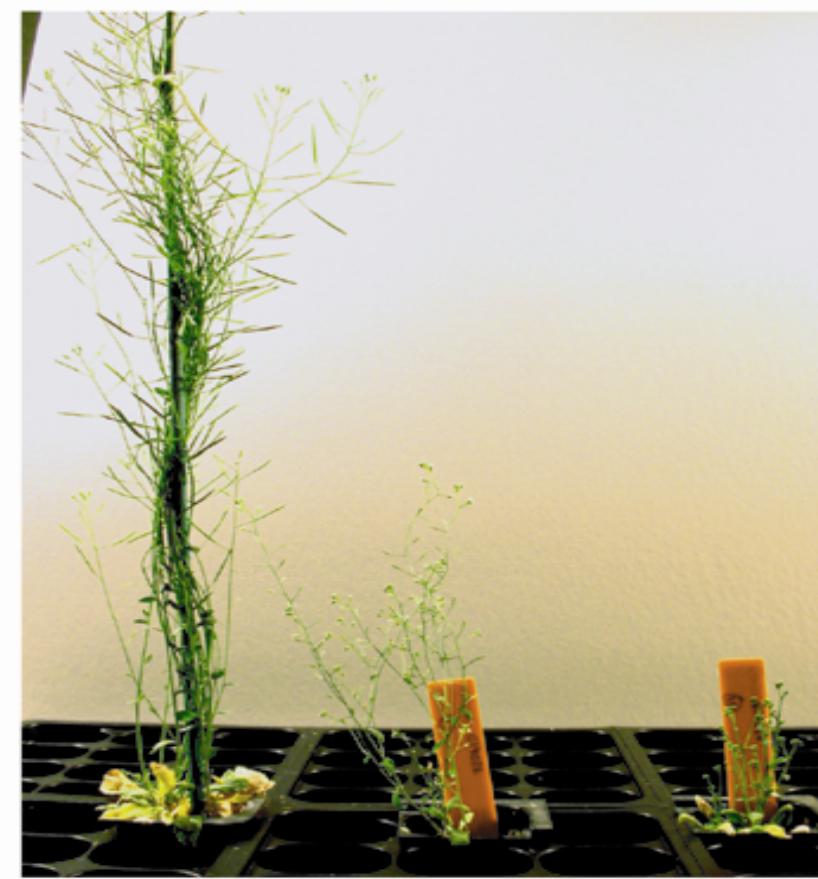


Phenotypic changes in mutant lines



At5g54690

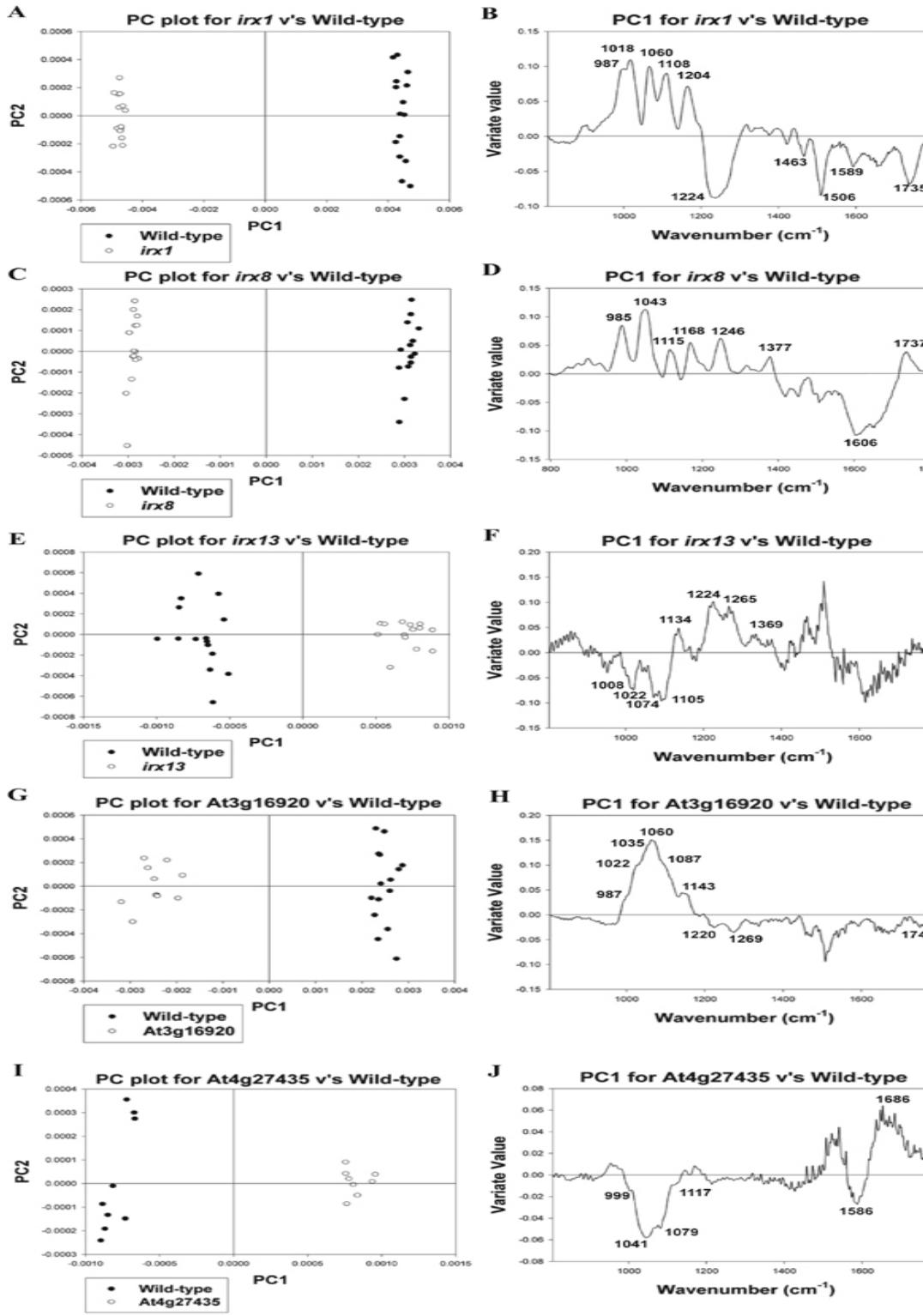
Wild-type



Wild-type

At5g54690_1

At5g54690_2

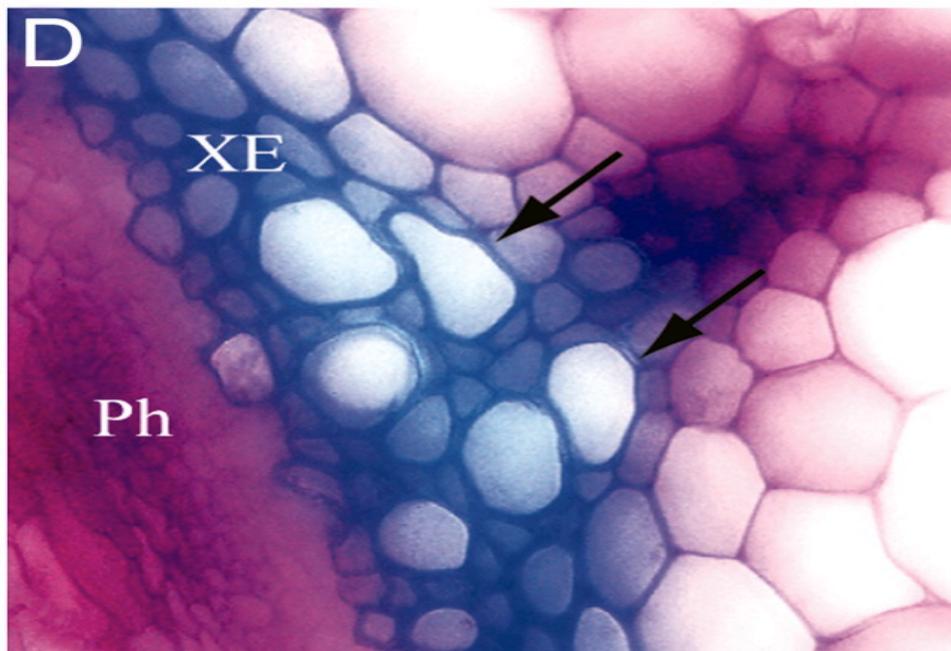
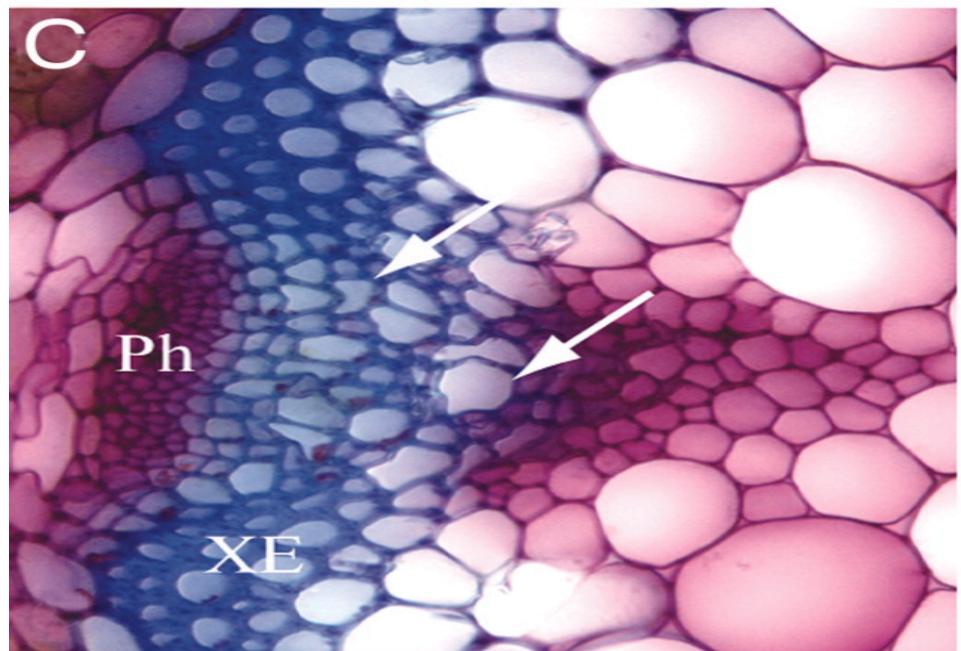
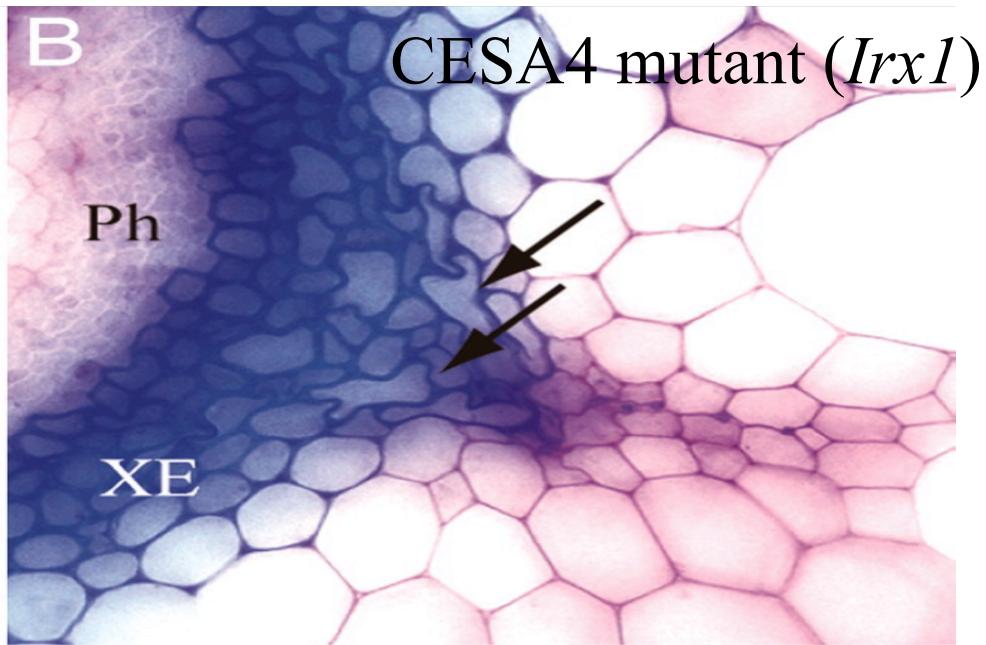
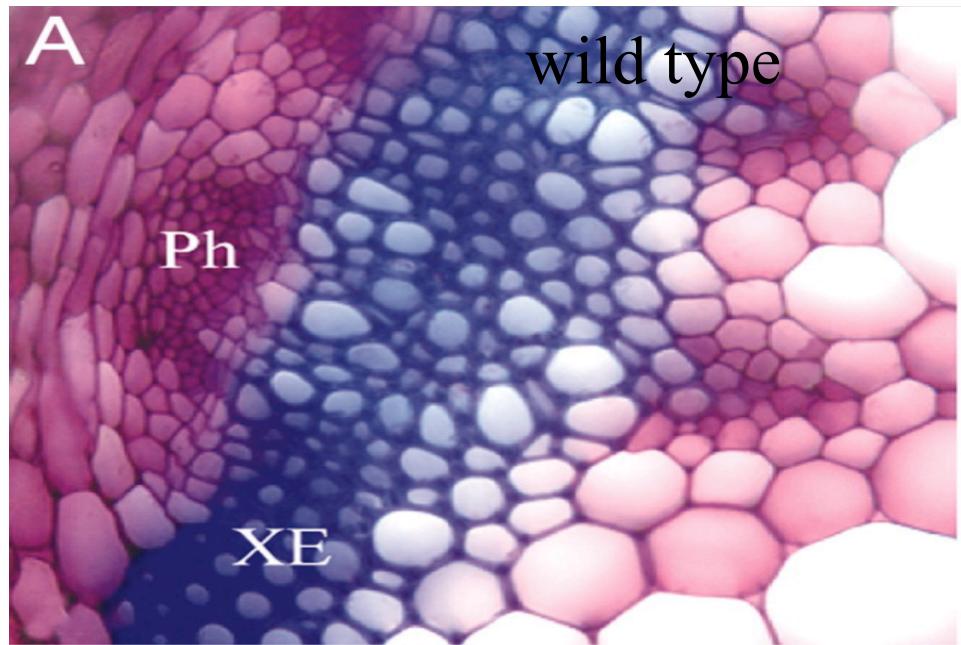


