Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression

Takeshi Obayashi and Kengo Kinoshita

Tohoku University, Japan

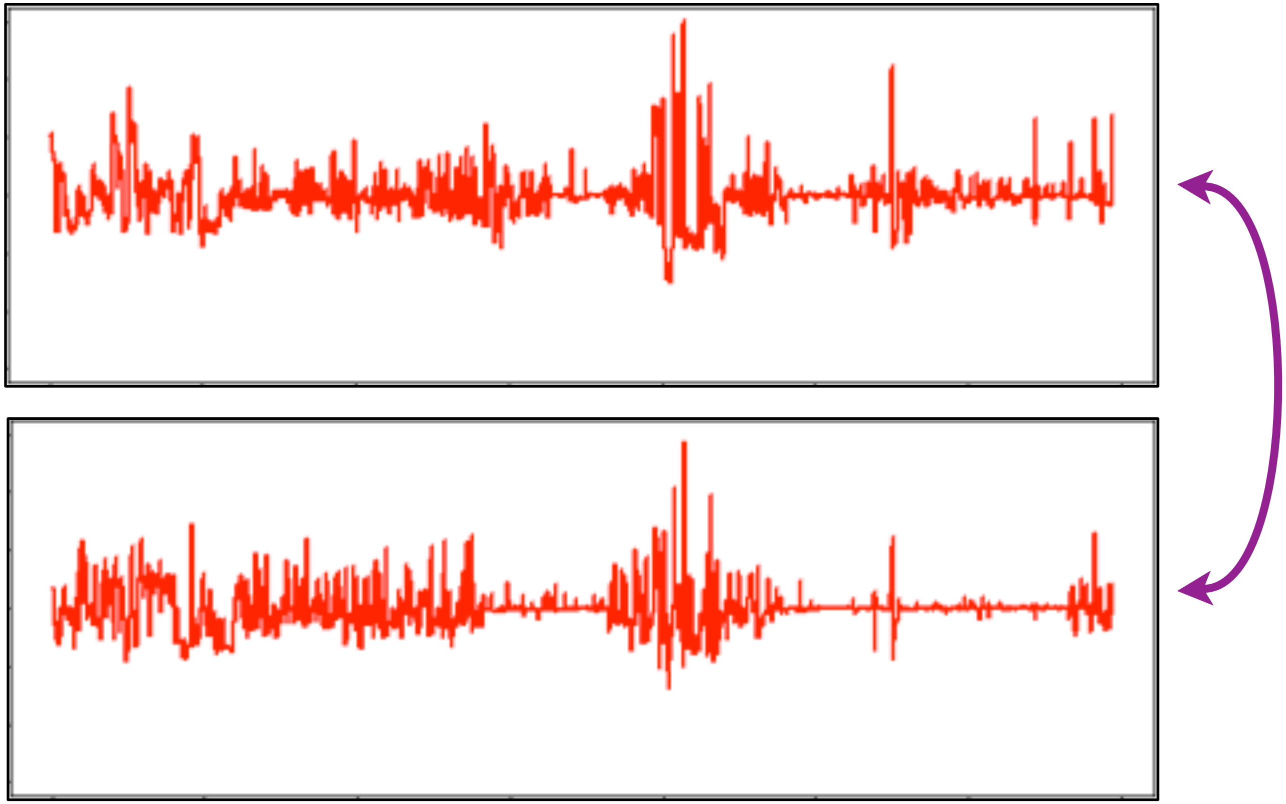
2010 / 3 / 10-12 AYRCOB @ Tainan

Topics about gene coexpression

- What is gene coexpression?
- How can we measure?

Gene coexpression

Pattern of expression changes



Gene A

Gene B

Coexpress
(r=0.7)

Microarray samples stored in public database

Gene coexpression is a similarity of gene expression patterns.

Coexpressed gene database

for Plants

ATTED-II Home

ATTED-II; *Arabidopsis thaliana* trans-factor and cis-element prediction database

last update; Mar, 18, 2008

Home:

- Recent update

[Mar, 18, 2008] Network drawing rule was changed for gene page.
[Mar, 18, 2008] Score for coexpressed gene list was changed to MR from Correlation value.
[Mar, 18, 2008] Coexpressed gene list was changed to include possibly cross-hybridized probes.
[Mar, 18, 2008] Public annotations were updated.
[Feb, 29, 2008] Google Maps representation of coexpressed gene network can be marked for multiple genes. [example]
[Dec, 20, 2007] PPI information includes PPIs for orthologs in other species.
[more]

- Aim of ATTED-II

ATTED-II provides co-regulated gene relationships in *Arabidopsis thaliana* to estimate gene functions.

- Basic contents of ATTED-II
  - Co-expressed genes
    - Data source; publicly available microarray data (58 experiments, 1388 slides) collected by AtGenExpress.
    - Calculation procedure; Pearson's correlation coefficients. [more]
    - Examples;

http://atted.jp

(Since 2003)

- Arabidopsis, rice

for Animals

COXPRESdb version 4.0

COXPRESdb
provides co-regulated gene relationships to estimate gene functions

Search
Obtaining data tables of coexpression or PPI
- GeneTable
- EdgeAnnotation
- TissueDataExtractor
- BLAST
- PPIsearch

Draw
Drawing gene-to-gene relationships with PNG or Postscript outputs
- NetworkDrawer
- HCluster
- CoexViewer
- ExViewer
- GOPathDrawer

Browse
Whole list of human tissues for gene network.

Bulk download
- Gene coexpression
- Subcellular localization

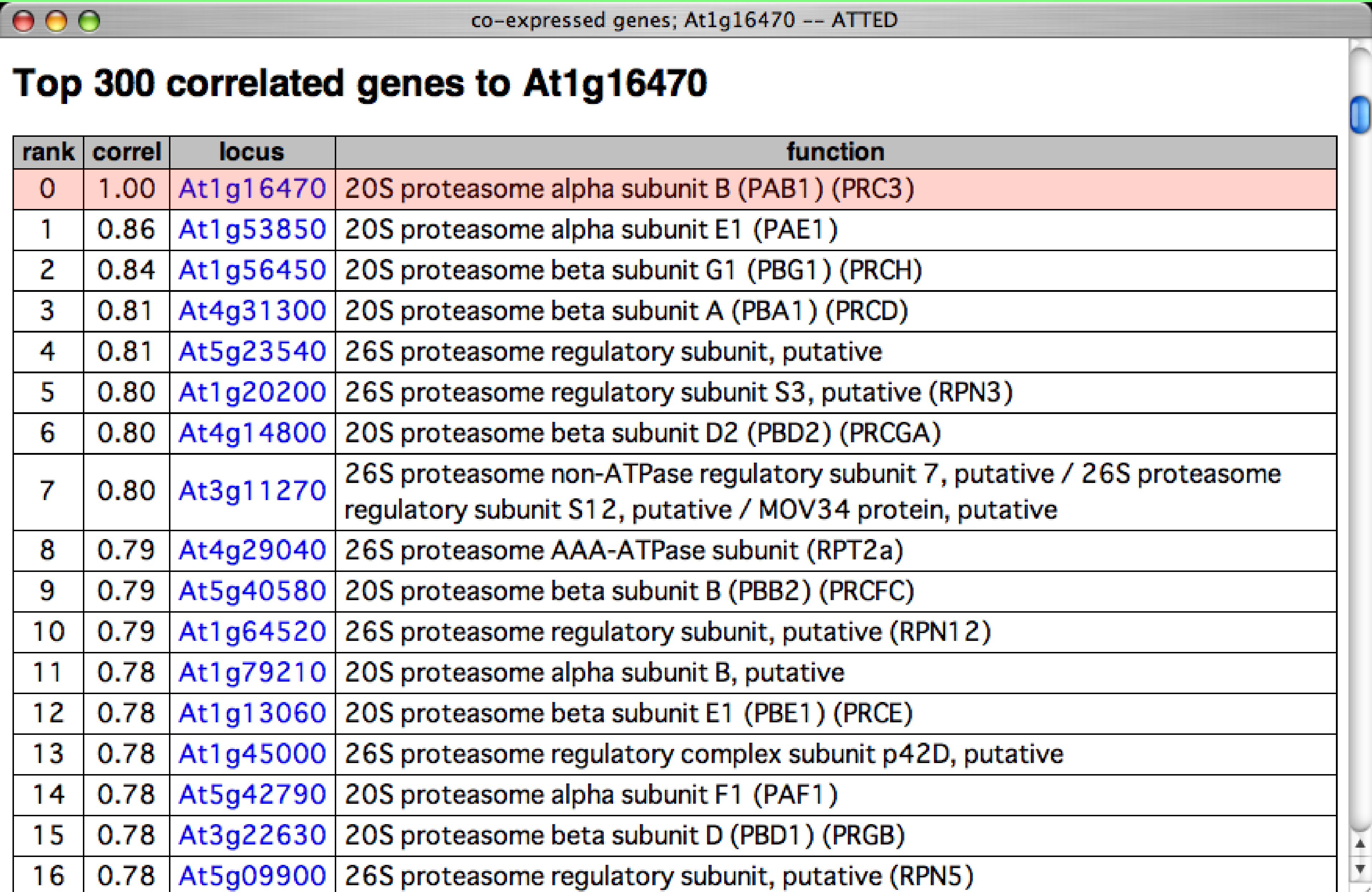
What's new

http://coxpresdb.jp

(Since 2007)

