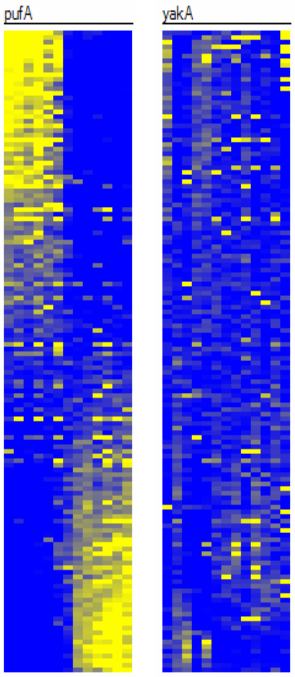


Računska fenomika

Blaž Zupan

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Department of Molecular and Human Genetics
Baylor College of Medicine, Houston, TX

Outline

- Phenotype
- Phenotypes in functional genomics
  - analysis of epistasis
- Toward global-scale phenotypes
- Utility of global-scale phenotypes and systems biology

Phenotype

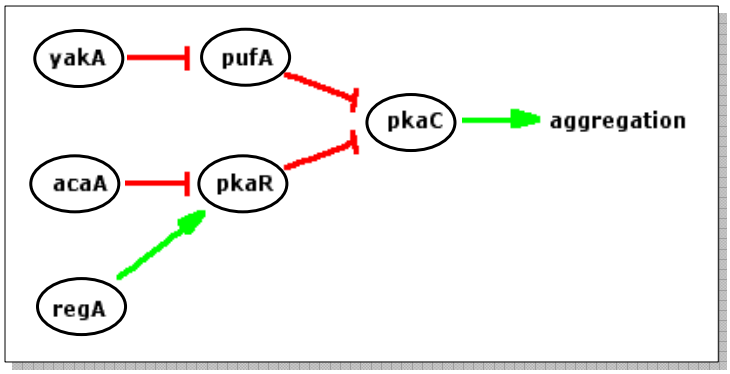
The phenotype of an individual organism is either its total physical appearance and constitution or a **specific manifestation of a trait**, such as size, eye color, or behavior that varies between individuals.

en.wikipedia.org

Because phenotypes are much easier to observe than genotypes (it doesn't take chemistry or sequencing to determine a person's eye color), **classical genetics uses phenotypes to deduce the functions of genes**.

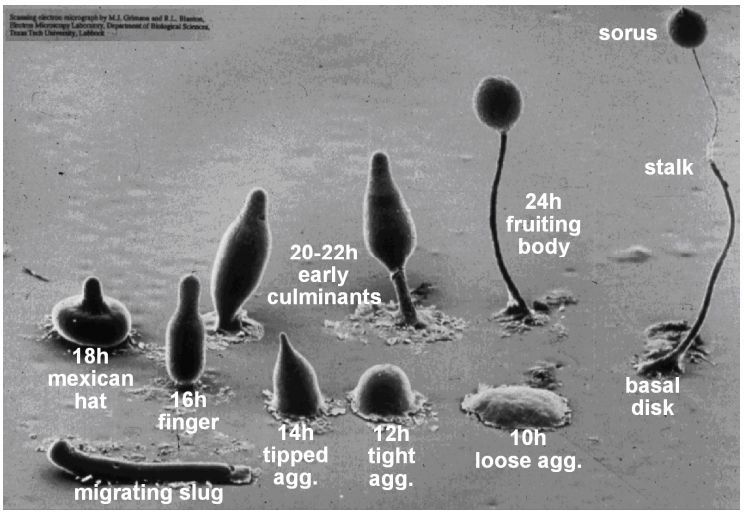
en.wikipedia.org

Gene Regulation Networks

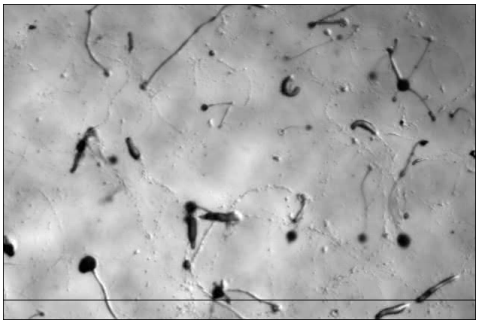


Aggregation of *D. discoideum*

Development Cycle of *Dictyostelium discoideum*

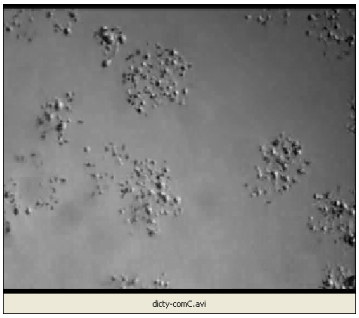


Dicty: Wild Type & Mutant



A full time course of wild type Dicty development showing aggregation, slug migration and fruiting body formation.

