Reconstruction of the Global Human Metabolic Network Based on Build-35 and Bibliomic Data

Bernhard Palsson

Dept of Bioengineering

UCSD

Outline

- Systems biology: the process
- Network reconstruction:
  - Bottom-up vs. top down
- BiGG data bases
- Determining functional states
- Lessons learned from small genomes
- Bottom-up reconstruction of the human metabolic map
- Three initial uses of the human map

Systems Biology: The Process

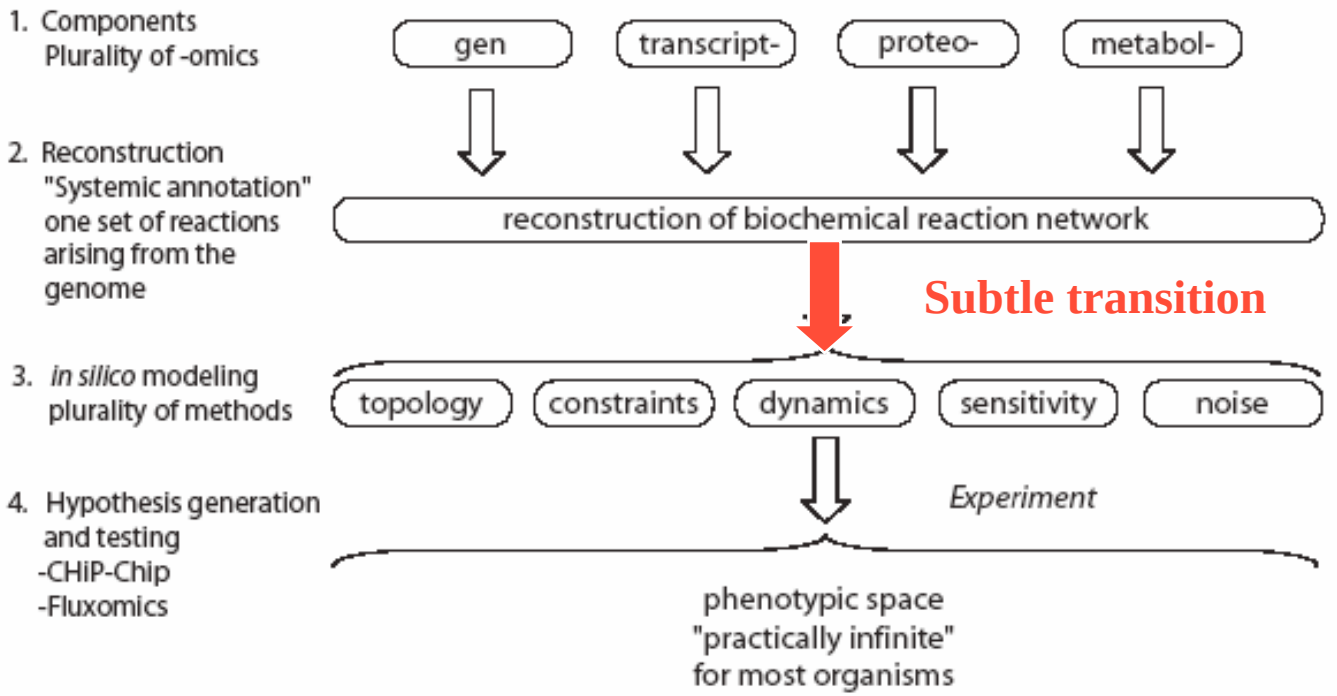
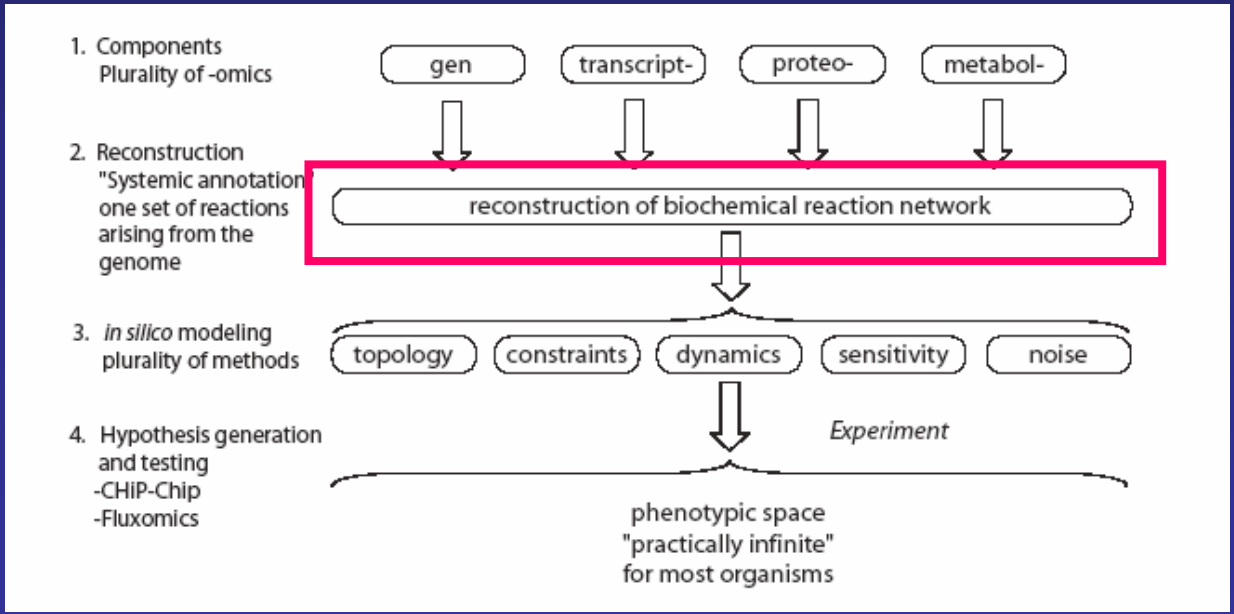
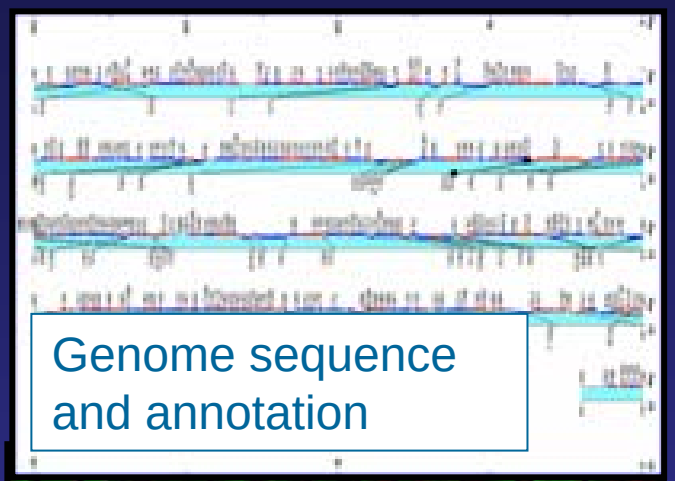