FTIR Analysis

Fourier Transform Infrared Spectroscopy is a technique which is used to obtain an infrared spectrum of absorption or emission of a solid, liquid or gas.

Cell wall analyses by FTIR. (A-F) Principal components plots for mutant [*irx1* (*CESA4* mutant), *irx8* (SALK_014026, At5g54690), *irx13* (SALK_046976, At5g03170), At3g16920 (SALK_055713), and At4g27435 (SALK_137109)] vs. wild-type spectra.

Spectra for all mutants show a clear separation from wild-type spectra based on principal component (PC) 1. In all cases, the mutant spectra show differences from wild-type spectra in the carbohydrate fingerprint regions that correspond to the deformations of cellulosic and noncellulosic polymers.



At5g54690 knockout (*Irx8*)

At5g03170 knockout (*Irx13*)

Related studies

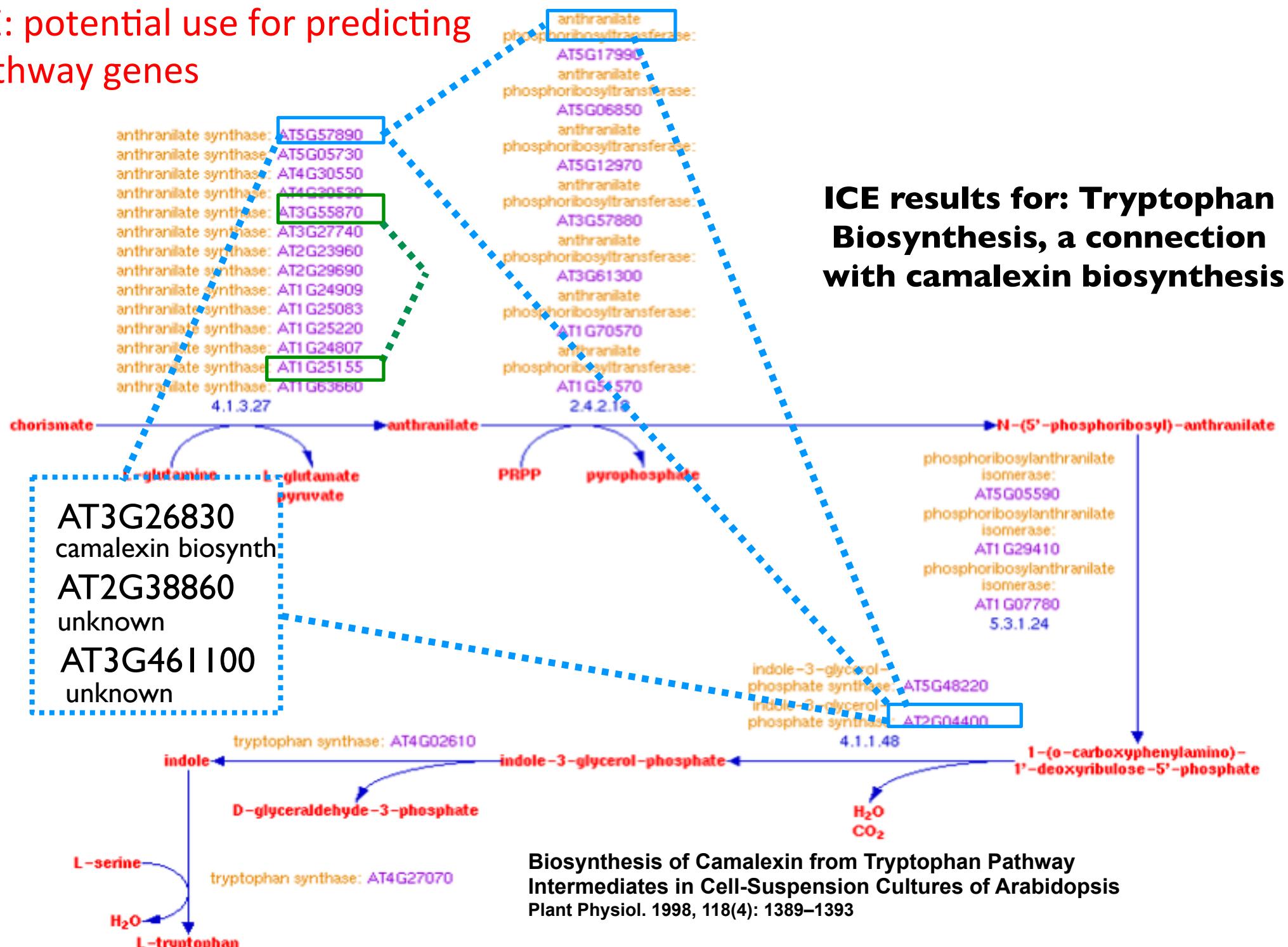
- 1. Among the 25 genes, 16 are the same as those genes identified by Dr. Simon Turner via forward genetics**

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *The Plant cell* 17(8): 2281-95.

- 2. *Irx8* (At5g54690) protein has been shown to target to Golgi, where glucuronoxylans (GXs) are synthesized. *Irx8* is implicated in affecting GX synthesis. GX, cellulose, and lignin are three major components of secondary cell walls in woody plants.**

Pena et al, *Arabidopsis irregular xylem8 and irregular xylem9: Implications for the Complexity of Glucuronoxylan Biosynthesis* *The Plant Cell* 19:549-563 (2007)

ICE: potential use for predicting pathway genes



Summary for coexpression network:

1. Co-expression network can be used to predict genes involved in various pathways and biological processes. It is especially suitable for predicting functionally associated genes, metabolic pathway genes, structural genes that are usually located at the low hierarchical terminals of gene regulatory networks. These genes are often co-regulated, which is why they are often highly correlated.
2. Due to the large number of terminal genes, the identification of true genes is often interfered by noise. The prediction of pathway genes in many cases is not accurate. However, it indeed works in some cases.

What is gene collaborative network ?