A full time course of a mutant (*comC*-) that aggregates with highly branched aggregation streams but fails to maintain the integrity of the aggregate. The aggregates fall apart (individual cells seen leaving the aggregate).

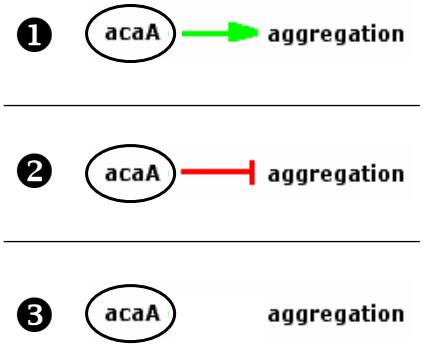


Experimental Data

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++

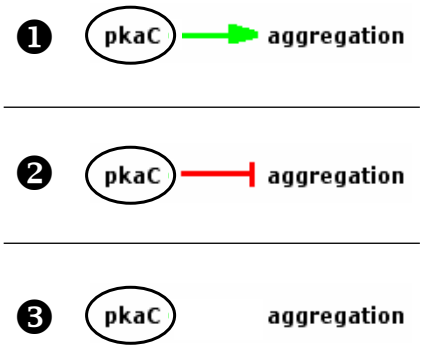
From Data to Networks

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++

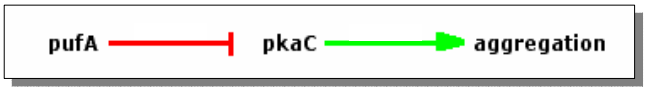


From Data to Networks

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++



Epistasis

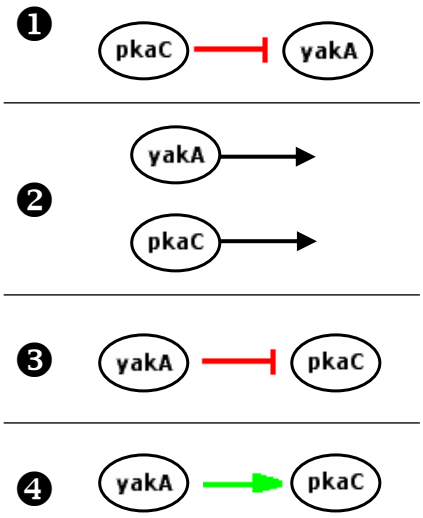


From Wikipedia:

Epistasis is the masking of the phenotypic effects of one gene by alleles of another at another locus. A gene is epistatic when its presence suppresses the effect of the other gene. Epistatic genes are also known as inhibiting genes due to their effects on other genes which are described as hypostatic.

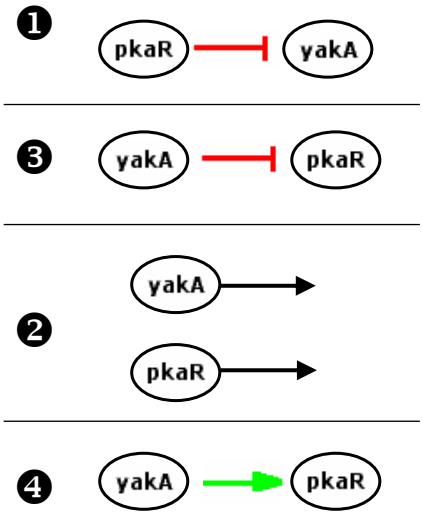
From Data to Networks

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++



From Data to Networks

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++



Reasoning Patterns

Epistasis

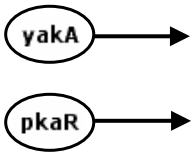
IF two different mutations (in genes A, B) result in two different phenotypes AND the phenotype of the double gene mutation is identical to the phenotype of the gene B mutation, THEN gene B is epistatic to gene A.

Influence

IF a mutation in a gene changes the phenotype relative to an otherwise identical strain, THEN the gene influences the biological process.

