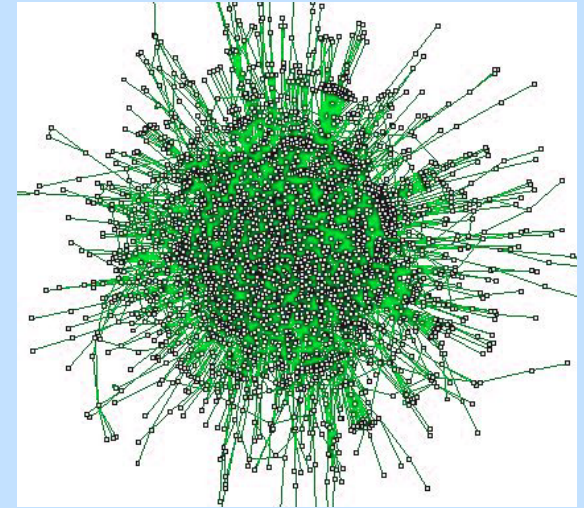


# Workshop on Analysis of Biological Networks



---

Roded Sharan & Adam Feinberg  
Spring 2017

# Goal

- Implement a large project in a team of 2-3.
- General topic: Predicting genetic interactions
- Apply to real data to evaluate performance and utility
- Advance research!

# Time Table

1. Introduction to genetic interaction & prediction
2. Introduction to ML; project suggestions
  - Choice of projects & partners.
  - June 8: presentations of assigned projects and progress reports.
  - University deadline (1.10): final submission.

# Basic genetics

- Simple protocol for finding a function of a gene:
  - Generate a strain without this gene (a **knock-out** strain)
  - If the strain is not viable – the gene is probably very important (an **essential** gene)
  - If the strain is sick – the sickness may teach you something
  - If the strain is healthy – find a condition in which it is sick (e.g., heat shock), it may teach you about the genes function (e.g., protein folding chaperone)

# Yeast Knock-Out library (YKO)

- 96% of knock-outs of yeast genes were done by 2002
- Each gene replaced by a molecular “barcode”
- Only 18.7% of the genes were **essential**
- 15% of non-lethal mutants grow slowly on rich medium (0.12-0.9 of the optimal growth rate)
- Conclusion – individual deletion most of the genes (roughly 70%) does not cause any phenotype!

# Why are so many genes redundant?

- In evolution genes that don't have a contribution to the fitness of the organism are supposed to become extinct
- Explanations for “no phenotype on rich media”
  - Maybe the gene is required in **other conditions**
  - Maybe the gene has a duplicate/similar gene that **buffers** its deletion
  - Maybe the entire pathway the gene is in is **buffered by a different pathway**
  - Maybe the phenotype is something that is difficult to detect, but affects the population on the long run (e.g., by affecting variability)

# Genetic interaction

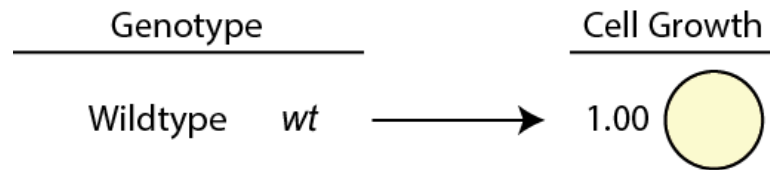
*A genetic interaction* is the interaction of two genetic perturbations in determining a phenotype.

*Synthetic lethality*: Two genes A,B are synthetic lethal if knockouts of A or B separately are viable, but knocking out both is lethal.

$$1 + 1 = 0$$

# Genetic Interactions




(unexpected genotype-phenotype relations)









# Genetic Interactions

(unexpected genotype-phenotype relations)

Genotype			Cell Growth	
Wildtype	<i>wt</i>	→	1.00	
Single-Gene Deletion	<i>aΔ</i>	→	0.70	
	<i>bΔ</i>	→	0.70	

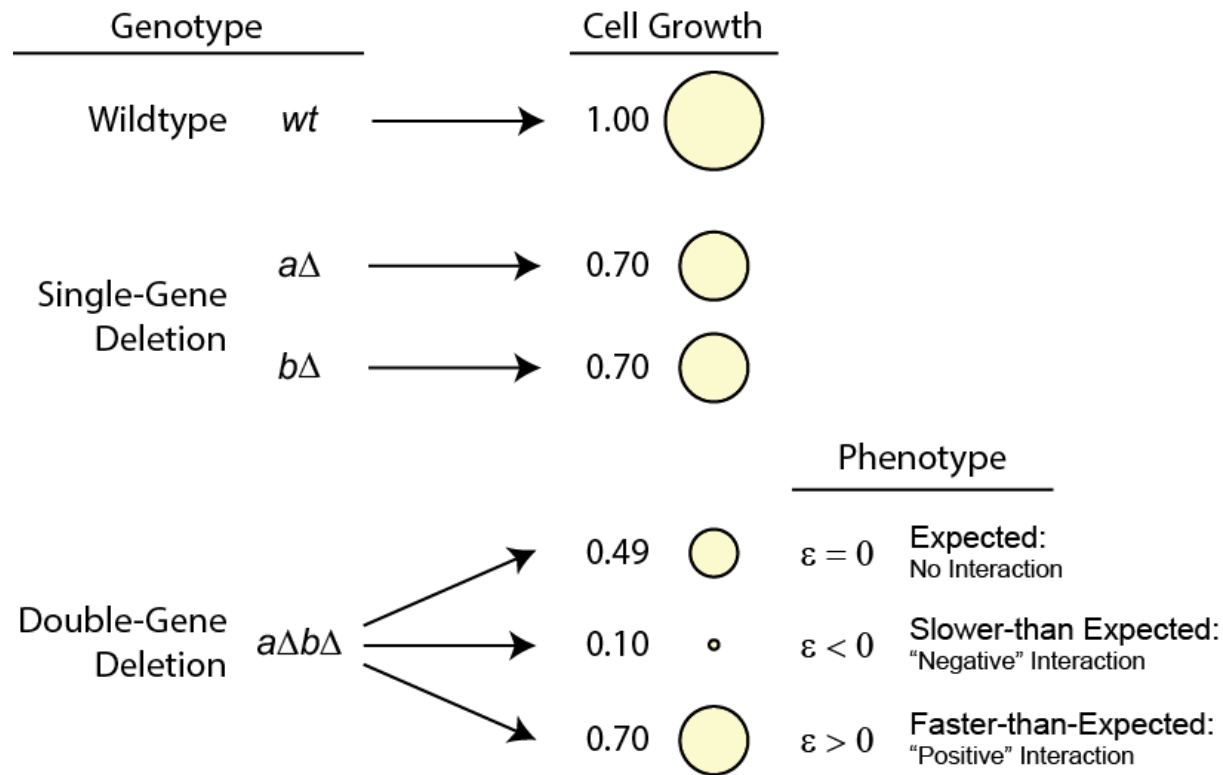
# Genetic Interactions

(unexpected genotype-phenotype relations)

Genotype		Cell Growth		Phenotype	
Wildtype	<i>wt</i>	→	1.00		
Single-Gene Deletion	<i>aΔ</i>	→	0.70		
	<i>bΔ</i>	→	0.70		
Double-Gene Deletion	<i>aΔbΔ</i>	↗	0.49		$\varepsilon = 0$ Expected: No Interaction