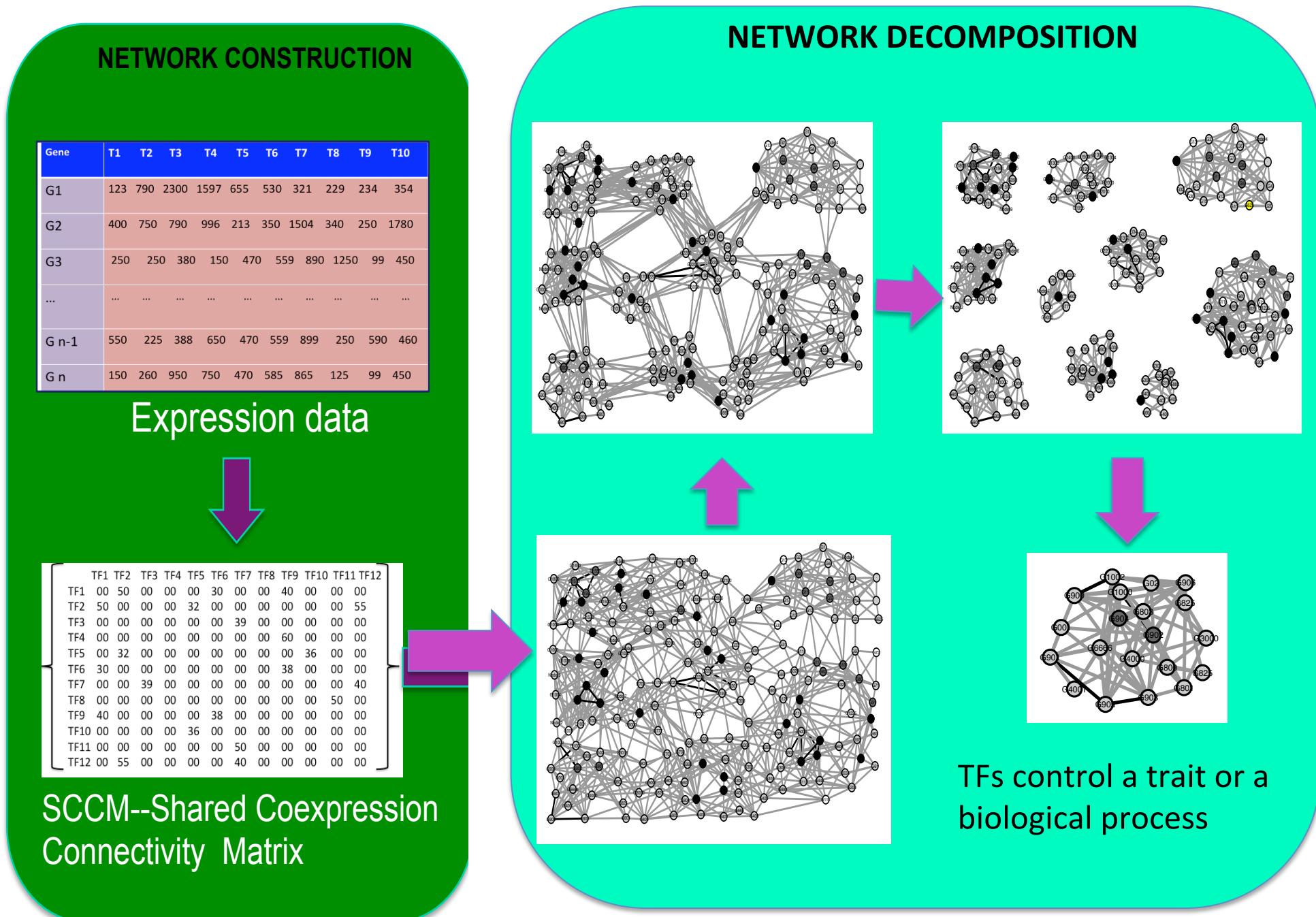
A **gene collaboration network (GCN)** is an undirected graph, where each node corresponds to a regulatory gene (e.g. TF), and a pair of nodes is connected with an edge if there is a significant **collaboration** relationship between them.

Microarray or RNA-seq data

Gene	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
G1	123	790	2300	1597	655	530	321	229	234	354
G2	400	750	790	996	213	350	1504	340	250	1780
G3	250	250	380	150	470	559	890	1250	99	450
...
G n-1	550	225	388	650	470	559	899	250	590	460
G n	150	260	950	750	470	585	865	125	99	450

Hypothesis: Regulatory genes (e.g. TFs) controlling a biological process or complex trait are collaborative.

Identification of Regulatory Genes That Collectively Govern Complex Trait?

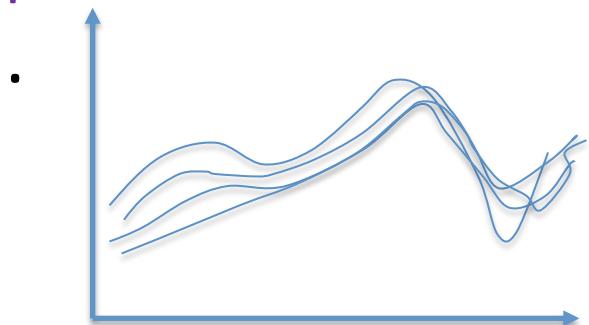


How can we identify regulatory genes that are collaborative?

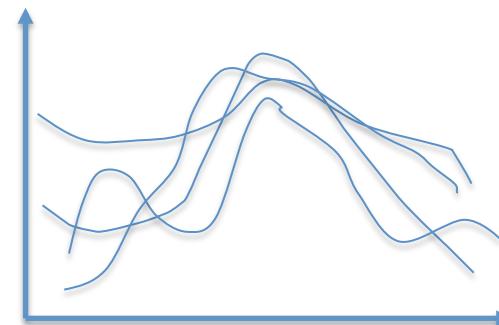
Hypothesis

Regulatory genes that collaboratively control the same complex trait are **coordinated in gene expression profiles**.

Coexpressed genes refer to those that have higher concordance in expression profiles, whereas coordinated genes refer to those that have relatively looser concordance in expression profiles.