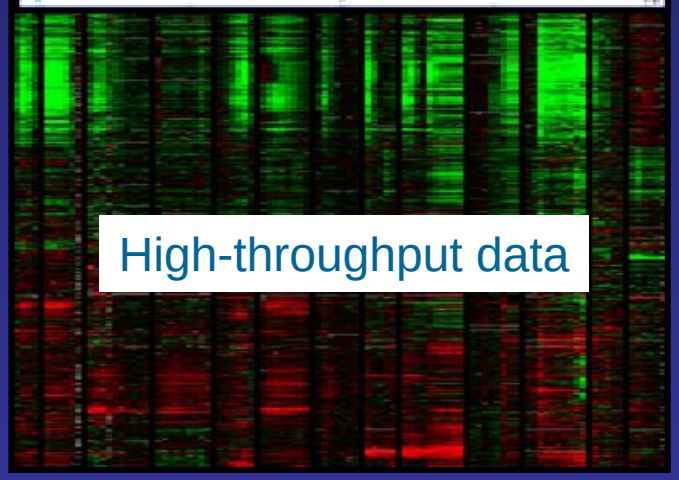
Figure 1.4: The four principal steps in the implementation of systems biology. Note that the second step is unique, while the others are diverse, and it is the interface between high-throughput data and *in silico* analysis.

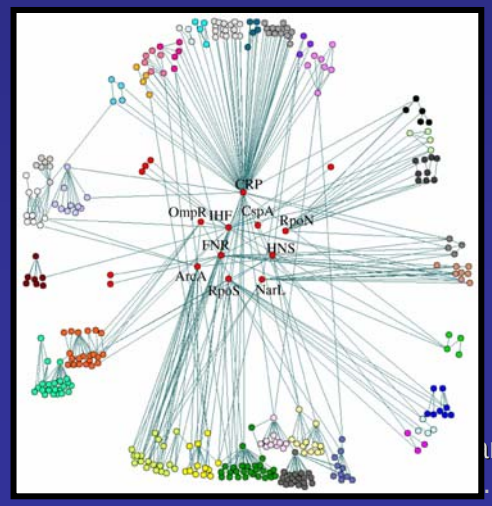
Network Reconstruction:
Bottom-up vs Top-down