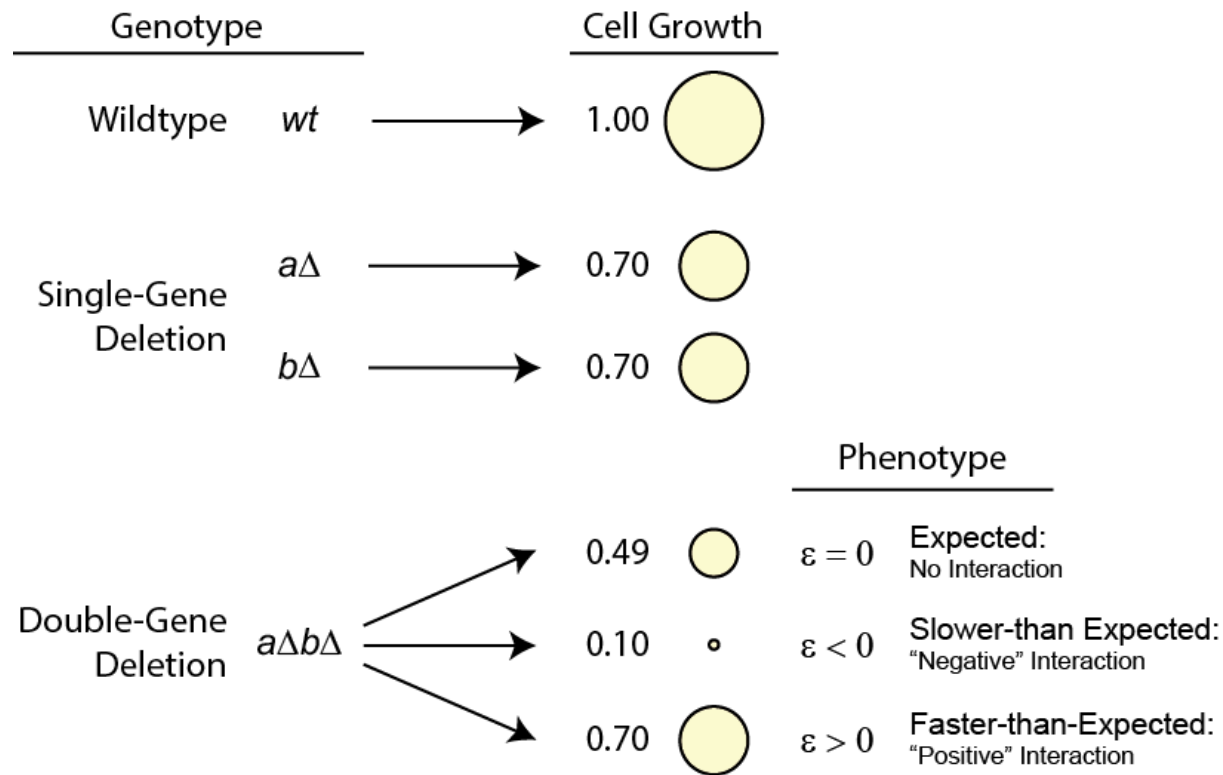
# Genetic Interactions

(unexpected genotype-phenotype relations)

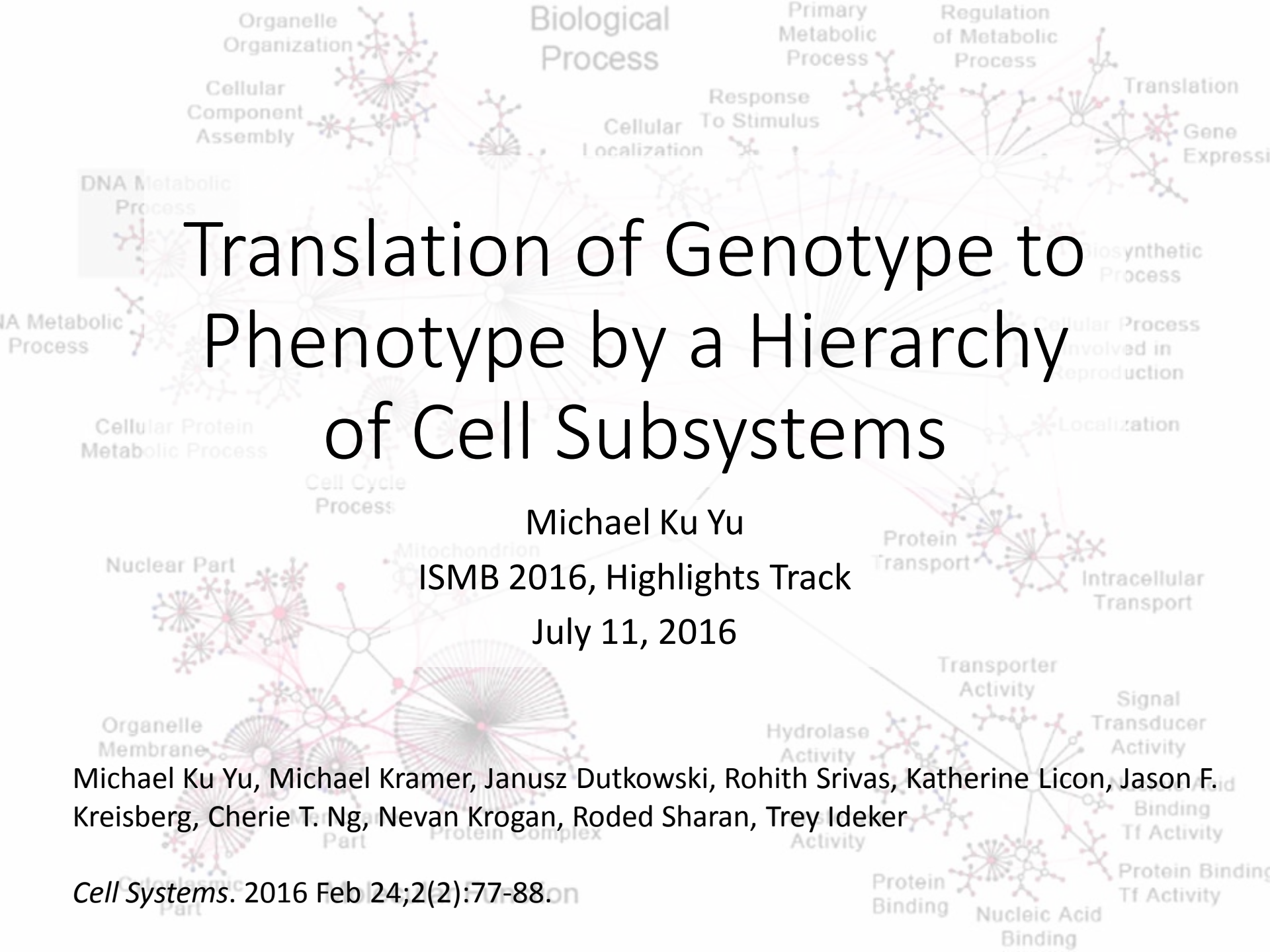


# Genetic Interactions

(unexpected genotype-phenotype relations)



- Genetic interactions measured for ~5 million pairs of genes in *S. cerevisiae*
- From Costanzo et al. *Science*, 2010. (One of the largest genotype-phenotype datasets)
- Given a pair of genes, predict if there is a genetic interaction.

A complex network diagram of biological processes, likely from a Gene Ontology (GO) enrichment analysis. The diagram consists of numerous nodes (represented by small star-like or circular clusters of colored dots) connected by thin lines. The nodes are arranged in a hierarchical and interconnected manner, with some larger, more central nodes and many smaller peripheral ones. The background is a light gray, and the network is rendered in various shades of gray, with some nodes highlighted in pink or blue. The overall structure suggests a dense web of biological relationships.

# Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems

Michael Ku Yu

ISMB 2016, Highlights Track

July 11, 2016

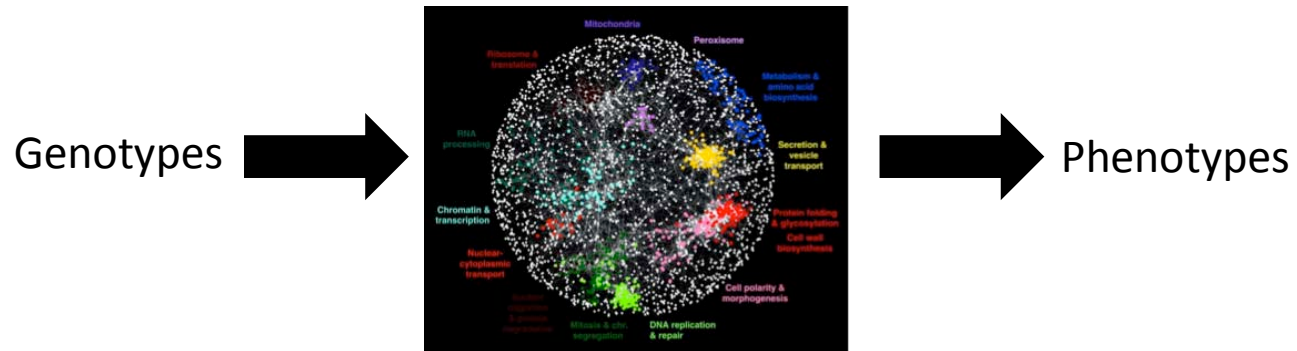
Michael Ku Yu, Michael Kramer, Janusz Dutkowski, Rohith Srivas, Katherine Licon, Jason F. Kreisberg, Cherie T. Ng, Nevan Krogan, Roded Sharan, Trey Ideker

*Cell Systems*. 2016 Feb 24;2(2):77-88.

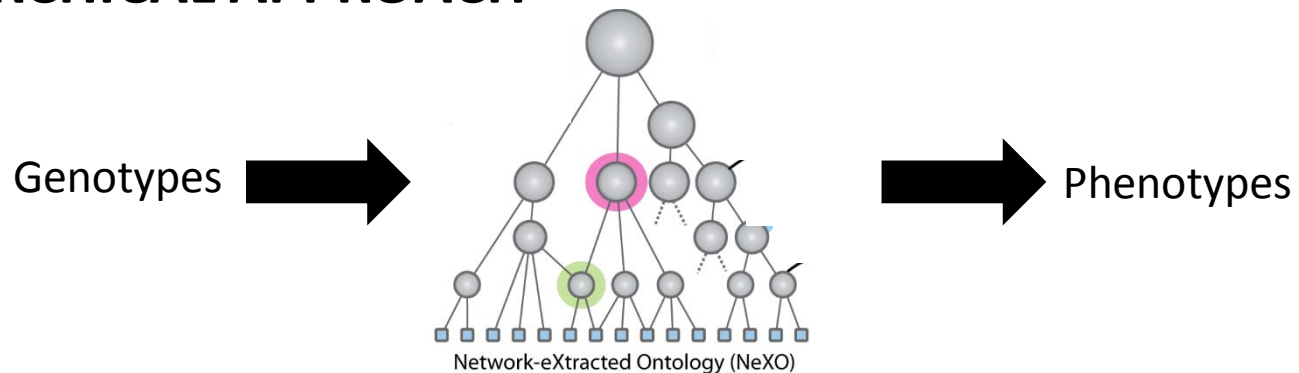
# PROBLEM

Given a training set of genotype-phenotype relations,  
predict the phenotypes of new genotypes

## NETWORK/PATHWAY APPROACH



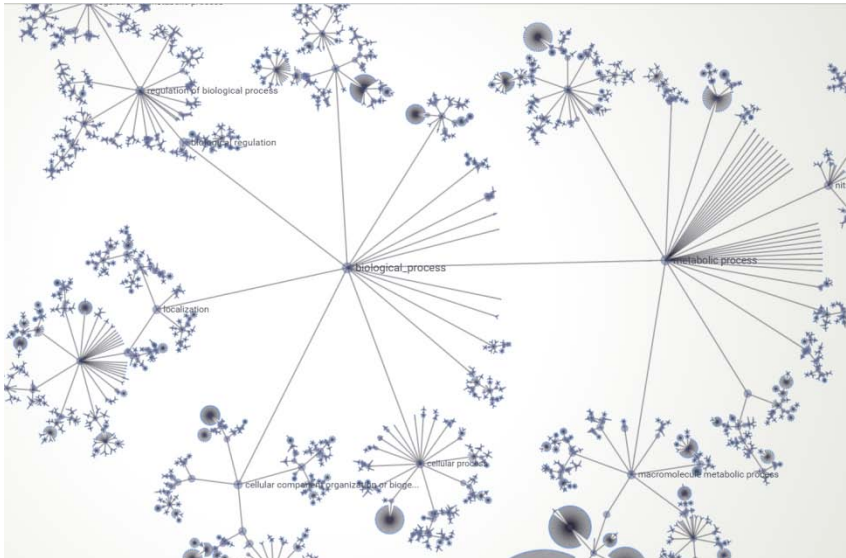
## HIERARCHICAL APPROACH



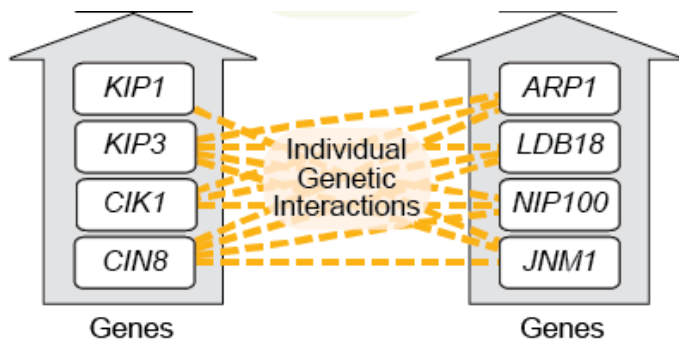
# Hierarchical Map of Cell Structure

## The Gene Ontology (GO)

- Manually curated from literature
- >41,000 terms, millions of gene annotations, dozens of species