Parallelism

IF mutations in either gene have an effect on the biological process AND the phenotype of the double mutant is different from either mutation alone THEN two genes are in parallel genetic paths



Inference of Genetic Networks

genetic experiments

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++

❶

pattern-based inference

ID	Gene	Influence	Bio. Entity
A3	yakA	->	aggregation
A4	pufA	-|	aggregation
A5	pkaR	-|	aggregation
A6	pkaC	->	aggregation
A7	acaA	->	aggregation
A8	regA	-|	aggregation
A9	regA	-|	pkaC
A10	yakA	-|	pufA
A11	pkaR	||	yakA

❸

relations btw genes and outcomes

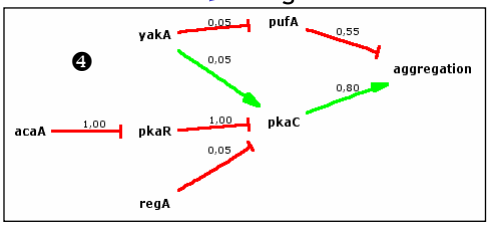
constraint satisfaction

genetic network

prior knowledge

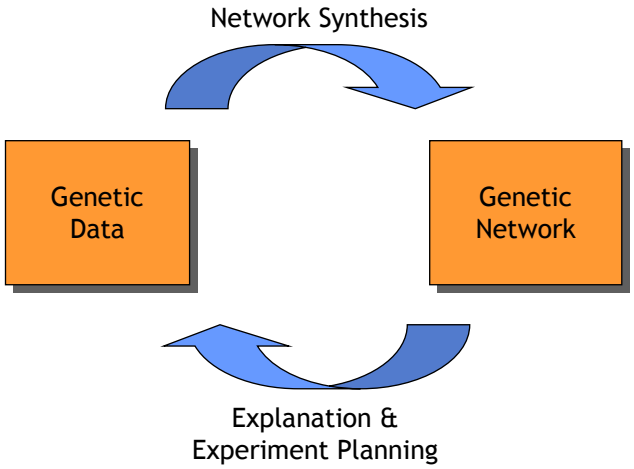
ID	Gene	Influence	Bio. Entity
P1	pkaR	-|	pkaC
P2	acaA	-|	pkaR

❷

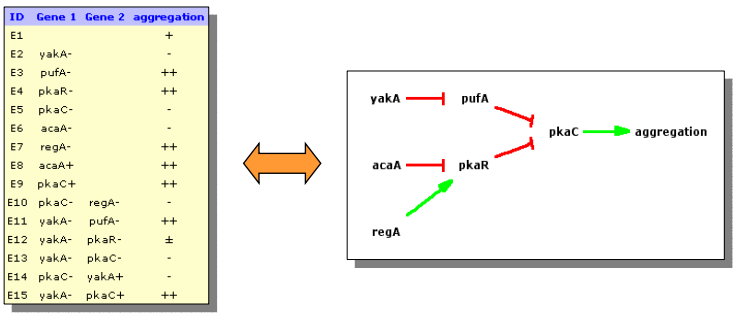


Zupan et al. *Bioinformatics* 2003.
Juvan et al. *Nucl Acids Res* 2005.

www.genepath.org

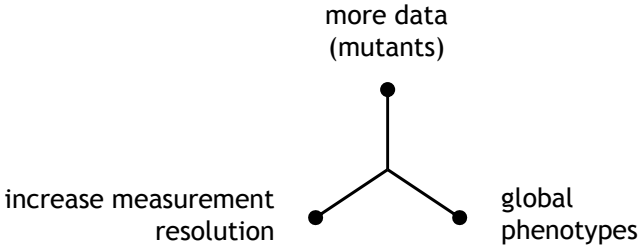


Zupan et al. *Bioinformatics* 2003.
Juvan et al. *Nucl Acids Res* 2005.



ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++

Systems Biology?



Phenotype

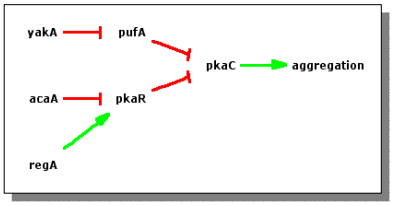
The phenotype of an individual organism is either its **total physical appearance and constitution** or a specific manifestation of a trait, such as size, eye color, or behavior that varies between individuals.

en.wikipedia.org

"New" Phenotypes

In whole genome studies typical of systems biology classical morphological phenotypes should be complemented with **surrogates that can encode the state of the entire organism** and provide increased resolution.