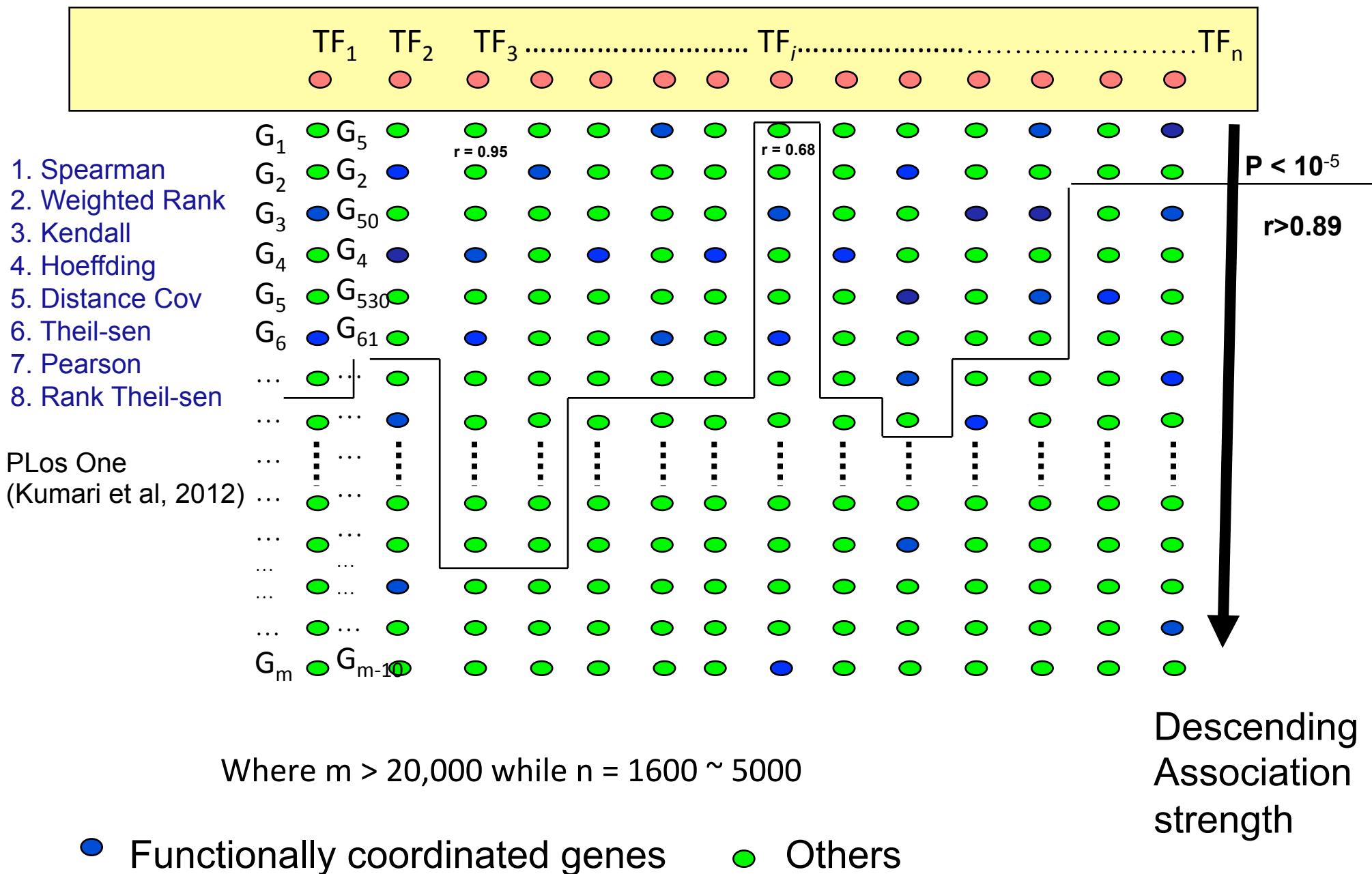


Coexpressed genes

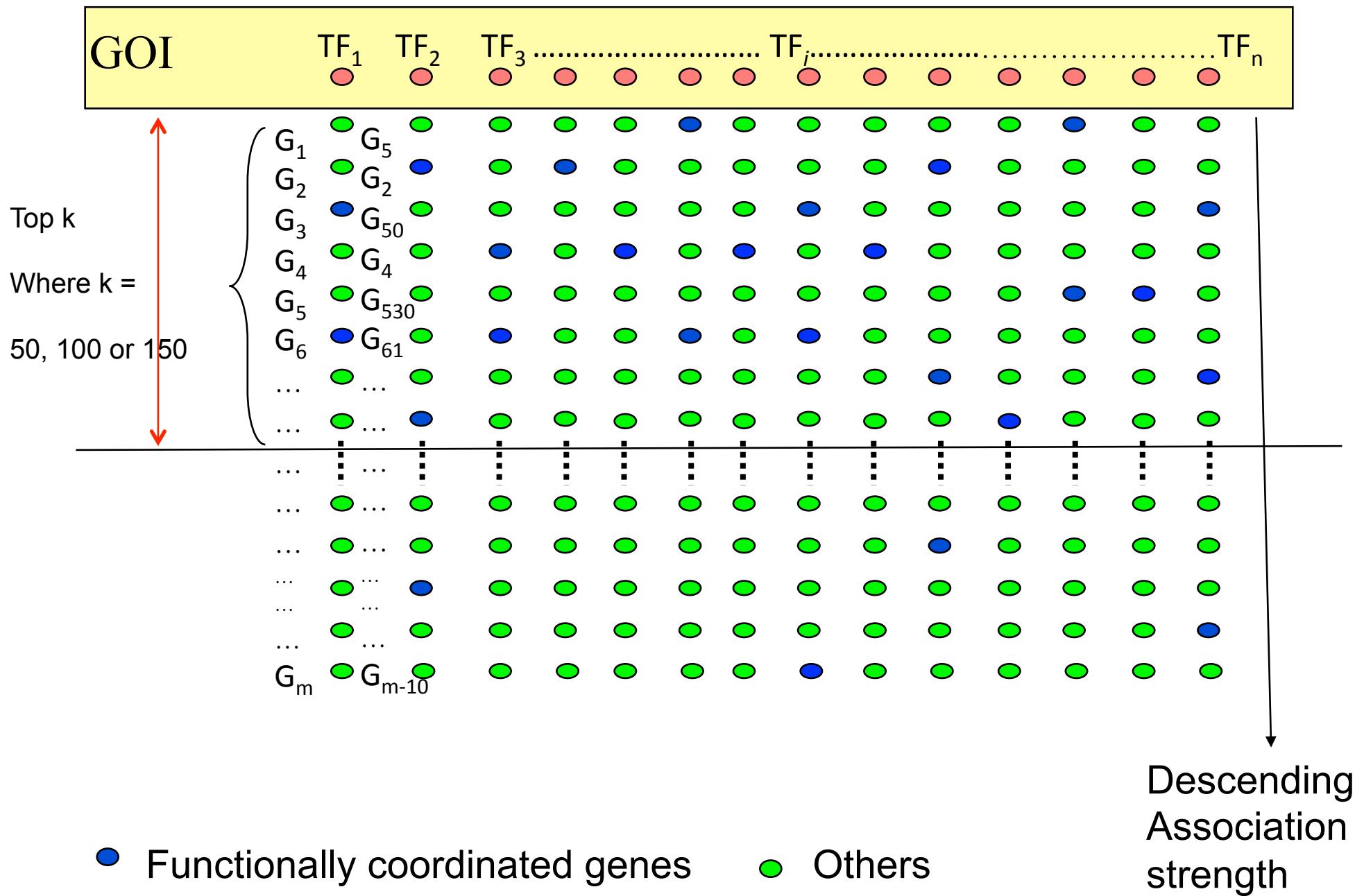


Collaborative genes

Genome-wide coexpression analysis

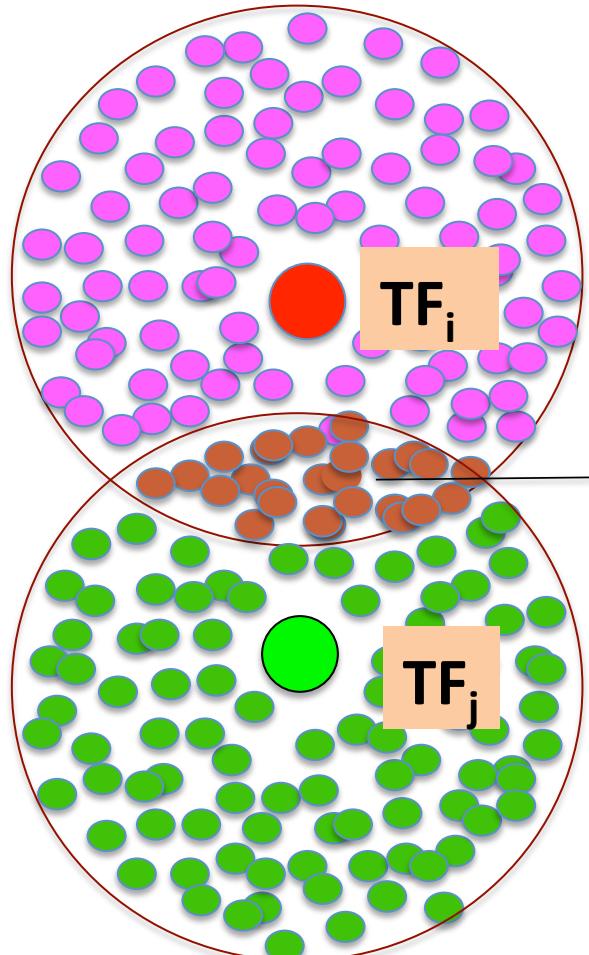


Genome-wide coexpression analysis



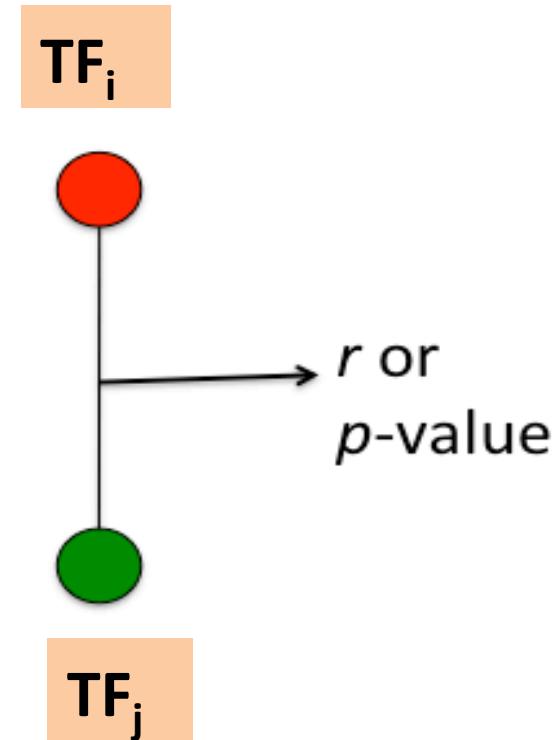
How to measure the loosely coordination between two TFs?

100 genes that are
most tightly coexpressed to TF_1



$$n_c$$

Number of
shared genes



100 genes that are
most tightly coexpressed to TF_j

Kumari S, Nie J, Chen H-S, Ma H, et al. (2012) Evaluation of Gene Association Methods for Coexpression Network Construction and Biological Knowledge Discovery. PLoS ONE 7(11): e50411. doi:10.1371/journal.pone.0050411
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0050411>

Shared Coexpression Connectivity Matrix (SCCM)

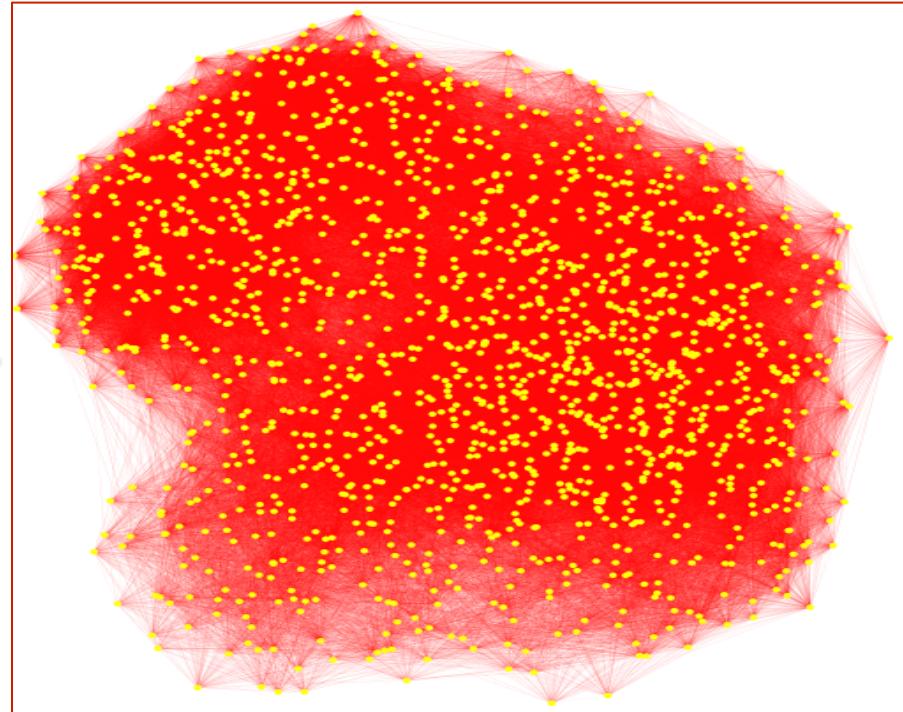
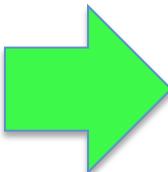
TFs

	AATF	ABCA11	ABCA11	ABCA11	ABT1	ACVR2A	ADNP	ADPGK	AEBP1	AEBP2	AFF1
AATF	100	0	0	0	3	0	0	0	0	0	0
ABCA11	0	100	0	13	0	0	0	0	0	0	0
ABCA11	0	0	100	11	0	0	0	0	0	0	0
ABCA11	0	13	11	100	0	0	0	0	0	0	0
ABT1	3	0	0	0	100	0	0	0	0	0	11
ACVR2A	0	0	0	0	0	100	8	8	0	0	0
ADNP	0	0	0	0	0	8	100	0	0	0	0
ADPGK	0	0	0	0	0	8	0	100	0	0	0
AEBP1	0	0	0	0	0	0	0	0	100	0	0
AEBP2	0	0	0	0	0	0	0	0	0	100	0
AFF1	0	0	0	0	11	0	0	0	0	0	100
AFF2	0	0	0	0	3	0	0	0	0	0	11
AFF3	0	0	0	0	0	0	0	0	0	0	0
AHR	0	0	0	0	0	0	0	0	0	0	0
AIP	3	0	0	0	13	0	0	0	0	0	15
AIP	0	0	0	0	0	0	0	0	0	4	0
AIRE	0	0	0	0	0	0	0	0	0	0	0
AK096318	0	0	0	0	0	0	0	0	0	0	0
AK126463	0	0	0	0	0	0	0	0	5	0	0
AL834146	0	0	0	0	0	4	11	3	0	0	0
ALF	0	0	0	0	0	0	0	0	0	0	0
ALX3	0	0	0	0	0	0	0	0	6	0	0
ALX4	0	0	0	0	0	0	0	0	0	0	0
AMOT	0	0	0	0	0	0	0	0	0	0	0
ANKRD30A	0	0	0	0	0	0	0	0	0	0	0
ANP32A	0	0	0	0	0	0	0	0	0	0	0
APBA2BP	0	0	0	0	0	0	0	0	0	0	0
APEX1	0	0	0	0	0	0	0	0	0	0	0
AR	0	0	0	0	0	0	0	0	0	0	0
ARC	0	0	0	0	0	0	0	0	0	0	0
ARID1A	0	0	0	0	0	0	17	0	0	0	0
ARID1B	0	0	0	0	8	0	0	0	0	0	19