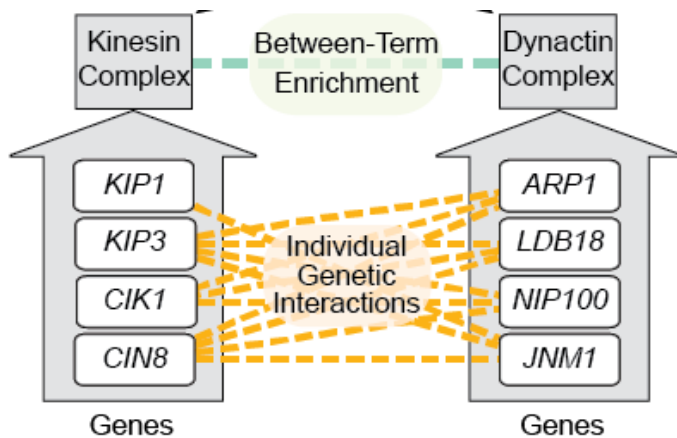


# Hierarchical Interpretation of Genetic Interactions

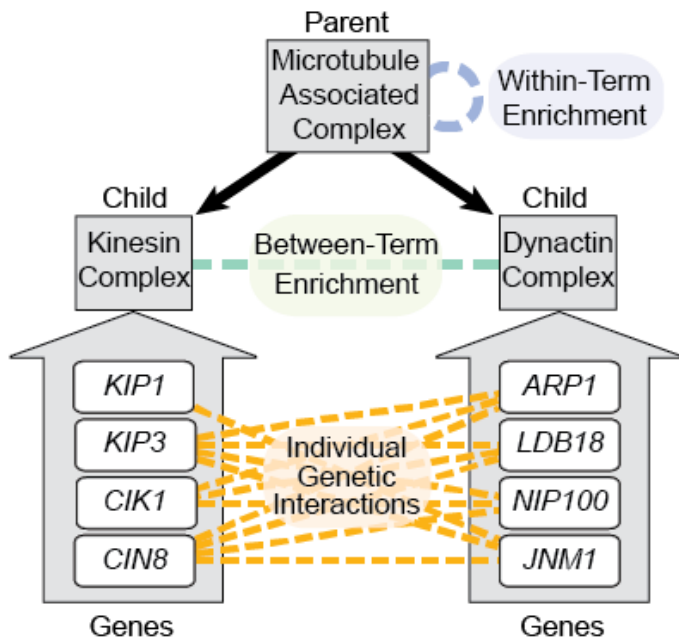




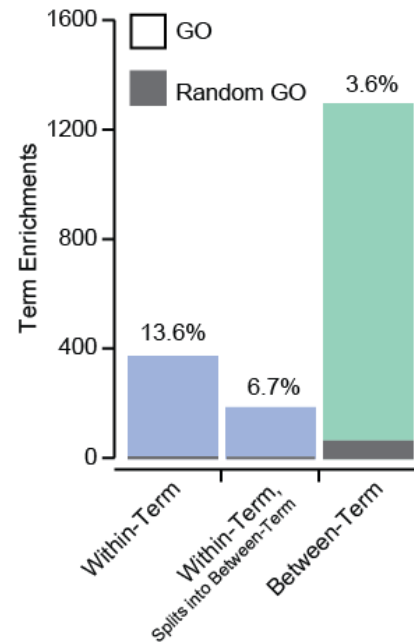
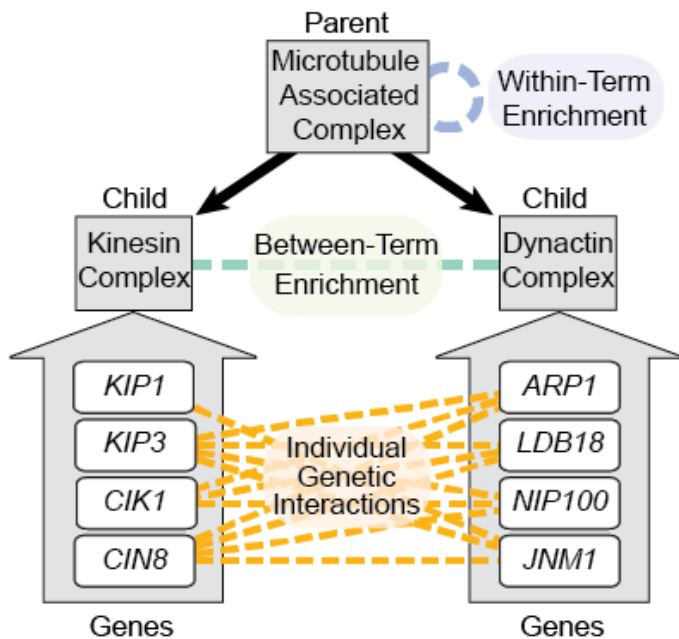
# Hierarchical Interpretation of Genetic Interactions



# Hierarchical Interpretation of Genetic Interactions

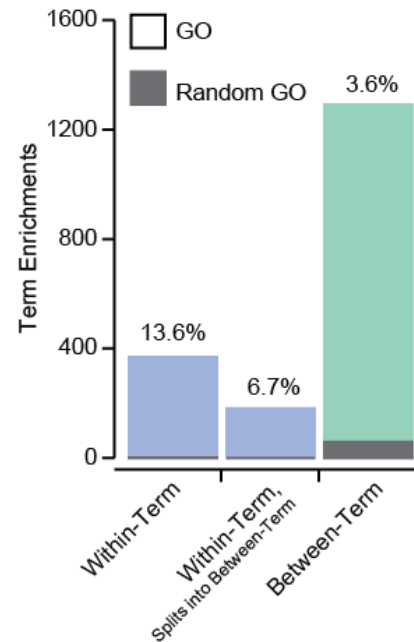
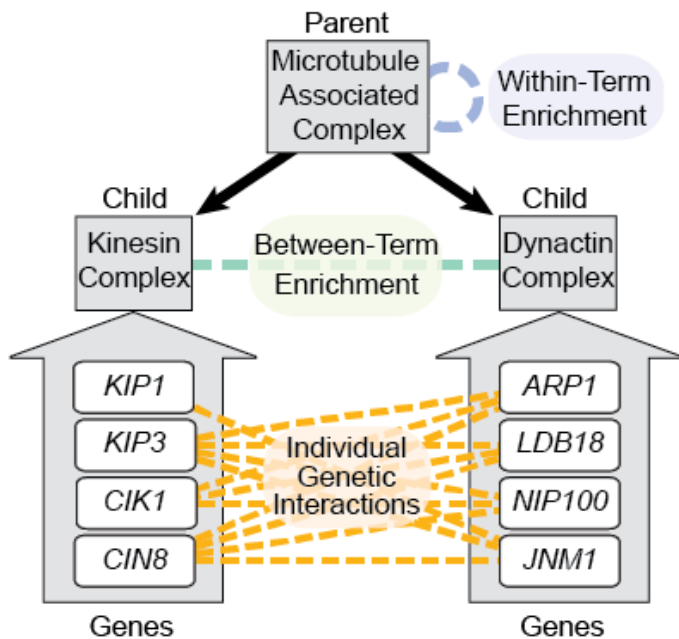


# Hierarchical Interpretation of Genetic Interactions

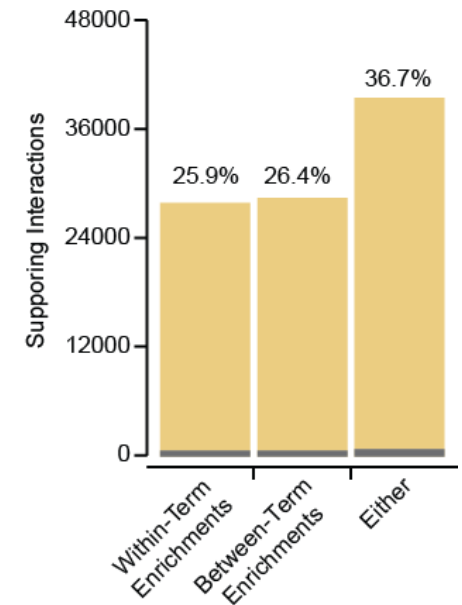


>400 within-term &  
>1,300 between-term  
enrichments

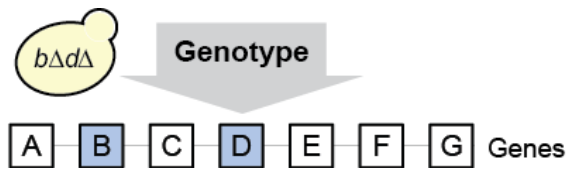
# Hierarchical Interpretation of Genetic Interactions

