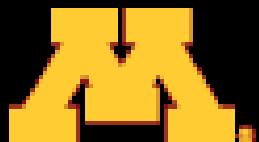
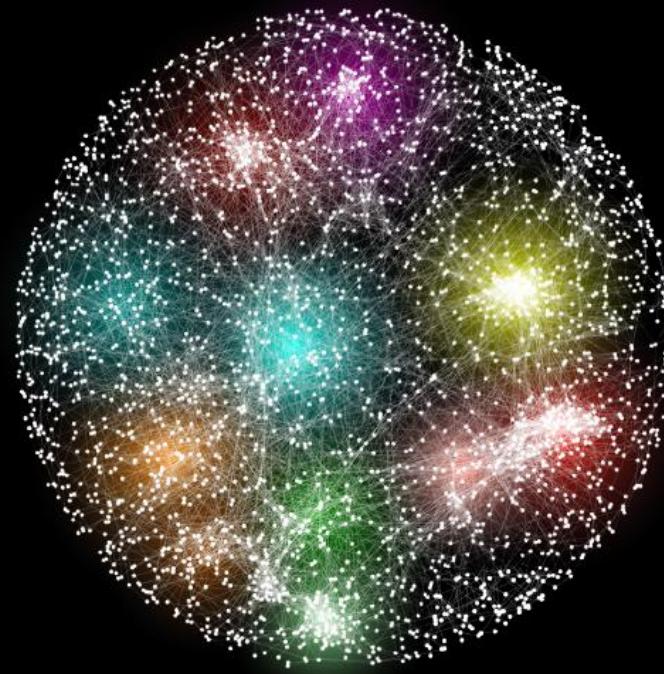


Using networks to understand biology from yeast to man



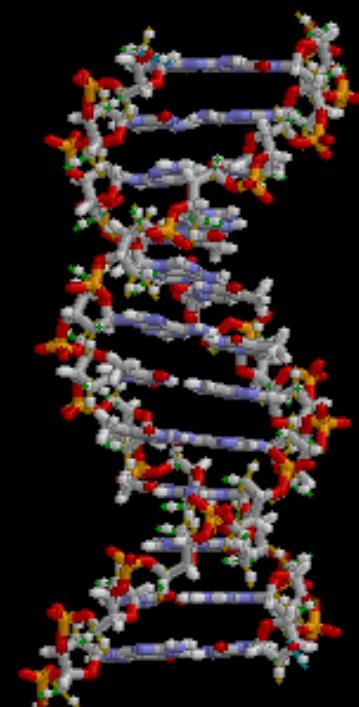
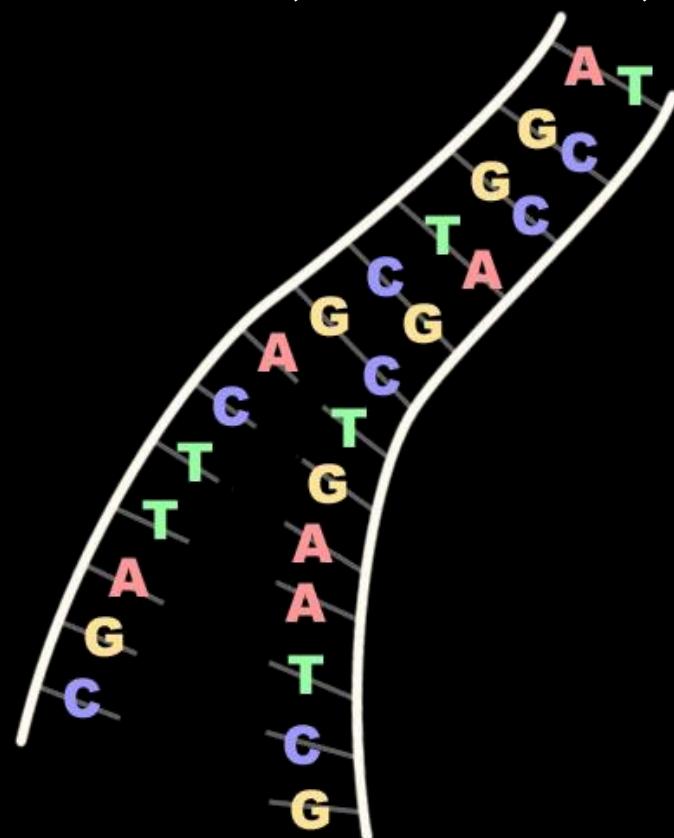
Chad Myers
Department of Computer Science and Engineering
University of Minnesota, Twin Cities

Talk Overview

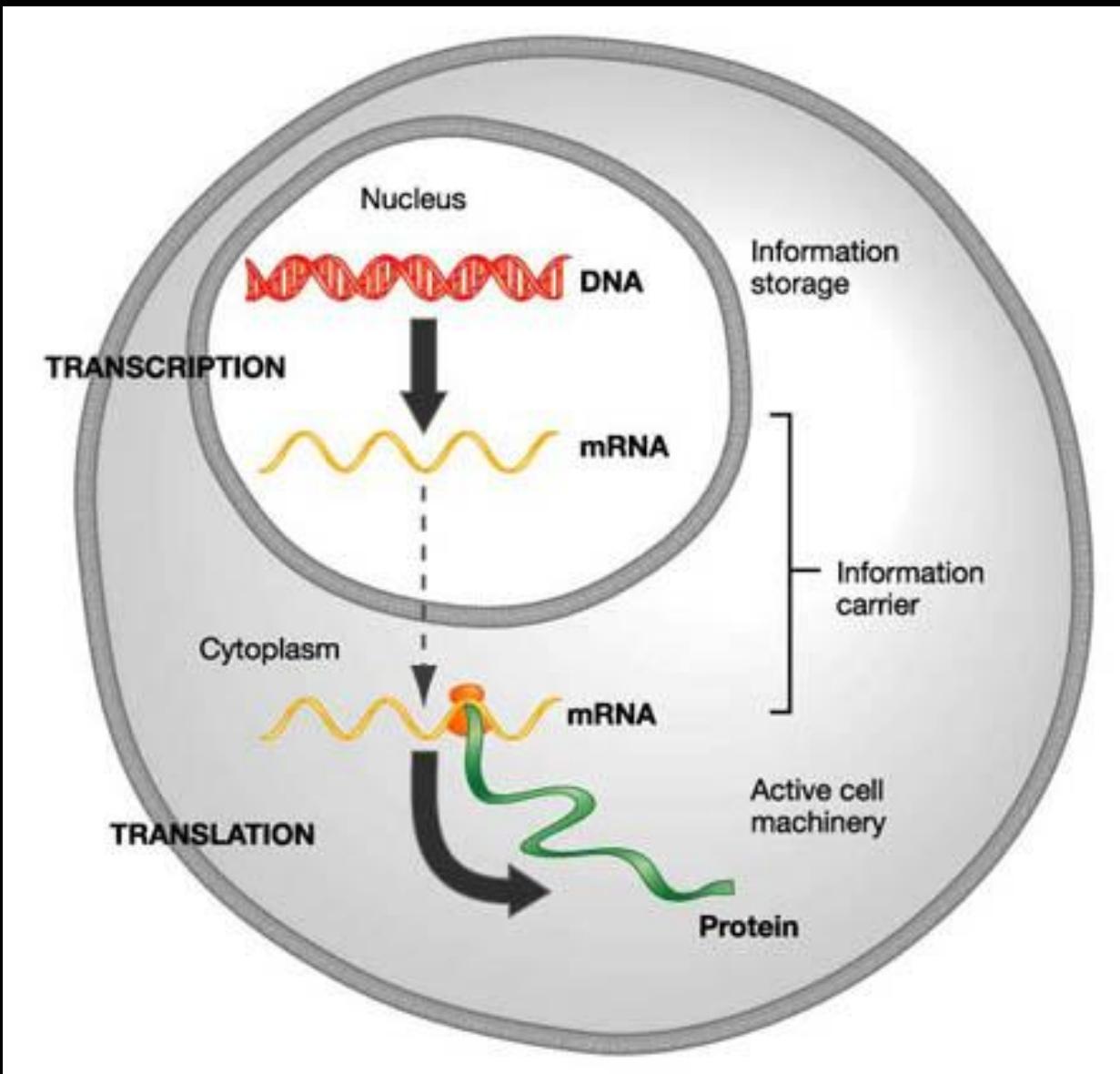
- Crash course in molecular biology
- The big questions in computational/systems biology
- An example problem in network biology from my lab
 - Understanding how genes interact using network data from yeast
- General thoughts on research in comp bio

A crash course in molecular biology

- DNA: Deoxyribonucleic Acid
 - Instructions, written in a 4-letter alphabet
 - Adenine, Thiamine, Cytosine, Guanine



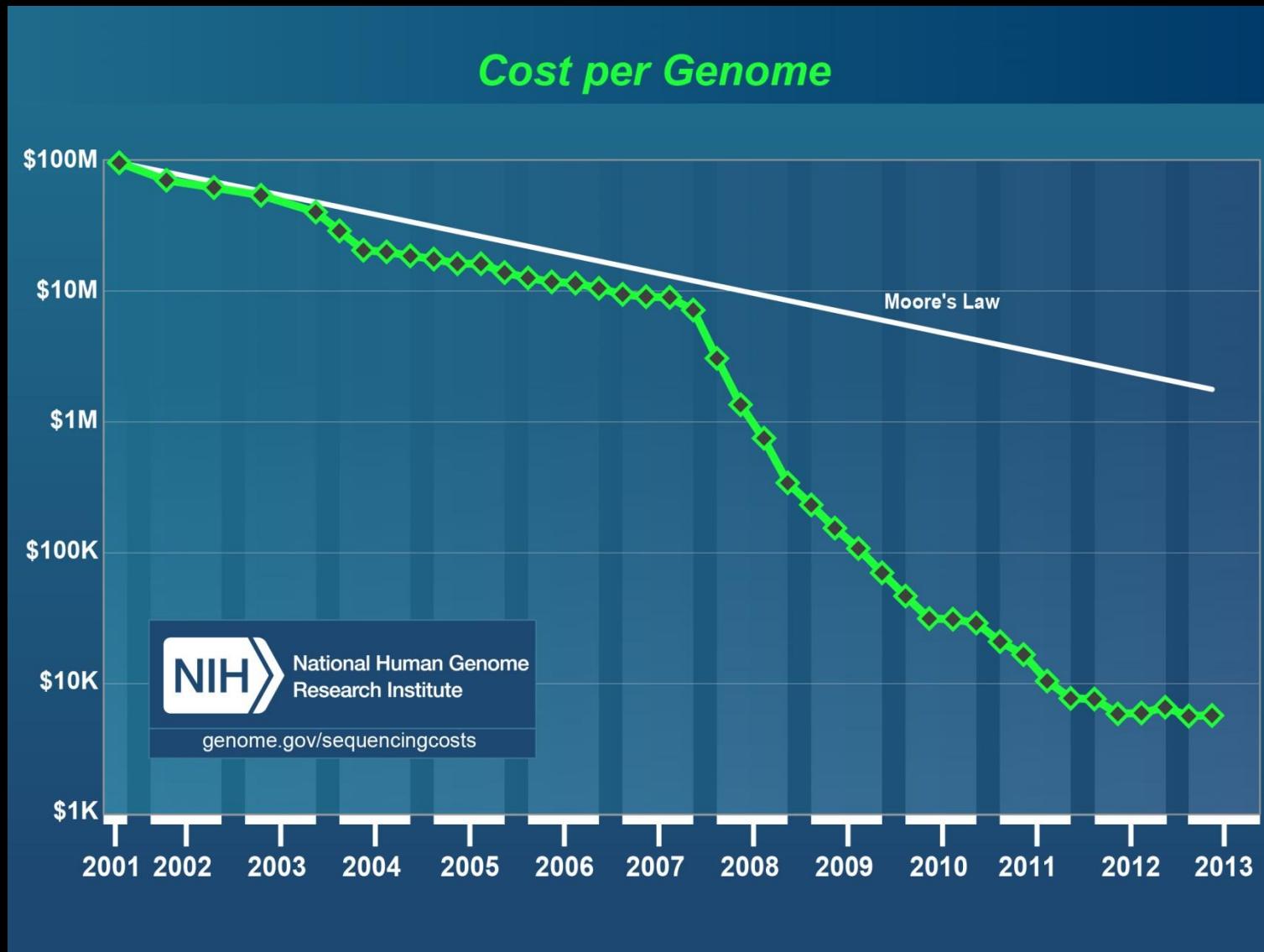
A crash course in molecular biology



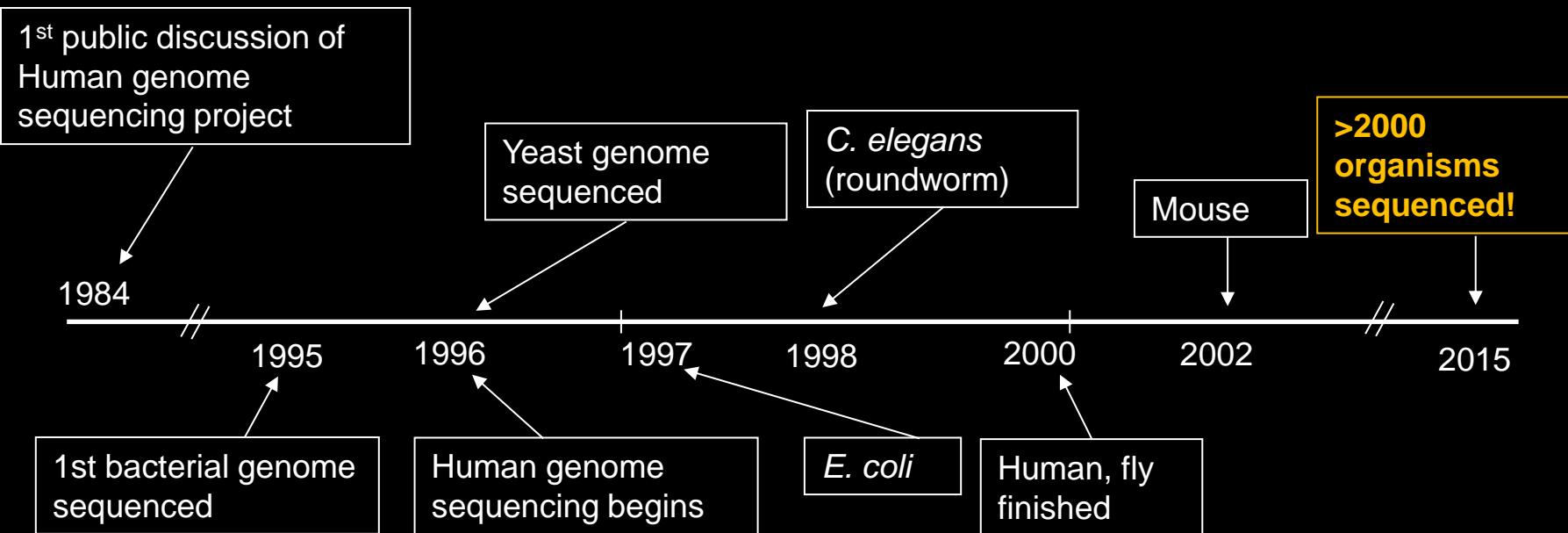
Information storage (DNA)
→ cellular machinery (proteins)

(Almost) all organisms have this basic process in common

The genome sequencing revolution

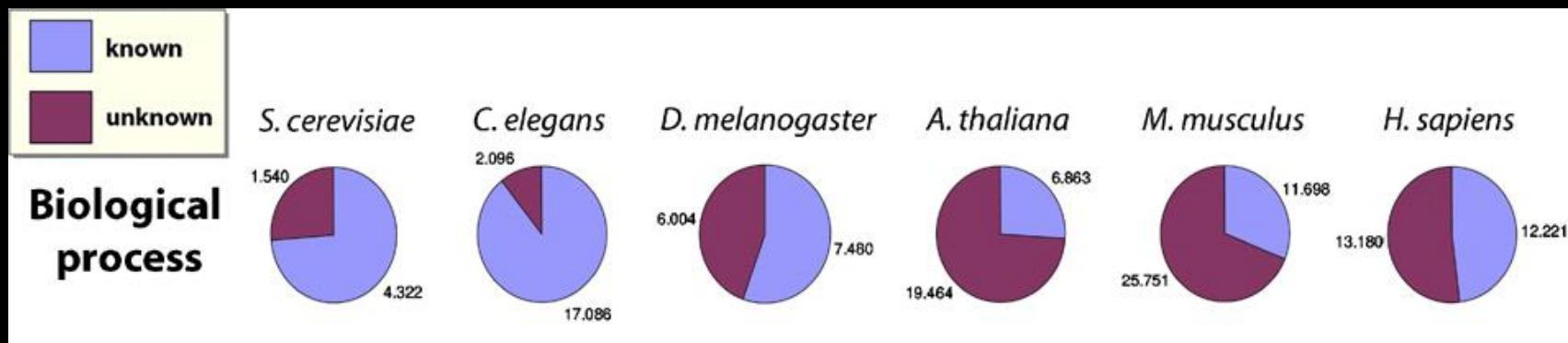


Revolutionizing genome sequencing technology



What's left to do?

With rapid sequencing technology, and many complete genomes sequenced, are we done? **NO!**



- the genome sequence is really just a “parts list”
- understanding the cell requires learning what each part does (e.g. which other parts it interacts with, which function(s) it carries out)

The big questions in computational/systems biology

- What are all of the parts? What is their function? When are they used?
- How do the various parts interact to carry out biological processes/form systems?
- How do genomes vary (evolve) across larger time scales?
- How do genomes vary across individuals? Which variants relate to which traits (e.g. disease)?