ID	Gene 1	Gene 2	aggregation
E1			+
E2	yakA-		-
E3	pufA-		++
E4	pkaR-		++
E5	pkaC-		-
E6	acaA-		-
E7	regA-		++
E8	acaA+		++
E9	pkaC+		++
E10	pkaC-	regA-	-
E11	yakA-	pufA-	++
E12	yakA-	pkaR-	±
E13	yakA-	pkaC-	-
E14	pkaC-	yakA+	-
E15	yakA-	pkaC+	++

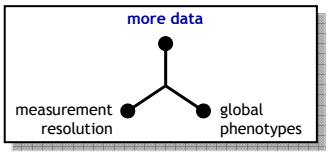


Systems Biology?

more data
(mutants)

increase
measurement
resolution

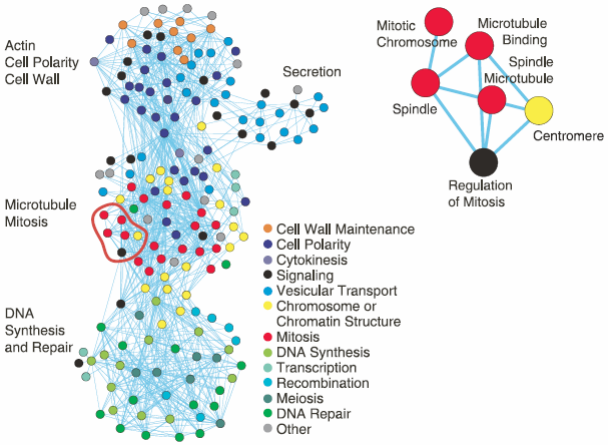
use global
phenotypes



SCIENCE VOL 303 6 FEBRUARY 2004

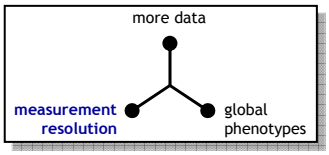
Global Mapping of the Yeast Genetic Interaction Network

Amy Hin Yan Tong,[1,2]* Guillaume Lesage,[3]* Gary D. Bader,[4] Huiming Ding,[1] Hong Xu,[1,2] Xiaofeng Xin,[1,2] James Young,[6] Gabriel F. Berriz,[7] Renee L. Brost,[1] Michael Chang,[5] YiQun Chen,[1] Xin Cheng,[1] Gordon Chua,[1] Helena Friesen,[2] Debra S. Goldberg,[7] Jennifer Haynes,[2] Christine Humphries,[2] Grace He,[1] Shamiza Hussein,[3] Lizhu Ke,[1] Nevan Krogan,[1,2] Zhijian Li,[1,2] Joshua N. Levinson,[3] Hong Lu,[1] Patrice Ménard,[3] Christella Munyana,[3] Ainslie B. Parsons,[1,2] Owen Ryan,[1] Raffi Tonikian,[1,2] Tania Roberts,[5] Anne-Marie Sdicu,[3] Jesse Shapiro,[3] Bilal Sheikh,[1] Bernhard Suter,[8] Sharyl L. Wong,[7] Lan V. Zhang,[7] Hongwei Zhu,[1] Christopher G. Burd,[9] Sean Munro,[10] Chris Sander,[4] Jasper Rine,[8] Jack Greenblatt,[1,2] Matthias Peter,[11] Anthony Bretscher,[6] Graham Bell,[3] Frederick P. Roth,[7] Grant W. Brown,[5] Brenda Andrews,[2]† Howard Bussey,[3]† Charles Boone[1,2]†



A genetic interaction network containing ~1000 genes and ~4000 interactions was mapped by crossing mutations in 132 different query genes into a set of ~4700 viable gene yeast deletion mutants and scoring the double mutant progeny for fitness defects.

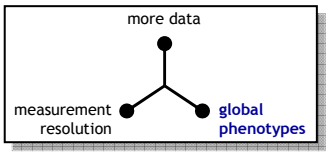
132 x 4.700 = 620.400



PROPHECY
PROfiling of PHEnotypic Characteristics in Yeast

http://prophecy.lundberg.gu.se
Blomberg Lab
Göteborg University

ORF	Name	Reference
YCR104w	PAU3	
YEL033w	YEL033w	
YEL048c	YEL048c	
YMR029c	FAR8	
YMR246w	FAA4	



Phenotype

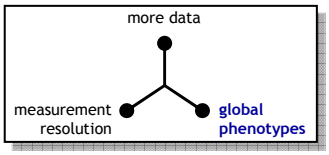
The phenotype of an individual organism is either its **total physical appearance and constitution** or a specific manifestation of a trait, such as size, eye color, or behavior that varies between individuals.