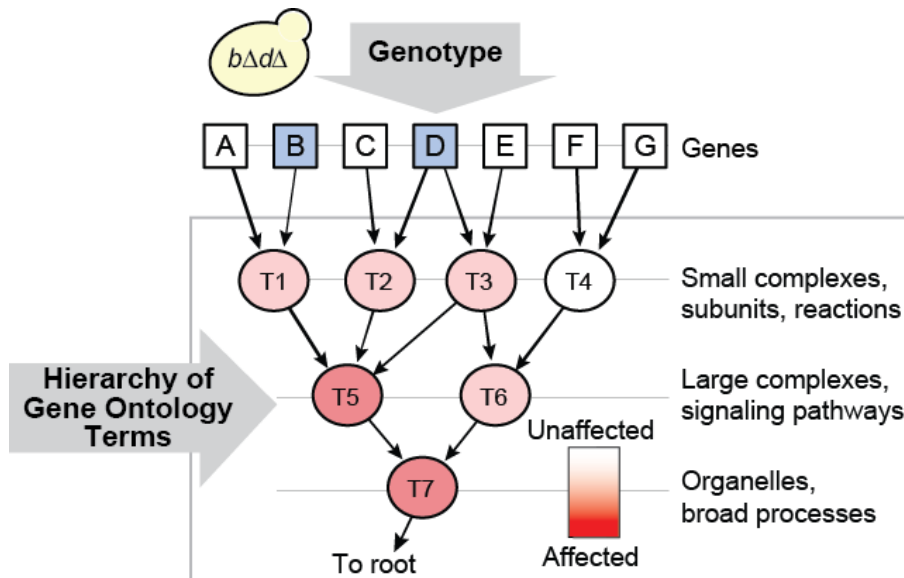


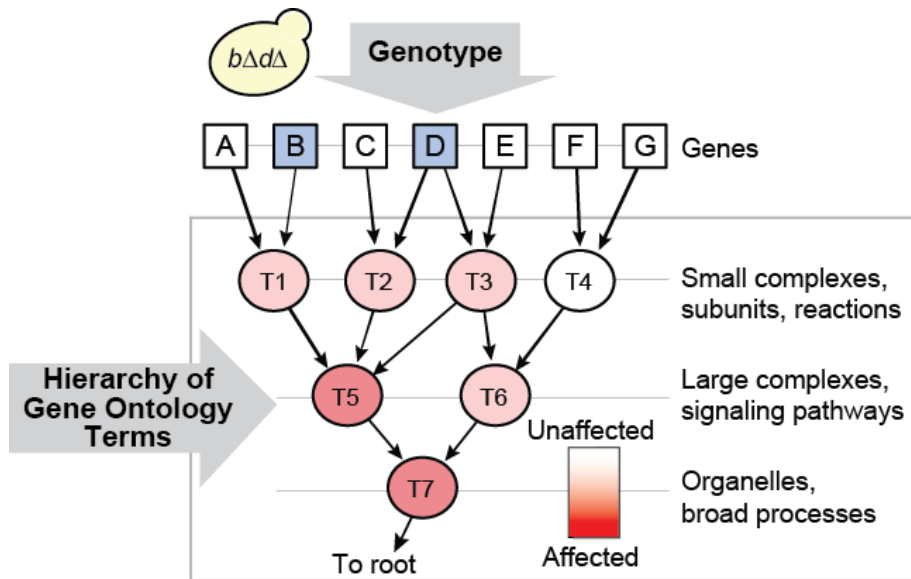
>400 within-term &  
>1,300 between-term  
enrichments



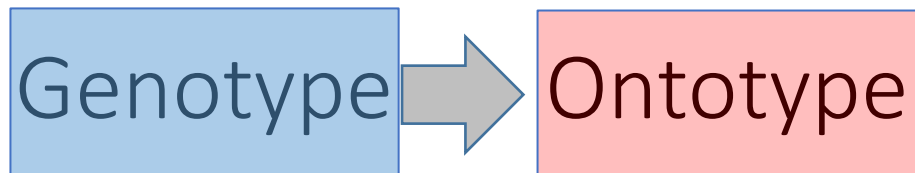
Enrichments explain  
36.7% of genetic  
interactions

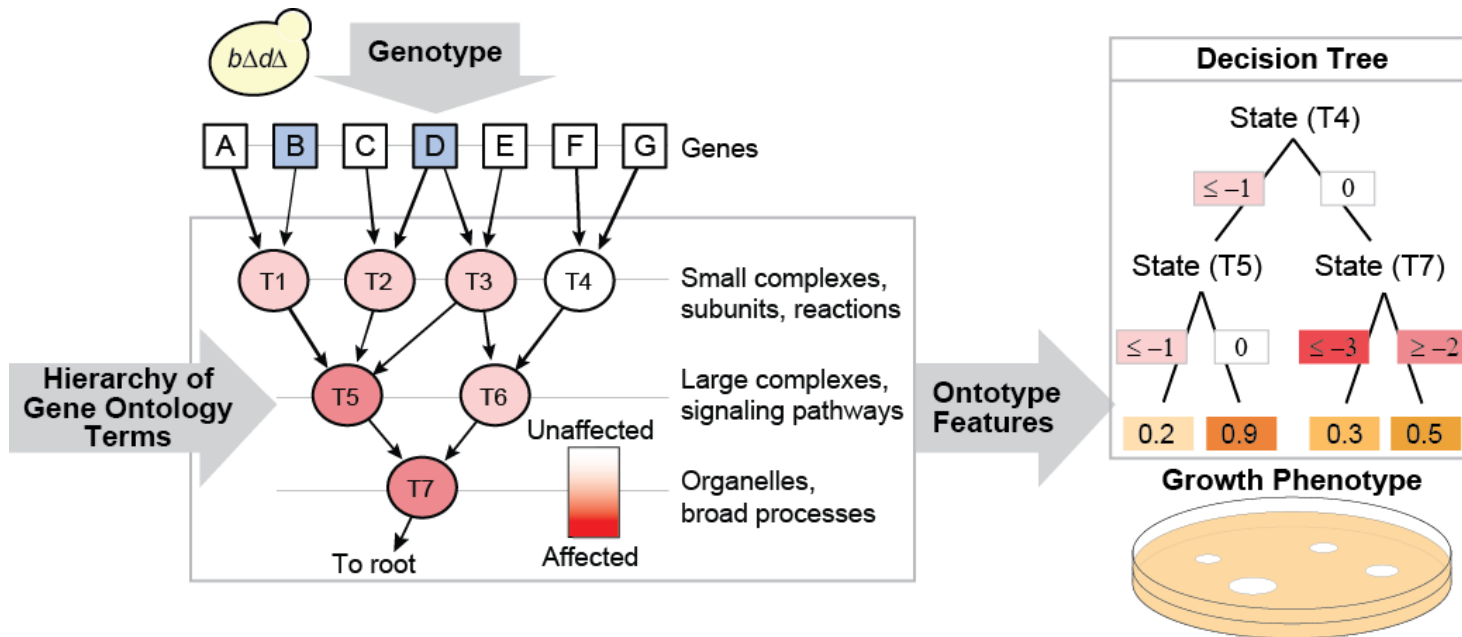




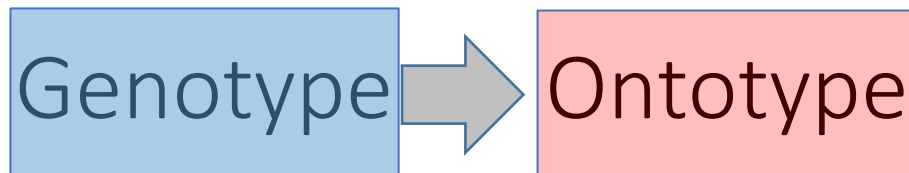


Genotype	Ontotype: T1	T2	T3	T4	T5	T6	T7
$b\Delta d\Delta$	-1	-1	-1	0	-2	-1	-2
$a\Delta d\Delta$	-1	-1	-1	0	-2	-1	-2
$f\Delta$	0	0	0	-1	0	-1	-1
$a\Delta c\Delta e\Delta$	-1	-1	-1	0	-3	-1	-3
$e\Delta g\Delta$	0	0	-1	-1	-1	-2	-2