- human, mouse, rat, chicken, zebrafish, fly, nematode

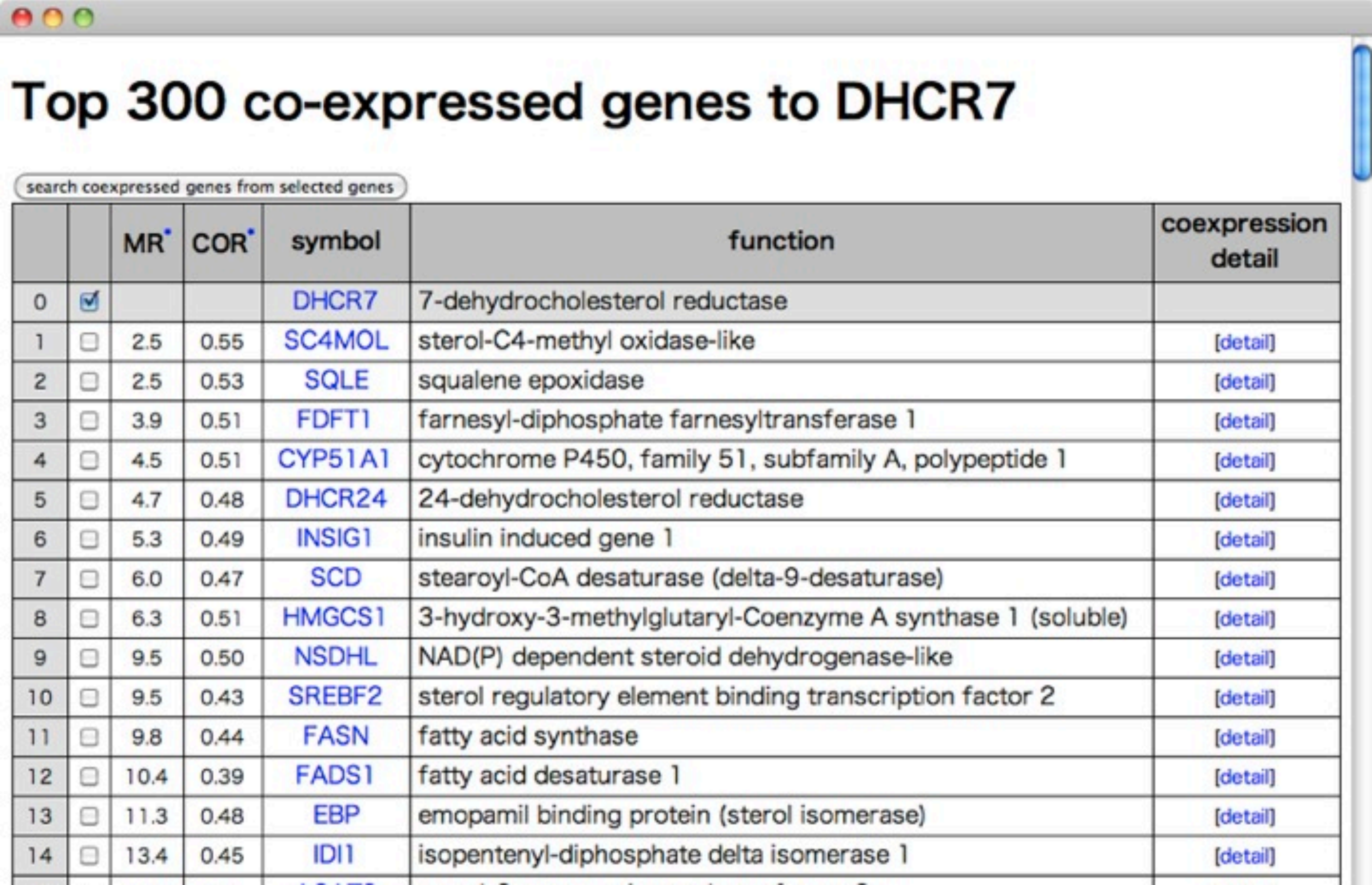
Example for complex protein



Top 300 correlated genes to At1g16470

rank	correl	locus	function
0	1.00	At1g16470	20S proteasome alpha subunit B (PAB1) (PRC3)
1	0.86	At1g53850	20S proteasome alpha subunit E1 (PAE1)
2	0.84	At1g56450	20S proteasome beta subunit G1 (PBG1) (PRCH)
3	0.81	At4g31300	20S proteasome beta subunit A (PBA1) (PRCD)
4	0.81	At5g23540	26S proteasome regulatory subunit, putative
5	0.80	At1g20200	26S proteasome regulatory subunit S3, putative (RPN3)
6	0.80	At4g14800	20S proteasome beta subunit D2 (PBD2) (PRCGA)
7	0.80	At3g11270	26S proteasome non-ATPase regulatory subunit 7, putative / 26S proteasome regulatory subunit S12, putative / MOV34 protein, putative
8	0.79	At4g29040	26S proteasome AAA-ATPase subunit (RPT2a)
9	0.79	At5g40580	20S proteasome beta subunit B (PBB2) (PRCFC)
10	0.79	At1g64520	26S proteasome regulatory subunit, putative (RPN12)
11	0.78	At1g79210	20S proteasome alpha subunit B, putative
12	0.78	At1g13060	20S proteasome beta subunit E1 (PBE1) (PRCE)
13	0.78	At1g45000	26S proteasome regulatory complex subunit p42D, putative
14	0.78	At5g42790	20S proteasome alpha subunit F1 (PAF1)
15	0.78	At3g22630	20S proteasome beta subunit D (PBD1) (PRGB)
16	0.78	At5g09900	26S proteasome regulatory subunit, putative (RPN5)