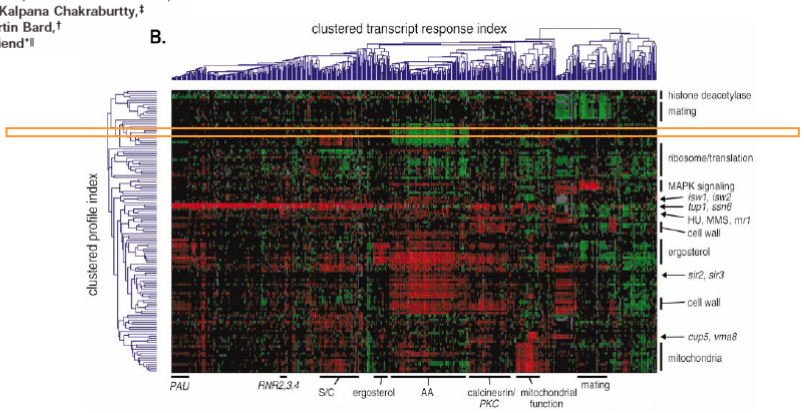
en.wikipedia.org

Cell, Vol. 102, 109–126, July 7, 2000, Copyright ©2000 by Cell Press

Functional Discovery via a Compendium of Expression Profiles

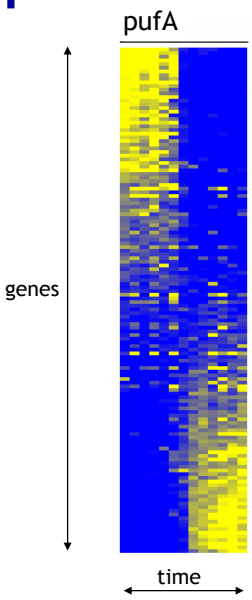
Timothy R. Hughes,*# Matthew J. Marton,*# Allan R. Jones,* Christopher J. Roberts,* Roland Stoughton,* Christopher D. Armour,* Holly A. Bennett,* Ernest Coffey,* Hongyue Dai,* Yudong D. He,* Matthew J. Kidd,* Amy M. King,* Michael R. Meyer,* David Slade,* Pek Y. Lum,* Sergey B. Stepaniants,* Daniel D. Shoemaker,* Daniel Gachotte,† Kalpana Chakraburtty,‡ Julian Simon,§ Martin Bard,† and Stephen H. Friend*‖



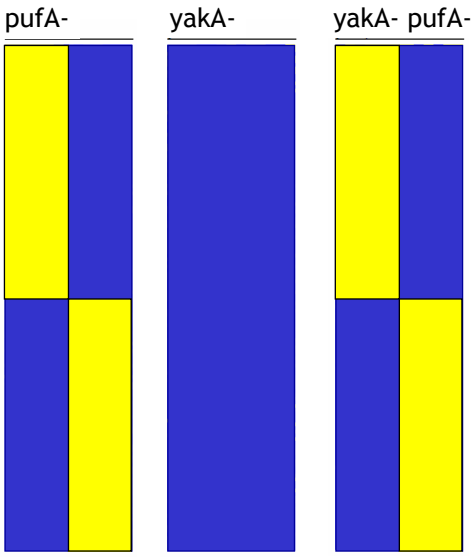


(B) Two-dimensional agglomerative hierarchical clustering of 127 experiments and 568 genes, selected to include only experiments with 2 or more genes up- or downregulated greater than 3-fold, and significant at $P \leq 0.01$; and only genes that are up- or downregulated at greater than 3-fold, and at $P \leq 0.01$, in 2 or more experiments. *PAU, PAU* gene family; S/C, stress and carbohydrate metabolism; AA, amino acid biosynthesis; *PKC*, responsive to protein kinase C. See Supplemental Data for a version with all gene and experiment names.