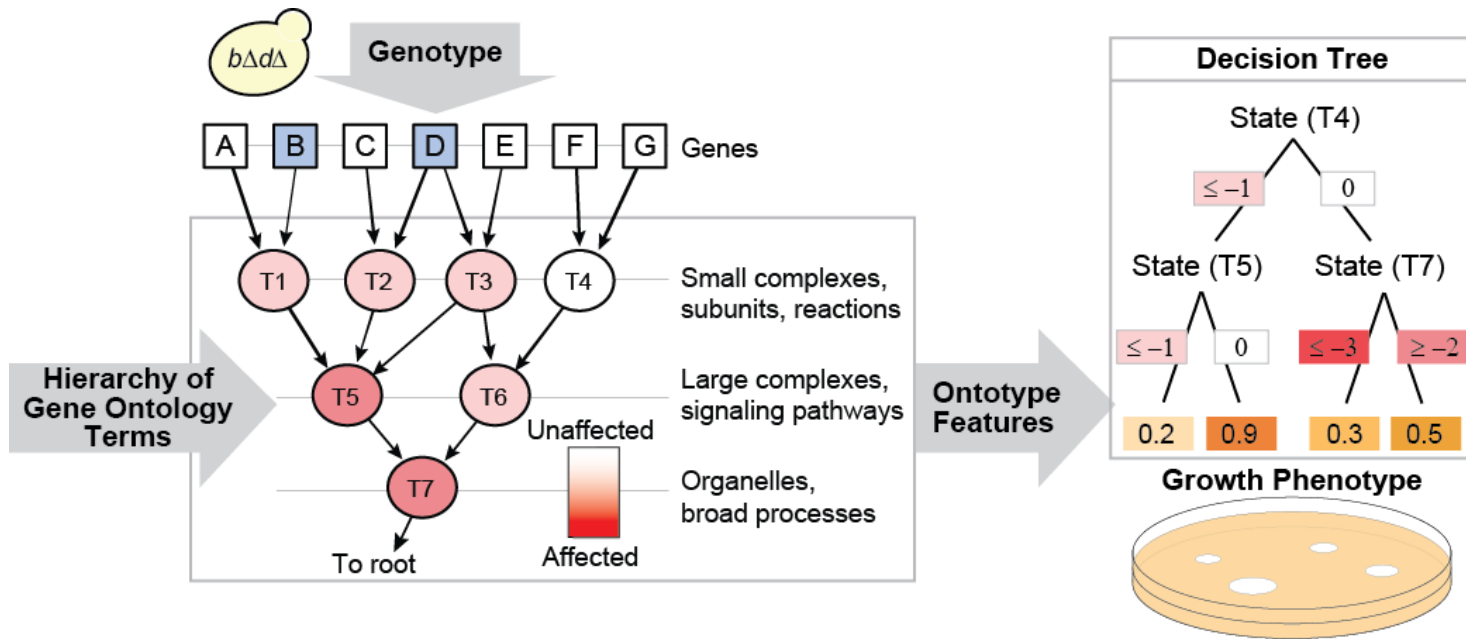




Genotype	Ontotype: T1	T2	T3	T4	T5	T6	T7
<i>bΔdΔ</i>	-1	-1	-1	0	-2	-1	-2
<i>aΔdΔ</i>	-1	-1	-1	0	-2	-1	-2
<i>fΔ</i>	0	0	0	-1	0	-1	-1
<i>aΔcΔeΔ</i>	-1	-1	-1	0	-3	-1	-3
<i>eΔgΔ</i>	0	0	-1	-1	-1	-2	-2

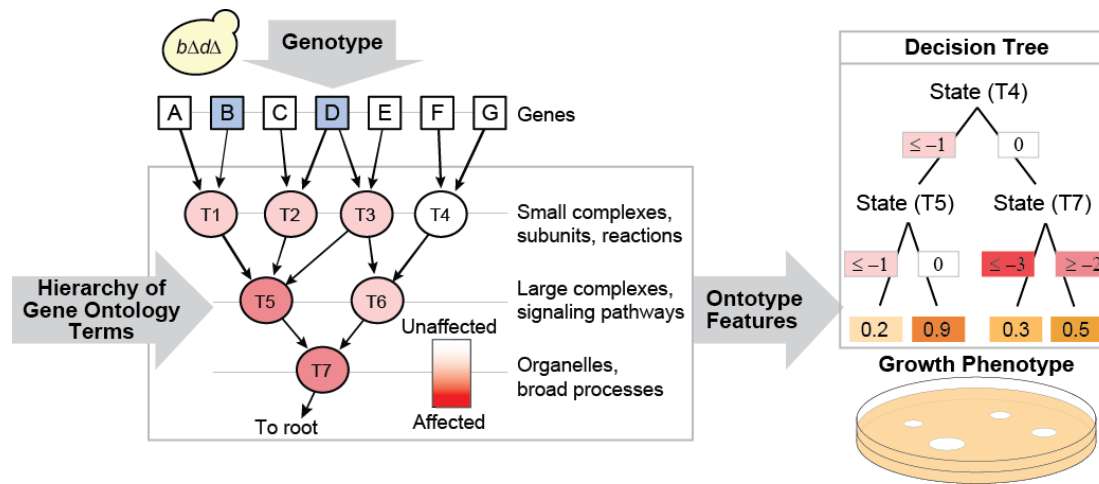






Genotype	Ontotype: T1 T2 T3 T4 T5 T6 T7							Phenotype
<i>bΔdΔ</i>	-1	-1	-1	0	-2	-1	-2	0.5
<i>aΔdΔ</i>	-1	-1	-1	0	-2	-1	-2	0.5
<i>fΔ</i>	0	0	0	-1	0	-1	-1	0.9
<i>aΔcΔeΔ</i>	-1	-1	-1	0	-3	-1	-3	0.3
<i>eΔgΔ</i>	0	0	-1	-1	-1	-2	-2	0.2





