The analysis of genetic interactions from gene expression data

Robert Castelo

robert.castelo@upf.edu @robertclab

Dept. of Experimental and Health Sciences

Universitat Pompeu Fabra

Barcelona

joint work with

Alberto Roverato

alberto.roverato@unibo.it

Dept. of Statistical Sciences

University of Bologna

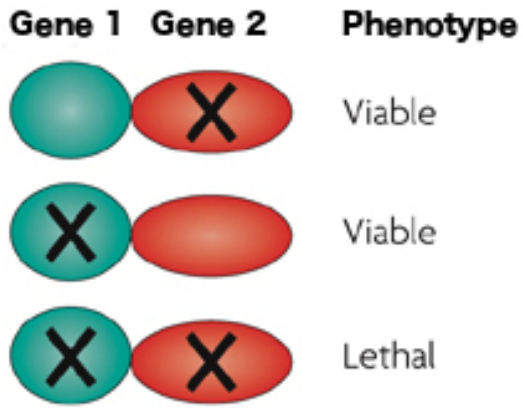
European Bioconductor Meeting

Cambridge, UK, December 5-6th, 2017

Funding: TIN2015-71079-P (MINECO/FEDER, UE)

Genetic interactions

- Most single-gene mutants (one gene deleted at a time) in yeast are viable (Winzeler *et al.*, 1999).
- Human genomes carry on average about 100 LoF variants with 20 genes completely inactivated (MacArthur *et al.*, 2012).
- A *genetic interaction* between two genes occurs when a change in fitness of a double-mutant significantly deviates from the expected change resulting from the combination of the two single mutant fitness effects.



upf.

Genetic interactions - Quantitative interaction profiles

Searching genetic interactions with high-throughput double-mutant screens.

The Genetic Landscape of a Cell

Michael Costanzo,[1,2*] Anastasia Baryshnikova,[1,2*] Jeremy Bellay,[3] Yungil Kim,[3] Eric D. Spear,[4] Carolyn S. Sevier,[4] Huiming Ding,[1,2] Judice L.Y. Koh,[1,2] Kiana Toufighi,[1,2] Sara Mostafavi,[1,5] Jeany Prinz,[1,2] Robert P. St. Onge,[6] Benjamin VanderSluis,[3] Taras Makhnevych,[7] Franco J. Vizeacoumar,[1,2] Solmaz Alizadeh,[1,2] Sondra Bahr,[1,2] Renee L. Brost,[1,2] Yiqun Chen,[1,2] Murat Cokol,[8] Raamesh Deshpande,[3] Zhijian Li,[1,2] Zhen-Yuan Lin,[9] Wendy Liang,[1,2] Michaela Marback,[1,2] Jadine Paw,[1,2] Bryan-Joseph San Luis,[1,2] Ermira Shuteriqi,[1,2] Amy Hin Yan Tong,[1,2] Nydia van Dyk,[1,2] Iain M. Wallace,[1,2,10] Joseph A. Whitney,[1,5] Matthew T. Weirauch,[11] Guoqing Zhong,[1,2] Hongwei Zhu,[1,2] Walid A. Houry,[7] Michael Brudno,[1,5] Sasan Ragibizadeh,[12] Balázs Papp,[13] Csaba Pál,[13] Frederick P. Roth,[8] Guri Giaever,[2,10] Corey Nislow,[1,2] Olga G. Troyanskaya,[14] Howard Bussey,[15] Gary D. Bader,[1,2] Anne-Claude Gingras,[9] Quaid D. Morris,[1,2,5] Philip M. Kim,[1,2] Chris A. Kaiser,[4] Chad L. Myers,[3†] Brenda J. Andrews,[1,2†] Charles Boone[1,2†]

A genome-scale genetic interaction map was constructed by examining 5.4 million gene-gene pairs for synthetic genetic interactions, generating quantitative genetic interaction profiles for ~75% of all genes in the budding yeast, *Saccharomyces cerevisiae*. A network based on genetic interaction profiles reveals a functional map of the cell in which genes of similar biological processes cluster together in coherent subsets, and highly correlated profiles delineate specific pathways to define gene function. The global network identifies functional cross-connections between all bioprocesses, mapping a cellular wiring diagram of pleiotropy. Genetic interaction degree correlated with a number of different gene attributes, which may be informative about genetic network hubs in other organisms. We also demonstrate that extensive and unbiased mapping of the genetic landscape provides a key for interpretation of chemical-genetic interactions and drug target identification.