Transcriptional Phenotypes



Epistasis by Transcriptional Phenotypes



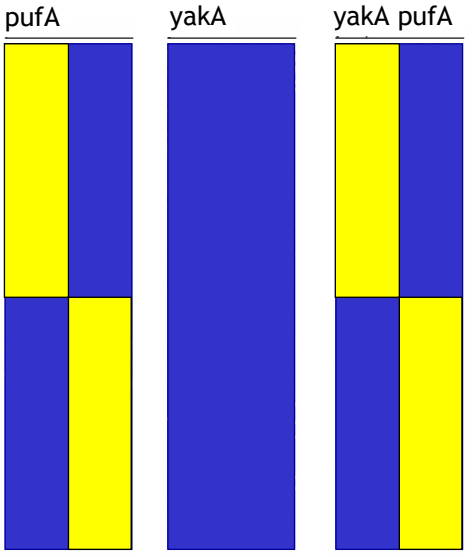
❶ pufA → yakA

❷ yakA

pufA

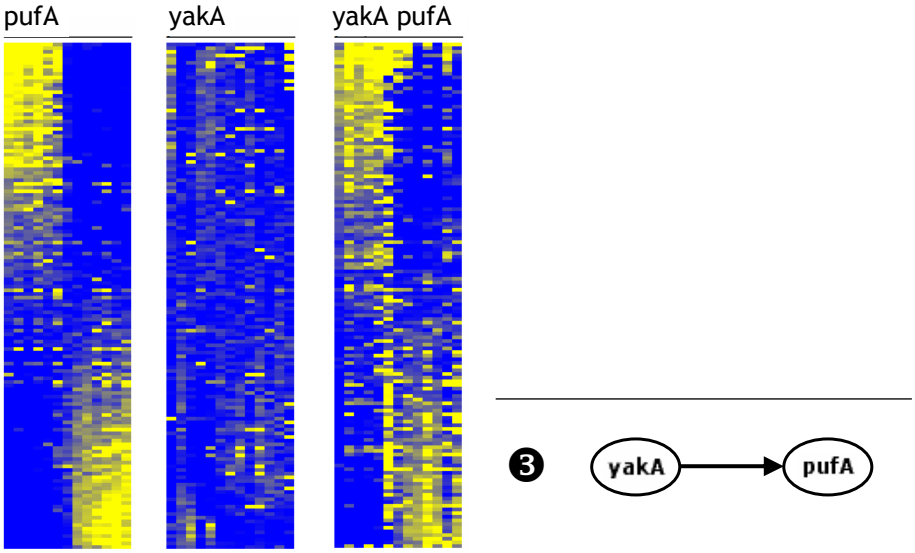
❸ yakA → pufA

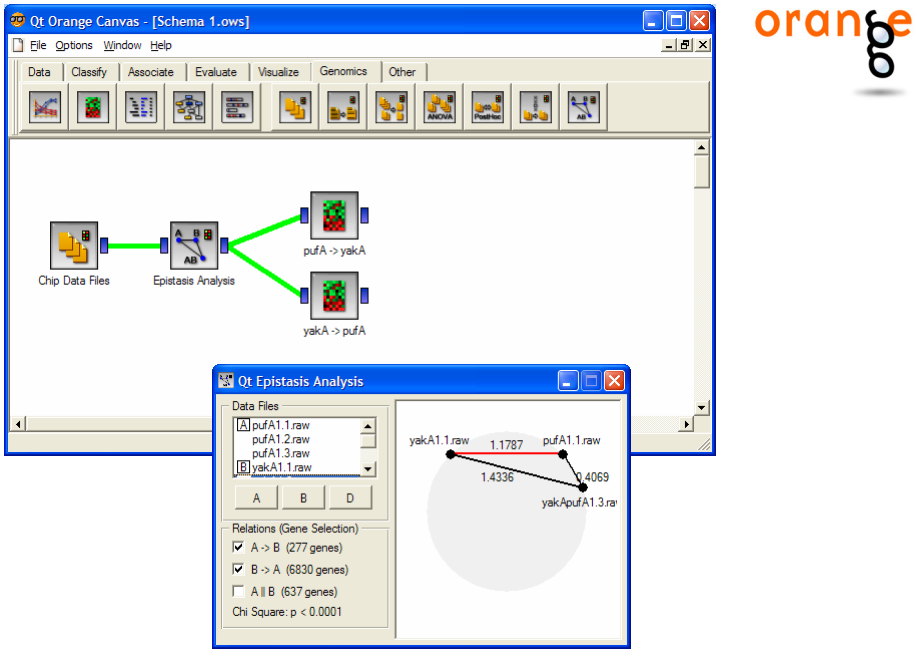
Epistasis by Transcriptional Phenotypes



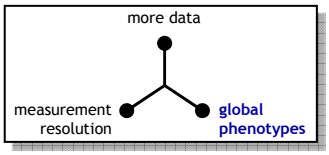
❸ yakA → pufA

Epistasis by Transcriptional Phenotypes

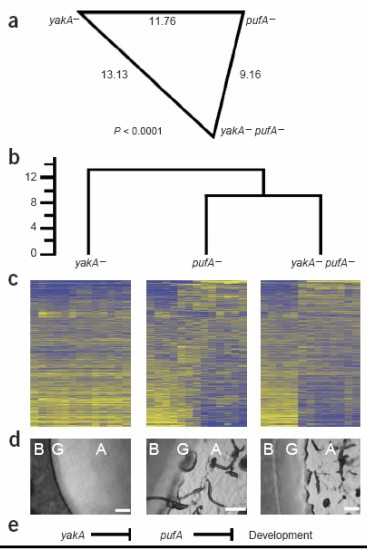




Epistasis Analysis with Global Transcriptional Phenotypes



Van Driessche et al.
Nat Genetics 2005.



Universal epistasis analysis

Timothy R Hughes

Epistasis analysis is a foundation in the analysis of genetic networks, but in complex or poorly defined processes, defining the phenotype can be an insurmountable hurdle. A new study shows that microarray expression profiles can be used as a 'phenotype' for epistasis analysis of the development of a multicellular organism, offering a potentially universal solution to this problem.

NATURE GENETICS | VOLUME 37 | NUMBER 5 | MAY 2005