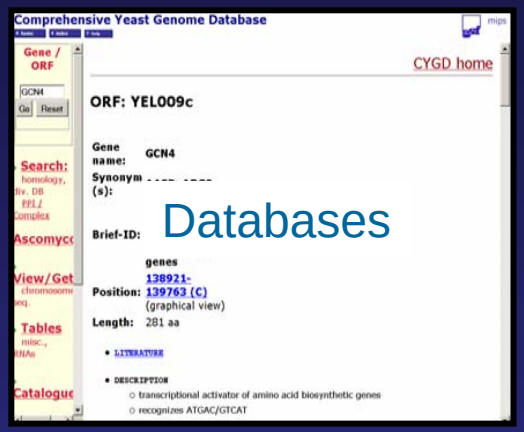
Top-Down Reconstruction

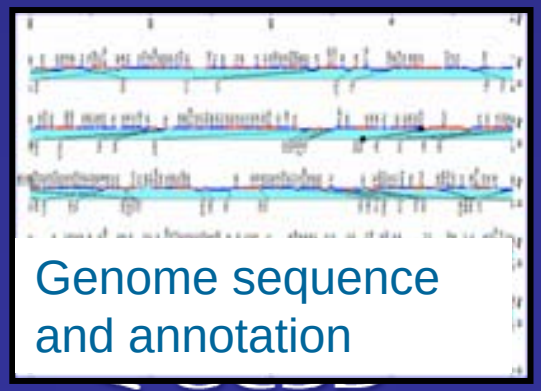


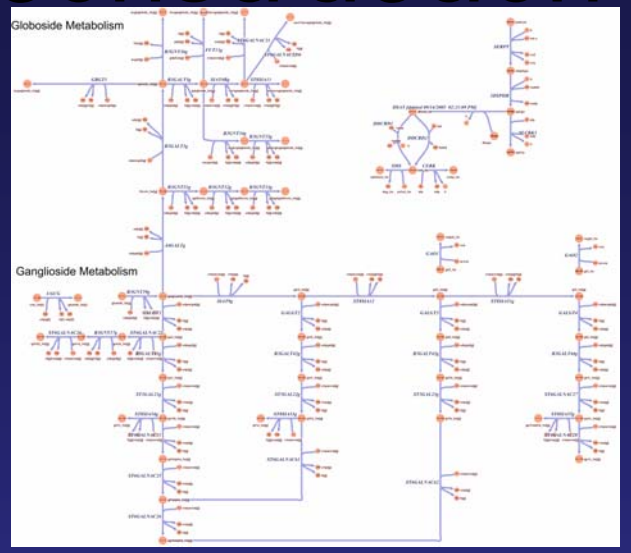


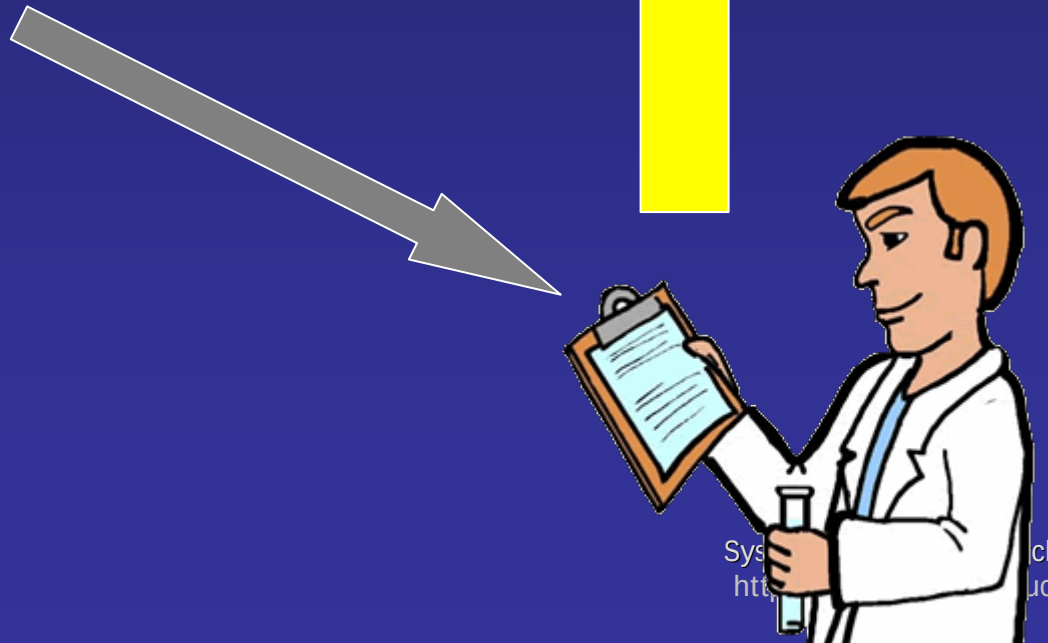


Bottom-Up Reconstruction









Top Down vs Bottom Up

TOP DOWN

- Aims to be comprehensive
- Results are often inconsistent
- Final conclusions may be 'soft'
- "discovery"

BOTTOM UP

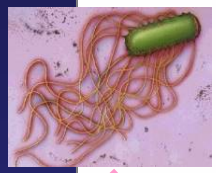
- Aims to be accurate, well-defined interactions
- Results are self consistent
- Conclusions result from underlying chemical basis
- "knowledge"(BiGG)

UCSD

Godel's theorem

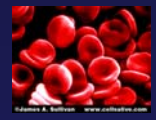
Genome-scale Metabolic Reconstructions – SBRG