Costanzo *et al. Science*, 327:425-431, 2010.

A global genetic interaction network maps a wiring diagram of cellular function

Michael Costanzo,[1*] Benjamin VanderSluis,[2,3*] Elizabeth N. Koch,[2*] Anastasia Baryshnikova,[4*] Carles Pons,[2*†] Guihong Tan,[1*] Wen Wang,[2] Matej Usaj,[1] Julia Hanchard,[1,5] Susan D. Lee,[6] Vicent Pelechano,[7‡] Erin B. Styles,[1,5] Maximilian Billmann,[8] Jolanda van Leeuwen,[1] Nydia van Dyk,[1] Zhen-Yuan Lin,[9] Elena Kuzmin,[1,5] Justin Nelson,[2,10] Jeff S. Piotrowski,[1,11§] Tharan Srikumar,[12||] Sondra Bahr,[1] Yiqun Chen,[1] Raamesh Deshpande,[2] Christoph F. Kurat,[1¶] Sheena C. Li,[1,11] Zhijian Li,[1] Mojca Mattiazzi Usaj,[1] Hiroki Okada,[13] Natasha Pascoe,[1,5] Bryan-Joseph San Luis,[1] Sara Sharifpoor,[1] Emira Shuteriqi,[1] Scott W. Simpkins,[2,10] Jamie Snider,[1] Harsha Garadi Suresh,[1] Yizhao Tan,[1] Hongwei Zhu,[1] Noel Malod-Dognin,[14] Vuk Janjic,[15] Natasa Przulj,[14,16] Olga G. Troyanskaya,[3,4] Igor Stagljar,[1,5,17] Tian Xia,[2,18] Yoshikazu Ohya,[13] Anne-Claude Gingras,[5,9] Brian Raught,[12] Michael Boutros,[8] Lars M. Steinmetz,[7,19] Claire L. Moore,[6] Adam P. Rosebrock,[1,5] Amy A. Caudy,[1,5] Chad L. Myers,[2,10#] Brenda Andrews,[1,5#] Charles Boone[1,5,11#]