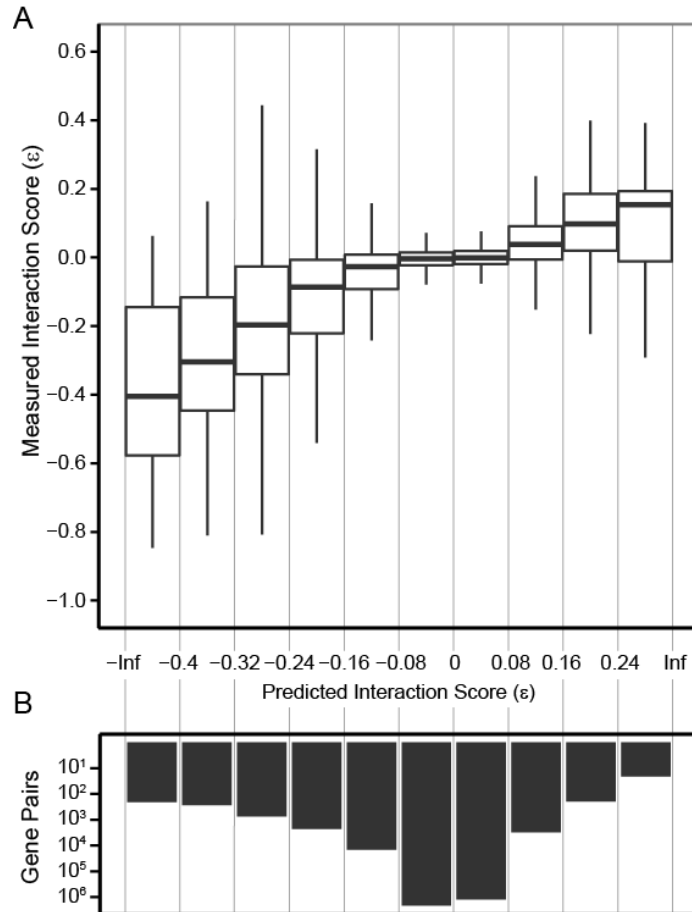
Genotype	Ontotype: T1 T2 T3 T4 T5 T6 T7							Phenotype
<i>bΔdΔ</i>	-1	-1	-1	0	-2	-1	-2	0.5
<i>aΔdΔ</i>	-1	-1	-1	0	-2	-1	-2	0.5
<i>fΔ</i>	0	0	0	-1	0	-1	-1	0.9
<i>aΔcΔeΔ</i>	-1	-1	-1	0	-3	-1	-3	0.3
<i>eΔgΔ</i>	0	0	-1	-1	-1	-2	-2	0.2



The ontotype is a multi-scale representation of the cell, inbetween genotype and phenotype

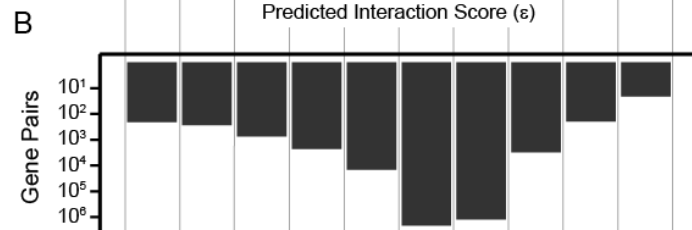
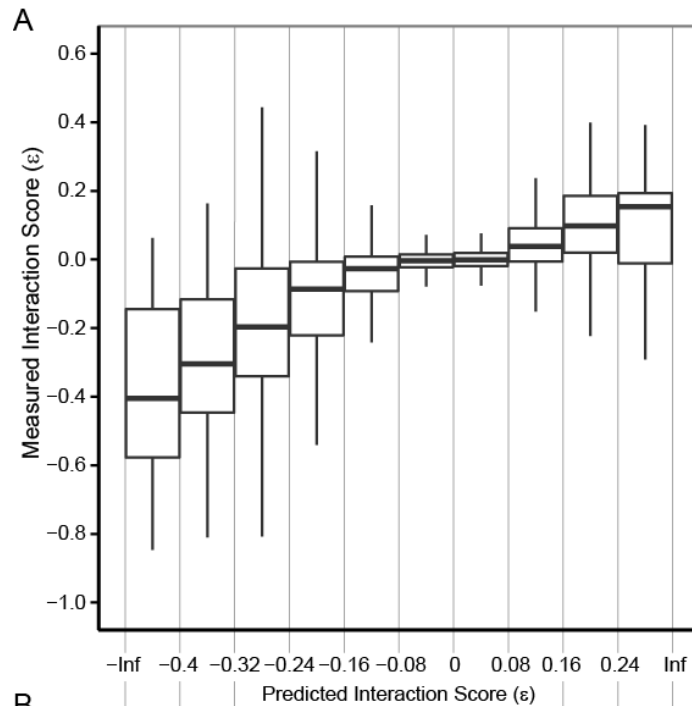
# Hierarchical model of yeast growth outperforms non-hierarchical methods

Predictions correspond to measurements



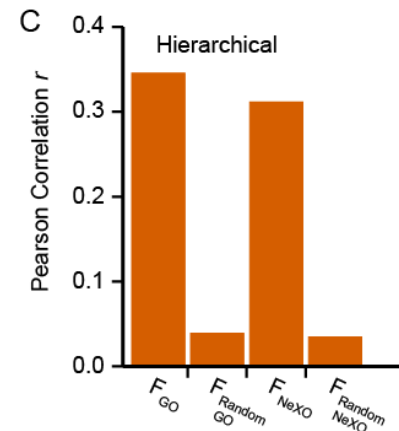
# Hierarchical model of yeast growth outperforms non-hierarchical methods

## Predictions correspond to measurements



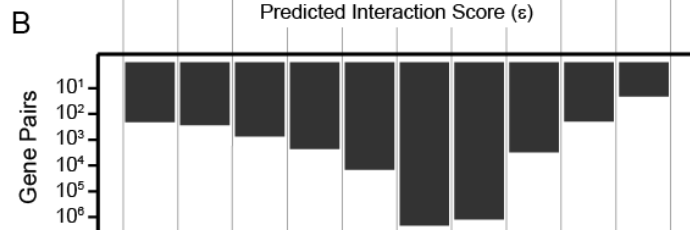
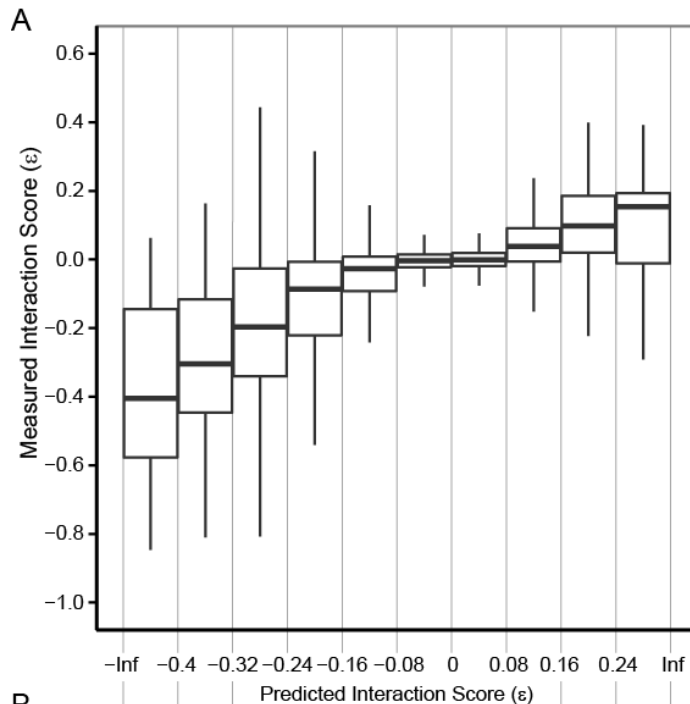
## Our hierarchical model outperforms ...