Computer science enables the investigation of all of these!

My lab's focuses

Gene expression variation in maize



Regulatory networks in *C. elegans*



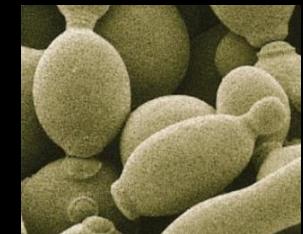
Network analysis & algorithms



Genetic interactions in *Arabidopsis*

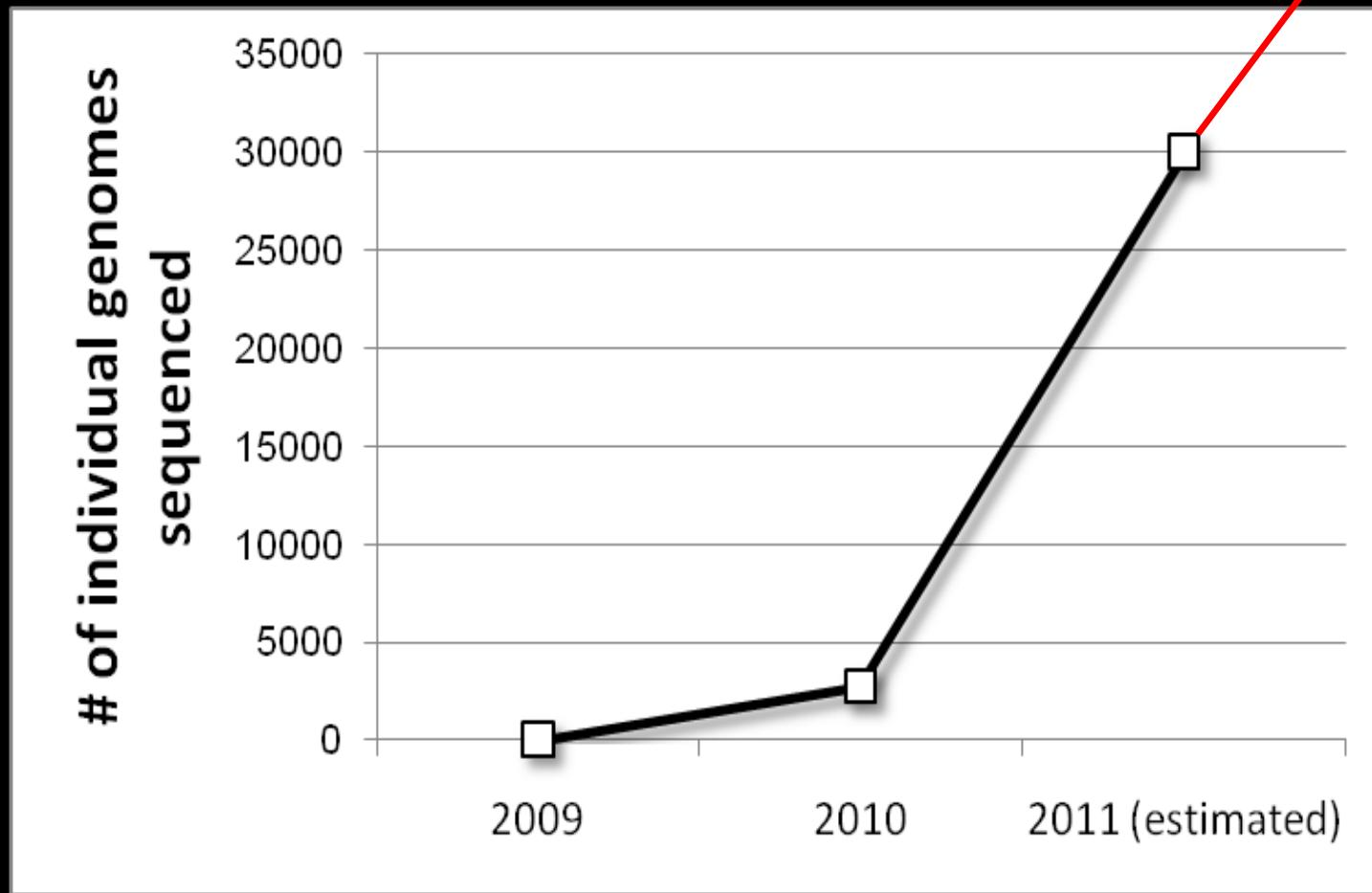


Targeted cancer therapeutics



Genetic interactions
Chemical genomics in yeast

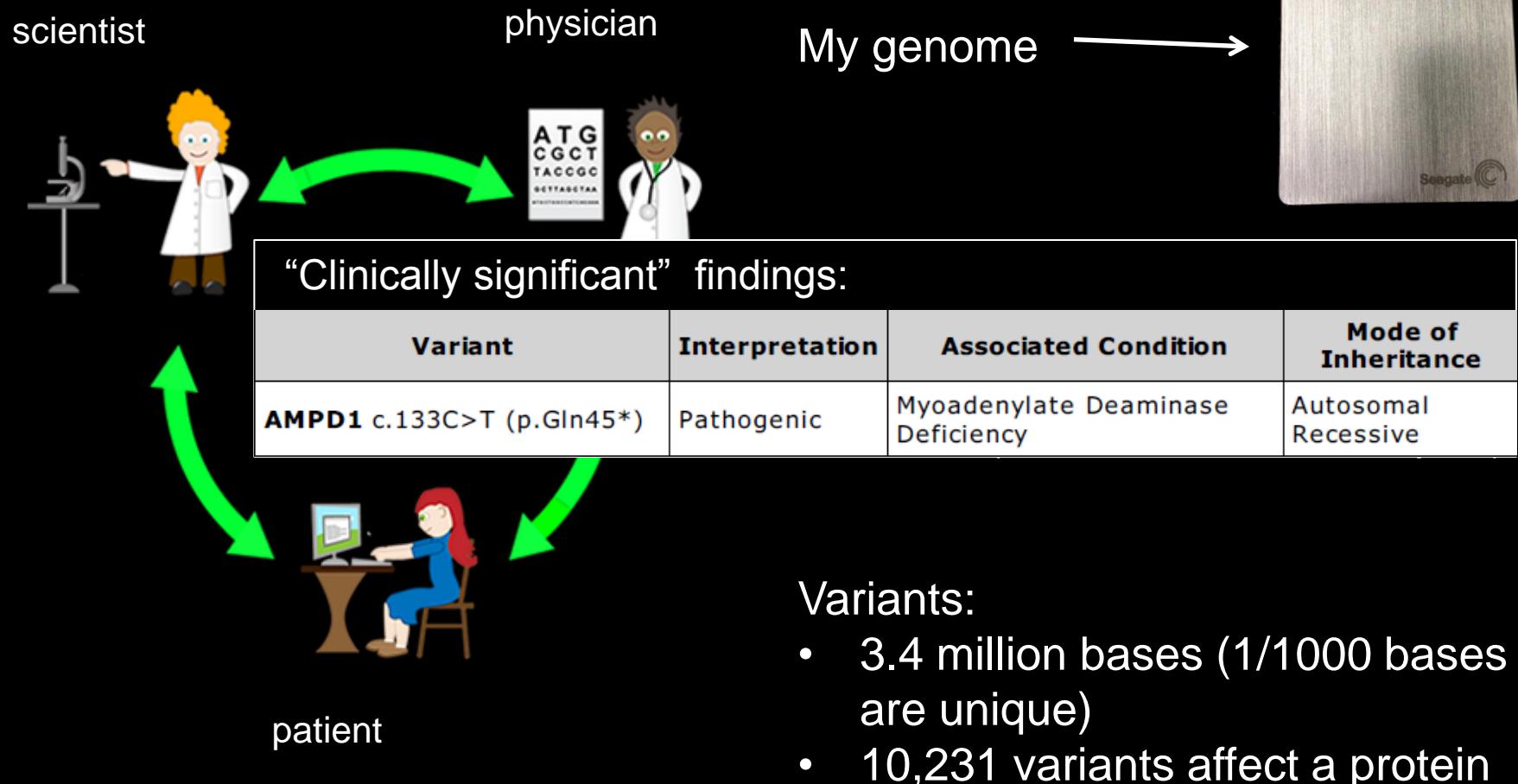
The genome sequencing revolution: individual human genomes



Data from:

Human genome: Genomes by the thousand. Nature 467, 1026-1027 (2010)
Yngvadottir *et al.* Genome Biology 2009 **10**:237 doi:10.1186/gb-2009-10-9-237

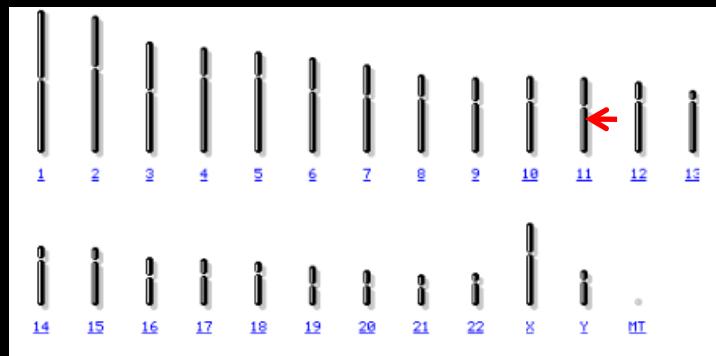
Genomics gets personal



Genomics gets personal

#SNP ID	chromosome	position	genotype	Name	Confidence ▲	Outcome
rs4477212	1	72017	AA	Alcohol Flush Reaction	★★★★	Does Not Flush
rs3094315	1	742429	AG	Bitter Taste Perception	★★★★	Can Taste
rs3131972	1	742584	AG	Earwax Type	★★★★	Wet
rs12124819				Eye Color	★★★★	Likely Blue
rs11240777				Hair Curl 	★★★★	Slightly Curlier Hair on Average
rs6681049				Lactose Intolerance	★★★★	Likely Tolerant
rs4970383				Malaria Resistance (Duffy Antigen)	★★★★	Not Resistant
rs4475691				Male Pattern Baldness 	★★★★	Increased Odds
rs7537756				Muscle Performance	★★★★	Likely Sprinter
rs13302982				Non-ABO Blood Groups	★★★★	See Report
rs1110052				Norovirus Resistance	★★★★	Not Resistant
rs2272756				Resistance to HIV/AIDS	★★★★	Not Resistant
rs3748597				Smoking Behavior	★★★★	If a Smoker, Likely to Smoke More
rs13303106						
rs28415373						
rs13303010						
rs6696281						
rs28391282						
rs2340592						
rs13303118						
rs6665000						
rs2341362						
rs9777703						
rs1891910						
rs9697457						
rs35940137						
rs3128117						
rs2465126						
rs2341365						
rs15842						
rs6657048	1	947503	CC			
rs2710888	1	949705	CC			
rs3128126	1	952073	AA			
rs2710875	1	967643	TT			
rs2465136	1	980280	TT			
rs2488991	1	984254	TT			
rs7526076	1	988258	AG			
rs3934834	1	995669	CT			
rs3766192	1	1007060	CC			
rs3766191	1	1007450	CT			
rs9442372	1	1008567	AA			
rs10907177	1	1011209	AG			
rs3737728	1	1011278	AG			
rs10907178	1	1011446	AC			

I should have been a sprinter??



Alpha-actinin-3 (ACTN3)

(actin-binding protein
expressed only in
fast twitch fibers)

Who	Genotype	What It Means
Mom	CC	Two working copies of alpha-actinin-3 in fast-twitch muscle fiber. Many world-class sprinters and some endurance athletes have this genotype.
Me	CT	One working copy of alpha-actinin-3 in fast-twitch muscle fiber. Many world-class sprinters and some endurance athletes have this genotype.
Dad	TT	No working copies of alpha-actinin-3 in fast-twitch muscle fiber. Few world-class sprinters have this genotype, but many world-class endurance athletes do.

Most variants result in relatively minor increase in disease risk

Elevated Risk 					
Name	Confidence	Your Risk	Avg. Risk	Compared to Average	
Gallstones 	★★★★	11.1%	7.0%	1.58x	
Age-related Macular Degeneration	★★★★	9.5%	7.0%	1.35x	
Restless Legs Syndrome	★★★★	2.5%	2.0%	1.25x	
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.4%	0.4%	1.21x	
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.3%	0.2%	1.22x	
Scleroderma (Limited Cutaneous Type)	★★★★	0.08%	0.07%	1.24x	

Decreased Risk 					
Name	Confidence	Your Risk	Avg. Risk	Compared to Average	
Alzheimer's Disease 	★★★★	4.9%	7.2%	0.69x	
Melanoma	★★★★	2.2%	2.9%	0.75x	
Rheumatoid Arthritis	★★★★	1.0%	2.4%	0.43x	
Ulcerative Colitis	★★★★	0.6%	0.8%	0.72x	
Exfoliation Glaucoma	★★★★	0.2%	0.7%	0.22x	
Type 1 Diabetes	★★★★	0.2%	1.0%	0.15x	
Crohn's Disease	★★★★	0.09%	0.53%	0.18x	
Celiac Disease	★★★★	0.07%	0.12%	0.58x	

“Phantom” heritability

- Variation in height is ~80% heritable
- known genetic loci (~40) only explain ~5% of heritable variance

Disease	Number of loci	Proportion explained
Age-related macular degeneration ⁷²	5	50%
Crohn's disease ²¹	32	20%
Systemic lupus erythematosus ⁷³	6	15%
Type 2 diabetes ⁷⁴	18	6%
HDL cholesterol ⁷⁵	7	5.2%
Height ¹⁵	40	5%
Early onset myocardial infarction ⁷⁶	9	2.8%
Fasting glucose ⁷⁷	4	1.5%

Manolio et al. Nature 461, 747-753(8 October 2009)



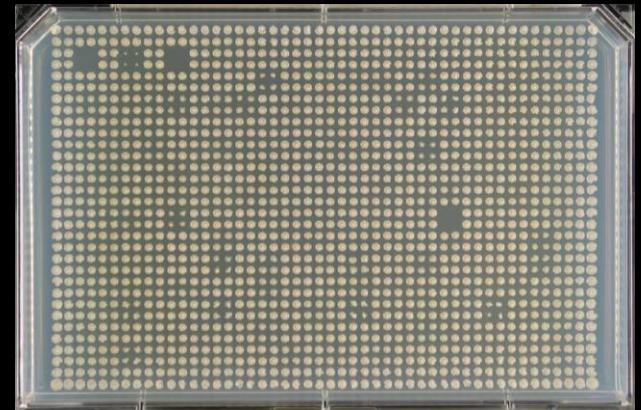
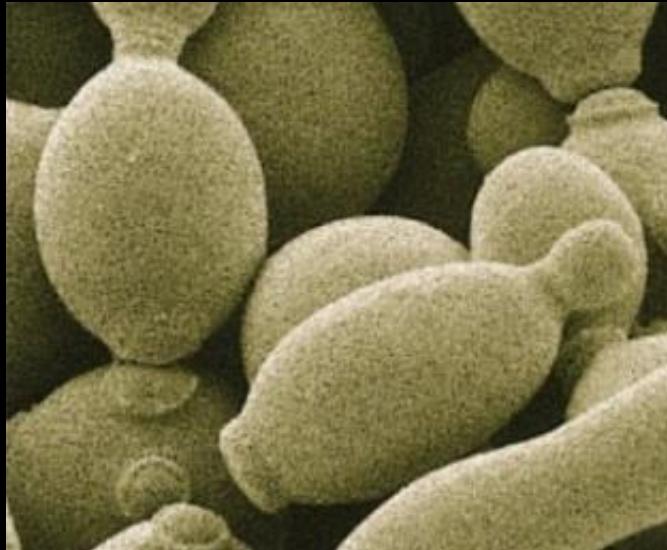
© BBC Publicity

<http://www.dailymail.co.uk/health/article-1316471/Genes-decide-tall-short.html>

Maybe combinations of genetic variants are the cause?