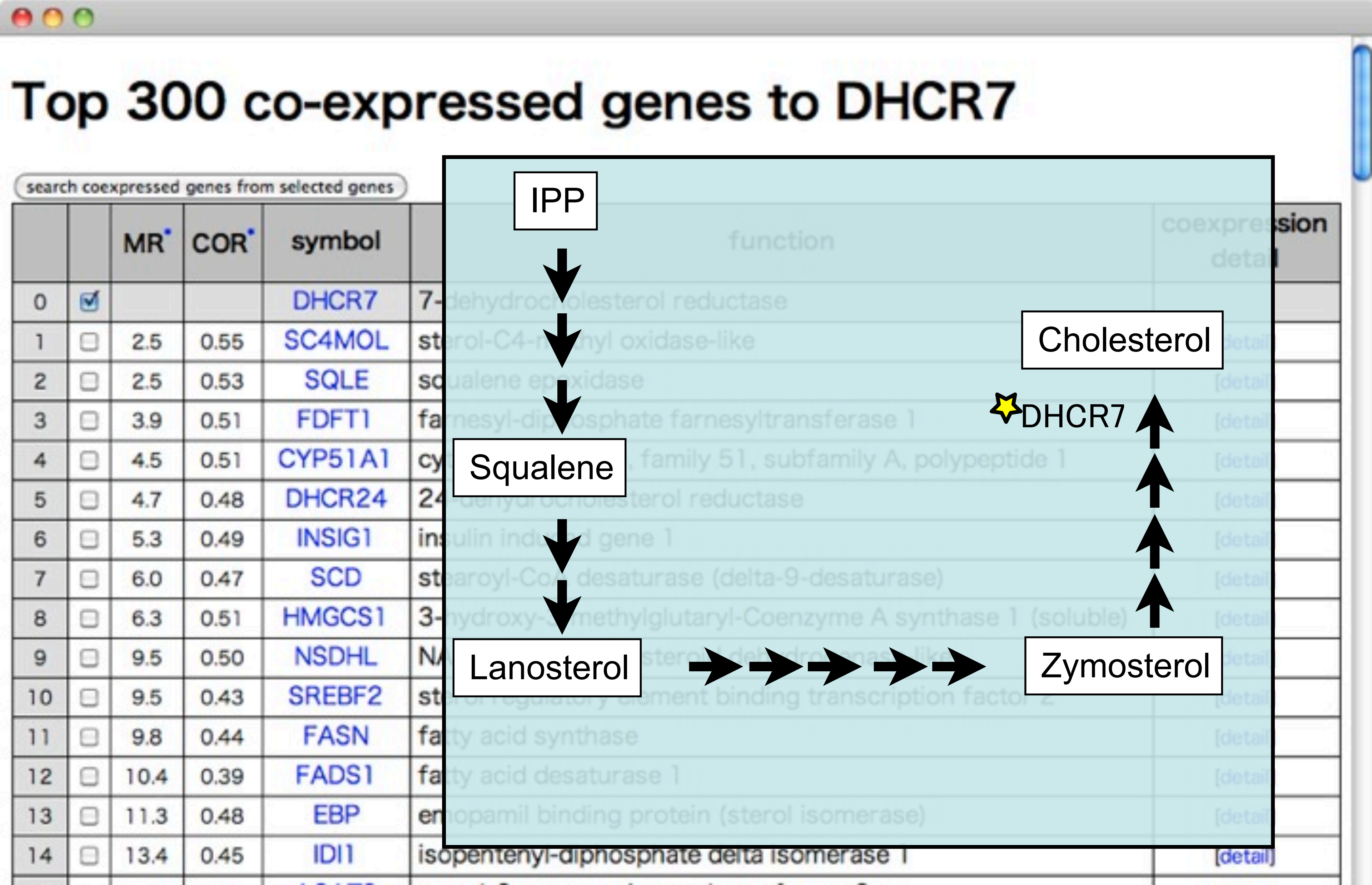
Example for metabolic enzyme



Top 300 co-expressed genes to DHCR7

		MR	COR	symbol	function	coexpression detail
0	☑			DHCR7	7-dehydrocholesterol reductase	
1	☐	2.5	0.55	SC4MOL	sterol-C4-methyl oxidase-like	[detail]
2	☐	2.5	0.53	SQLE	squalene epoxidase	[detail]
3	☐	3.9	0.51	FDFT1	farnesyl-diphosphate farnesyltransferase 1	[detail]
4	☐	4.5	0.51	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1	[detail]
5	☐	4.7	0.48	DHCR24	24-dehydrocholesterol reductase	[detail]
6	☐	5.3	0.49	INSIG1	insulin induced gene 1	[detail]
7	☐	6.0	0.47	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	[detail]
8	☐	6.3	0.51	HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	[detail]
9	☐	9.5	0.50	NSDHL	NAD(P) dependent steroid dehydrogenase-like	[detail]
10	☐	9.5	0.43	SREBF2	sterol regulatory element binding transcription factor 2	[detail]
11	☐	9.8	0.44	FASN	fatty acid synthase	[detail]
12	☐	10.4	0.39	FADS1	fatty acid desaturase 1	[detail]
13	☐	11.3	0.48	EBP	emopamil binding protein (sterol isomerase)	[detail]
14	☐	13.4	0.45	IDI1	isopentenyl-diphosphate delta isomerase 1	[detail]

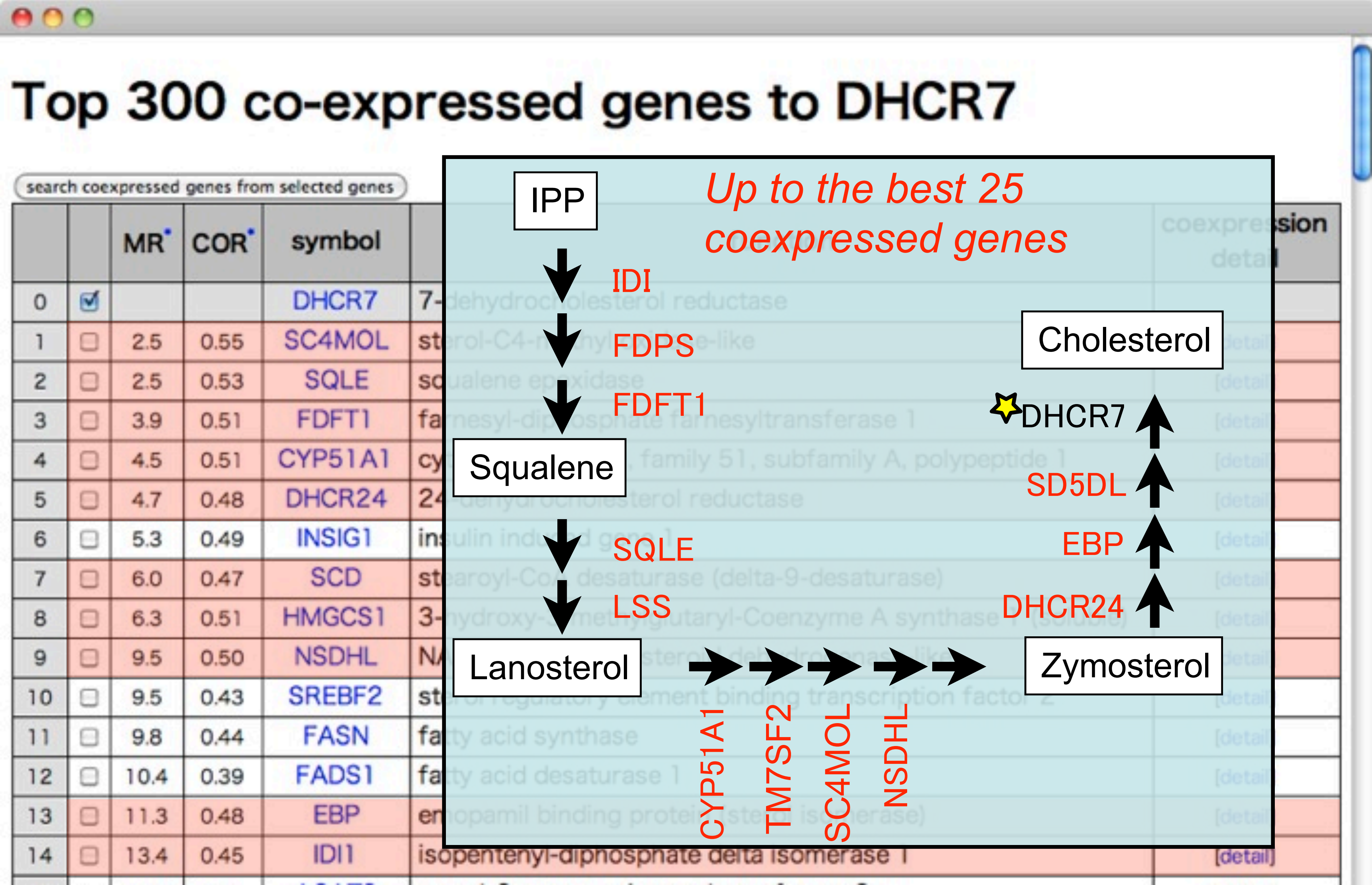
Example for metabolic enzyme



Top 300 co-expressed genes to DHCR7

		MR	COR	symbol	function	coexpression detail
0	☑			DHCR7	7-dehydrocholesterol reductase	
1	☐	2.5	0.55	SC4MOL	sterol-C4-methyl oxidase-like	[detail]
2	☐	2.5	0.53	SQLE	squalene epoxidase	[detail]
3	☐	3.9	0.51	FDFT1	farnesyl-diphosphate farnesyltransferase 1	[detail]
4	☐	4.5	0.51	CYP51A1	cy , family 51, subfamily A, polypeptide 1	[detail]
5	☐	4.7	0.48	DHCR24	24-dehydrocholesterol reductase	[detail]
6	☐	5.3	0.49	INSIG1	insulin induced gene 1	[detail]
7	☐	6.0	0.47	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	[detail]
8	☐	6.3	0.51	HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	[detail]
9	☐	9.5	0.50	NSDHL	NA sterol dehydrogenase-like	[detail]
10	☐	9.5	0.43	SREBF2	sterol regulatory element binding transcription factor 2	[detail]
11	☐	9.8	0.44	FASN	fatty acid synthase	[detail]
12	☐	10.4	0.39	FADS1	fatty acid desaturase 1	[detail]
13	☐	11.3	0.48	EBP	emopamil binding protein (sterol isomerase)	[detail]
14	☐	13.4	0.45	IDI1	isopentenyl-diphosphate delta isomerase 1	[detail]

Example for metabolic enzyme



Top 300 co-expressed genes to DHCR7

		MR	COR	symbol		coexpression detail
0	☑			DHCR7	7-dehydrocholesterol reductase	
1	☐	2.5	0.55	SC4MOL	sterol-C4-methyl oxidase-like	[detail]
2	☐	2.5	0.53	SQLE	squalene epoxidase	[detail]
3	☐	3.9	0.51	FDFT1	farnesyl-diphosphate farnesyltransferase 1	[detail]
4	☐	4.5	0.51	CYP51A1	cy , family 51, subfamily A, polypeptide 1	[detail]
5	☐	4.7	0.48	DHCR24	24-dehydrocholesterol reductase	[detail]
6	☐	5.3	0.49	INSIG1	insulin induced gene 1	[detail]
7	☐	6.0	0.47	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	[detail]
8	☐	6.3	0.51	HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	[detail]
9	☐	9.5	0.50	NSDHL	NA steroid dehydrogenase-like	[detail]
10	☐	9.5	0.43	SREBF2	sterol regulatory element binding transcription factor 2	[detail]
11	☐	9.8	0.44	FASN	fatty acid synthase	[detail]
12	☐	10.4	0.39	FADS1	fatty acid desaturase 1	[detail]
13	☐	11.3	0.48	EBP	emopamil binding protein (sterol isomerase)	[detail]
14	☐	13.4	0.45	IDI1	isopentenyl-diphosphate delta isomerase 1	[detail]