### 1. Same model using a random hierarchy



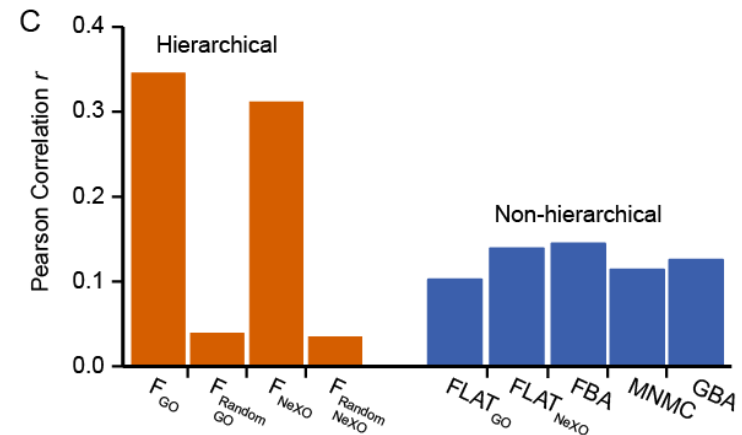
# Hierarchical model of yeast growth outperforms non-hierarchical methods

## Predictions correspond to measurements



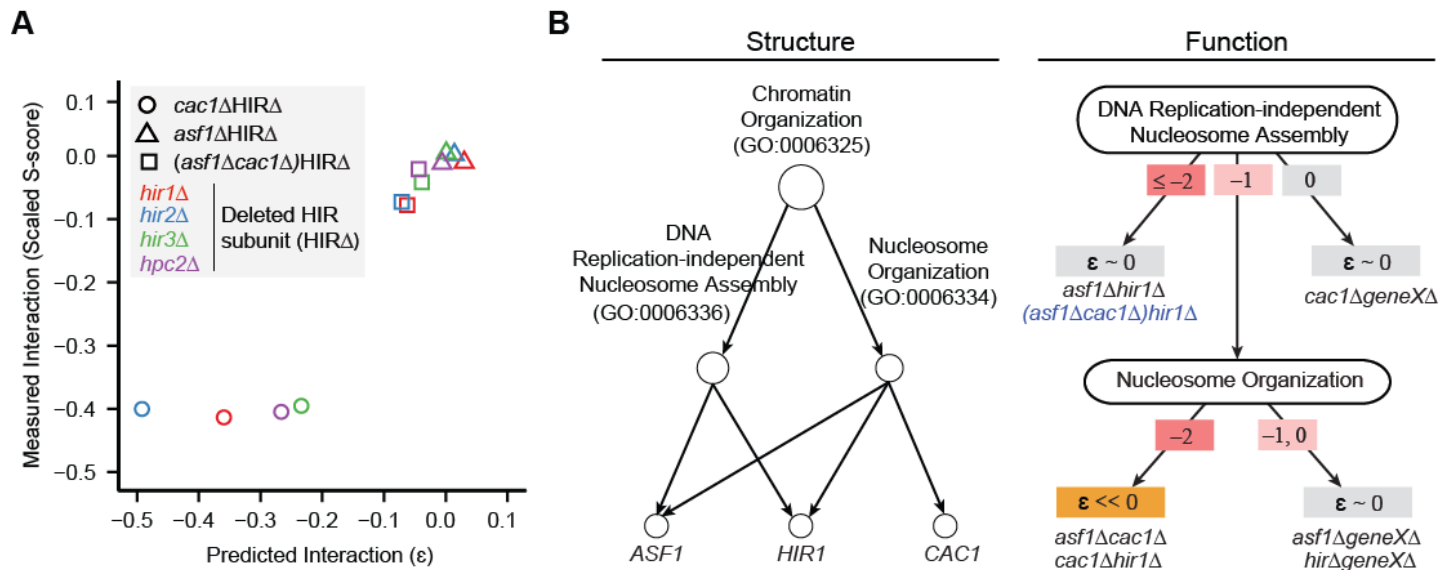
## Our hierarchical model outperforms ...

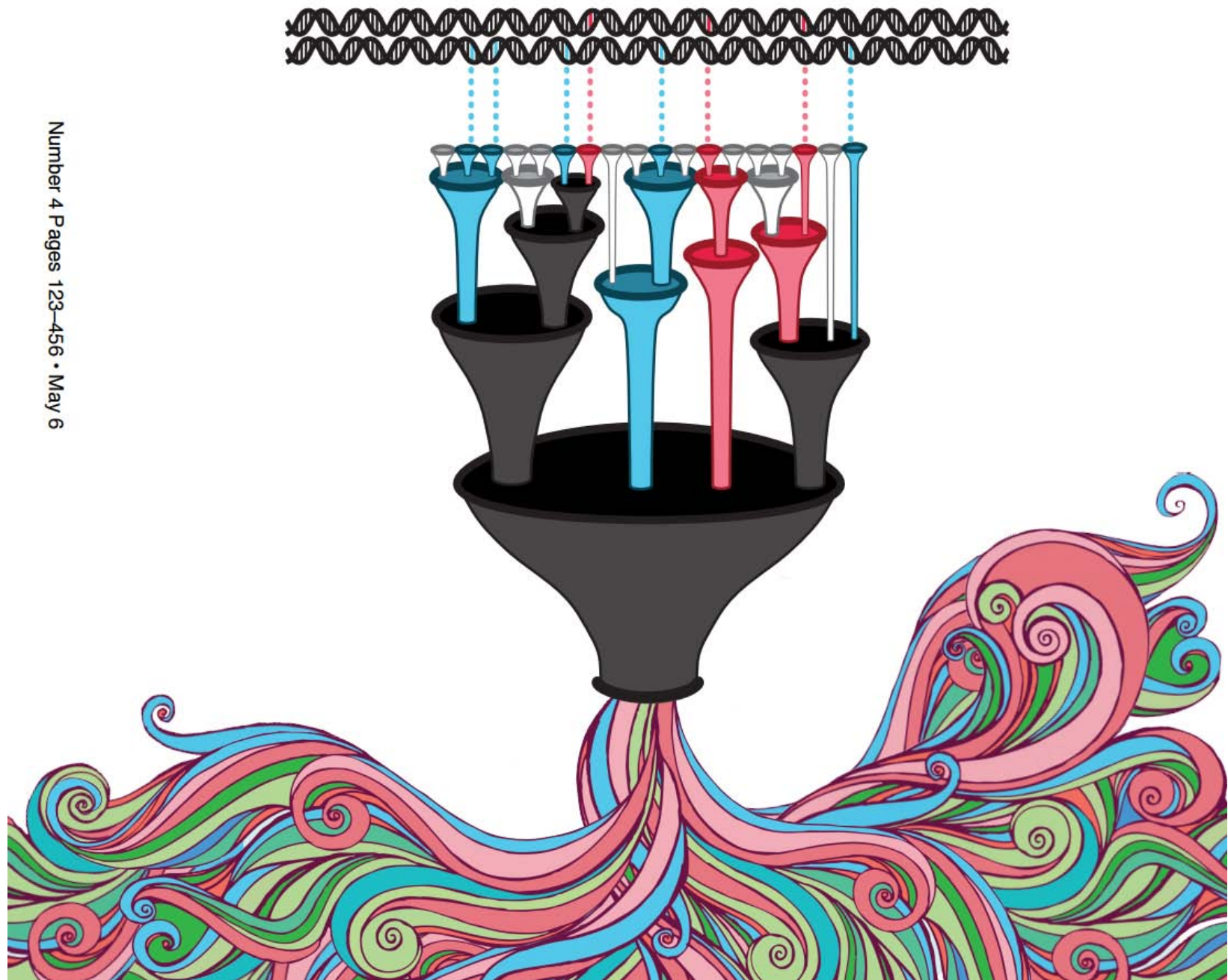
1. Same model using a random hierarchy
2. Previous, non-hierarchical methods



# Towards complex genotypes

- Haber et al. *Cell Rep*, 2014
  - Screened double and triple deletion mutants
  - Deleted ASF1 and/or CAC1
  - Deleted another gene in the HIR complex (HIR1, HIR2, HIR3, HPC2)
- Double mutants  $cac1\Delta HIR\Delta$  grow slower than expected, but not triple mutants  $asf1\Delta cac1\Delta HIR\Delta$





# Suggested projects

Prediction & evaluation: implement and compare existing algorithms for GI prediction; develop transfer learning.

Applications: adapt and apply GI prediction to predict gene essentiality in different backgrounds, drug side effects, patient survival and response to treatments.