S. typhimurium
- 898 Reactions
- 826 Genes



S. aureus
- 640 Reactions
- 619 Genes

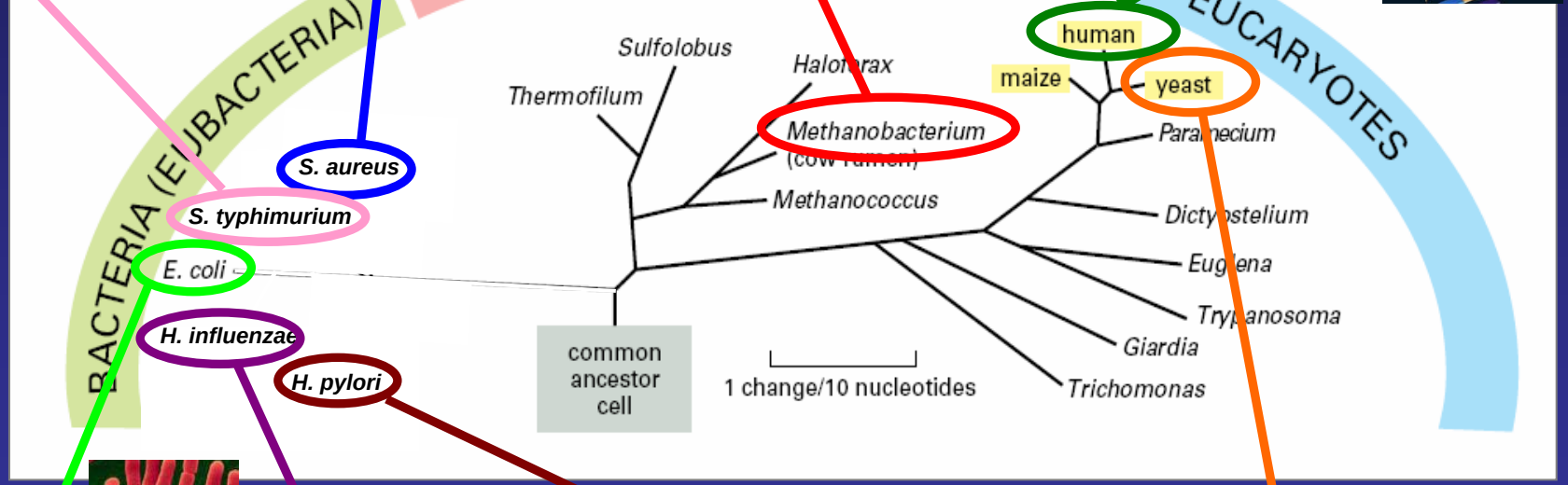
M. barkeri
- 619 Reactions
- 692 Genes



RBC
- 39 Rxns

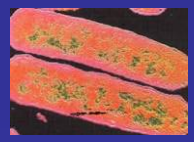
Mitoc.
- 218 Rxns

H. sapiens
- 3311 Reactions
- 1496 Genes

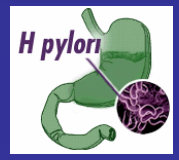


E. coli
- 2035 Reactions
- 1243 Genes

H. influenzae
- 472 Reactions
- 376 Genes



H. pylori
- 558 Reactions
- 341 Genes



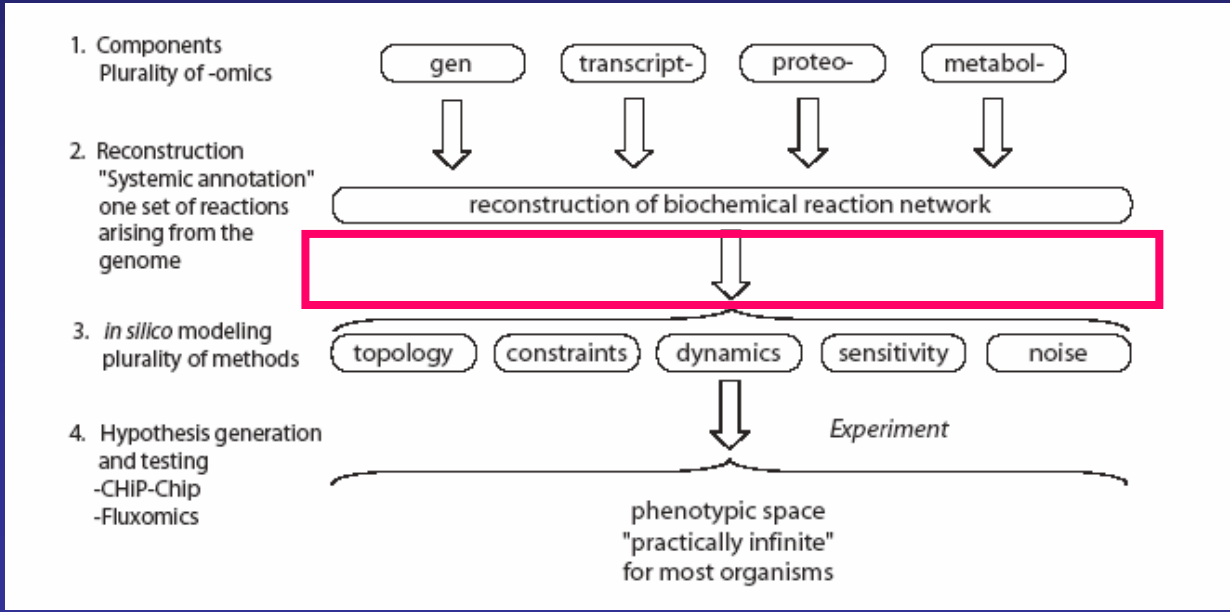
S. cerevisiae
- 1402 Reactions
- 910 Genes

UCSD

Systems Biology Research Group
http://systemsbiology.ucsd.edu

BiGG data bases:

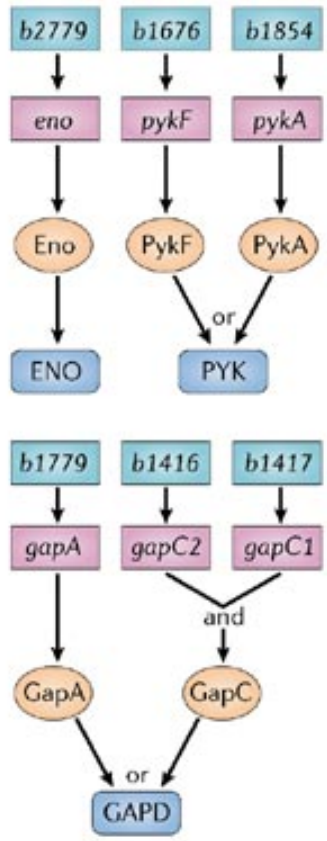
Mathematical Representation of a bottom-up reconstruction

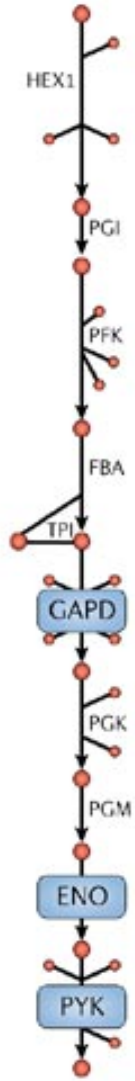


BiGG data bases and constraint based analysis

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	*glk*
PGI	[c]G6P ↔ F6P	*pgi*
PFK	[c]ATP + F6P → ADP + FDP + H	*pfkA, pfkB*
FBA	[c]FDP ↔ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ↔ 3PG + ATP	*pgk*
PGM	[c]3PG ↔ 2PG	*gpmA, gpmB*
ENO	[c]2PG ↔ H_2O + PEP	*eno*
PYK	[c]ADP + H + PEP → ATP + PYR	*pykA, pykF*

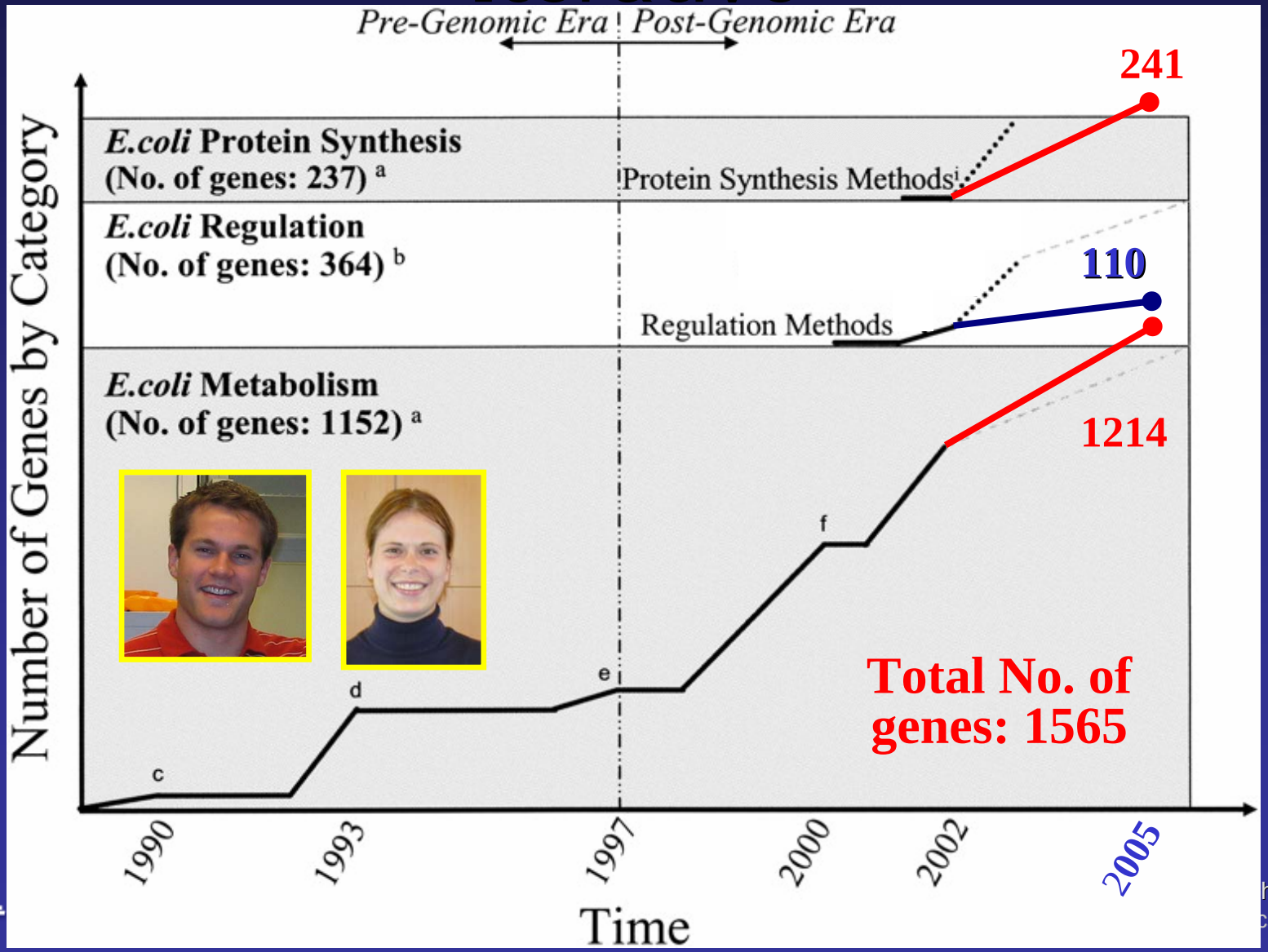
	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
ATP	−1	0	−1	0	0	0	1	0	0	1
GLC	−1	0	0	0	0	0	0	0	0	0
ADP	1	0	1	0	0	0	−1	0	0	−1
G6P	1	−1	0	0	0	0	0	0	0	0
H	1	0	1	0	0	1	0	0	0	−1
F6P	0	1	−1	0	0	0	0	0	0	0
FDP	0	0	1	−1	0	0	0	0	0	0
DHAP	0	0	0	1	−1	0	0	0	0	0
G3P	0	0	0	1	1	−1	0	0	0	0
NAD	0	0	0	0	0	−1	0	0	0	0
PI	0	0	0	0	0	−1	0	0	0	0
13DPG	0	0	0	0	0	1	−1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0
3PG	0	0	0	0	0	0	1	−1	0	0
2PG	0	0	0	0	0	0	0	1	−1	0
PEP	0	0	0	0	0	0	0	0	1	−1
H_2O	0	0	0	0	0	0	0	0	1	0
PYR	0	0	0	0	0	0	0	0	0	1