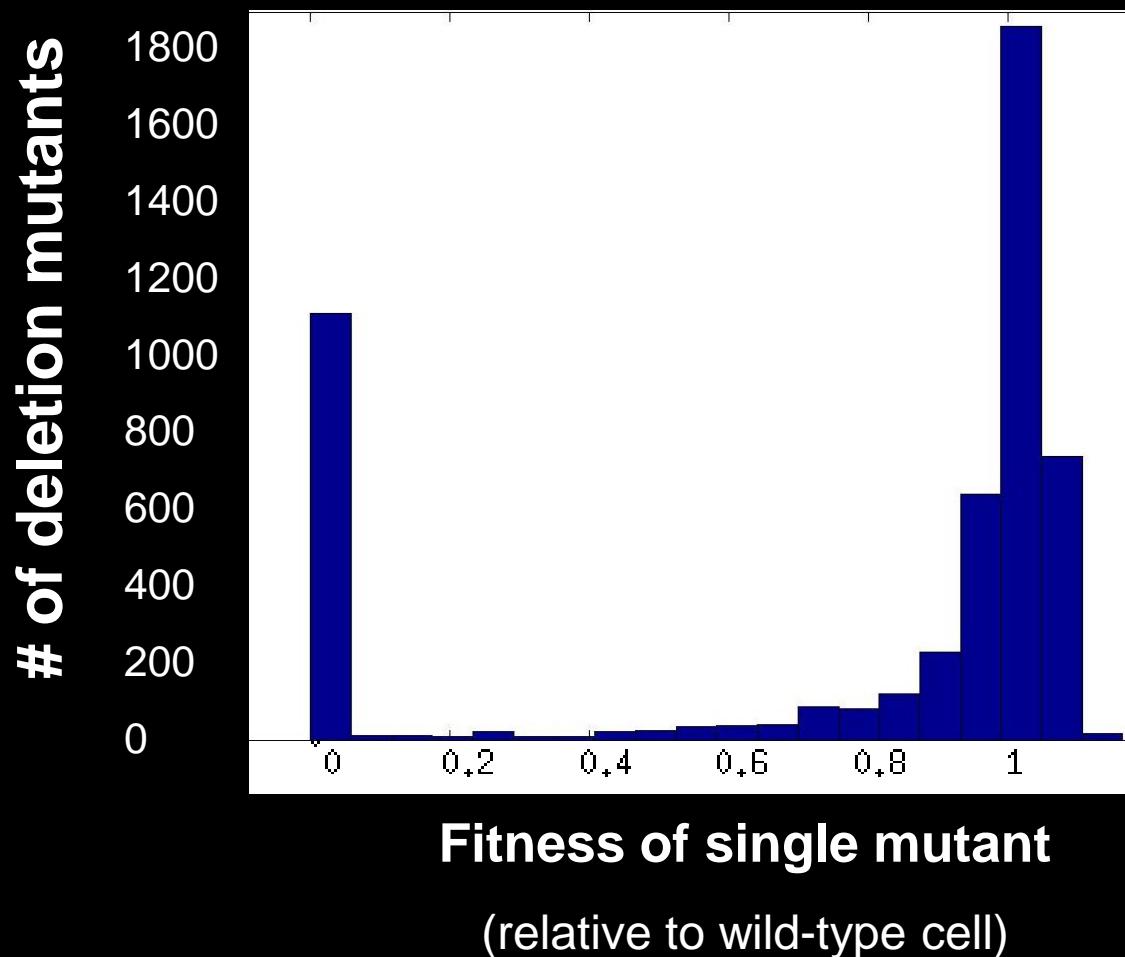
A simpler organism, simpler phenotype

Baker's yeast (*Saccharomyces cerevisiae*)



How harmful are single gene deletions to a yeast cell?

Fitness distribution of all 6000 yeast deletion mutants



One interesting outcome of combining gene deletions

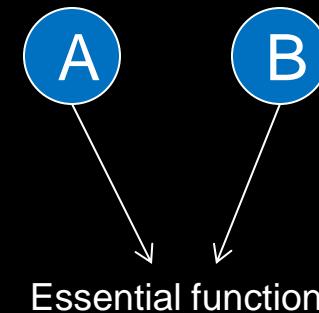
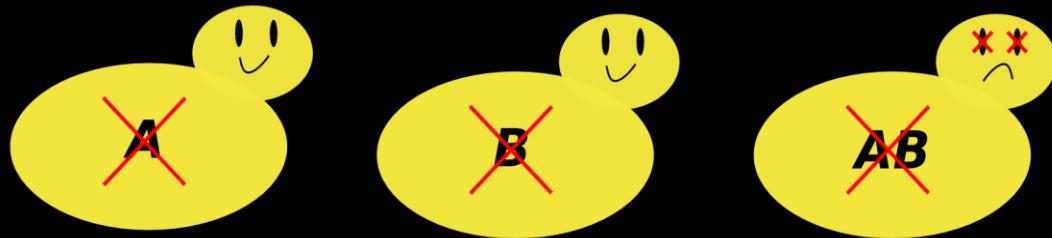
A	$a\Delta$	A	$a\Delta$
→	→	→	→
→	→	→	→
B	B	$b\Delta$	$b\Delta$

Wild-type

Viable

Viable

Lethal



“synthetic lethality”

Or, more generally, “genetic interaction”

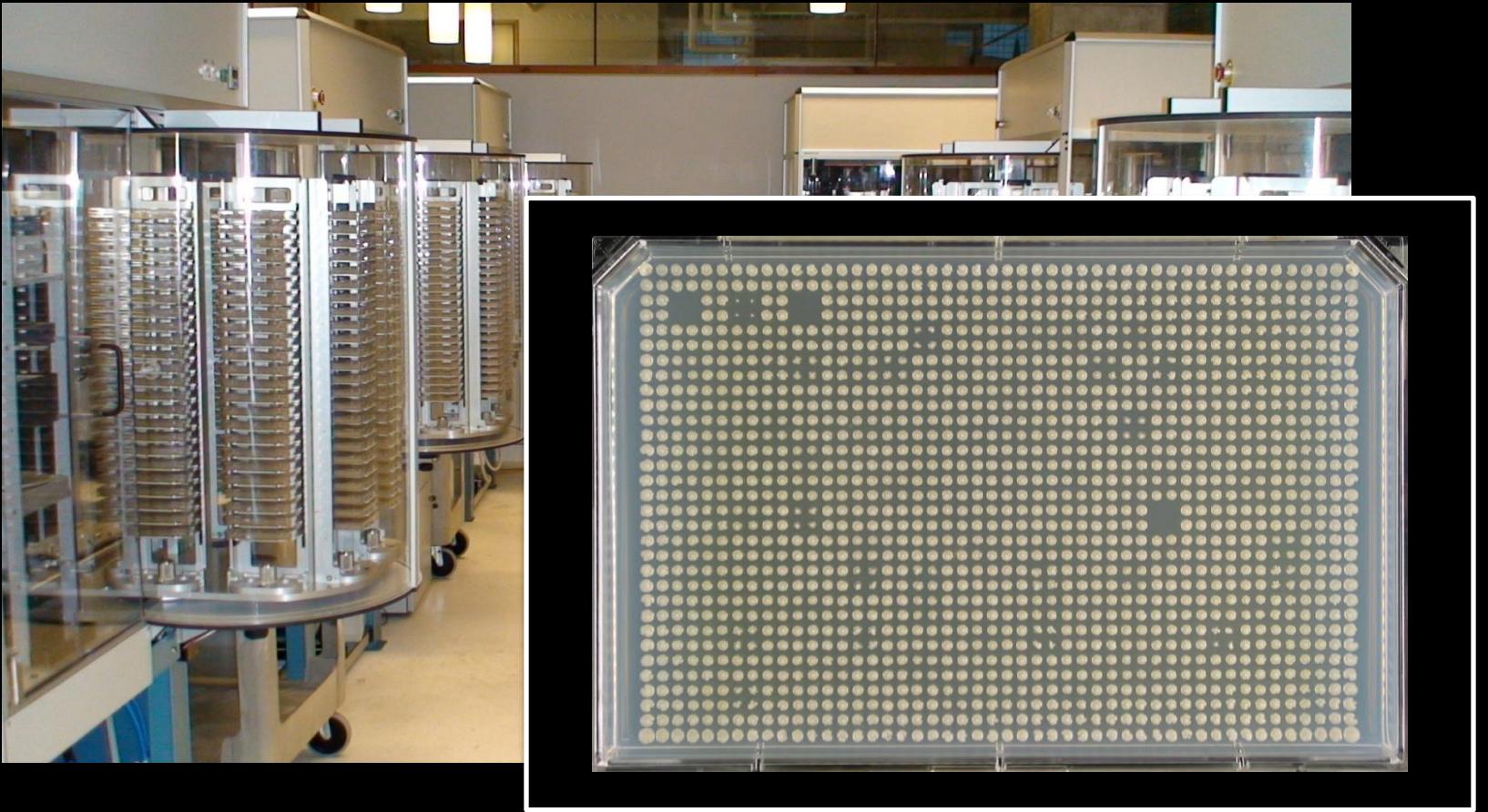
The problem with combinatorial perturbation

Consider yeast, ~6000 genes

- ~18 million pairs
- $\sim 10^{10}$ triples

Need automated, high throughput approach!

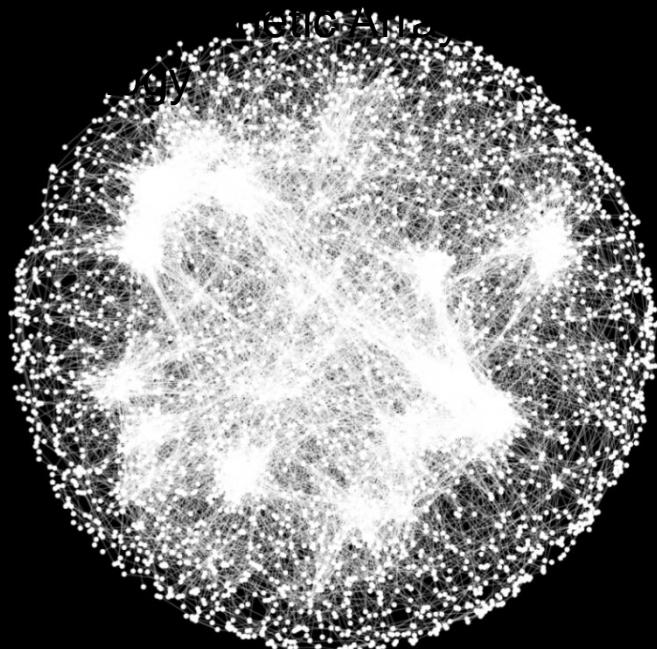
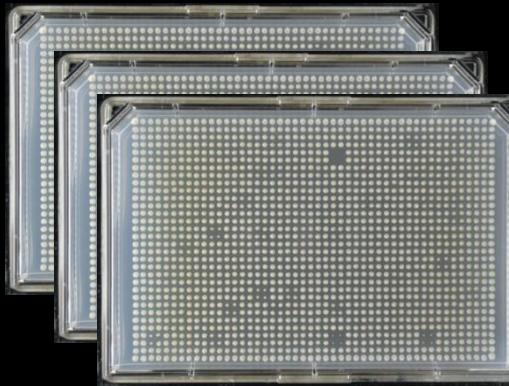
Robots to the rescue!



Tong, A.H.Y. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364-2368 (2001).

Charlie Boone, Brenda Andrews (U. Toronto)

A near complete genetic interaction map in yeast



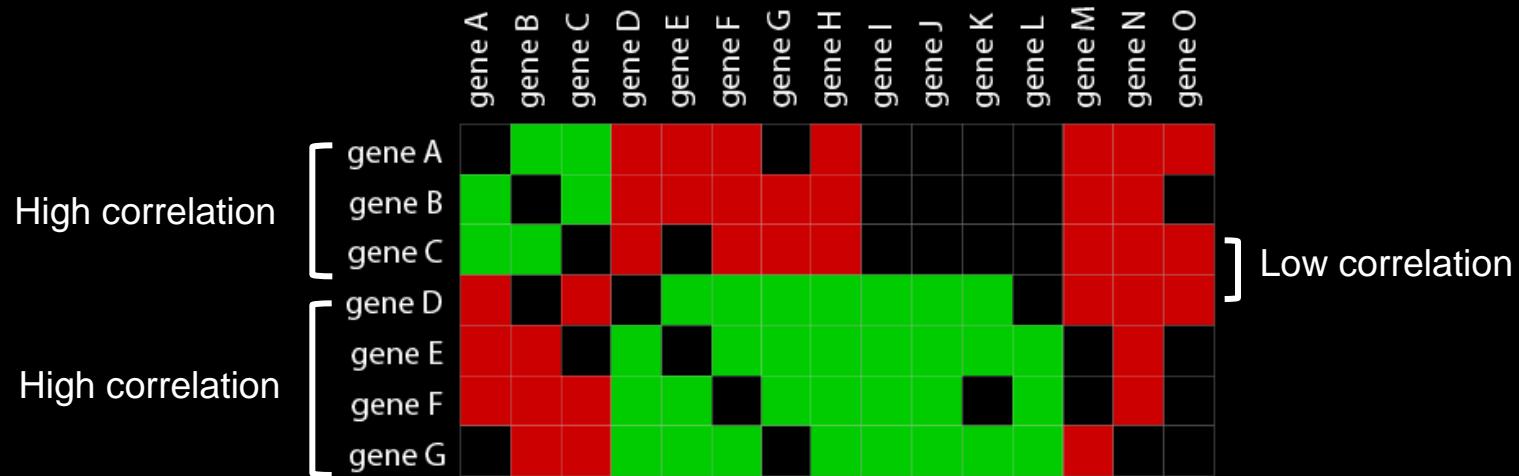
~23 million double mutants screened for genetic interactions (~90% of possible gene pairs)

~1 million genetic interactions discovered

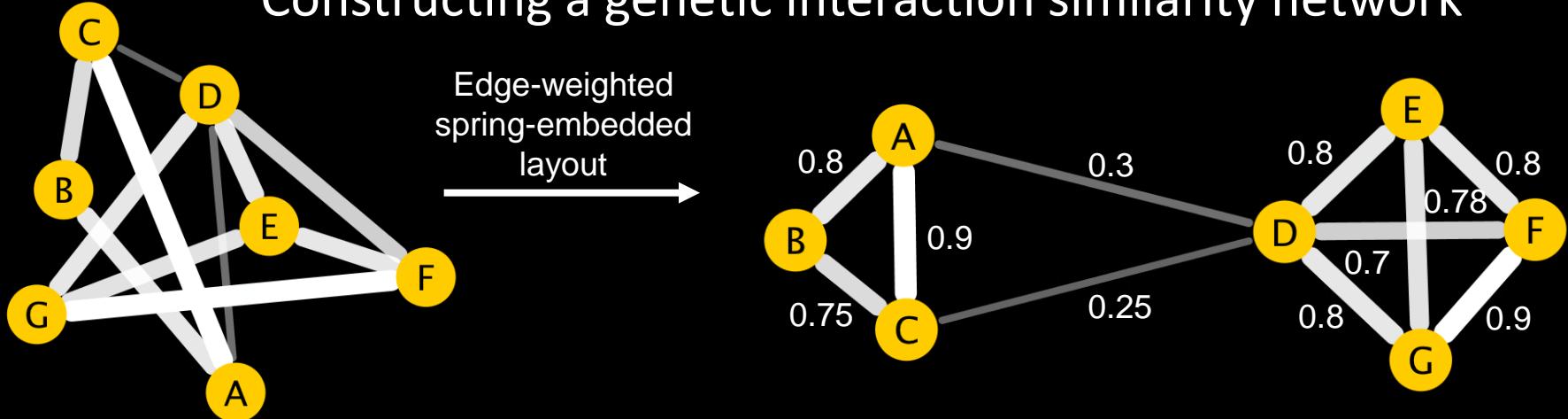
- 550, 000 negative interactions
- 350, 000 positive interactions

Costanzo et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* 2016.