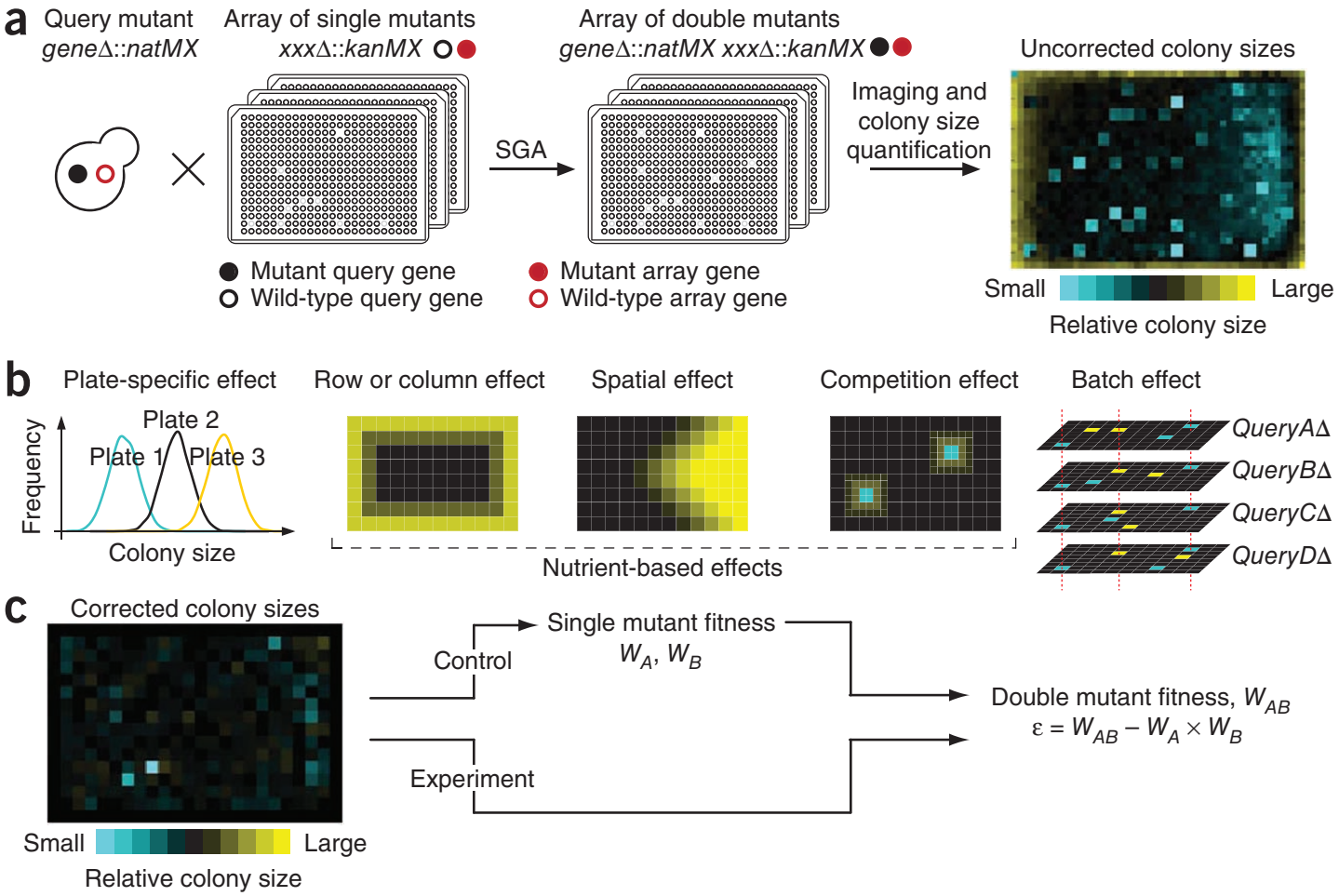
We generated a global genetic interaction network for *Saccharomyces cerevisiae*, constructing more than 23 million double mutants, identifying about 550,000 negative and about 350,000 positive genetic interactions. This comprehensive network maps genetic interactions for essential gene pairs, highlighting essential genes as densely connected hubs. Genetic interaction profiles enabled assembly of a hierarchical model of cell function, including modules corresponding to protein complexes and pathways, biological processes, and cellular compartments. Negative interactions connected functionally related genes, mapped core bioprocesses, and identified pleiotropic genes, whereas positive interactions often mapped general regulatory connections among gene pairs, rather than shared functionality. The global network illustrates how coherent sets of genetic interactions connect protein complex and pathway modules to map a functional wiring diagram of the cell.

Costanzo *et al. Science*, 353:1381-1396, 2016.

upf.

Genetic interactions - SGA scores

A so-called *Synthetic Genetic Array (SGA) score* measures experimentally the difference between the observed fitness of a double-mutant and the fitness expected from the combination of two single-gene mutants.



Genetic interactions - Limitations of double-mutant screens

- High-throughput double-mutant screens grow quadratically in the number of genes and an exhaustive exploration becomes infeasible in plants or animals.

- Genetic interactions may occur conditionally to their cellular context. Our experimental condition (cellular state, tissue, etc.) of interest may be very different from the ones used in publicly available double-mutant screens.