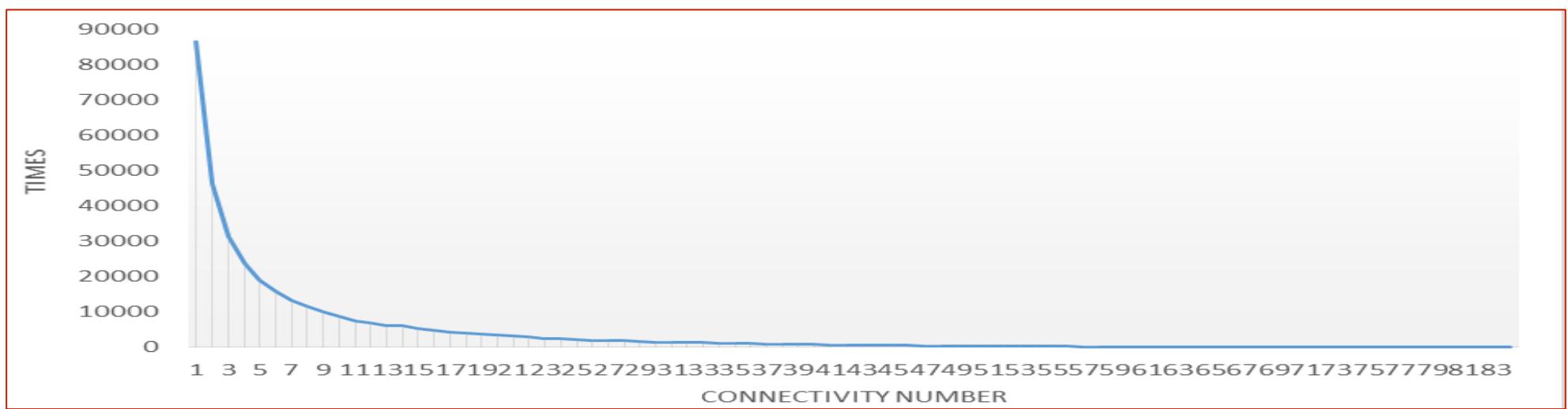
Collaborative Network of TF

	AATF	ABCA11	ABCA11	ABCA11	ABT1	ACVR2A	ADNP	ADPGK	AEBP1	AEBP2	AFF1
AATF	100	0	0	0	3	0	0	0	0	0	0
ABCA11	0	100	0	13	0	0	0	0	0	0	0
ABCA11	0	0	100	11	0	0	0	0	0	0	0
ABCA11	0	13	11	100	0	0	0	0	0	0	0
ABT1	3	0	0	0	100	0	0	0	0	0	11
ACVR2A	0	0	0	0	0	100	8	8	0	0	0
ADNP	0	0	0	0	0	8	100	0	0	0	0
ADPGK	0	0	0	0	0	8	0	100	0	0	0
AEBP1	0	0	0	0	0	0	0	0	100	0	0
AEBP2	0	0	0	0	0	0	0	0	0	100	0
AFF1	0	0	0	0	11	0	0	0	0	0	100
AFF2	0	0	0	0	3	0	0	0	0	0	11
AFF3	0	0	0	0	0	0	0	0	0	0	0
AHR	0	0	0	0	0	0	0	0	0	0	0
AIP	3	0	0	0	13	0	0	0	0	0	15
AIP	0	0	0	0	0	0	0	0	0	4	0
AIRE	0	0	0	0	0	0	0	0	0	0	0
AK096318	0	0	0	0	0	0	0	0	0	0	0
AK126463	0	0	0	0	0	0	0	0	0	5	0
AL834146	0	0	0	0	0	4	11	3	0	0	0
ALF	0	0	0	0	0	0	0	0	0	0	0
ALX3	0	0	0	0	0	0	0	0	6	0	0
ALX4	0	0	0	0	0	0	0	0	0	0	0
AMOT	0	0	0	0	0	0	0	0	0	0	0
ANKRD30A	0	0	0	0	0	0	0	0	0	0	0
ANP32A	0	0	0	0	0	0	0	0	0	0	0
APBA2BP	0	0	0	0	0	0	0	0	0	0	0
APEX1	0	0	0	0	0	0	0	0	0	0	0
AR	0	0	0	0	0	0	0	0	0	0	0
ARC	0	0	0	0	0	0	0	0	0	0	0
ARID1A	0	0	0	0	0	0	0	17	0	0	0
ARID1B	0	0	0	0	8	0	0	0	0	0	19

shared coexpression connectivity matrix (SCCM)



Collaborative network of all TFs

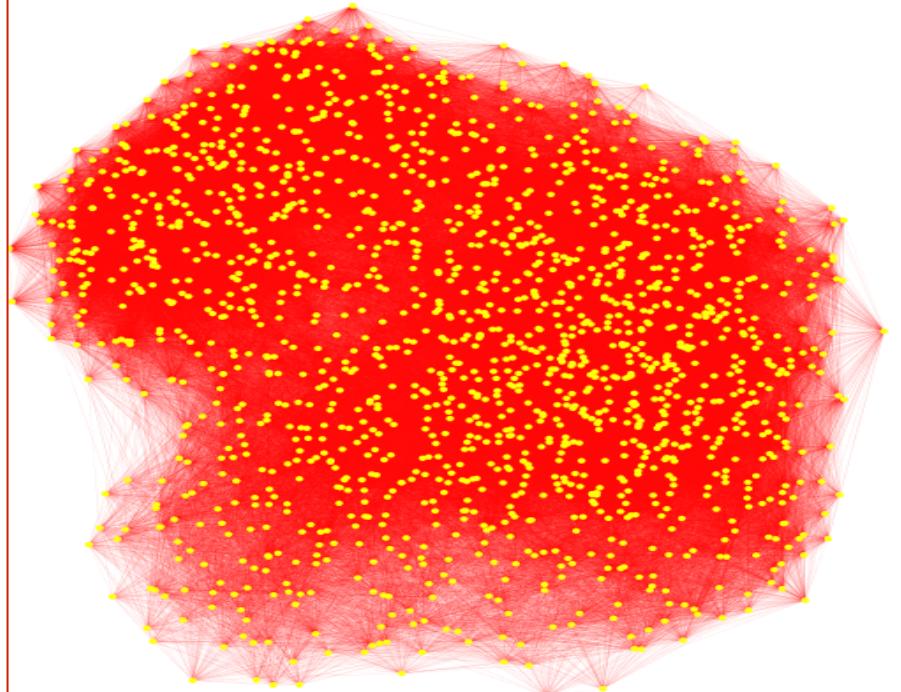
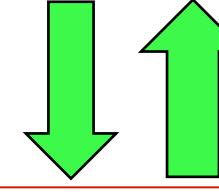
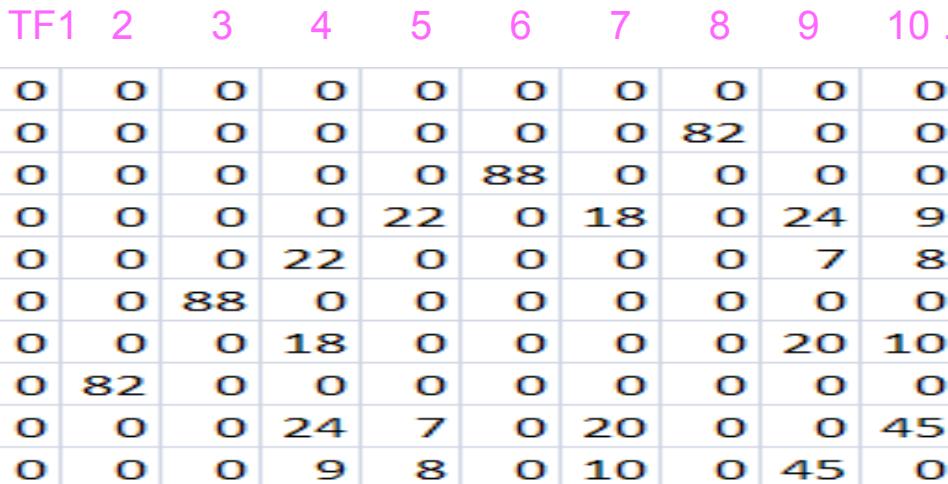


Gene	N6_R1	N6_R2	N12_R1	N12_R2	H24_R1	H24_R2	N48_R1	N48_R2	N96_R1	N96_R2	N94_R1	N94_R2	CTRL_R1
PgkHfr_1001_1.41_at	7.553895	7.388326	7.292131	7.446973	7.413151	7.288017	7.300384	7.449564	7.462835	7.240070	7.418269	7.446259	7.031911
PgkHfr_1002_1.41_at	3.5515519	3.8297352	3.8963538	3.8963538	3.7042045	3.69912479	3.7042045	3.873781	3.873781	3.7042045	3.7042045	3.7042045	3.0277639
PgkHfr_1003_1.41_at	9.736727	9.389712	9.5073374	6.6388644	8.31078	8.273278	8.273278	8.273278	8.273278	9.047749	9.192535	9.073054	8.390585
PgkHfr_1004_1.41_at	8.479371	8.0589591	8.0589591	8.0589591	8.0589591	8.1796352	8.1796352	8.1796352	8.1796352	8.0589591	8.0589591	8.0589591	8.058959
PgkHfr_1008_1.41_at	11.043653	11.0408655	11.227214	10.8505409	10.2673264	10.360719	11.7506269	10.0289713	9.54526729	9.3979947	9.5688636	9.65705946	9.0486651
PgkHfr_1009_1.41_at	7.0147048	8.840578	7.5098454	7.212688	7.0147048	7.385163	7.4045056	7.4475284	7.5162457	6.5270581	8.8426963	7.4475284	7.035943
PgkHfr_1010_1.41_at	1.4676576	1.1174528	1.2021702	1.7904957	1.4592645	1.576788	1.576788	1.576788	1.576788	1.2021702	1.2021702	1.2021702	1.3616707
PgkHfr_1011_1.41_at	1.5014096	1.081026	1.3901255	1.3801255	1.2801255	1.4084861	1.2712162	1.2694845	1.6986149	1.2705017	1.4473165	1.2745292	1.2056749
PgkHfr_1012_1.41_at	1.5767496	1.5993246	1.5993246	1.5993246	1.5993246	1.7030857	1.7030857	1.7030857	1.7030857	1.5993246	1.5993246	1.5993246	1.5993246
PgkHfr_1013_1.41_at	1.3194411	1.0287298	1.3101685	1.3843655	1.7165761	1.4023026	1.5646065	1.15181941	1.7019298	1.3867635	1.7553214	1.6149369	1.2048449
PgkHfr_1014_1.41_at	1.808886	1.4264626	1.7318408	1.7032369	1.4419117	1.9749469	1.9749469	1.9749469	1.9749469	1.4264626	1.4264626	1.4264626	1.3044884
PgkHfr_1015_1.41_at	1.3194411	1.0287298	1.3101685	1.3843655	1.7165761	1.4023026	1.5646065	1.15181941	1.7019298	1.3867635	1.7553214	1.6149369	1.2048449
PgkHfr_1016_1.41_at	1.808886	1.4264626	1.7318408	1.7032369	1.4419117	1.9749469	1.9749469	1.9749469	1.9749469	1.4264626	1.4264626	1.4264626	1.3044884
PgkHfr_1017_1.41_at	1.808886	1.4264626	1.7318408	1.7032369	1.4419117	1.9749469	1.9749469	1.9749469	1.9749469	1.4264626	1.4264626	1.4264626	1.3044884
PgkHfr_1018_1.41_at	1.021727	1.8546358	1.8033761	4.8691317	4.4981361	8.7426743	8.3464529	8.3464529	8.3464529	8.2506458	8.2506458	8.4691317	7.803563
PgkHfr_1019_1.41_at	5.6402934	5.7631207	5.7631207	5.7631207	5.7631207	5.6402934	5.6402934	5.6402934	5.6402934	5.7631207	5.7631207	5.7631207	5.838573
PgkHfr_1020_1.41_at	1.6740464	1.0484254	2.0791212	1.5409361	5.7833939	5.4342342	5.3283087	5.4051675	6.1078045	6.0040845	4.4521583	5.4958214	5.1949582
PgkHfr_1021_1.41_at	1.0041461	1.4884293	1.4884293	1.4884293	1.4884293	1.3339468	1.3339468	1.3339468	1.3339468	1.4884293	1.4884293	1.4884293	1.4884293
PgkHfr_1022_1.41_at	1.0041461	1.4884293	1.4884293	1.4884293	1.4884293	1.3339468	1.3339468	1.3339468	1.3339468	1.4884293	1.4884293	1.4884293	1.4884293
PgkHfr_1023_1.41_at	1.1719729	1.1569058	1.3794067	4.3893644	4.7812947	8.7289749	8.2874741	8.2874741	8.2874741	8.4089454	8.4089454	8.7142142	9.1949000
PgkHfr_1024_1.41_at	1.1719729	1.1569058	1.3794067	4.3893644	4.7812947	8.7289749	8.2874740	8.2874740	8.2874740	8.4089454	8.4089454	8.7142142	8.3782882
PgkHfr_1025_1.41_at	1.1719729	1.1569058	1.3794067	4.3893644	4.7812947	8.7289749	8.2874740	8.2874740	8.2874740	8.4089454	8.4089454	8.7142142	8.3782882
PgkHfr_1026_1.41_at	1.1719729	1.1569058	1.3794067	4.3893644	4.7812947	8.7289749	8.2874740	8.2874740	8.2874740	8.4089454	8.4089454	8.7142142	8.3782882
PgkHfr_1027_1.41_at	1.1719729	1.1569058	1.3794067	4.3893644	4.7812947	8.7289749	8.2874740	8.2874740	8.2874740	8.4089454	8.4089454	8.7142142	8.3782882
PgkHfr_1028_1.41_at	5.1015513	5.4894954	5.3980493	7.6481033	5.4894954	5.5474055	4.4745518	6.4101605	5.4894954	5.4894954	5.4894954	5.4894954	5.4894954
PgkHfr_1029_1.41_at	2.6310267	5.5831771	5.1450936	5.1450936	5.8242668	6.1151649	5.7541041	5.9021603	6.4487774	6.4420277	8.8620257	6.4521564	5.3849364
PgkHfr_1030_1.41_at	1.9047967	1.8270408	2.8735534	1.0233270	2.0118945	1.2101894	6.2881082	4.5403749	5.3018744	3.0719477	2.7589452	1.9549676	8.7841542
PgkHfr_1031_1.41_at	9.0790455	9.7802103	9.7802103	10.0083645	10.0070449	10.1405153	10.1231057	10.1231057	10.1231057	10.1231057	10.1231057	9.9515136	9.8472404
PgkHfr_1032_1.41_at	7.6032323	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262	7.7237262
PgkHfr_1033_2.41_at	2.0211868	2.1316286	2.1316286	2.1316286	2.1316286	2.0364058	2.0364058	2.0364058	2.0364058	2.1316286	2.1316286	2.1316286	2.1316286
PgkHfr_1034_2.41_at	2.0211868	2.1316286	2.1316286	2.1316286	2.1316286	2.0364058	2.0364058	2.0364058	2.0364058	2.1316286	2.1316286	2.1316286	2.1316286
PgkHfr_1035_2.41_at	2.0211868	2.1316286	2.1316286	2.1316286	2.1316286	2.0364058	2.0364058	2.0364058	2.0364058	2.1316286	2.1316286	2.1316286	2.1316286
PgkHfr_1036_2.41_at	2.0211868	2.1316286	2.1316286	2.1316286	2.1316286	2.0364058	2.0364058	2.0364058	2.0364058	2.1316286	2.1316286	2.1316286	2.1316286
PgkHfr_1037_2.41_at	9.390764	9.332941705	9.46859477	9.3381178	5.57482743	9.4549407	7.0094948	9.55731081	9.55731081	9.55731081	9.2308901	9.11755671	9.1833644