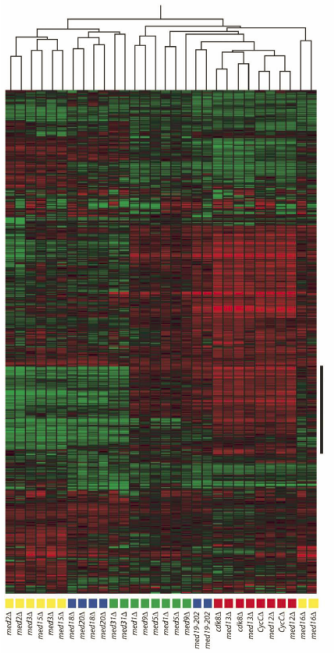


Figure 2. Expression Profiling Reveals Mediator Subunit Relationships

(A) Diagram of all 927 genes (clustered vertically) exhibiting significant changes specific to one or more of the deletion strains (clustered horizontally). Results are shown for the replicate microarray hybridizations individually. Change in gene expression compared to wt is depicted in red (up), green (down), or black (no change). The black bar indicates some of the genes that behave in opposite ways, depending on which subunit is deleted (Figure 3).

Molecular Cell, Vol. 19, 511–522, August 19, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.molcel.2005.06.033

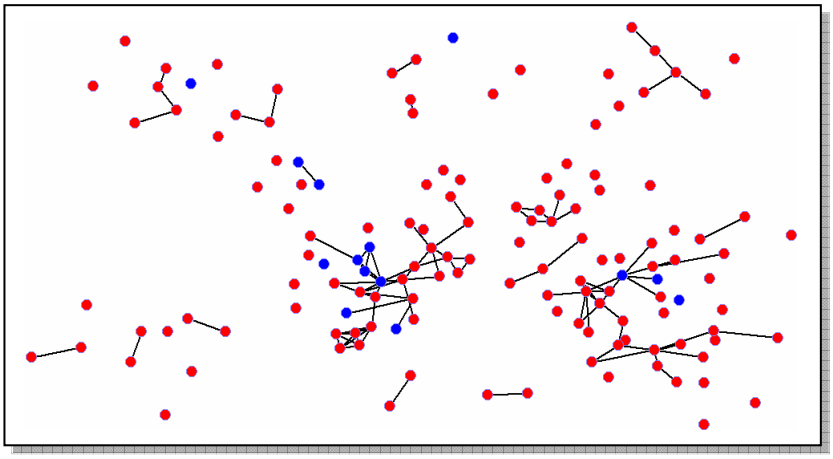
Mediator Expression Profiling Epistasis Reveals a Signal Transduction Pathway with Antagonistic Submodules and Highly Specific Downstream Targets

Jeroen van de Peppel,[1] Nienke Kettelarij,[1] Harm van Bakel,[2] Thessa T.J.P. Kockelkorn,[1] Dik van Leenen,[1] and Frank C.P. Holstege[1,*]