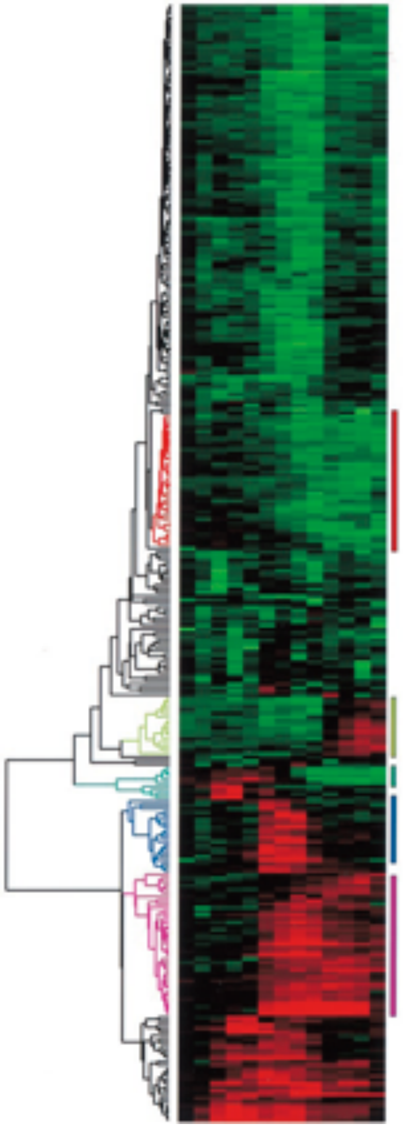
Up to the best 25 coexpressed genes

Topics about gene coexpression

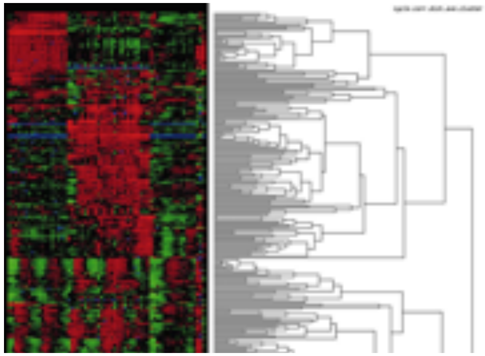
- What is gene coexpression?
- How can we measure?