Interaction profiles are predictive of gene function

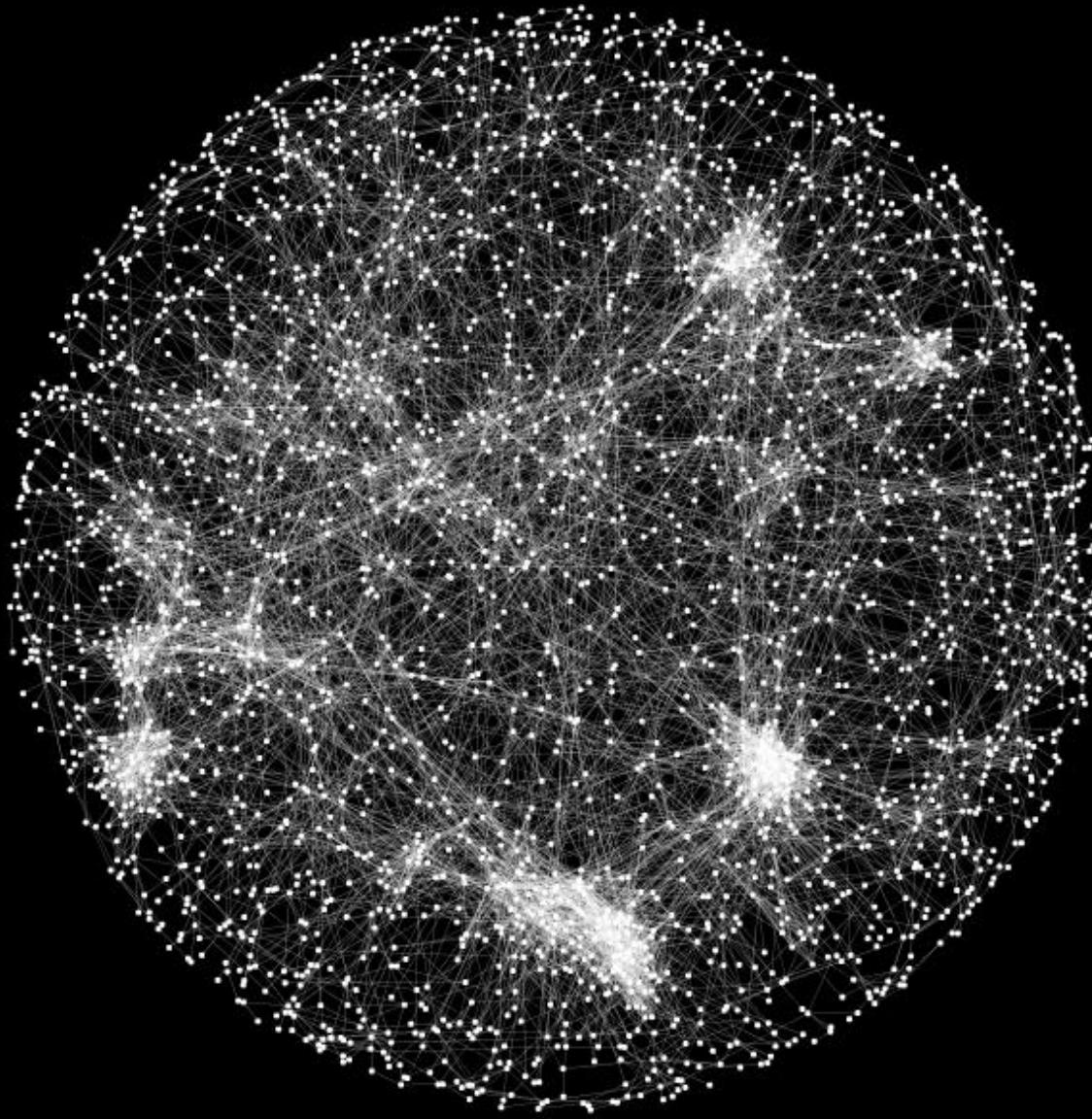


Constructing a genetic interaction similarity network

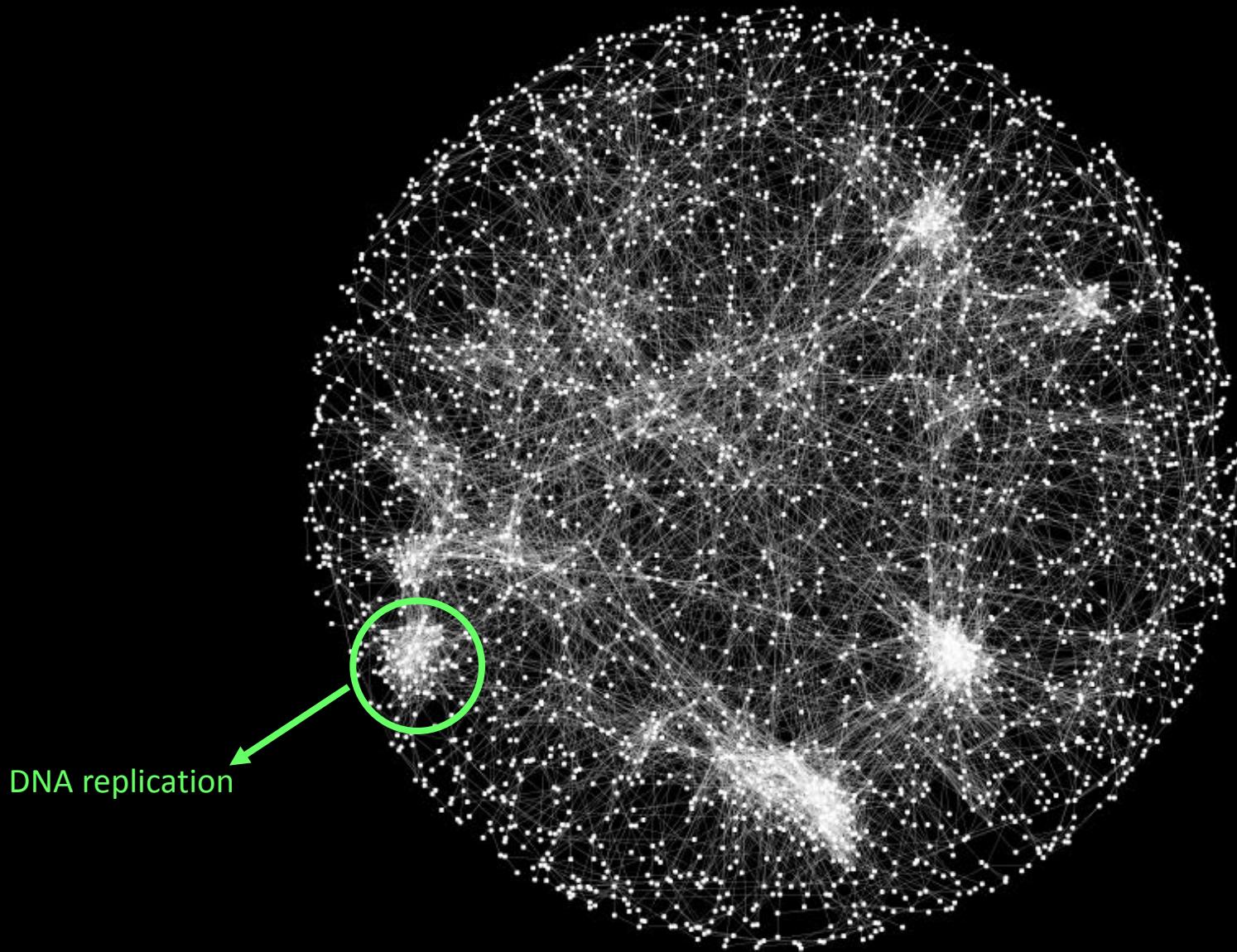


Similar genetic interaction profiles reveal global functional map

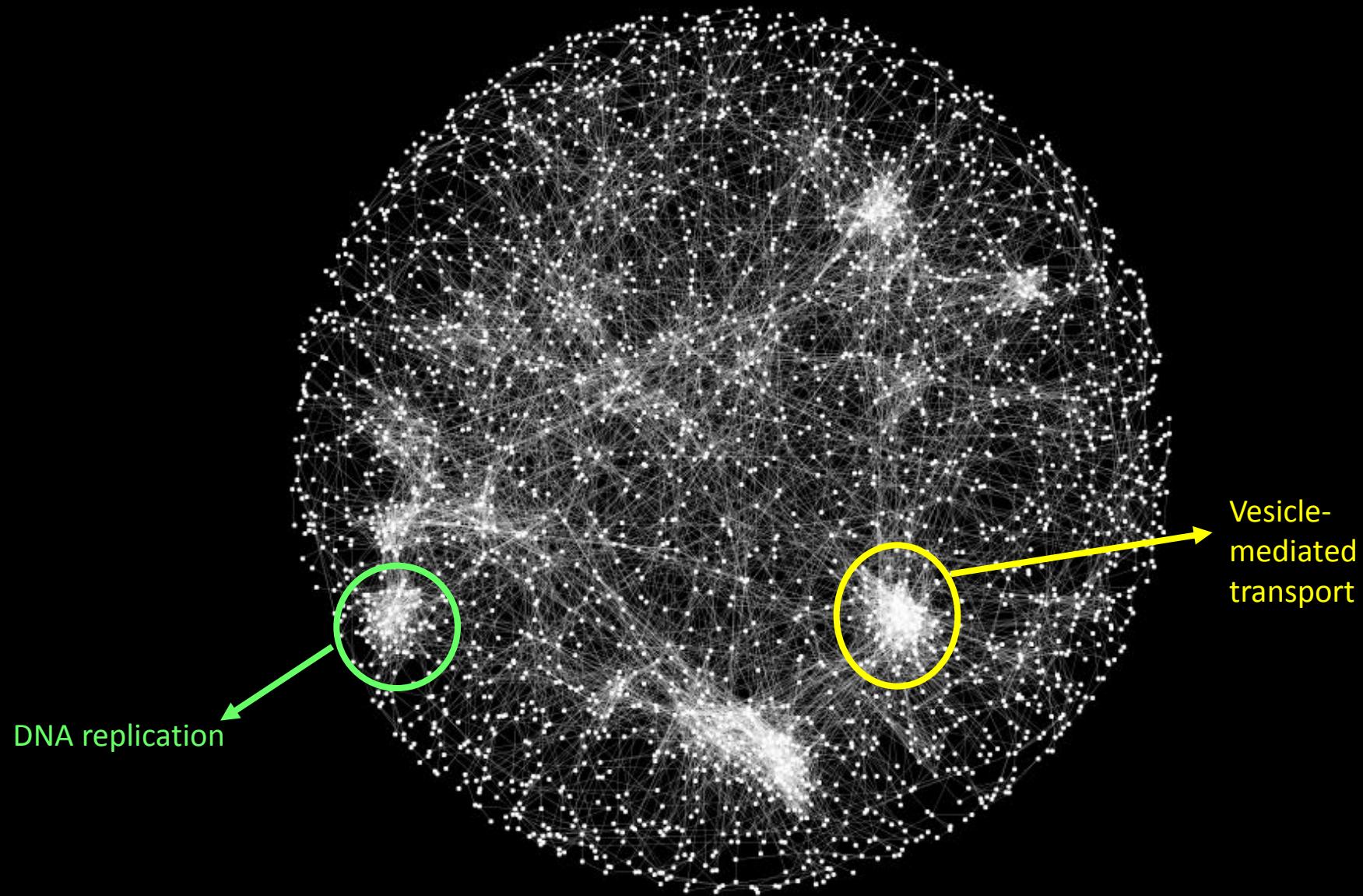
(connection → similar interaction profile)



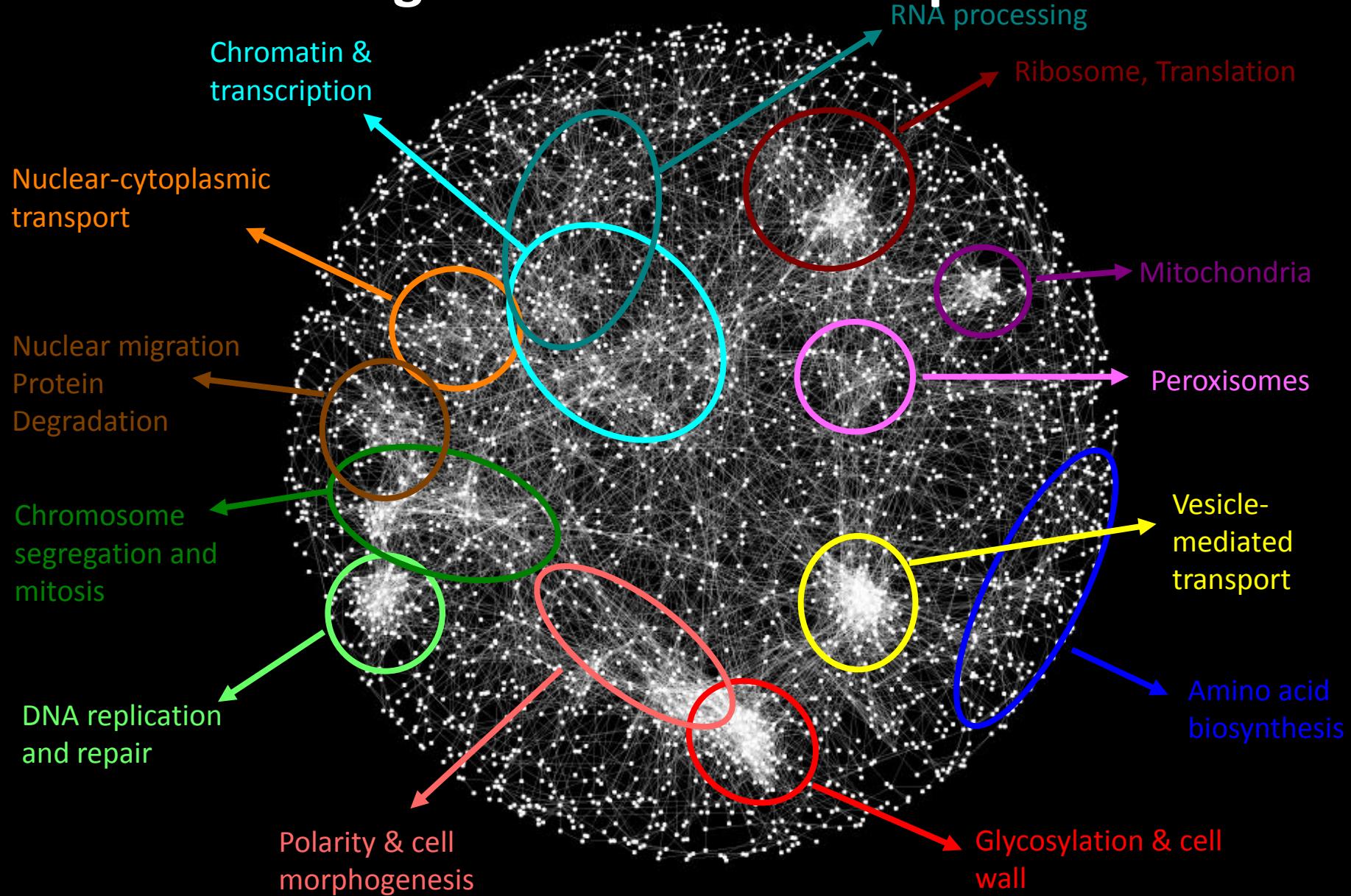
Similar genetic interaction profiles reveal global functional map



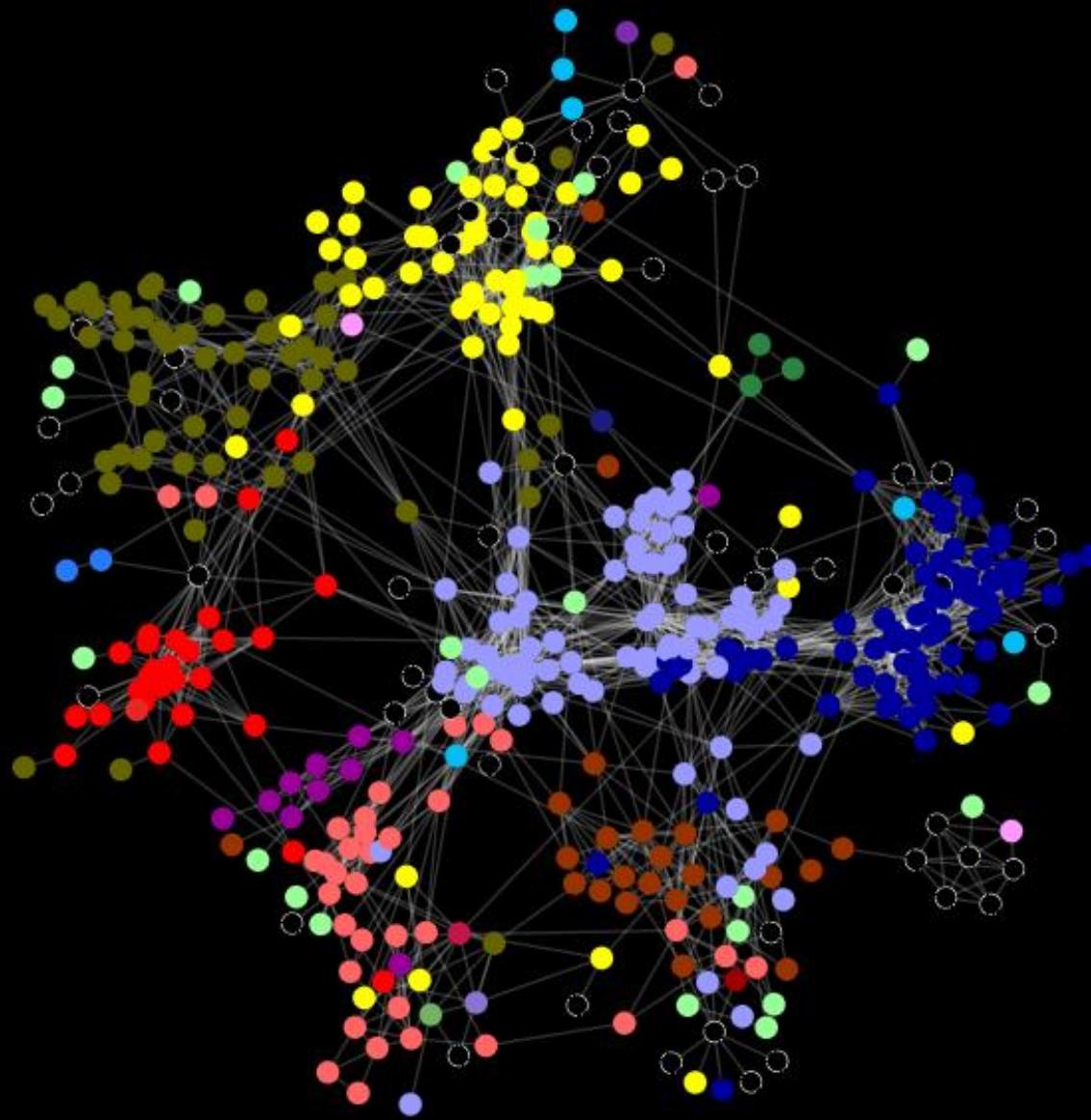
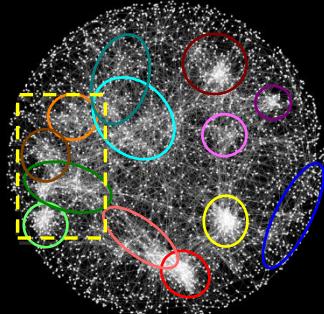
Similar genetic interaction profiles reveal global functional map