Human: 2712 x 3311

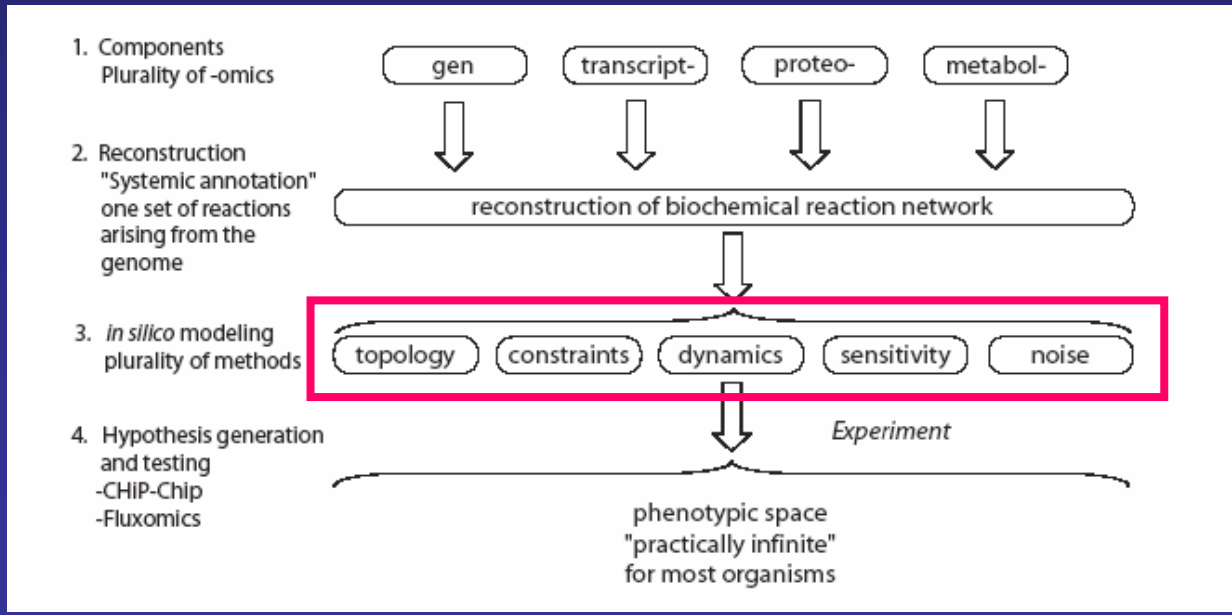
The reconstruction Process is Iterative



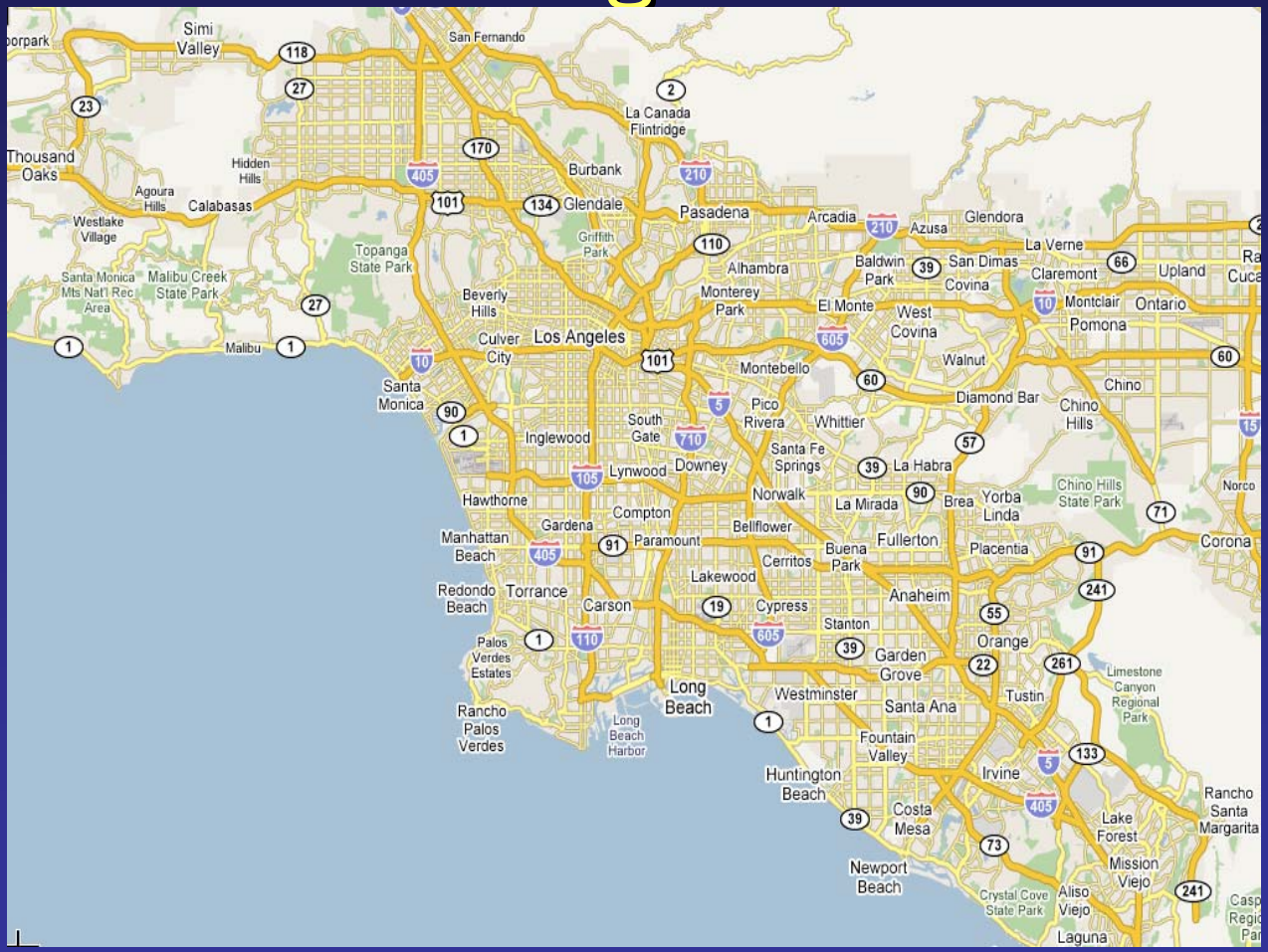
Growing scope of bottom up reconstructions:

- Signaling Networks
  - Jak-Stat Signaling. Papin et al Biophys J. **87**: 37-46(2004
  - Ca 250 reactions
- Transcriptional Regulatory networks
  - Logistic statements (not chemical rules), Gianchandani et al *PLoS Comput Biol*, 2006 Aug 11;2(8):e101
- Translation/transcription
  - The 'dogma' matrix, O(10,000 x 10,000), Thiele, in preparation)

Determining Network Functional States:
(BiGG query tools)