PCC is a de-facto standard measure

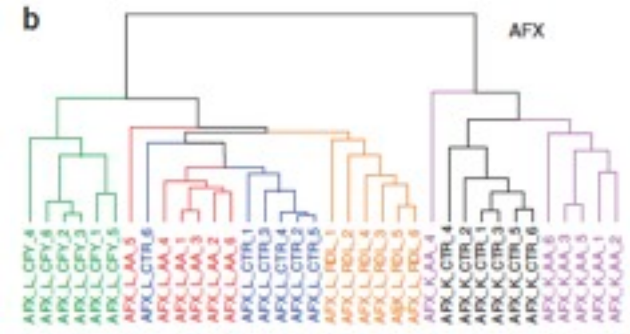
Eisen et al, 1998



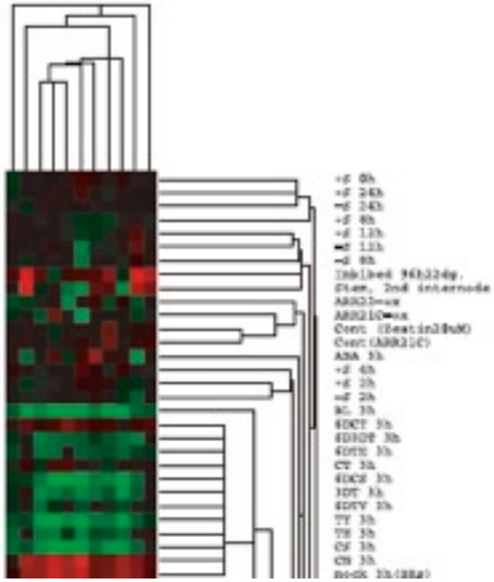
Brazma et al, 2000



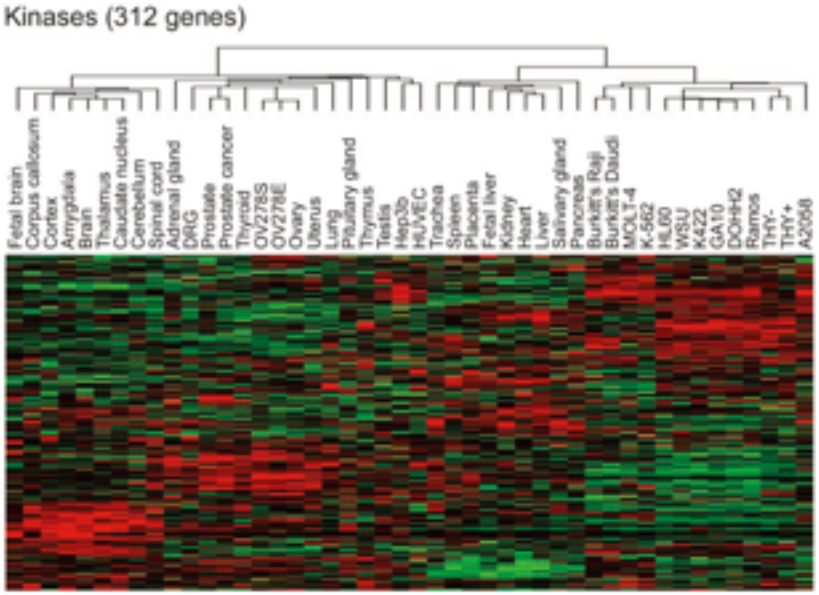
Guo et al, 2006



Goda et al, 2008

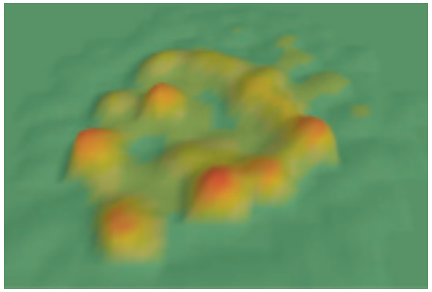


Su et al, 2002

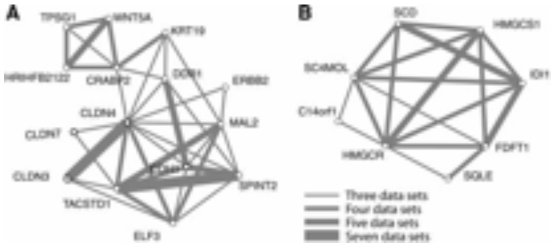


PCC is a de-facto standard measure

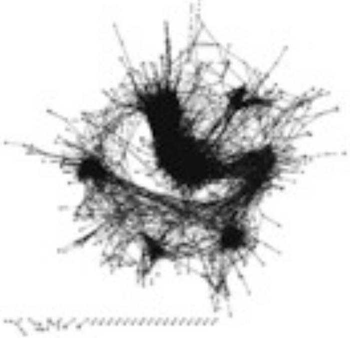
Stuart et al., 2003



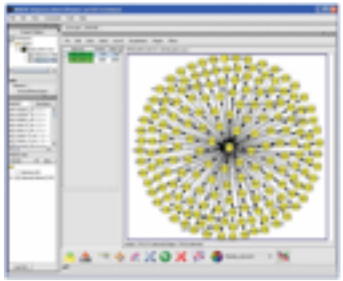
Lee et al., 2004



Jordan et al., 2004

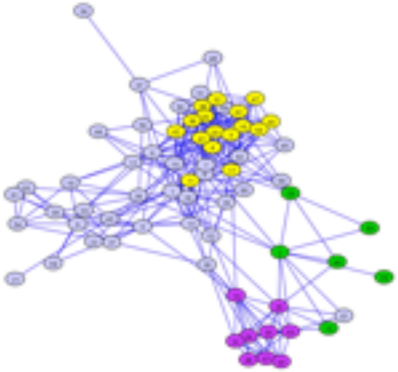


Margolin et al., 2006

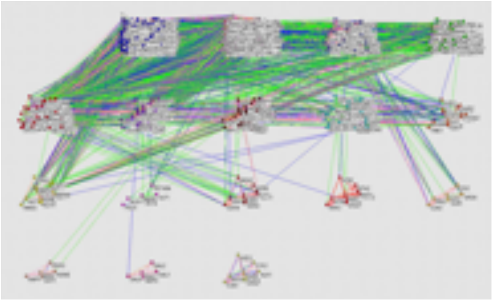


Wei et al., 2006

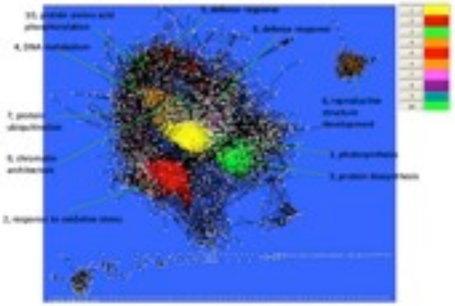
Mentzen and Wurtele, 2008