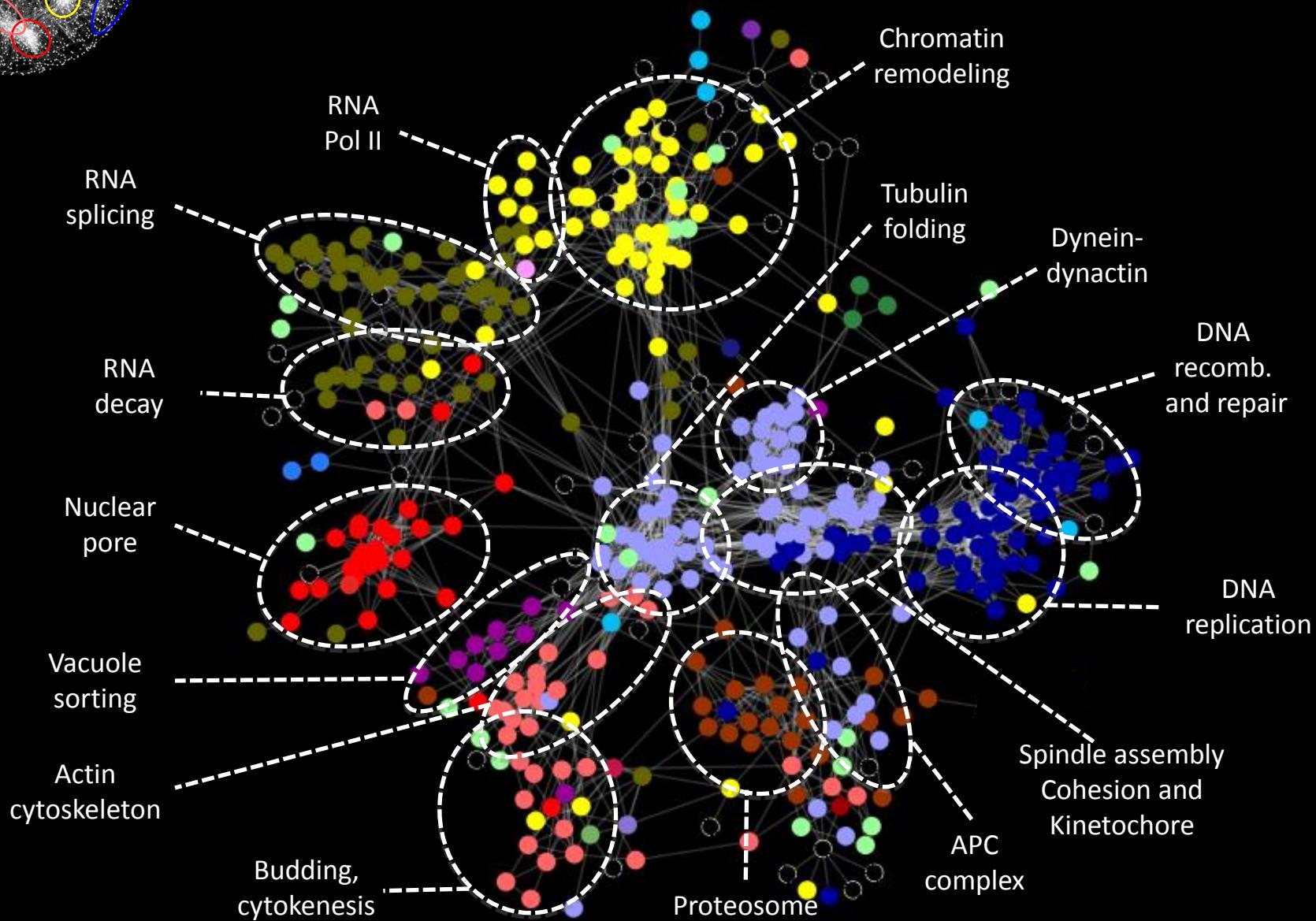
Similar genetic interaction profiles reveal global functional map



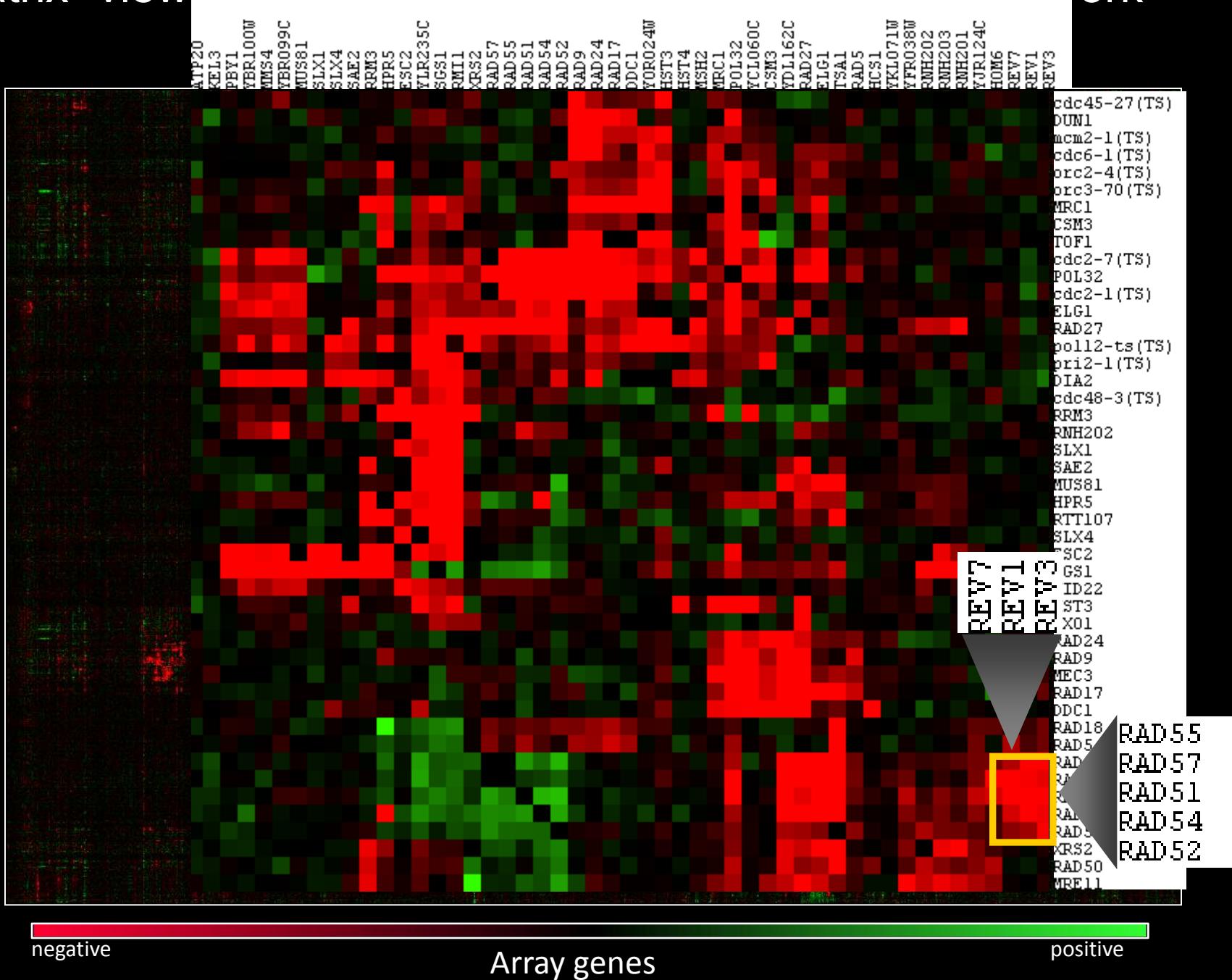
Similar genetic interaction profiles reveal global functional map



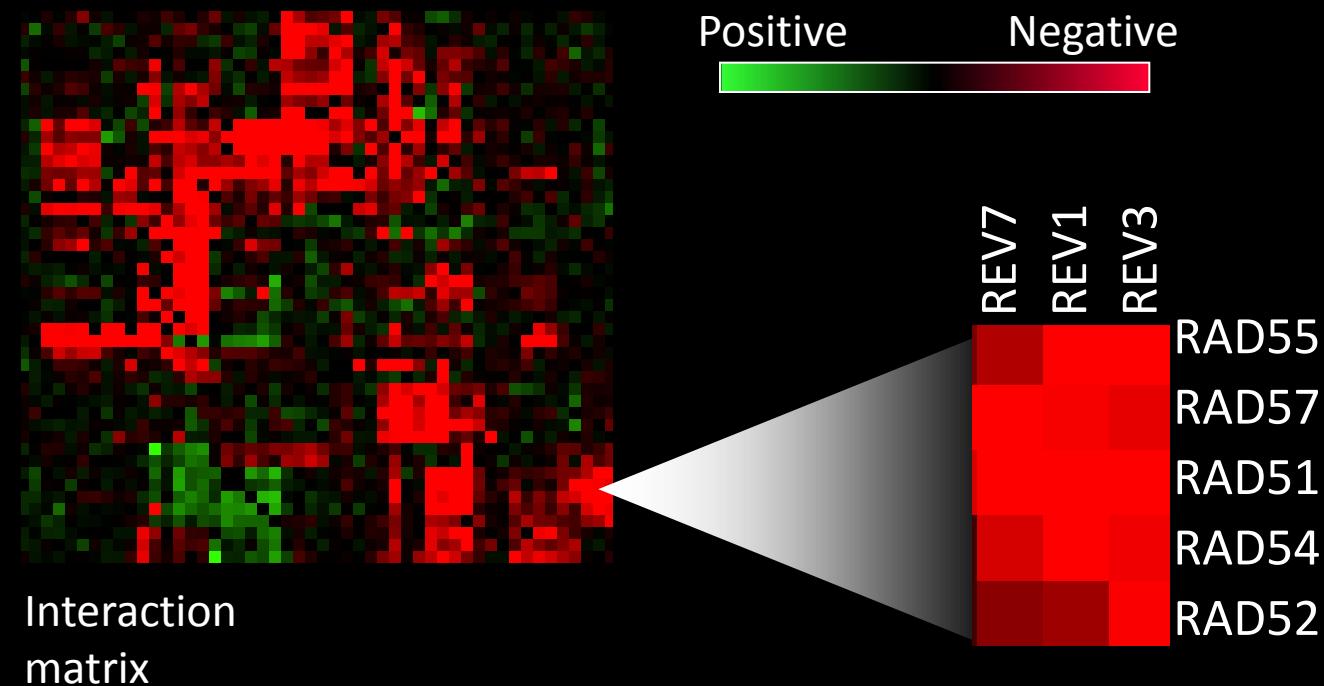
Similar genetic interaction profiles reveal global functional map



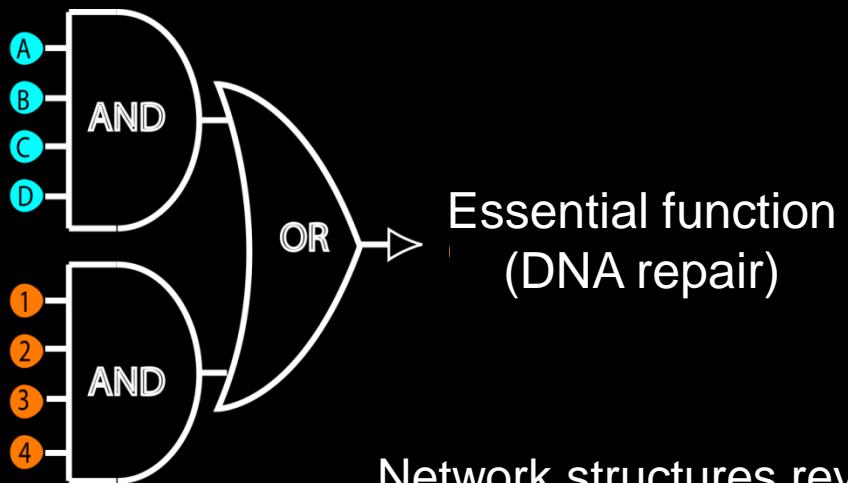
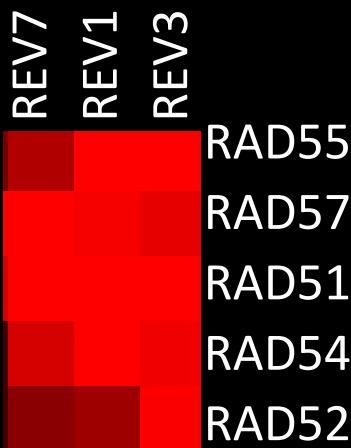
“Matrix” view of the global yeast genetic interaction network



What kind of local network structure leads to highly correlated interaction profiles?

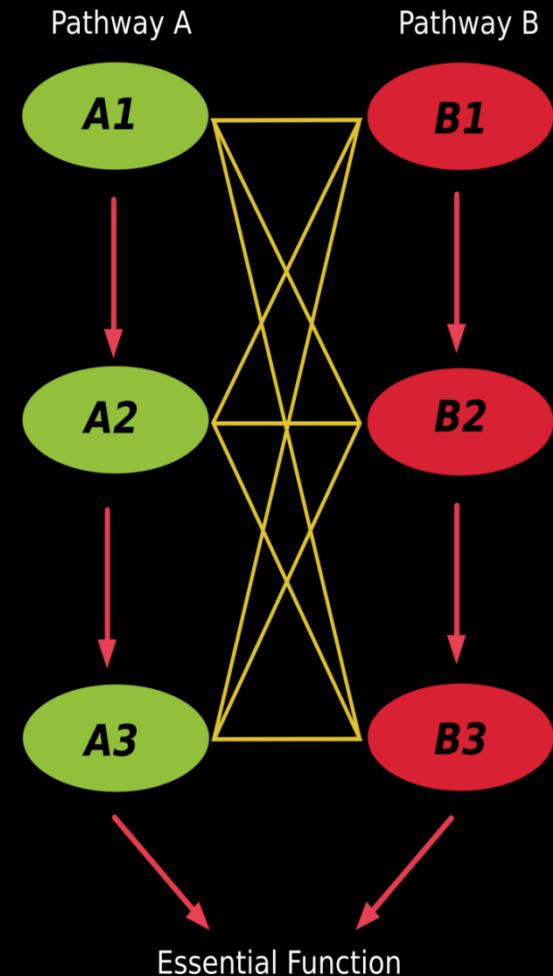


Local network structure

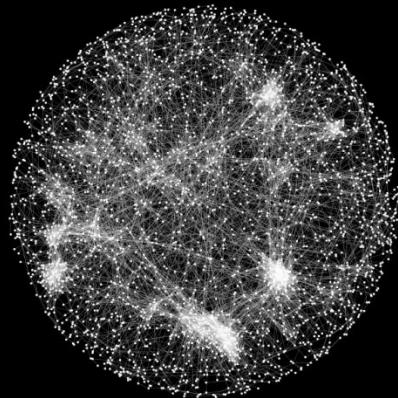


Network structures reveal:

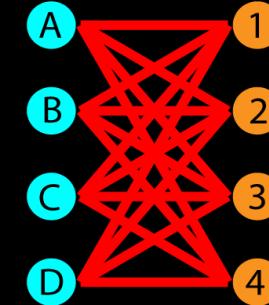
- gene modules
- redundancy often occurs at the *module* (not gene) level



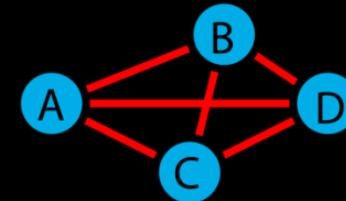
Exhaustive mining for local network structures



Exhaustive structure
mining algorithm



OR



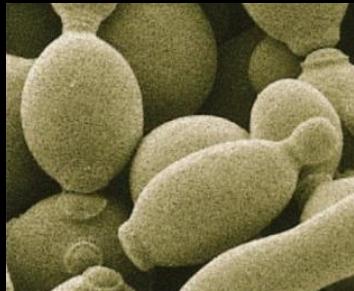
Results (negative interactions):

Real network: 10,459 blocks

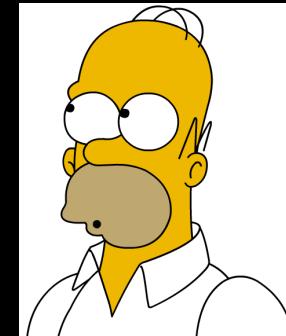
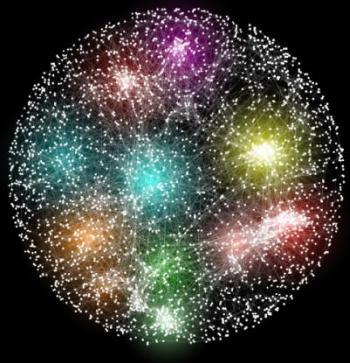
Randomized networks: ~20
blocks

($\geq 3 \times 3$ interaction blocks)

How can yeast help us understand/treat human disease?



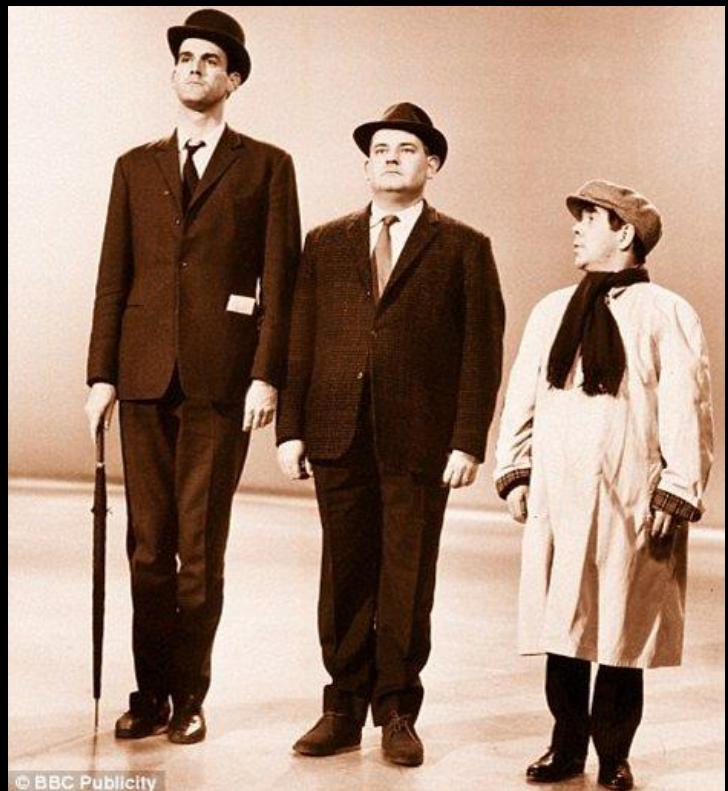
?