Computational Approaches

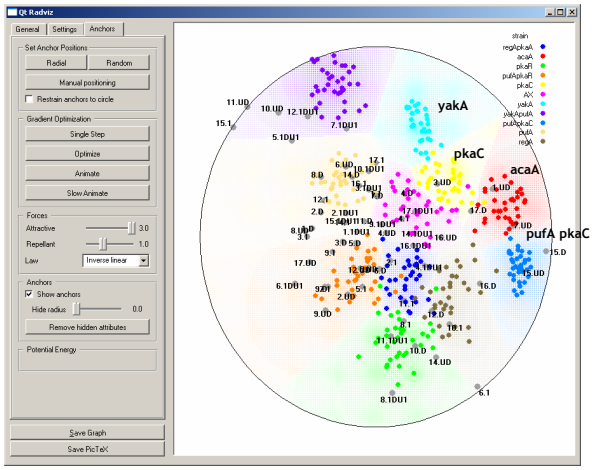
- Distance-based methods
  - Clustering
  - MDS
  - …
- Feature subset selection and construction
  - unsupervised
  - supervised (requires class labels, possibly with few class values)

Gene-Coexpression Networks



● endoplasmic reticulum

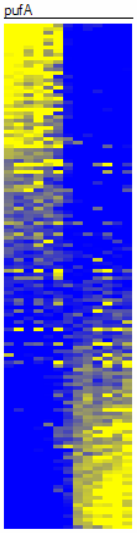
correlation > 0.76



The non-aggregating mutants pufA-pkaC-, pkaC-, yakA- and acaA- formed a cluster, pufA- and yakA-pufA- which aggregate normally, clustered separately, and the other clusters contained mutants with accelerated aggregation rates (the wild type was placed in the center).

Computational Phenomics

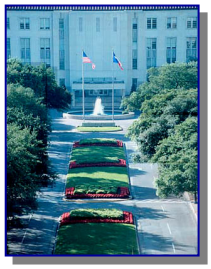
- Global (e.g. transcription) phenotypes are raw
- Comprise large number of measurements
- They need to be computationally processed and characterized with a smaller set of informative and possibly intuitive feature
- We refer to procedures that use them as **computational phenomics**



Thanks to …

AI Lab
Faculty of Comp & Info Sci
University of Ljubljana

- Tomaz Curk
- Janez Demsar
- Peter Juvan
- Gregor Leban
- Minca Mramor
- Ales Erjavec
- Ivan Bratko



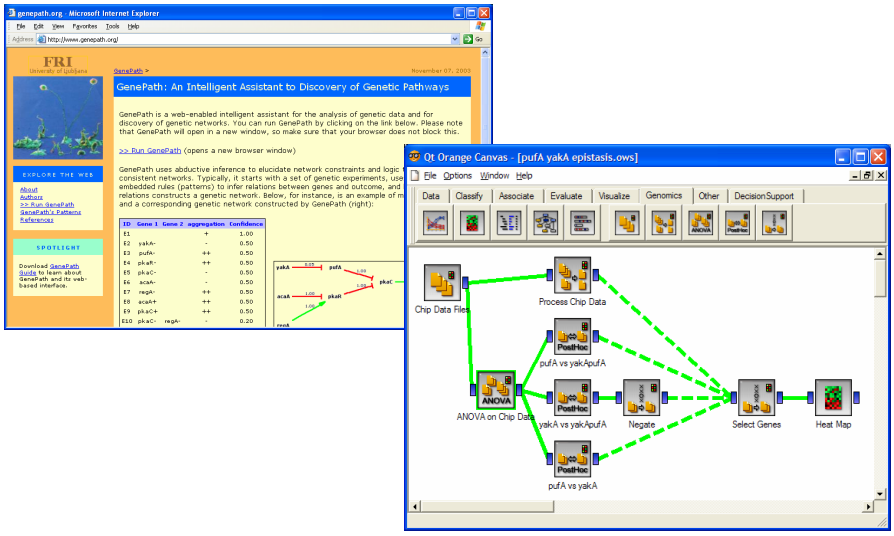
Dep. Molecular & Human Genetics, Baylor College of Medicine Houston, TX

- Gadi Shaulsky
- Adam Kuspa
- Nancy Van Driessche
- Ezgi Okyay

"Jožef Stefan" Institute

- Uros Petrovic

www.genepath.org



www.ailab.si/orange