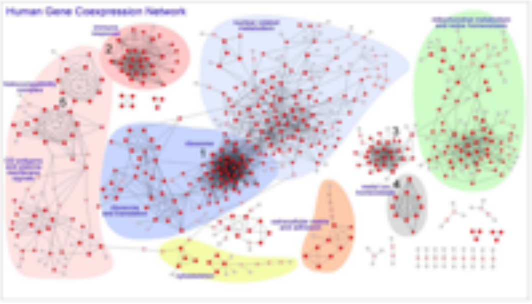
Zhu and Wu, 2008



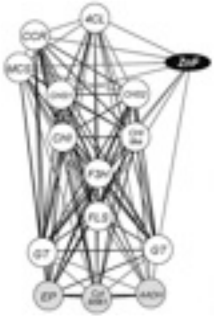
Mao et al., 2009



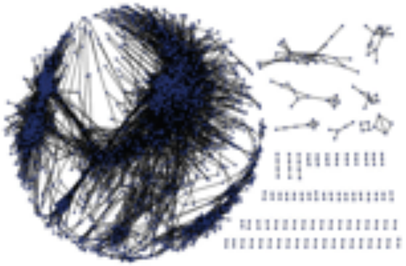
Prieto et al., 2008



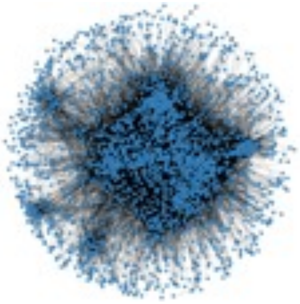
Ozaki et al., 2008



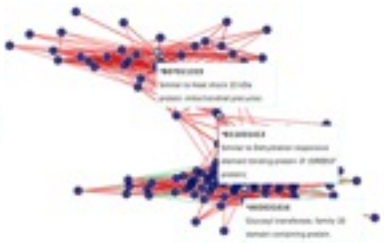
Atias et al., 2009



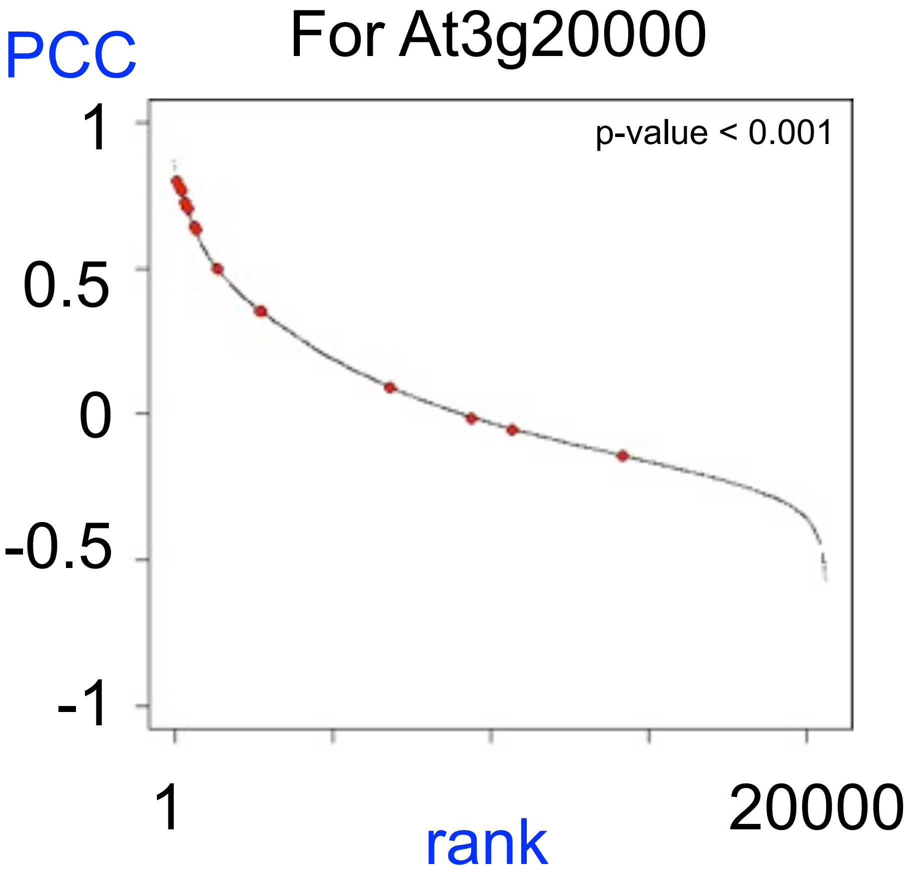
Nayak et al., 2009

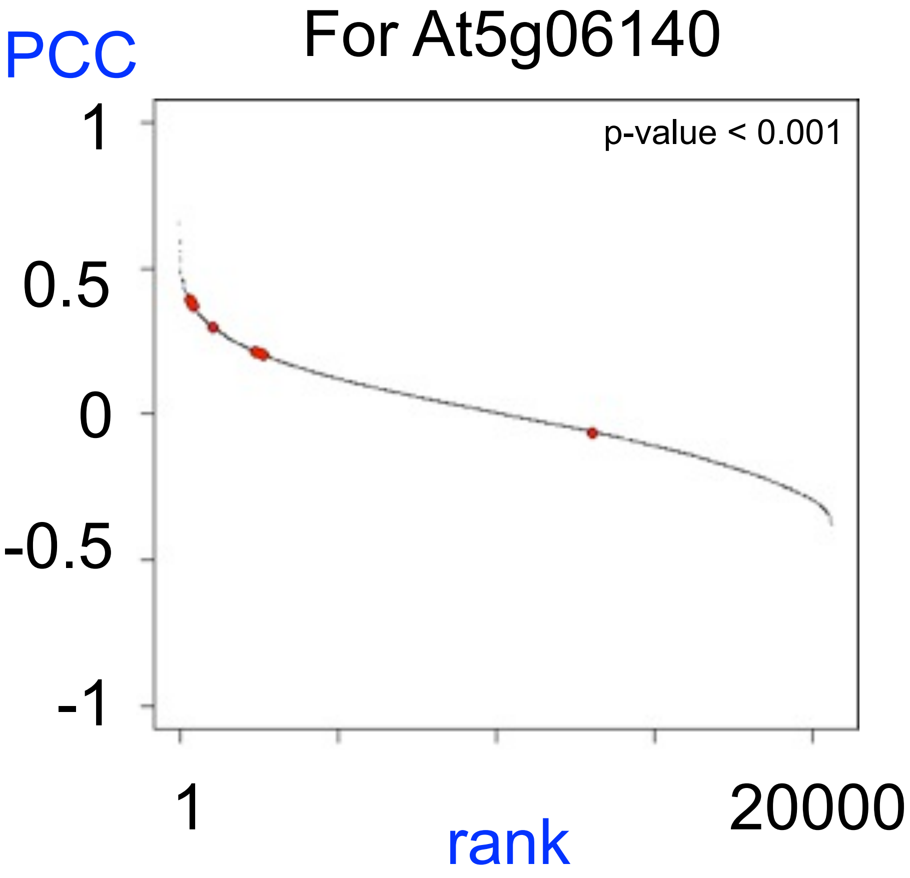


Lee et al., 2009



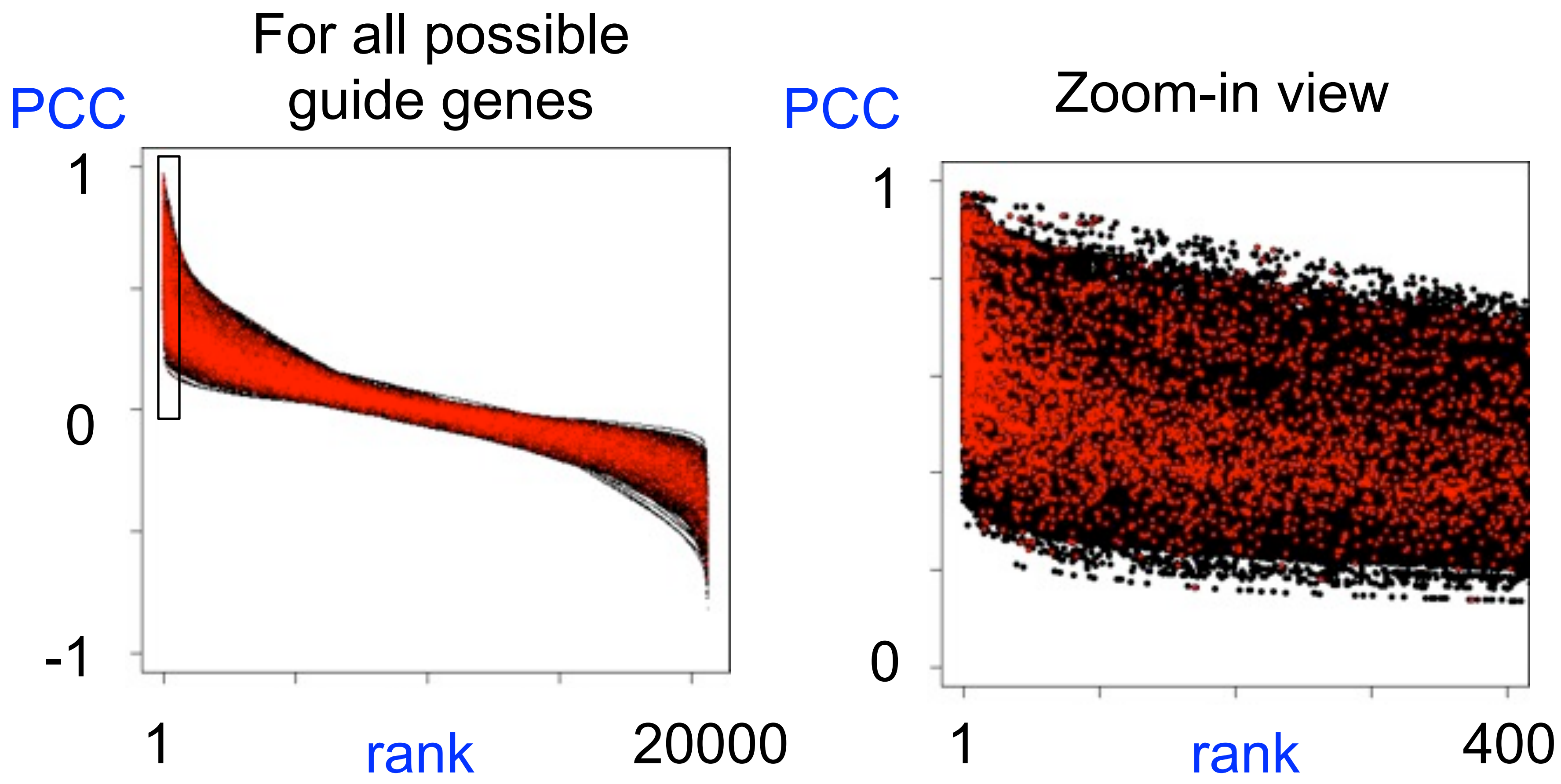
PCC value to predict gene function





Red dot: gene with same
GO BP with the guide gene

PCC value to predict gene function

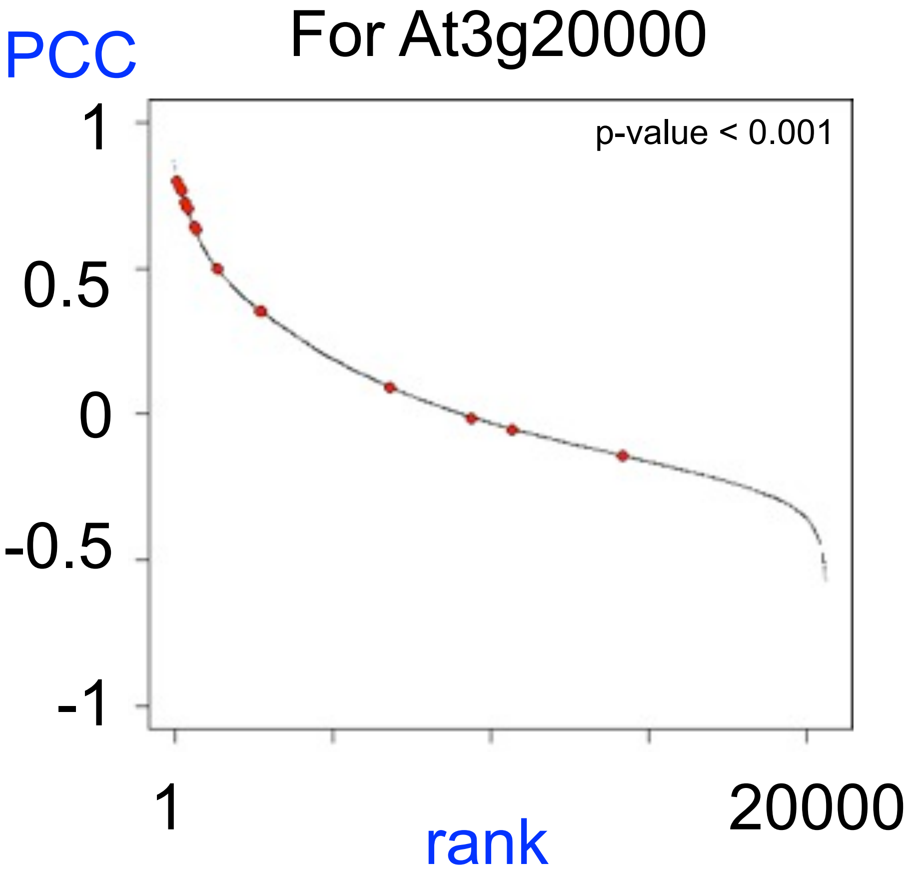


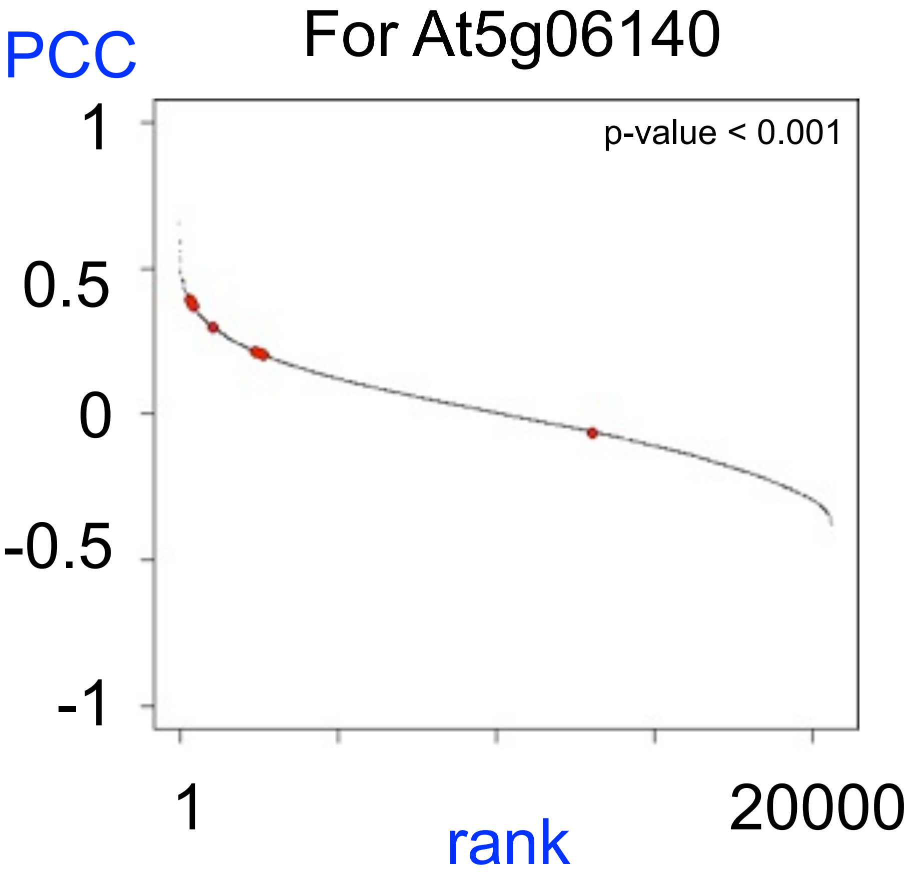