- Detecting interactions in genome-wide association studies
- Developing new cancer drugs

Reminder about “Phantom” heritability

- Variation in height is ~80% heritable
- known genetic loci (~40) only explain ~5% of heritable variance



<http://www.dailymail.co.uk/health/article-1316471/Genes-decide-tall-short.html>

Maybe combinations of genetic variants are the cause?

Lessons from yeast

Isolated interactions

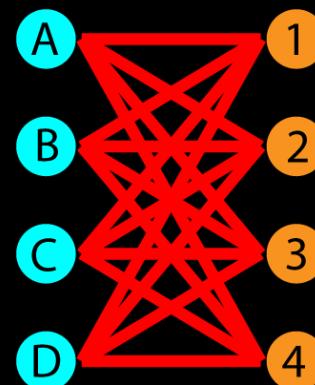


Estimated frequency:

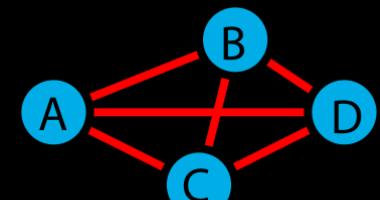
~25%

Structured

Between
Module

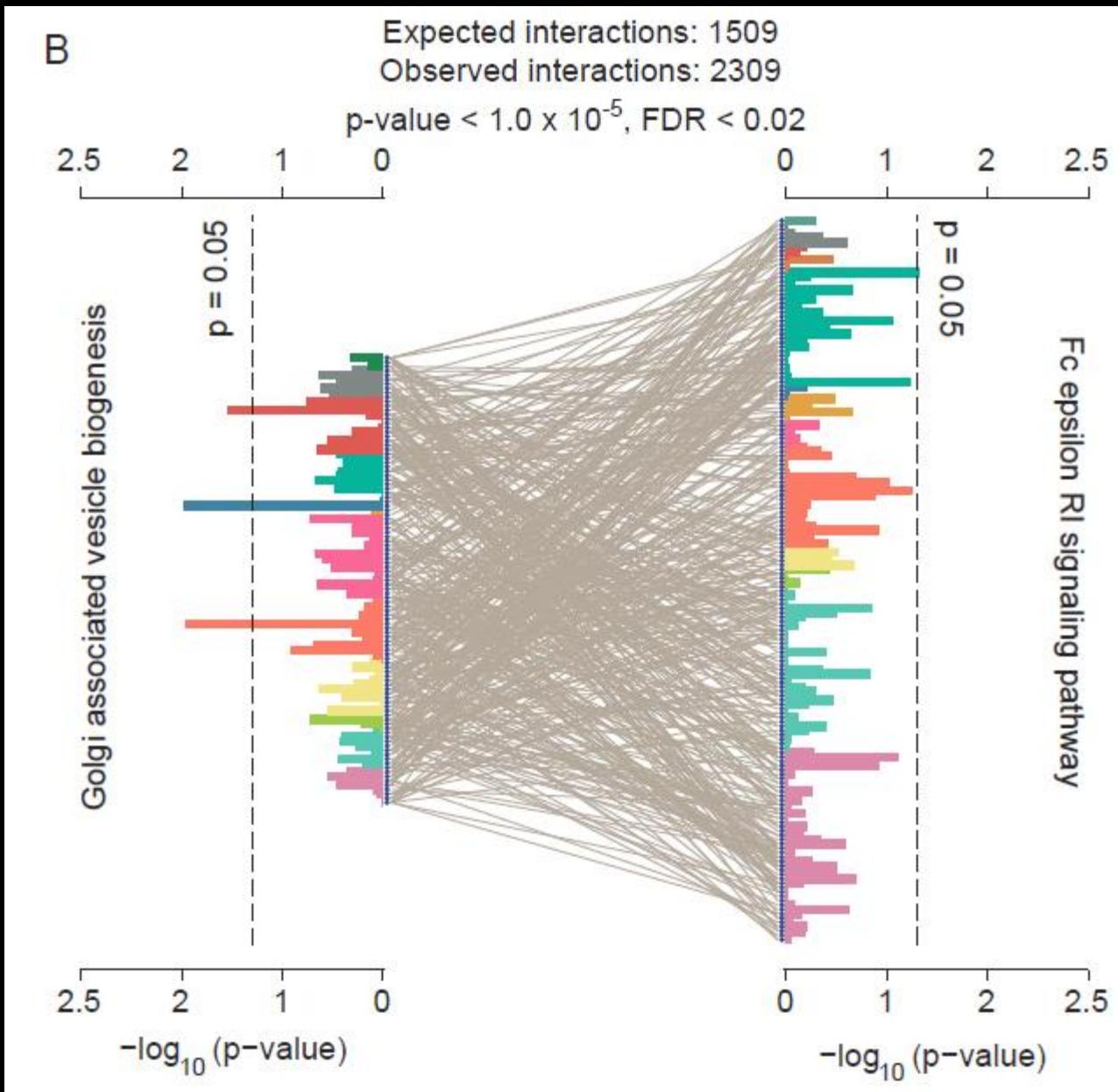


Within Module



~75%

Example interaction (Parkinson's disease)

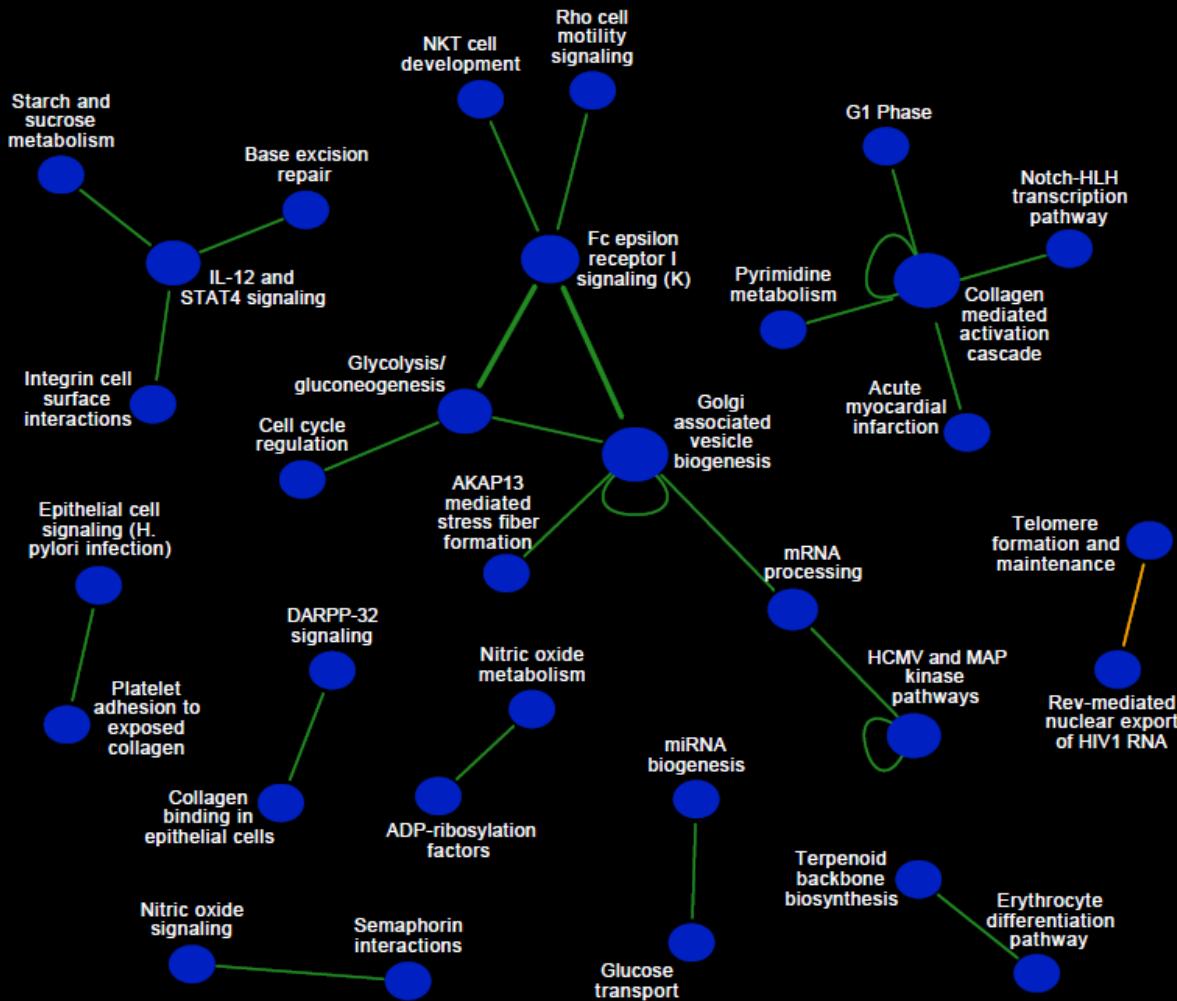


Parkinson's
disease
GWAS

Simón-Sánchez
et al., *Nature Genetics*, 2009

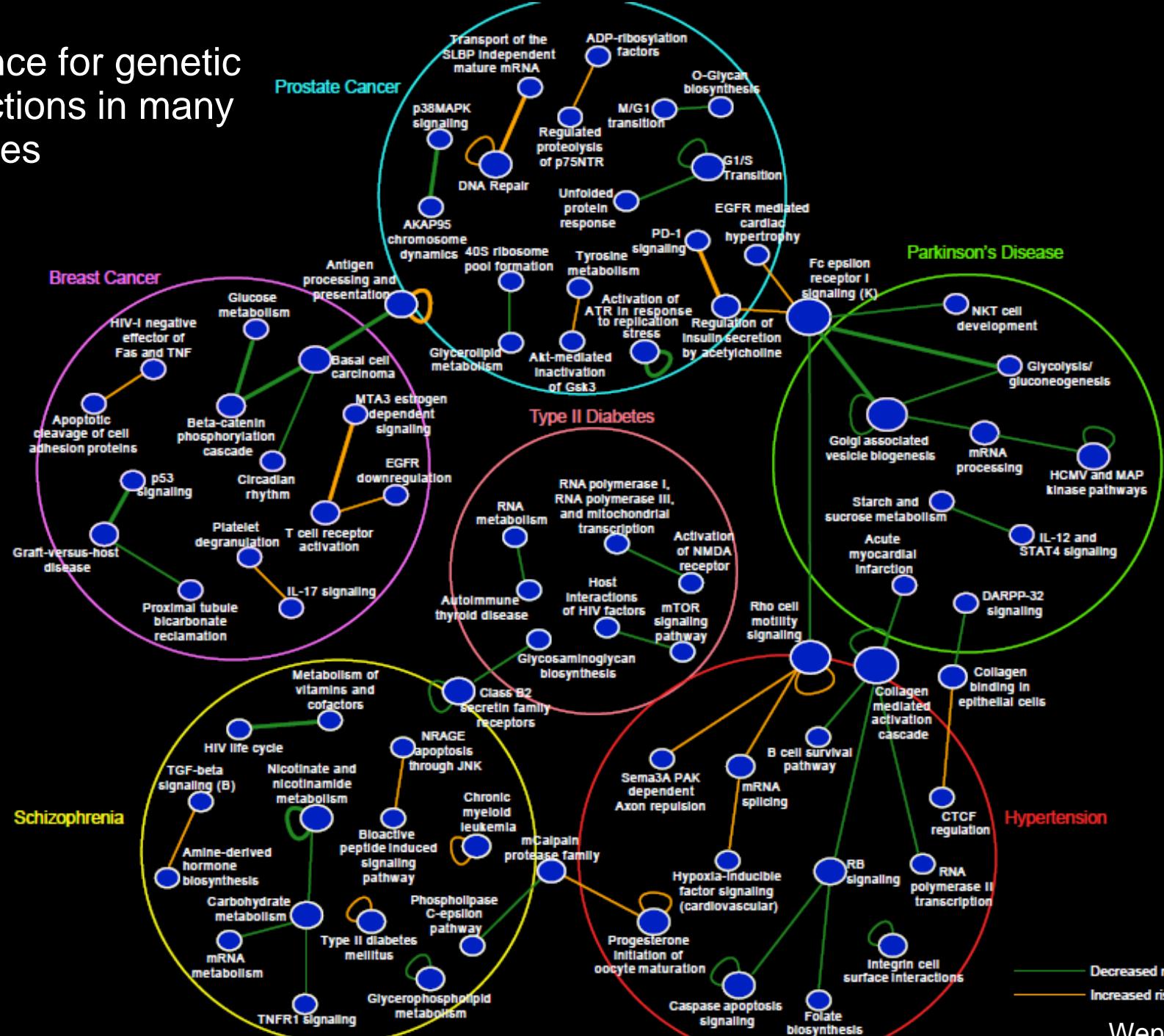
Wen Wang
Gang Fang
Vipin Kumar

Evidence for many genetic interactions in Parkinson's disease



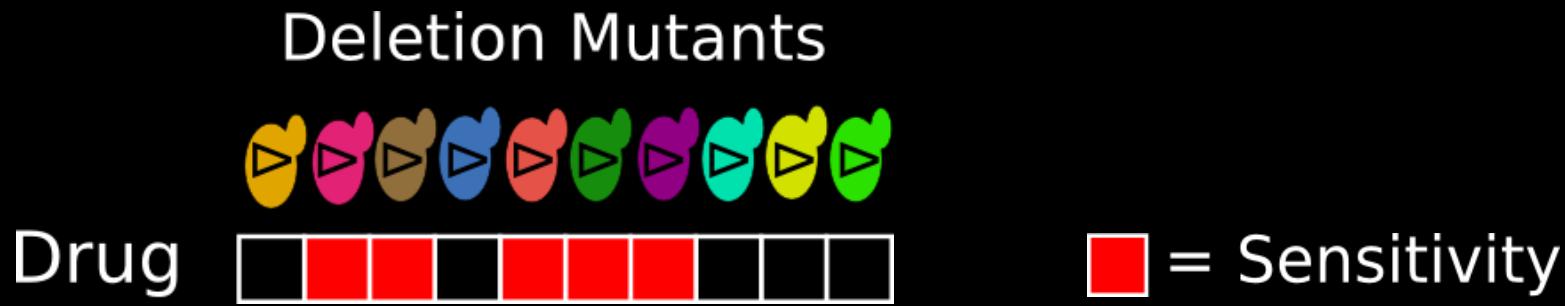
FDR<0.25:
23 interactions

Evidence for genetic interactions in many diseases



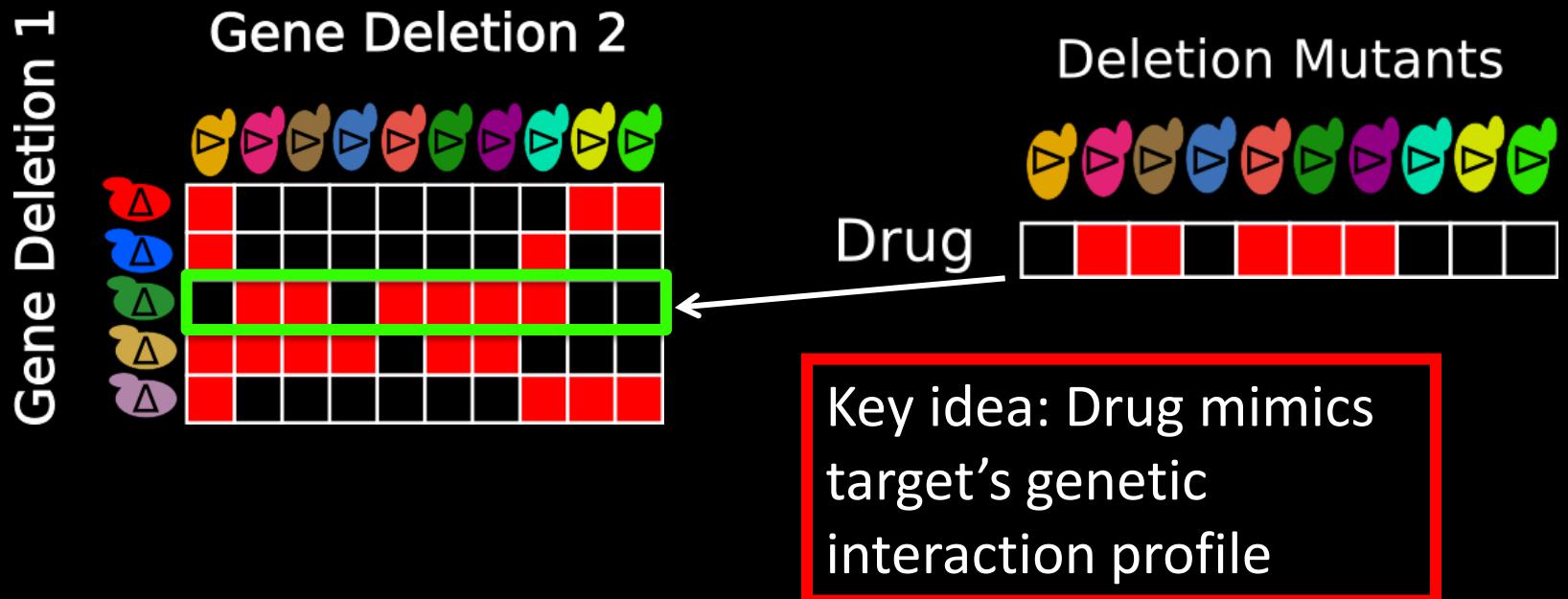
Large-scale chemical genomics:
Understanding how chemicals interact
with cells