How to decompose this collaborative network into subnetworks, each of which controls a biological process or complex trait?

Collaborative Network

100



TF	TF1	2	3	4	5	6	7	8	9	10	...
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	82	0	0	0
3	0	0	0	0	0	88	0	0	0	0	0
4	0	0	0	0	22	0	18	0	24	9	0
5	0	0	0	22	0	0	0	0	7	8	0
6	0	0	88	0	0	0	0	0	0	0	0
7	0	0	0	18	0	0	0	0	20	10	0
8	0	82	0	0	0	0	0	0	0	0	0
9	0	0	0	24	7	0	20	0	0	45	0
10	0	0	0	9	8	0	10	0	45	0	0

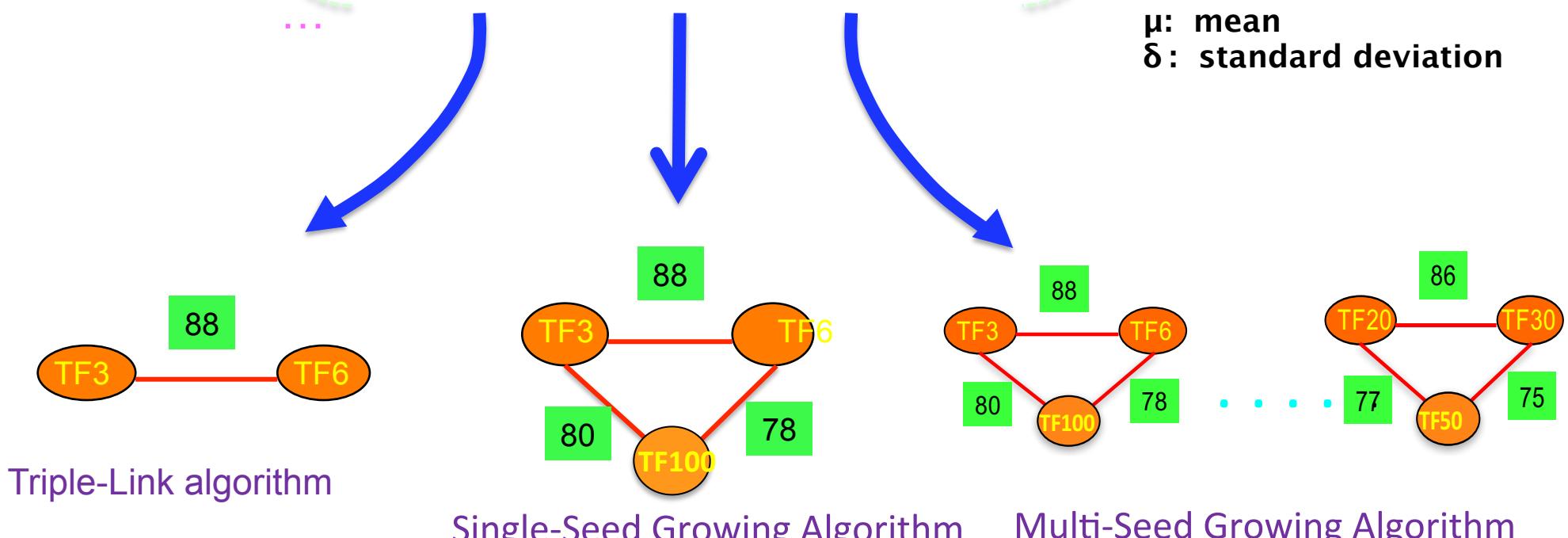
$$\alpha = \mu + \delta$$

$$\beta = \mu + 2\delta$$

$$\gamma = \mu + 3\delta$$

μ : mean

δ : standard deviation



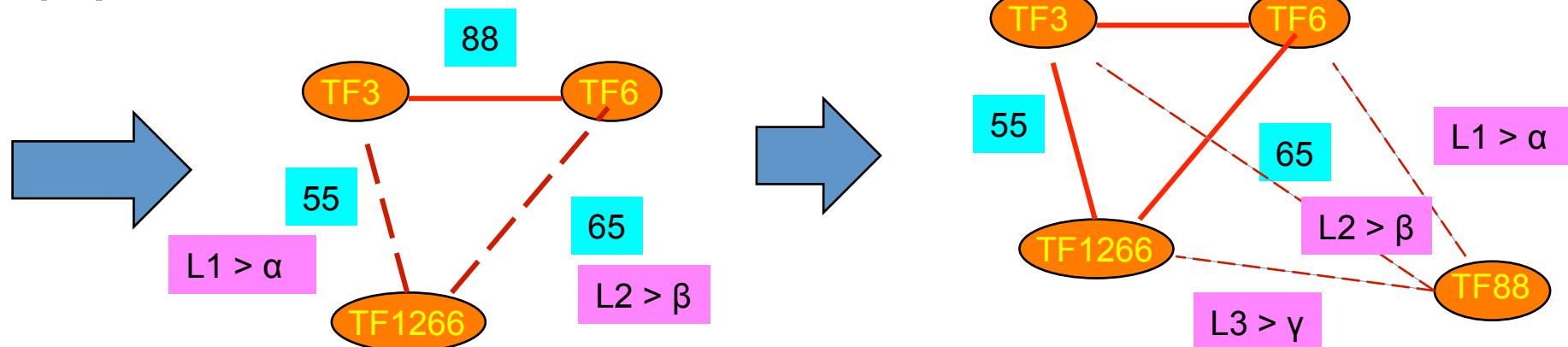
Triple-link algorithm

TF	TF1	2	3	4	5	6	7	8	9	10	...
1	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	82	0	0	
3	0	0	0	0	0	88	0	0	0	0	
4	0	0	0	0	22	0	18	0	24	9	
5	0	0	0	22	0	0	0	0	7	8	
6	0	0	88	0	0	0	0	0	0	0	
7	0	0	0	18	0	0	0	0	20	10	
8	0	82	0	0	0	0	0	0	0	0	
9	0	0	0	24	7	0	20	0	0	45	
10	0	0	0	9	8	0	10	0	45	0	
...											