For all possible guide genes

Zoom-in view

Red dot: gene with same GO BP with the guide gene

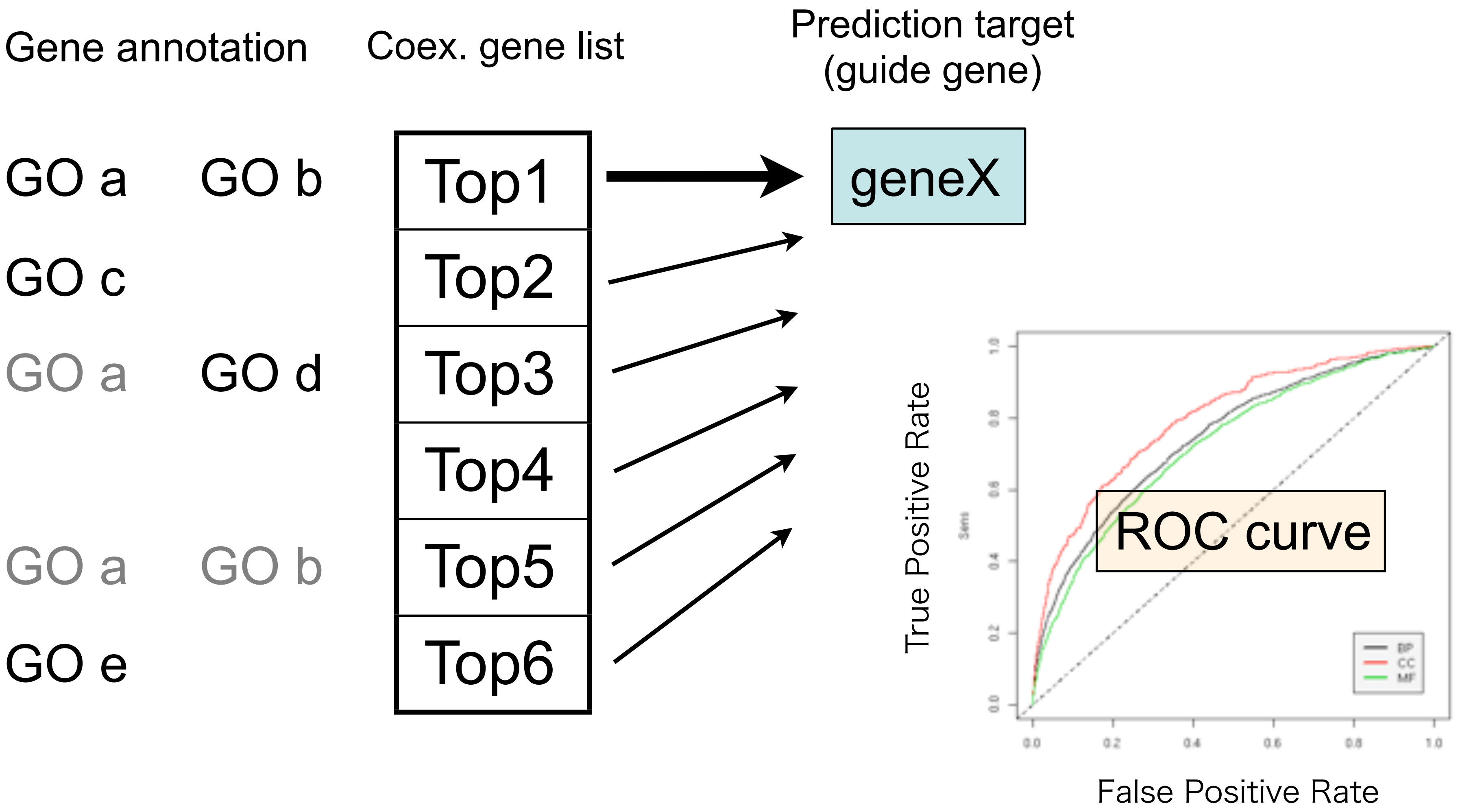
PCC value to predict gene function


For At3g20000


For At5g06140

Red dot: gene with same
GO BP with the guide gene

Outline of gene function prediction



Gene annotation

Coex. gene list

Prediction target
(guide gene)

GO a GO b

GO c

GO a GO d

GO a GO b

GO e

Top1

Top2

Top3

Top4

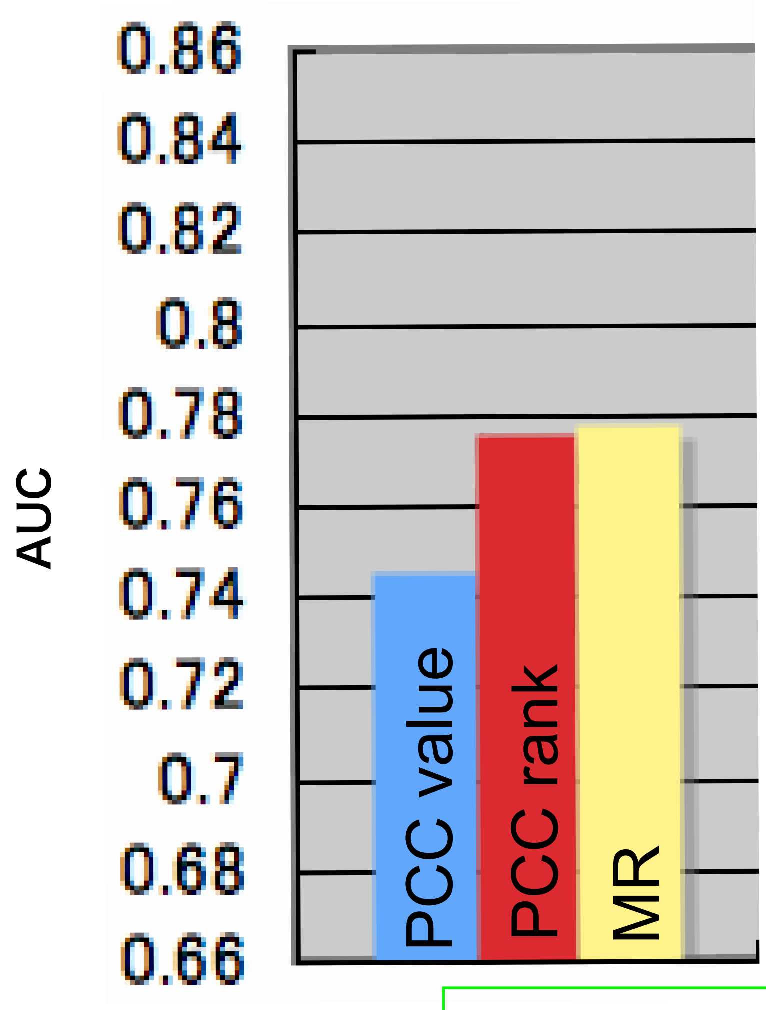
Top5

Top6

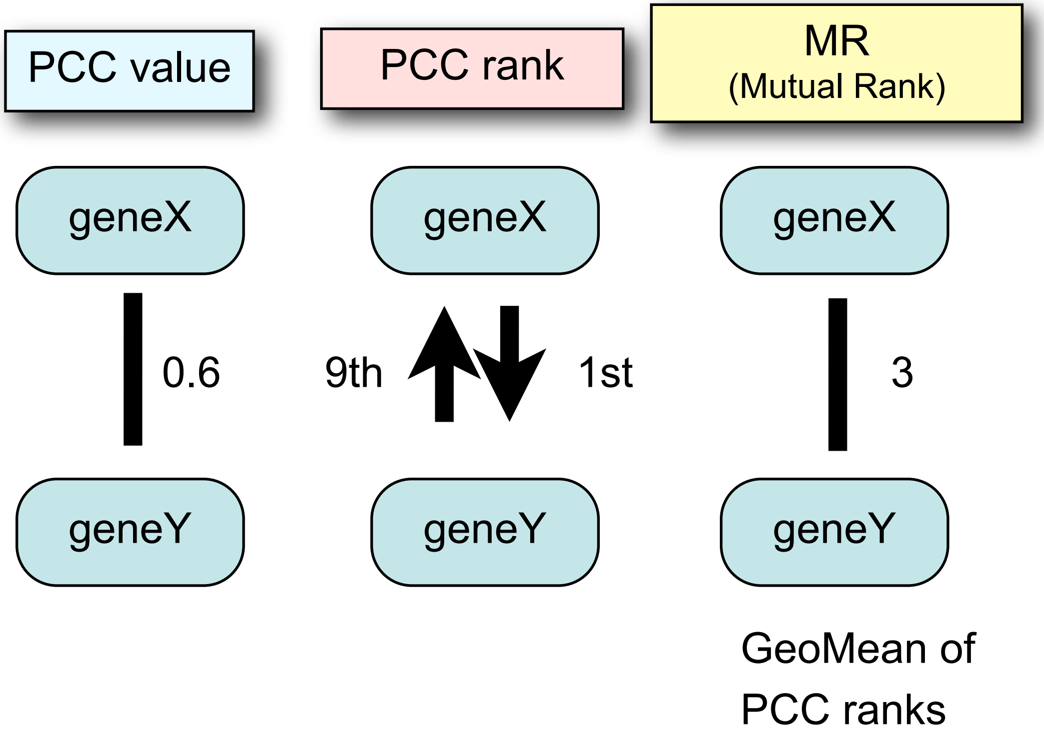
geneX

ROC curve

Result of gene function prediction



1 ; Perfect prediction
0.5; Random prediction



PCC value

geneX

0.6

geneY

PCC rank

geneX

9th

1st

geneY

MR
(Mutual Rank)