Chemical genetic profiling

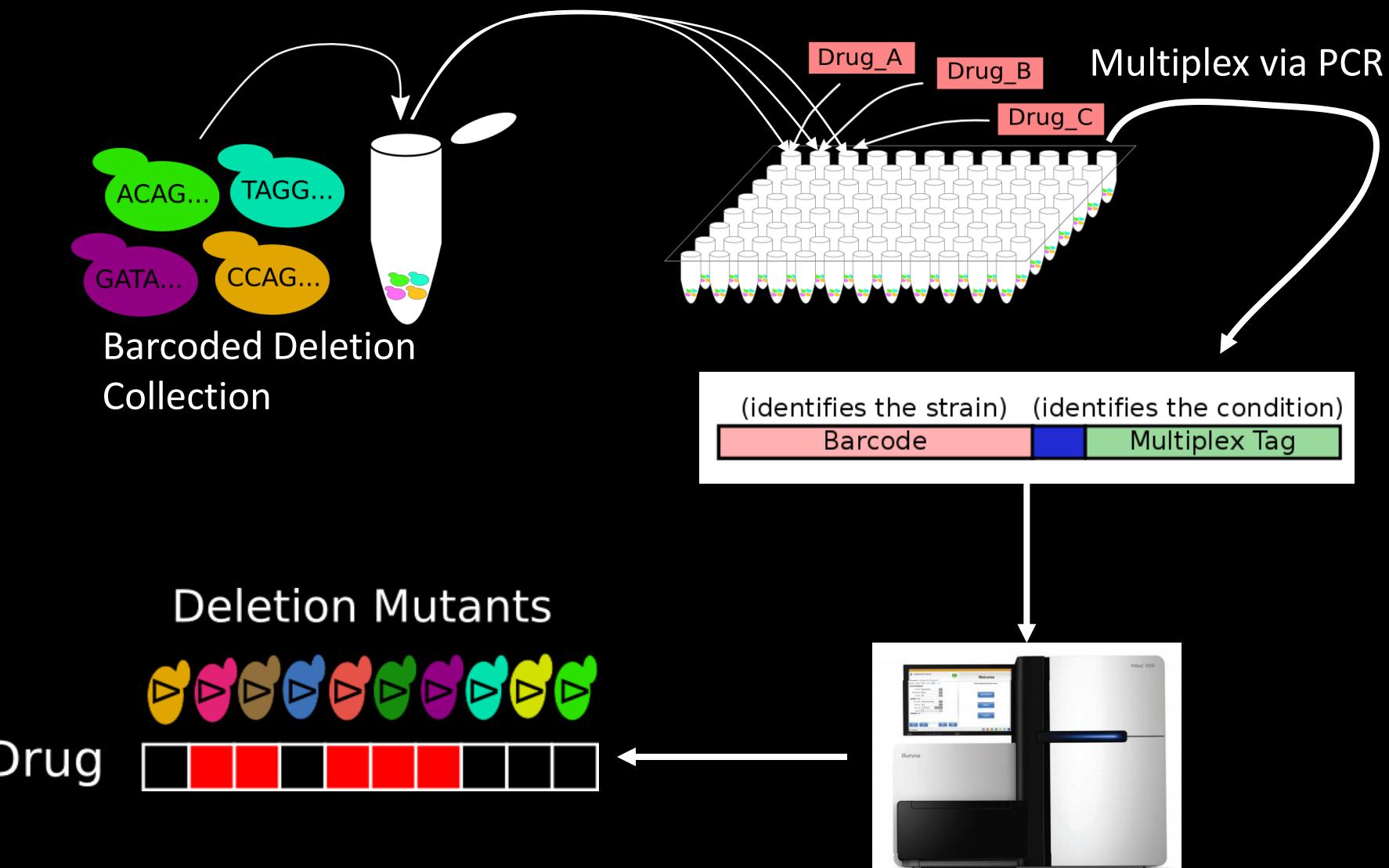


Gaiever/Nislow:
Hillenmeyer et al. 2008
Lee et al. 2014

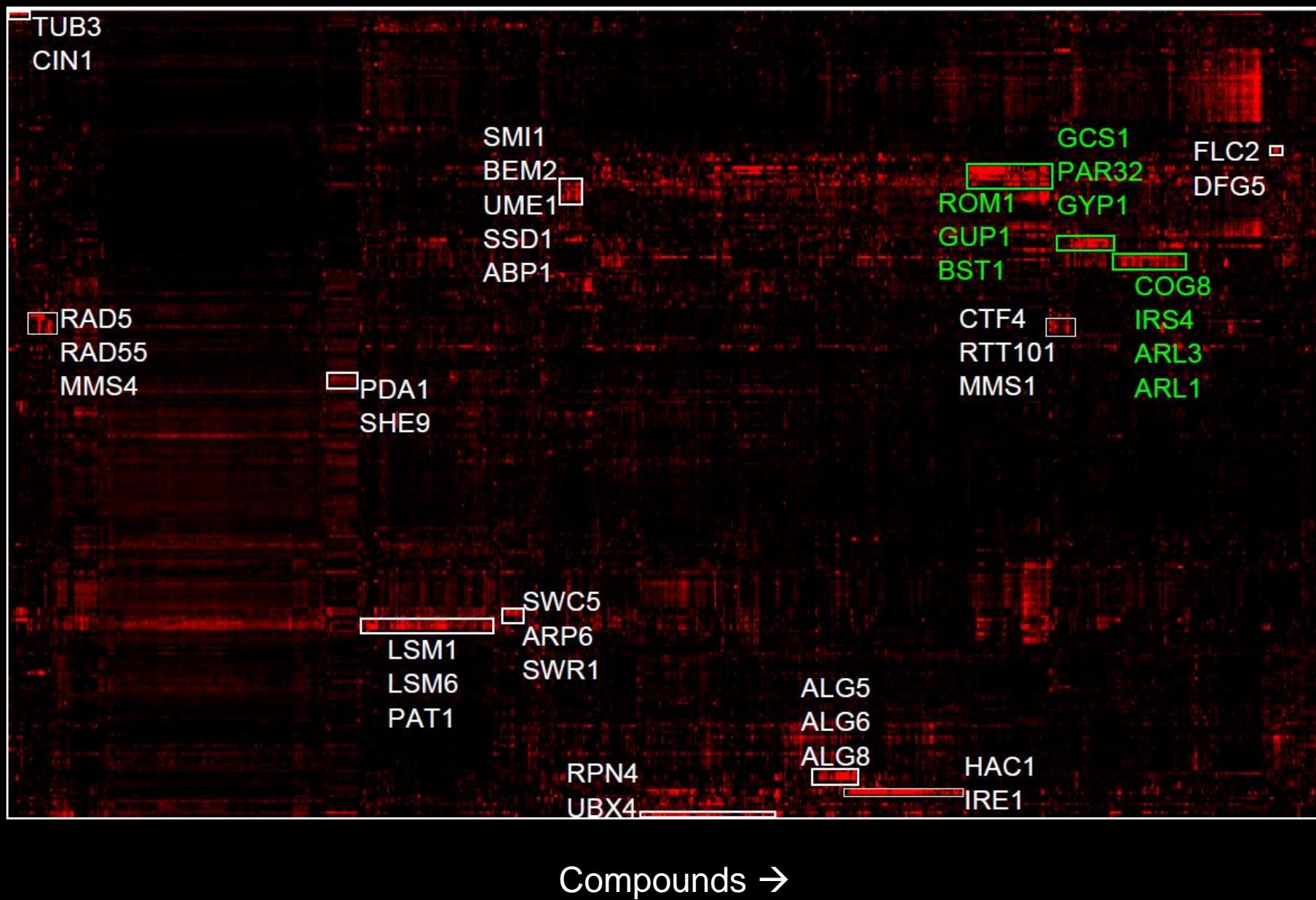
Combining gene-gene interactions and chemical-gene interactions