$$\alpha = \mu + \delta$$

$$\beta = \mu + 2\delta$$

$$\gamma = \mu + 3\delta$$



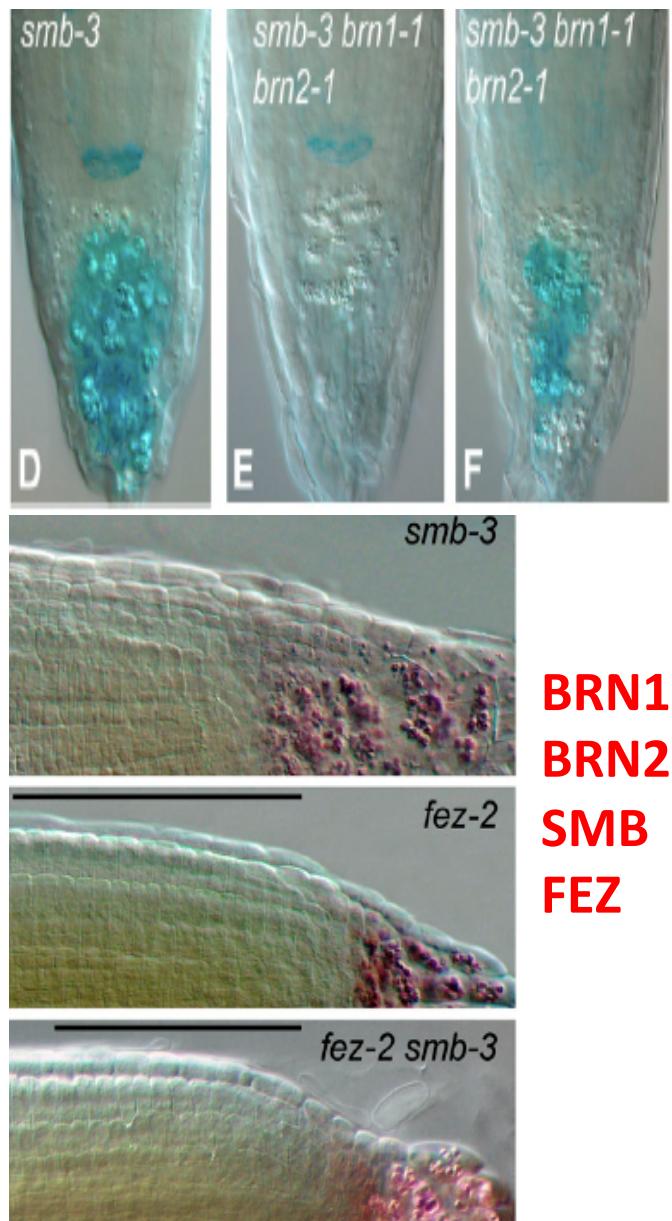
Test 1: Arabidopsis Data

Arabidopsis salt stress experiments

- 6 Microarray experiments of Arabidopsis
- 108 chips (downloaded from GEO)

Question: What TFs control root growth under salt stress condition in Arabidopsis?

Cluster 2: A cluster of TFs control root cap development (stem cells of roots)			
AT1G33280	BRN1	BRN1, SMB control root cap maturation	[66]
AT4G10350	BRN2	BRN2, SMB control root cap maturation	[66]
AT1G79580	SMB	FEZ and SMB control root stem cells	[67]
AT5G39820	ANAC094	Apical meristem protein, function unknown	[59]
AT1G26870	FEZ	FEZ and SMB control root stem cells in cap	[67]
AT1G74500	TOM7	Embryonic root initiation	[68]
AT3G27010	TCP20	Postembryonic cell division in root	[70]
AT2G30340	LBD13	Expressed in cells at the adaxial base of lateral roots	[104]
AT2G40470	LBD15	Expressed in cells at the adaxial base of lateral roots	[104]
AT1G51190	PLT2	Control root stem cell activity near cap	[69]
AT1G66350	RGL1	Root epidermal differentiation	[105]
AT2G37260	TTG2	Differentiation of trichomes and root hairless cells	[106]
AT5G57420	IAA33	IAA is involved in root development	[107, 108]
AT2G29060		scarecrow transcription factor family protein	



Test 2: Data Set from humans

What TFs are necessary for pluripotency maintenance in human stem cells?

Present knowledge about regulation of human stem cell pluripotency

NANOG + POU5F1 + SOX2

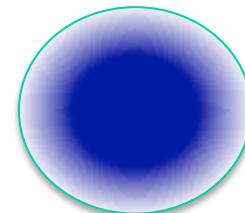
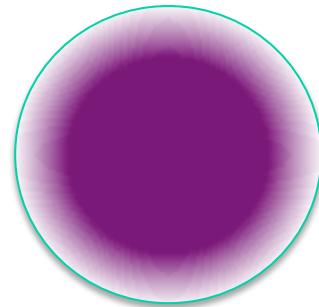
SALL2
PHC1
PRDM14
DNMT3A

CSRP1
NANOGP8
LOC653441
SOX3

LIN28
MYCL1
DNMT3B
RORB

Identified by many biologists from 1998 to 2008

The Power of TFs: Reprogramming in somatic cells



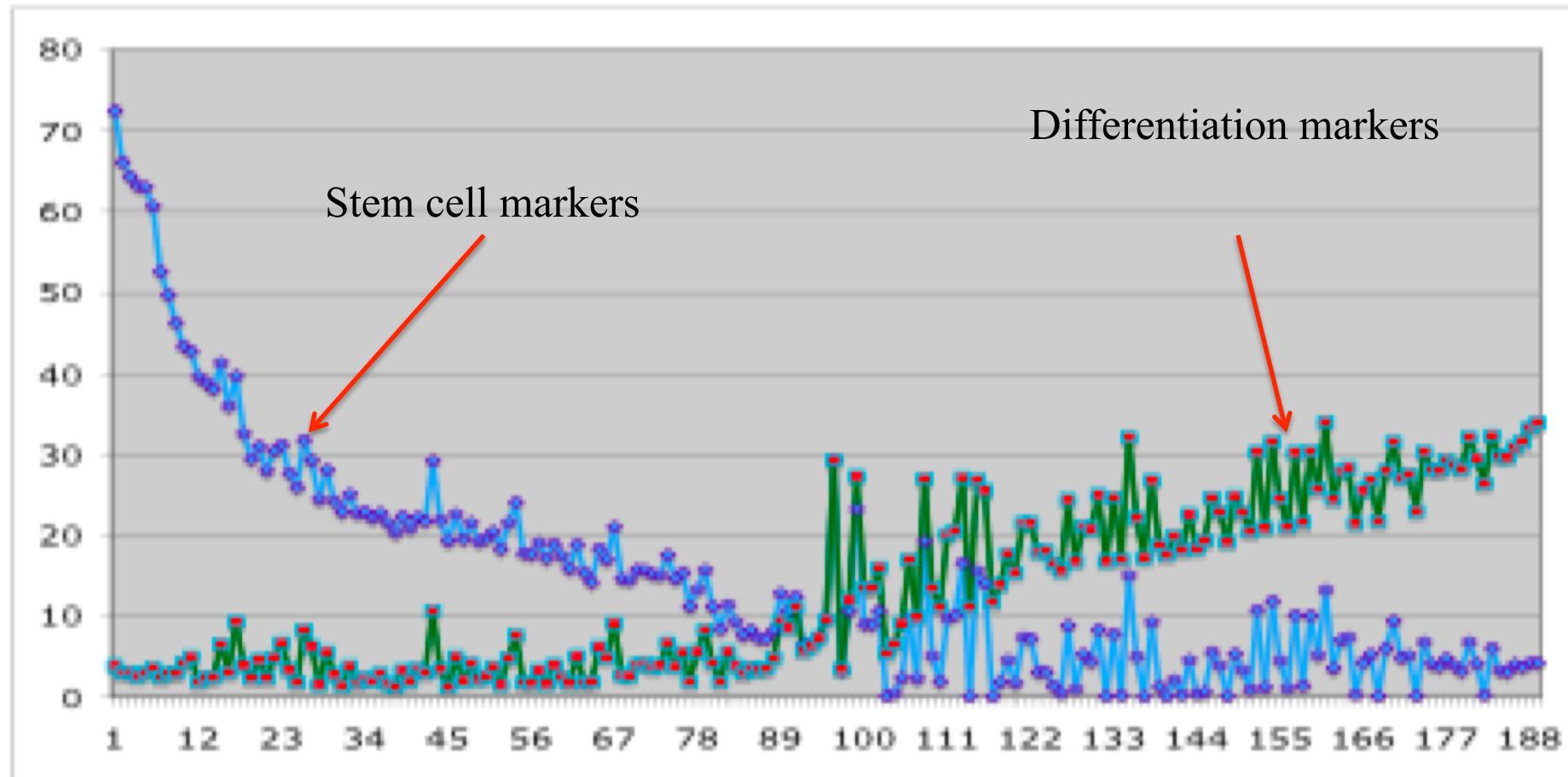
NANOG
POU5F1
SOX2

+ MYC2

Skin Cell

Induced Pluripotency Cell (IPC)

Input data: 189 microarray chips from human stem cells that underwent differentiation from 0 ~ 96 hours.



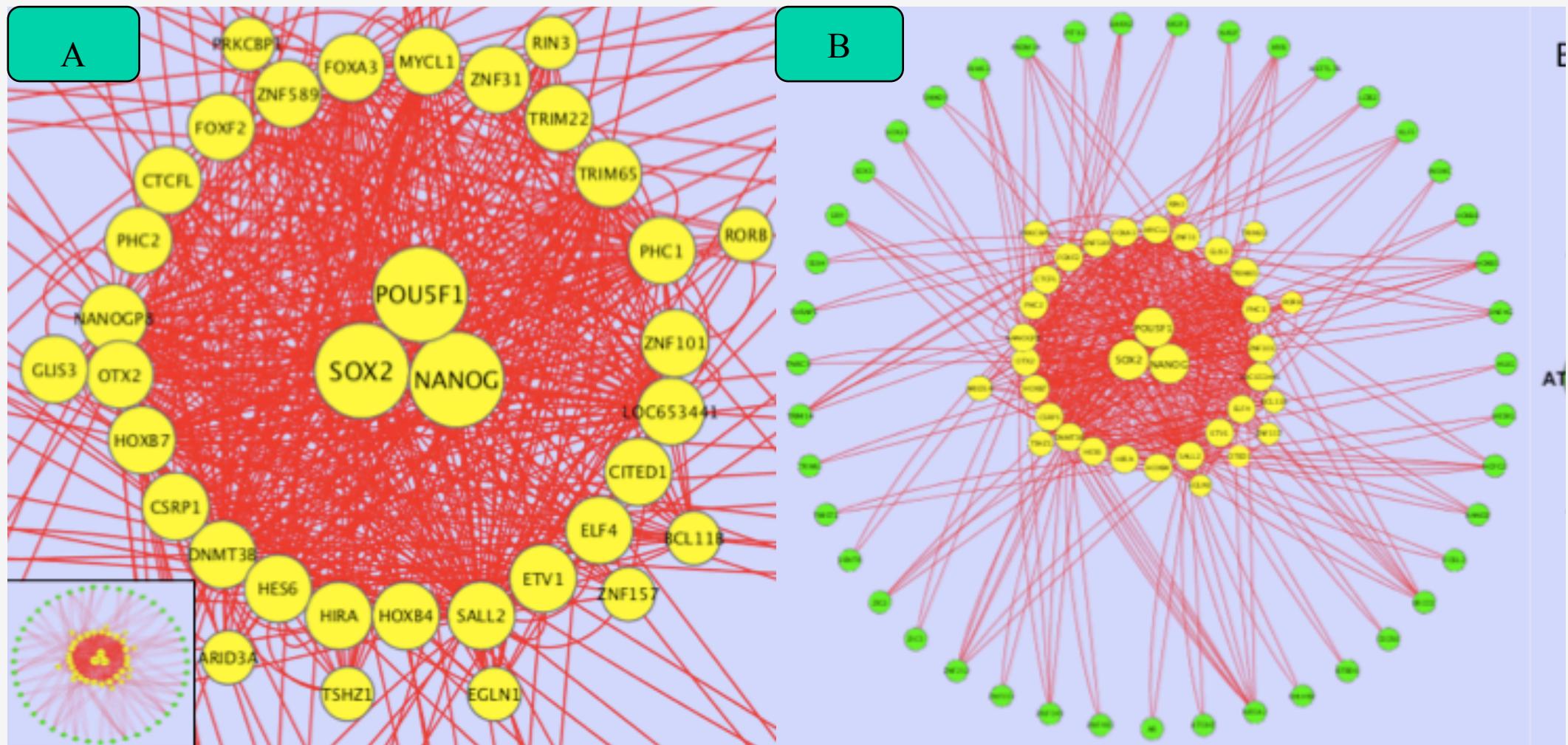
Microarray chips (189) provided by James Thomson @ U of Wisconsin