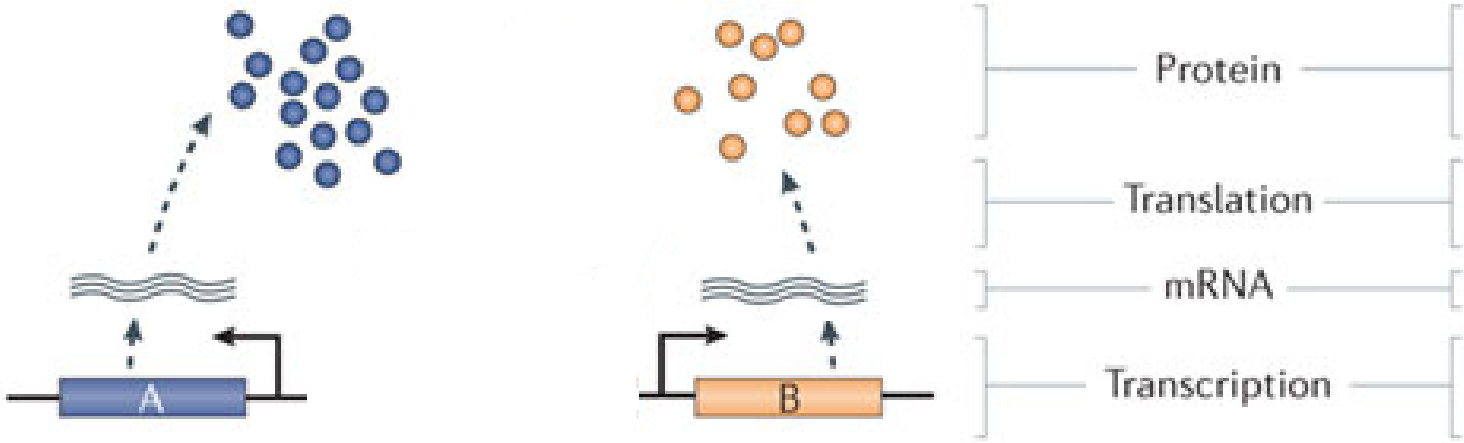
- One may attempt to *predict* genetic interactions on the basis of sequence and evolutionary features, and gene expression profiled on our condition of interest.

upf.

Genetic interactions - Buffering mechanisms

- One mechanism behind genetic interactions consists of a buffering relationship between two genes.

- One type of buffering releationship is the one by which two genes are positively *coexpressed*.

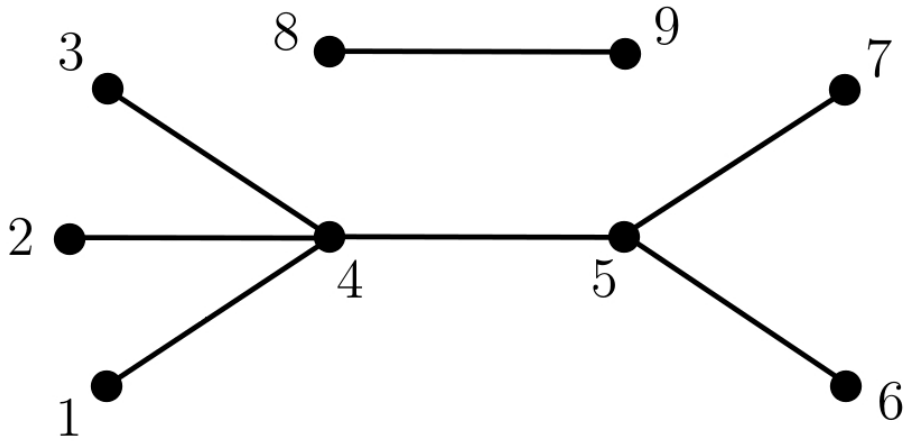


- Positive coexpression, however, occurs mostly between pairs of genes co-operating on the same cellular function. The coexpressed interacting pair should form an "important" edge within the cellular "network".

upf.

Standard coexpression: partial correlation

Build an inverse covariance matrix K of the following graph, such that **all** edges have identical non-zero partial correlation values.




```
> K
```

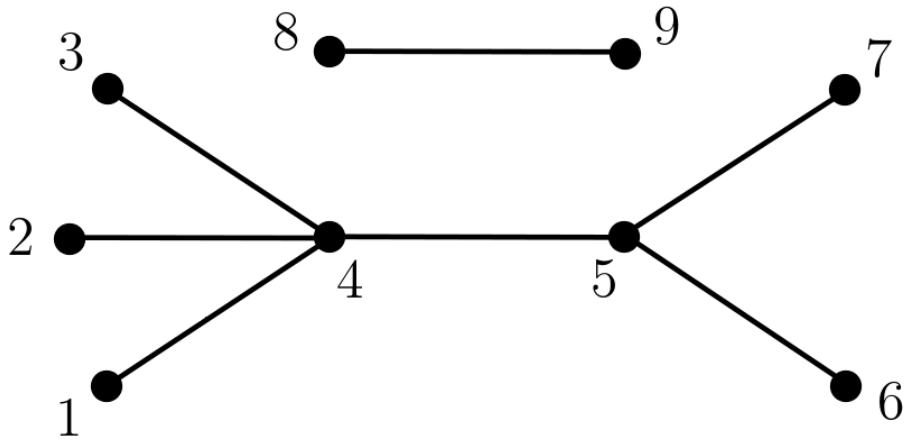
	1	2	3	4	5	6	7	8	9
1	1.0	0.0	0.0	-0.4	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	-0.4	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	1.0	-0.4	0.0	0.0	0.0	0.0	0.0
4	-0.4	-0.4	-0.4	1.0	-0.4	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	-0.4	1.0	-0.4	-0.4	0.0	0.0
6	0.0	0.0	0.0	0.0	-0.4	1.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	-0.4	0.0	1.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	-0.4
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	1.0