Barcode sequencing to generate chemical-genetic interaction profiles

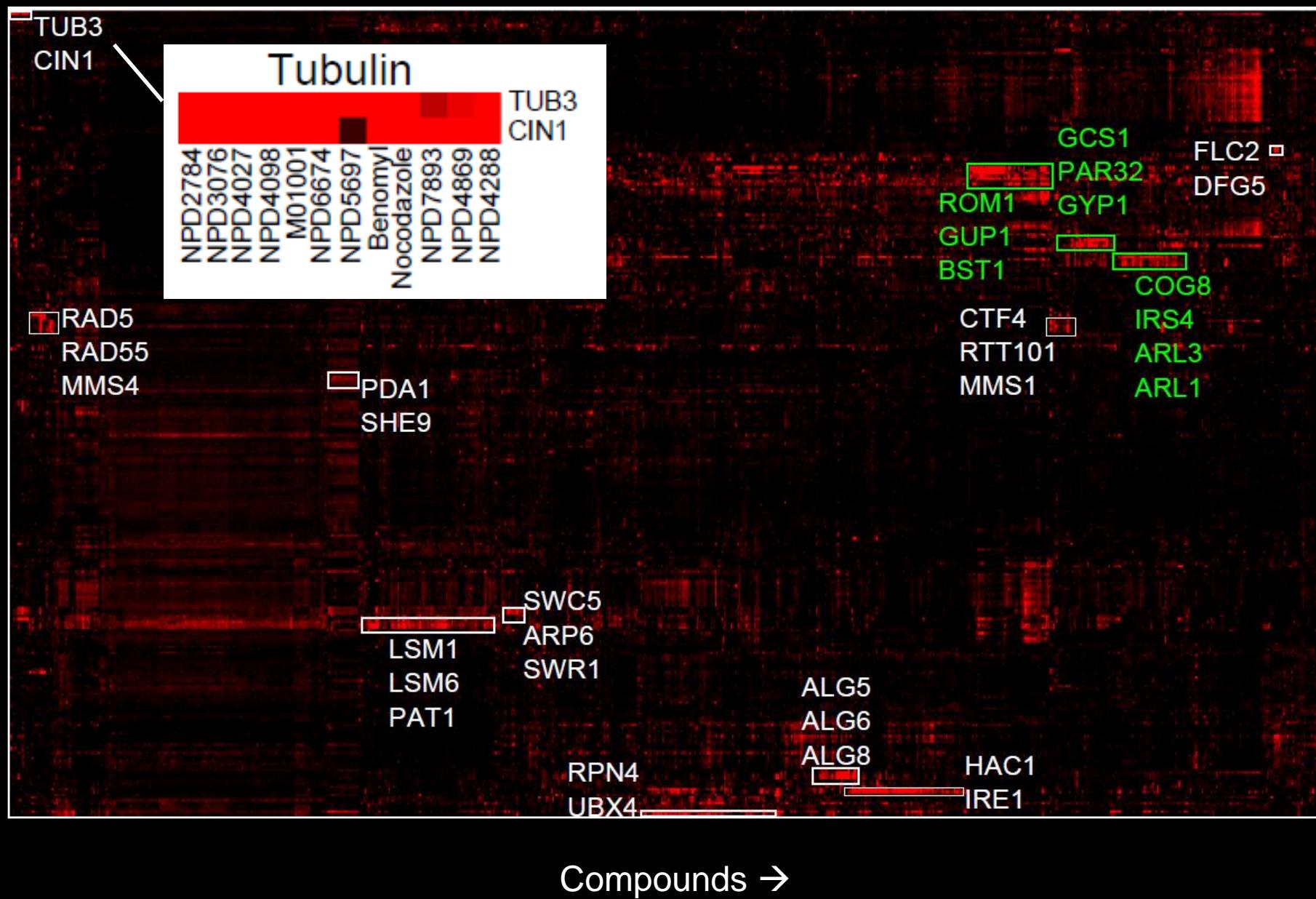


Chemical genetic interaction profiles



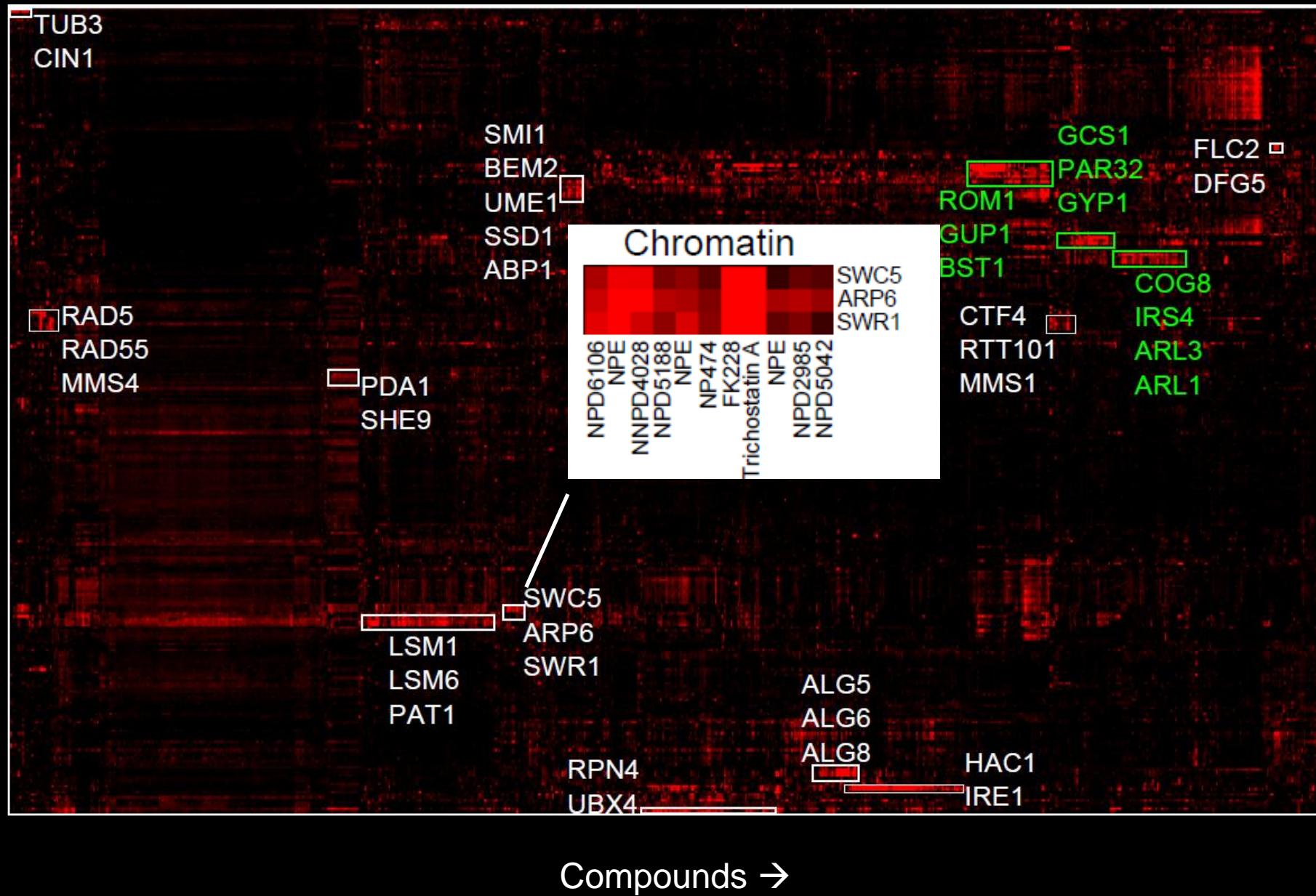
Chemical genetic interaction profiles

Genes →



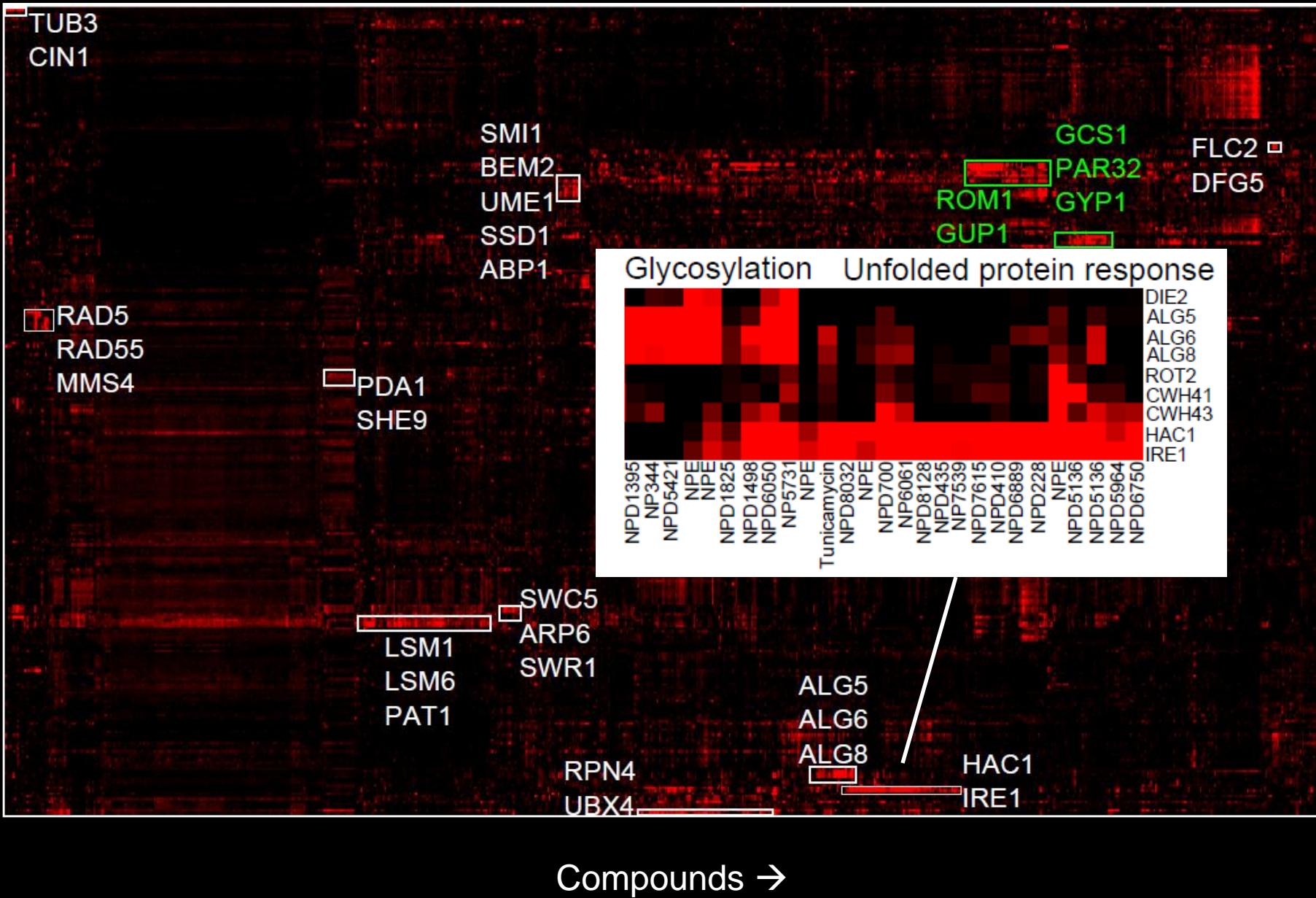
Chemical genetic interaction profiles

Genes →

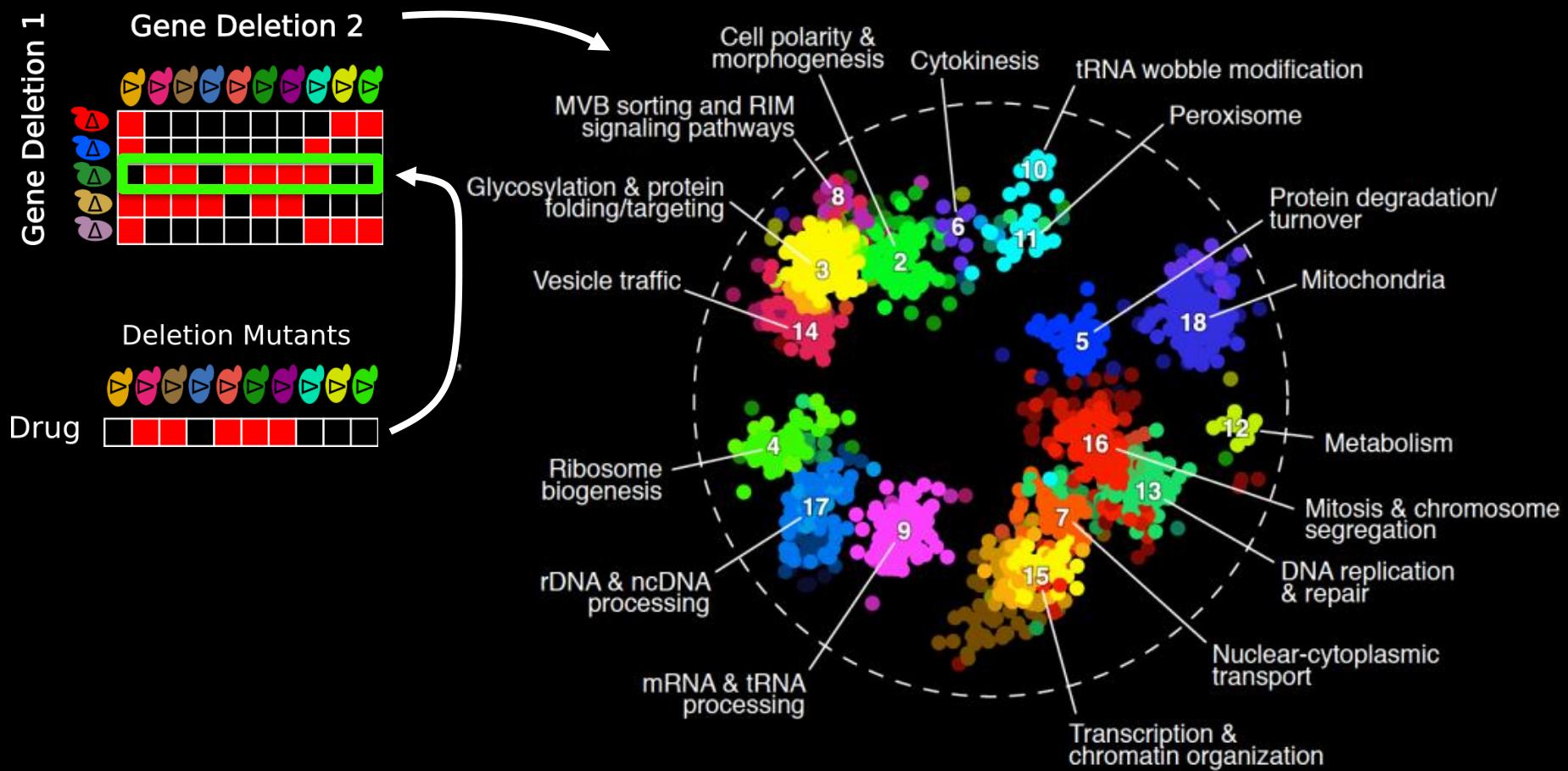


Chemical genetic interaction profiles

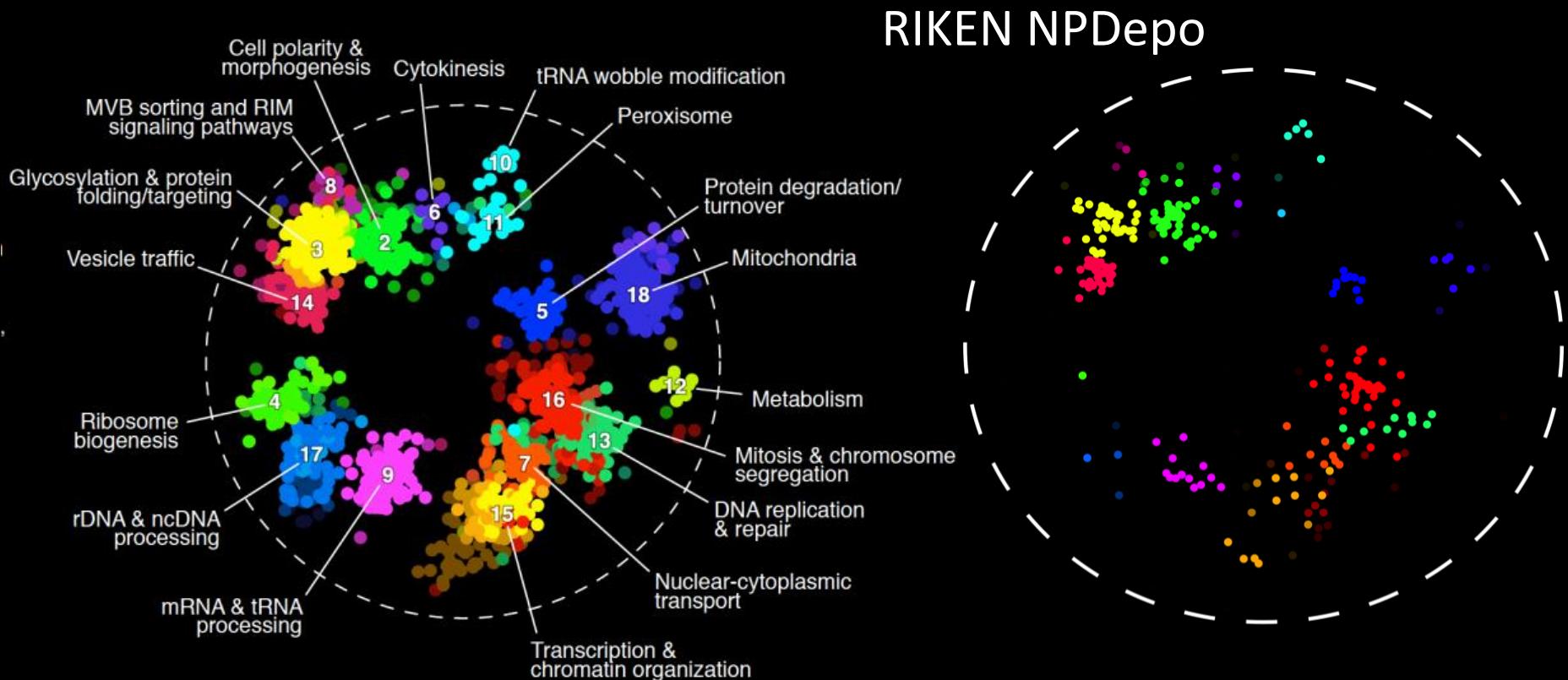
Genes →



Putting compounds on the functional map

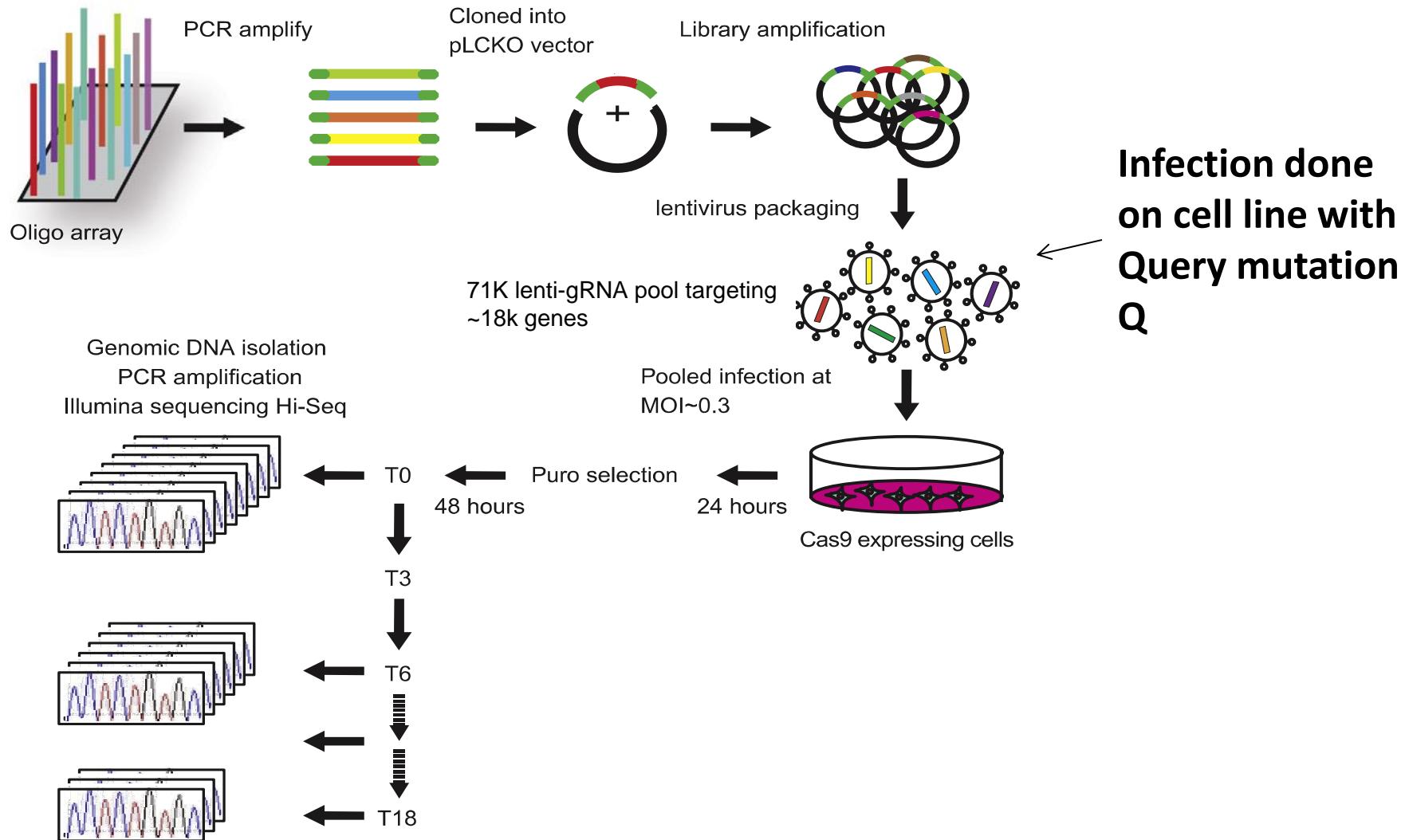


New molecular probes span a diversity of biological processes



Translating and applying genetic
perturbation approaches to human cells

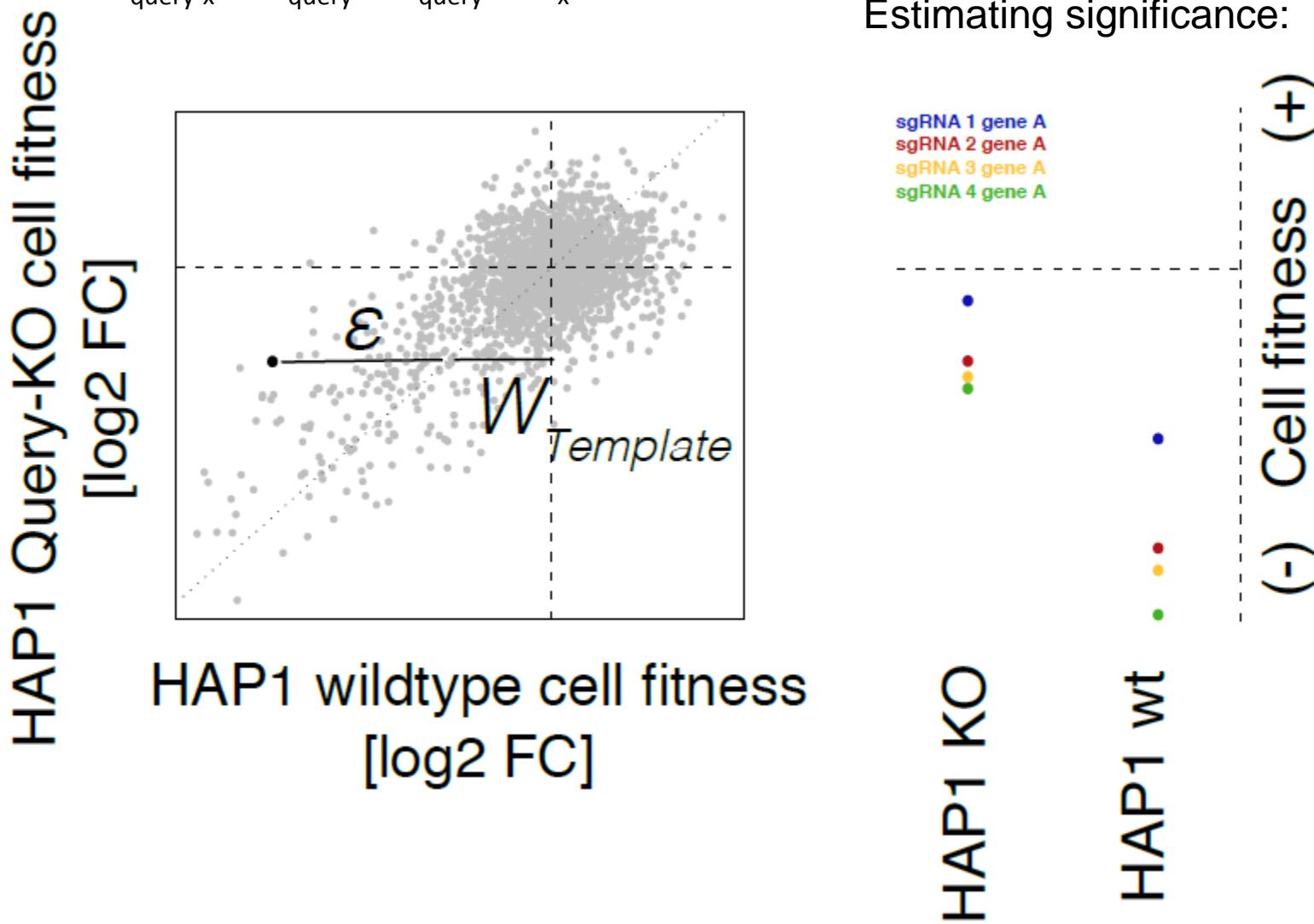
Genome-wide CRISPR screens in human cells



Scoring quantitative genetic interactions in human cells

$$\varepsilon_{\text{query}-x} = W_{\text{query}} - W_{\text{query}} * W_x$$

Estimating significance:



Computer science: the hub of “inter-disciplinary” science