geneX

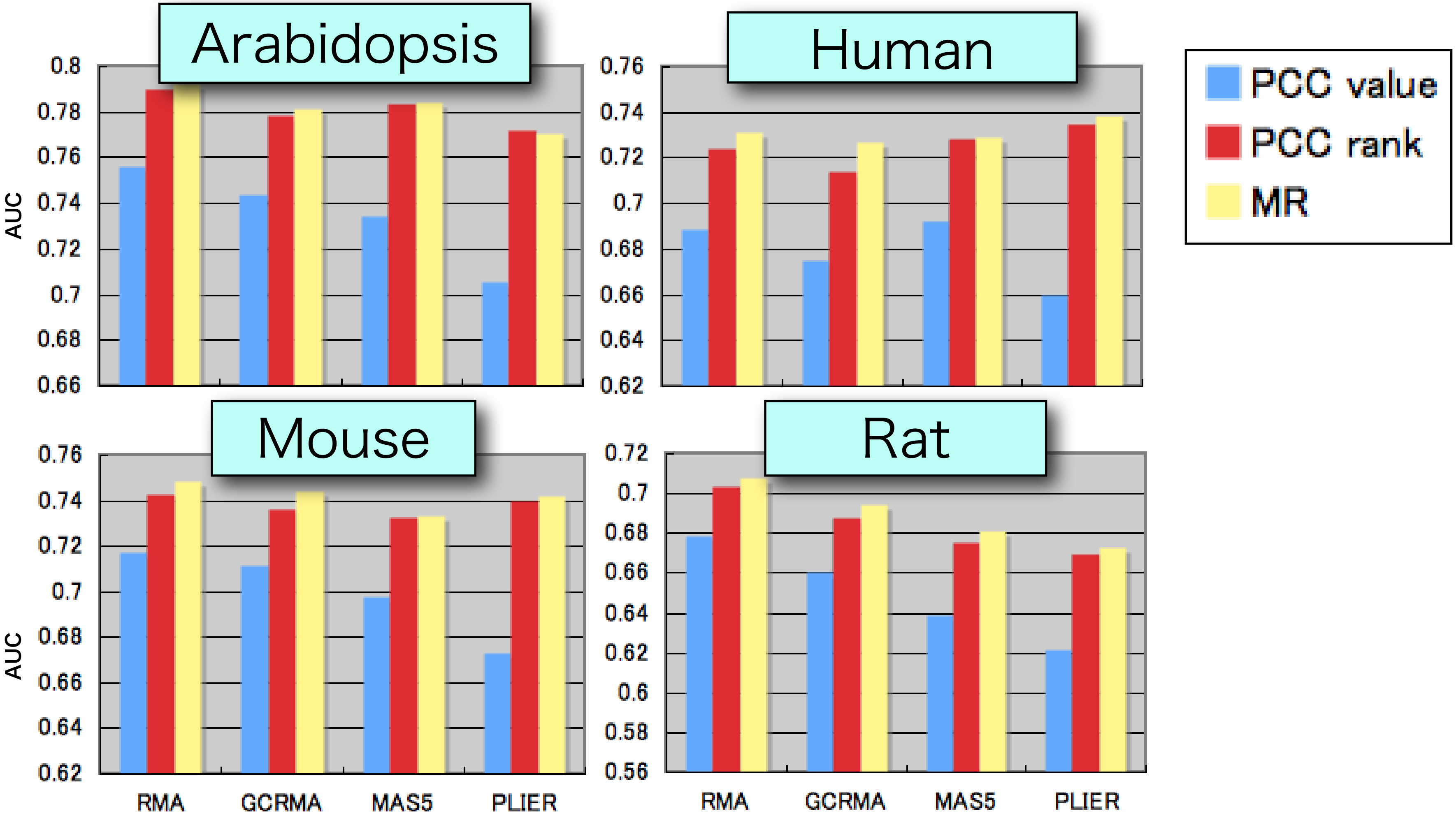
3

geneY

GeoMean of
PCC ranks

In Arabidopsis RMA,
PCC rank and MR are better than PCC value.

Universality of the rank-based measures



Same results were obtained for other data.

Coexpressed gene list from HIST1H2BM

Sorted by Cor

MR	**Cor**	Gene
5.4	**0.29**	HIST1H2AI
11.4	**0.27**	HIST1H1D
11.5	**0.27**	HIST1H3D
96.2	**0.27**	TMEM132A
114.7	**0.27**	KIAA1652
135.9	**0.27**	CHRND
130.9	**0.26**	GRIN3B
183.4	**0.26**	TES
189.3	**0.26**	SSX5
192.8	**0.26**	442503
201.3	**0.26**	SPDEF
208.1	**0.26**	RAP1A
208.4	**0.26**	hCG_1795560
217.4	**0.26**	APOC4
229.6	**0.26**	FOXN4
249.5	**0.26**	ATOH1
264.1	**0.26**	CXXC5

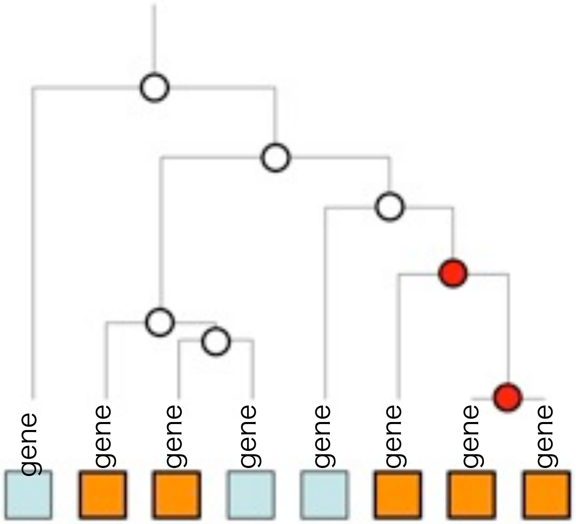
Sorted by MR

MR	Cor	Gene
5.4	0.29	HIST1H2AI
11.4	0.27	HIST1H1D
11.5	0.27	HIST1H3D
28.7	0.25	HIST1H2AJ
74	0.23	HIST1H2BB
79.8	0.23	HIST1H1E
96.2	0.27	TMEM132A
113	0.22	HIST1H4E
114.7	0.27	KIAA1652
124.5	0.24	HIST1H2AB
130.9	0.26	GRIN3B
135.9	0.27	CHRND
145.2	0.21	HIST1H4D
147.8	0.19	HIST1H3C
160.4	0.19	HIST1H3A
166	0.22	HIST1H3B
168.3	0.22	HIST1H2BI

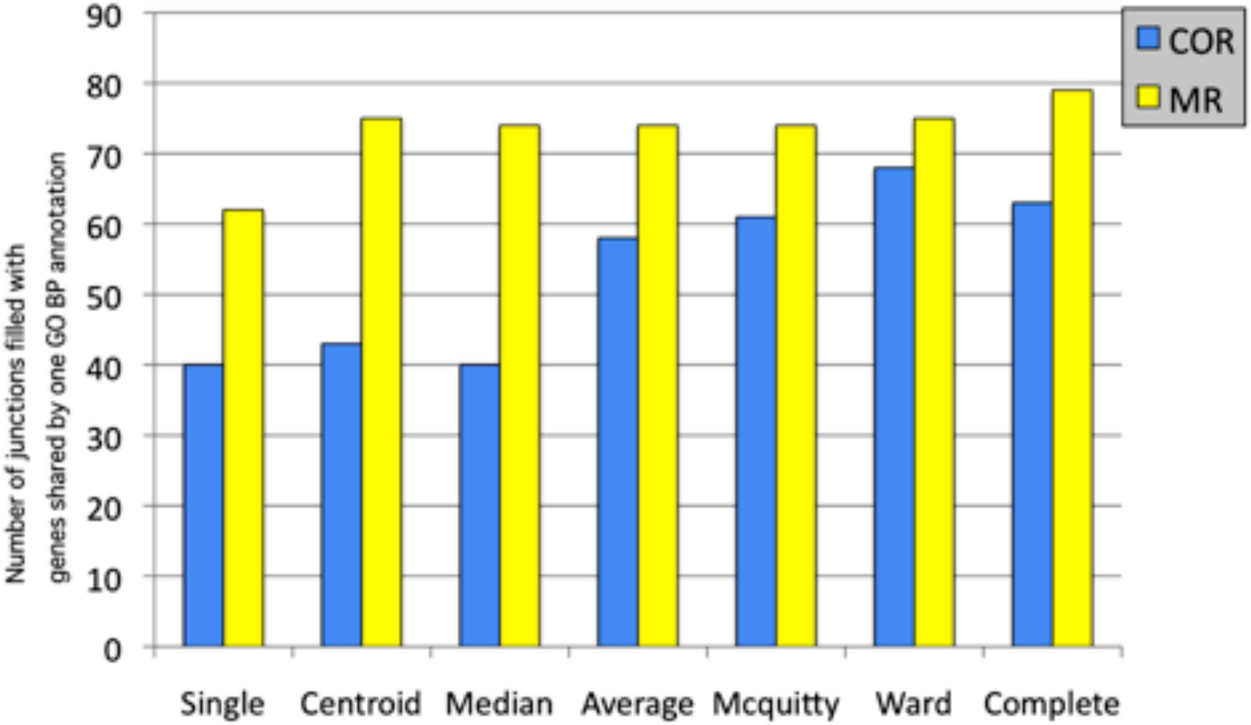
Hierarchical clustering with MR

Measure of goodness:
number of junctions filled with one GO BP annotation

Example



Result



MR was also an effective measure for hierarchical clustering.

Summary

- **Gene coexpression**

It is a similarity of gene expression patterns. It can be used to search functionally related genes.

- **Coexpression measure**

PCC value is universally used as a gene coexpression measure. However we found that PCC rank was more suitable to introduce common threshold for any guide gene.