- Non-zero partial correlations indicate what edges are present.
- **Problem:** Because all partial correlations are identical, all edges can be considered equally important under this correlation measure.

upf.

Standard coexpression: Pearson (marginal) correlation

Calculate the corresponding covariance matrix $\Sigma = K^{-1}$.




```
> S <- solve(K)
> round(cov2cor(S), digits=1)
```

	1	2	3	4	5	6	7	8	9
1	1.0	0.4	0.4	0.6	0.4	0.2	0.2	0.0	0.0
2	0.4	1.0	0.4	0.6	0.4	0.2	0.2	0.0	0.0
3	0.4	0.4	1.0	0.6	0.4	0.2	0.2	0.0	0.0
4	0.6	0.6	0.6	1.0	0.7	0.4	0.4	0.0	0.0
5	0.4	0.4	0.4	0.7	1.0	0.5	0.5	0.0	0.0
6	0.2	0.2	0.2	0.4	0.5	1.0	0.3	0.0	0.0
7	0.2	0.2	0.2	0.4	0.5	0.3	1.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.4
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.0

- Marginal (Pearson) correlations are not all of them identical between the different edges, and therefore, not equally important.
- **Problem:** Non-zero marginal (Pearson) correlations occur between every pair of connected vertices.

upf.

The networked partial correlation

Solution: The networked partial correlation (Roverato and Castelo, *J. R. Stat. Soc. Ser. C*, 66:647-665, 2017; DOI:10.1111/rssc.12166; #OA).

$$\psi_{xy\cdot(V\setminus\{x,y\})} = \frac{\rho_{xy\cdot V\setminus\{x,y\}}}{1-\rho^2_{(xy)(V\setminus\{x,y\})}}, \tag{1}$$

where,

- $\rho_{xy\cdot V\setminus\{x,y\}}$ is the **partial correlation** of X_x and X_y given $X_{V\setminus\{x,y\}}$ and captures the presence of the edge $\{x, y\}$in the network.

- $\rho_{(xy)(V\setminus\{x,y\})}$ is the **vector correlation coefficient** between $X_{\{x,y\}}$ and $X_{V\setminus\{x,y\}}$ and captures the strength of the association between $\mathbf{X}_{\{x,y\}}$ and the remaining genes $\mathbf{X}_{V\setminus\{x,y\}}$.

- If $\{x, y\}$ are disconnected from the rest of the network then $\psi_{xy\cdot(V\setminus\{x,y\})} = \rho_{xy\cdot V\setminus\{x,y\}}$.

upf.

The networked partial correlation

Using the R/BioC package qpgraph and the function qpPathWeight():