Determining Functional States: In silico modeling methods

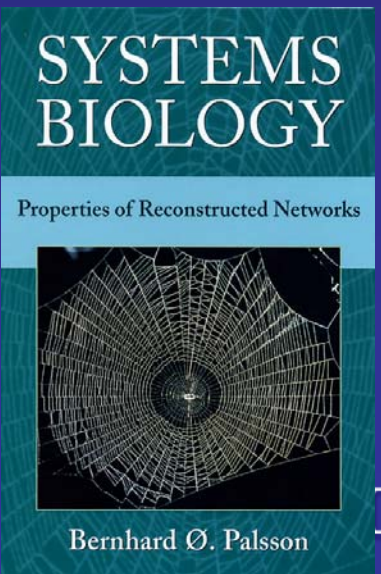


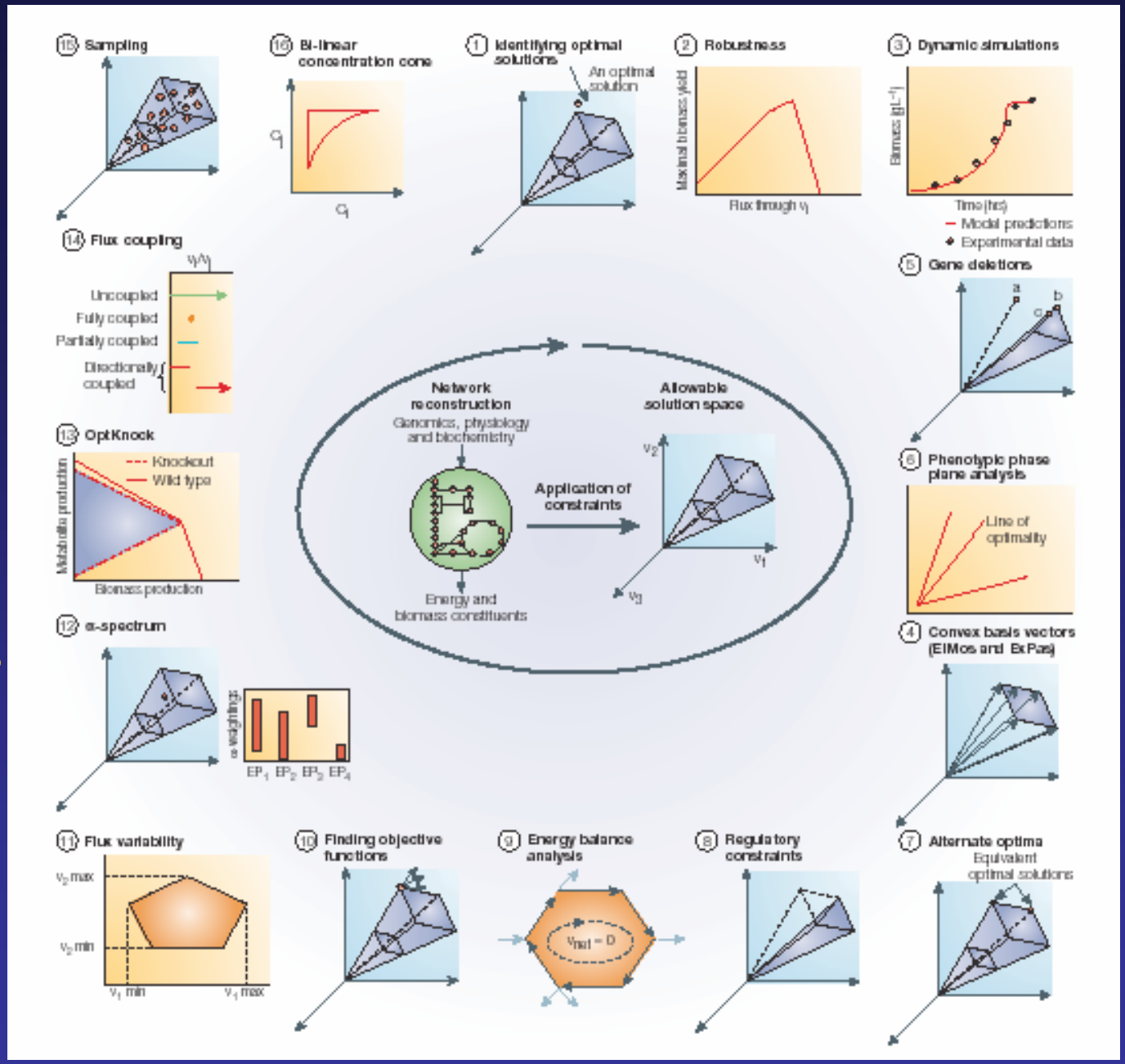
UCSD

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Genome-scale constraint-based modeling:

a rapidly growing





UCSD Extension

Price, et al Nat Rev Microbiol. Nov 2004

Using genome-scale networks to guide discovery:

Lessons from microbes (since 2000)

Ask not what you can do for a reconstruction but what a reconstruction can do for you