Subnetwork 1 : Output from TF-cluster

Genes	Symbol	Description	Evidence
NM_024865	NANOG	Nanog homeobox	[14]
BC099704	NANOGP8	Nanog homeobox pseudogene 8	Pseudogene with similarity to Nanog.
NM_003106	SOX2	SRY box 2	[14]
NM_002701	POU5F1	POU class 5 homeobox 1	[14]
NM_006892	DNMT3B	DNS methyltransferase 3 beta	[15, 16]
NM_004078	CSRP1	cysteine-rich protein	
NM_080618	CTCFL	CCCTC-binding factor (zinc finger protein)-like	[17]
NM_016089	ZNF589	Zinc finger 589	[18, 19]
NM_004426	PHC1	Polyhomeotic homolog 1	[20]
NM_005407	SALL2	SAL2 like	[18]
NM_018645	HES6	hairy and enhancer of split 6	
NM_173547	TRIM65	Tripartite motif containing 65	
NM_004427	PHC2	Polyhomeotic homolog 2	[21]
NM_032805	ZFP206	Zinc finger protein 206 (ZSCAN10)	[22, 23]
NM_001421	ELF4	ETS domain TF	
NM_003325	HIRA	HIR Histone Cell Cycle regulator	[24]
NM_033204	ZNF101	Zinc finger protein 101	
BC098403	ETV1	ets variant 1	[25, 26]
NM_006079	CITED2	Cbp/p300-interacting transactivator	[27, 28]
NM_021728	OTX2	orthodenticle homeobox 2	
NM_024015	HOXB4	homeobox B4	
NM_006074	TRIM22	tripartite motif-containing 22	[19]
XM_929986	LOC653441	similar to polyhomeotic 1-like	Gene with sequence similarity to PHC1
NM_004497	FOXA3	Forkhead box 3	

Among 24 predicted TFs, 17 have literature evidence (~70%)

The coordination between TFs within Subnetwork 1



- A. Intra-connections among the genes within Collaborative Subnetwork 1
- B. Comparison of intra-subnetwork connections with outward-connections

Test 3: Axolotl RNA-Seq

Axolotl transcriptome

S, Z, A, Digits,

30 day blastema

RNA-Seq (on GA IIx)

Eval for patterns in adult limb
and Blastema-specific genes

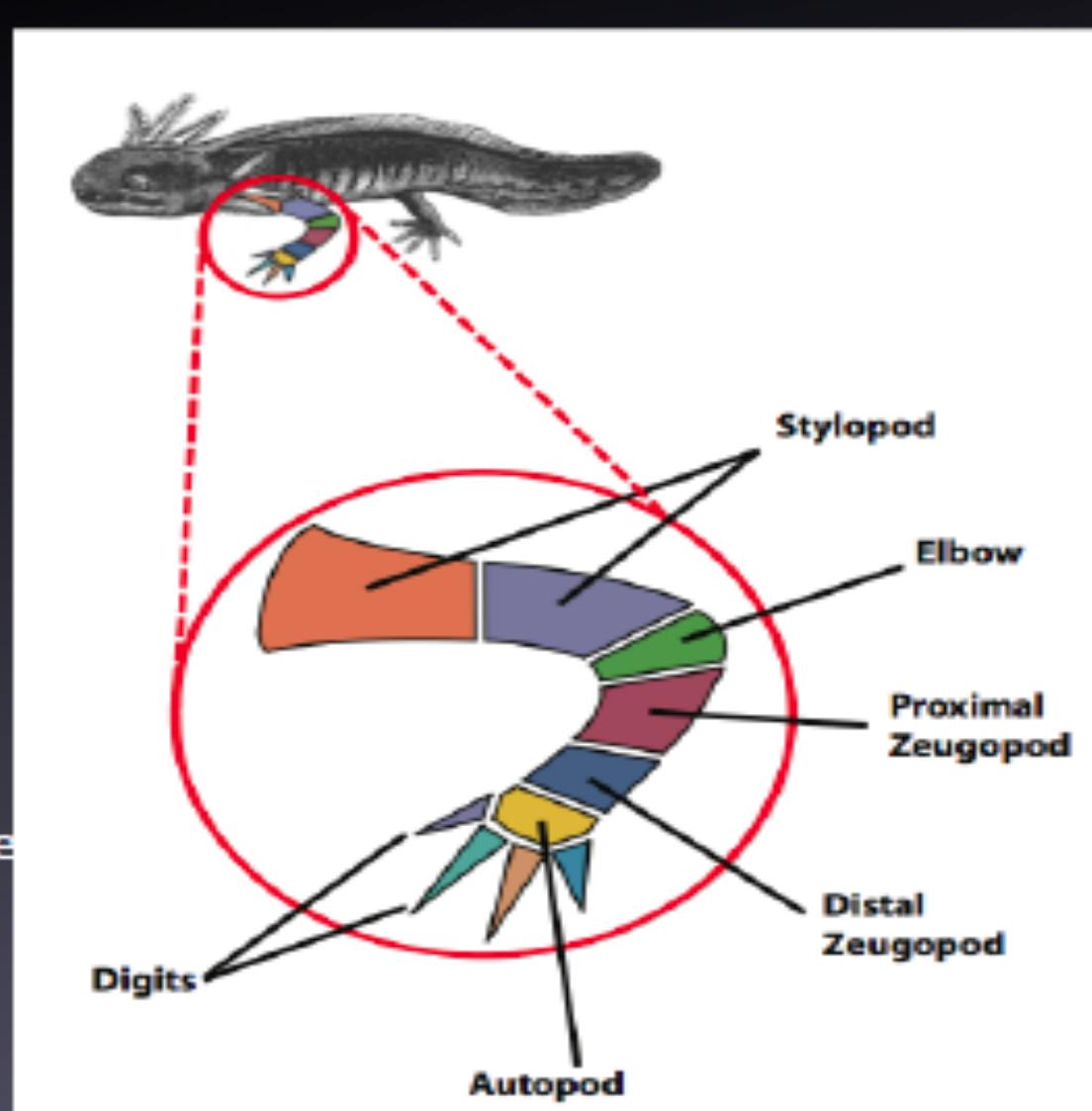
Challenges

No Genome,

Poor gene annotation

Genome 30GB,

Evol distance to Human/Mouse

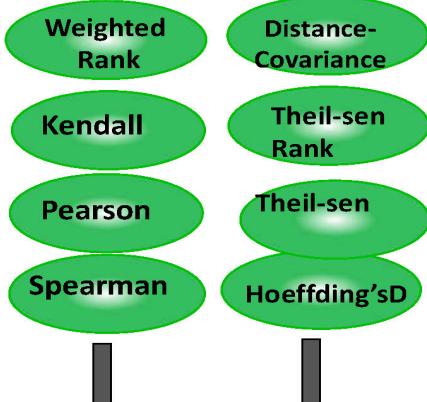
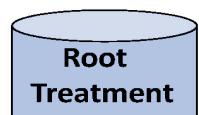
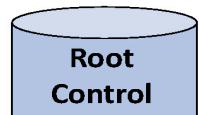
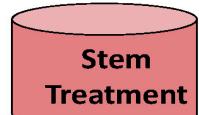


Cluster 2: Limb development

- 1. HES4: multicellular organismal development
- 2. HOXB1: anterior/posterior pattern formation
- 3. BARHL2: neuron differentiation
- 4. CDX2: caudal type homeobox 2
- 5. CIC
- 6. FOXH1 nodal expression; Lower-limb Deep Vein Thrombosis
- 7. FOXJ1: motile cilia, lupus erythematosus
- 8. FOXN1: skin, nail development
- 9. GLIS3: limb mesenchymal patterns
- 10. HES6: limb buds
- 11. HESX1: multicellular organismal development
- 12. HNF4A: limb mesoderm
- 13. HOXA5: skeletal and organ morphogenesis
- 14. HOXB6: Development
- 15. OLIG1: oligodendroglial differentiation
- 16. PAX2:
- 17. PLAGL1
- 18. PPARGC1B
- RIN3
- 20. SCML4
- 21. T: anterior/posterior pattern formation
tail morphogenesis notochord development
- 22. TAF6L
- 23. TAL2
- 24. TGIF2
- 25. TRAF4: development of the axial skeleton,
the closure of the neural tube
- 26. ZBTB16
- 27. ZBTB4
- 28. ZFPM2: hematopoiesis and cardiogenesis
- 29. ZNF154

Collaborative Network Pipeline

High-throughput
Gene Expression
Data



Genome-Wide
Coexpression Analysis

	TF ₁	TF ₂	TF ₃	...	TF _{m-1}	TF _m
G ₁	0.98	0.76	0.75	...	0.89	0.94
G ₂	0.78	0.83	0.92	...	0.96	0.45
G ₃	0.53	0.63	0.25	...	0.47	0.56
G ₄	0.68	0.73	0.36	...	0.21	0.15
.
G _{n-1}	0.56	0.93	0.74	...	0.36	0.96
G _n	0.38	0.36	0.53	...	0.74	0.46

Coexpression matrix

Ji et al. *Scientific Reports* (2017)

Networks comparison

Treatments

Control

Leaves

Stems

Roots

Tissue-specific
Key regulators

Treatment - specific
Key regulators

MSGAs

SSGA

Leaves

Stems

Roots

	TF ₁	TF ₂	TF ₃	...	TF _{m-1}	TF _m
TF ₁	100	56	0	...	0	25
TF ₂	56	100	35	...	32	0
TF ₃	0	35	100	...	85	0
TF ₄	0	0	0	...	0	0
.	0	.
.	65	.
.	0	.
TF _{m-1}	0	32	85	...	100	0
TF _m	25	0	0	...	0	100

Collaborative network

Triple_link

Collaborative
Subnetworks

Summary for Collaborative Gene Network:

1. Construction of collaborative subnetworks of TFs is novel approach for identifying collaborative TFs that govern the same biological process or complex trait using gene expression data. Compared to markers-assistant approaches for identifying genes that control a trait, collaborative subnetwork construction is a more direct one. It appears to be highly efficient because it is based on the footprints of regulatory events in gene expression profiles rather than the information encoded in DNA sequences, for instance, genome-wide gene association by SNPs.
2. As gene expression data explode, we believe construction of collaborative network will find its way into a wide range of applications for novel biological knowledge discovery.