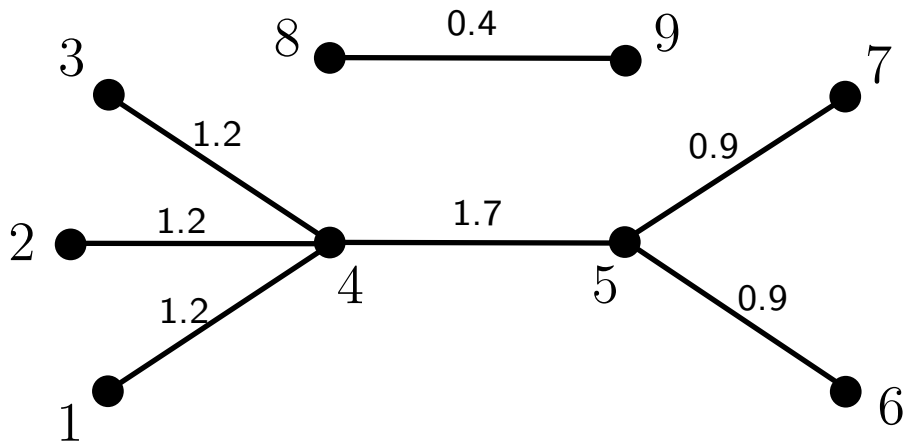
```
> edg
```

```
     [,1] [,2]
[1,]    1    4
[2,]    2    4
[3,]    3    4
[4,]    4    5
[5,]    5    6
[6,]    5    7
[7,]    8    9
```

```
> library(qpgraph)
> npc <- sapply(1:nrow(edg), function(i) qpPathWeight(S, edg[i, ]))
> npcmat <- K
> npcmat[edg] <- npcmat[cbind(edg[, 2], edg[, 1])] <- npc
```

```
> round(npcmat, digits=1)
```

```
   1   2   3   4   5   6   7   8   9
1 1.0 0.0 0.0 1.2 0.0 0.0 0.0 0.0 0.0
2 0.0 1.0 0.0 1.2 0.0 0.0 0.0 0.0 0.0
3 0.0 0.0 1.0 1.2 0.0 0.0 0.0 0.0 0.0
4 1.2 1.2 1.2 1.0 1.7 0.0 0.0 0.0 0.0
5 0.0 0.0 0.0 1.7 1.0 0.9 0.9 0.0 0.0
6 0.0 0.0 0.0 0.0 0.9 1.0 0.0 0.0 0.0
7 0.0 0.0 0.0 0.0 0.9 0.0 1.0 0.0 0.0
8 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.4
9 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.4 1.0
```



Modeling SGA scores with gene expression data

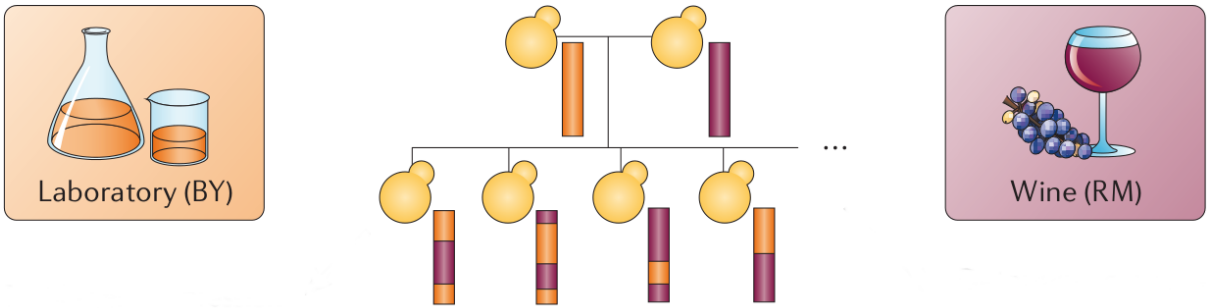
Profile gene expression on an experimental condition of interest and attempt to predict quantative interaction profiles from gene expression data.

The landscape of genetic complexity across 5,700 gene expression traits in yeast

Rachel B. Brem*† and Leonid Kruglyak*‡

*Division of Human Biology and ‡Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Communicated by Leland H. Hartwell, Fred Hutchinson Cancer Research Center, Seattle, WA, November 23, 2004 (received for review August 11, 2004)



Brem and Kruglyak (2005) crossed two strains of yeast (BY×RM) to produce $n = 112$ offspring whose gene expression were profiled using microarray chips.

Y	1	2	$\ldots$	n
g_1	y_{11}	y_{12}	$\cdots$	y_{2n}
g_2	y_{21}	y_{22}	$\cdots$	y_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
g_p	y_{p1}	y_{p2}	$\cdots$	y_{pn}

upf.

Modeling SGA scores with gene expression data

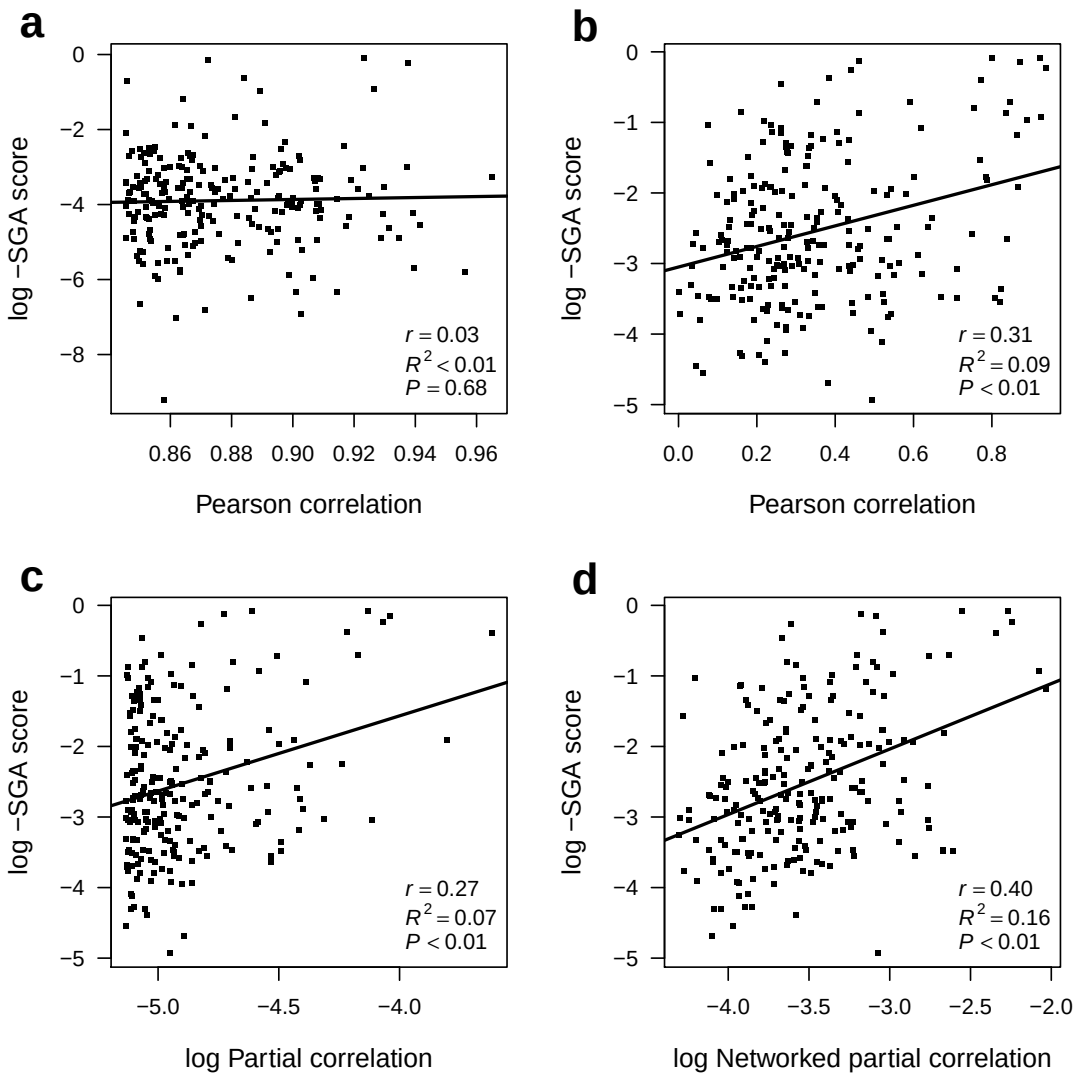
Networked partial correlations explain a larger fraction of the variability of SGA scores from Costanzo *et al.* (2010).

(a) Top-k largest absolute Pearson correlations.

(b) Pearson correlations of the significant non-zero partial correlations.

(c) Significant non-zero partial correlations.

(d) Networked partial correlations.



Concluding remarks

- The networked partial correlation (NPC) derives from the covariance decomposition over the paths of an undirected graph, which is the undirected counterpart of Sewall Wright's path analysis. Check out the paper if you want to find out how! (DOI:10.1111/rssc.12166).

- NPCs can explain more variability of quantitive interaction profiles than classical coexpression measures such as Pearson or partial correlations.

- Calculations on high-dimensional gene expression data are currently done using shrinkage estimates of partial correlations and sparse canonical correlations (CRAN GeneNet and PMA packages). An unitary approach to calculate NPCs may be more efficient (implementing it in `qpgraph`).

- Looking for graduate students willing to work in the subject! Get in touch if you're interested (robert.castelo@upf.edu).

upf.