- Substrate preferences
- The consequences of gene KOs
- Synthetic lethals
- Optimal growth rates
- Outcome of adaptive evolution
- Horizontal gene transfer
- Evolution to minimal genomes
- Metabolic engineering
- Gap filling/discovery of gene functions

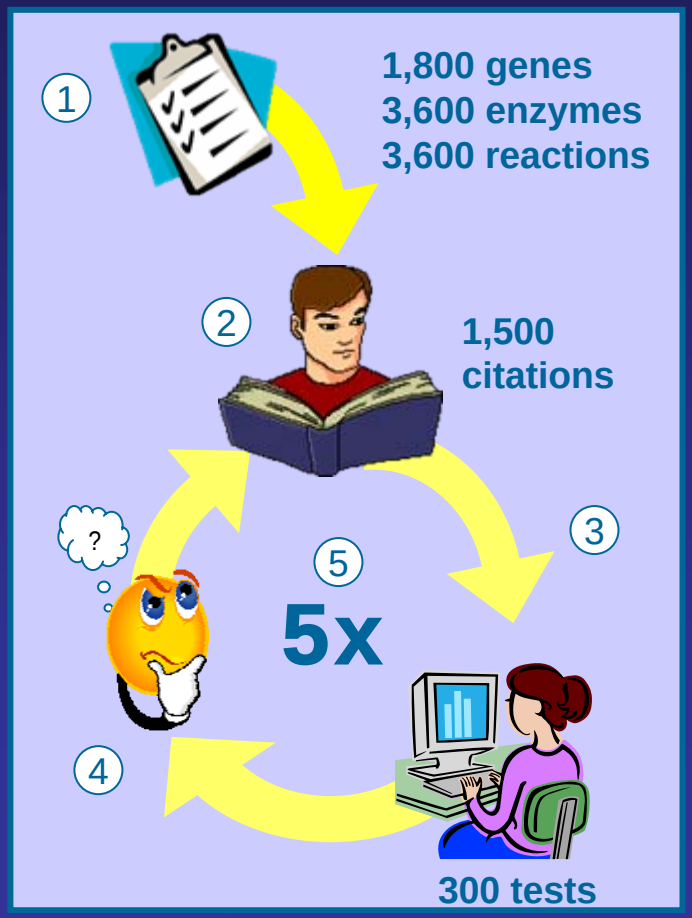
Reconstruction of the human metabolic map: genome-scale

UCSD

Systems Biology Research Group
http://systemsbiology.ucsd.edu

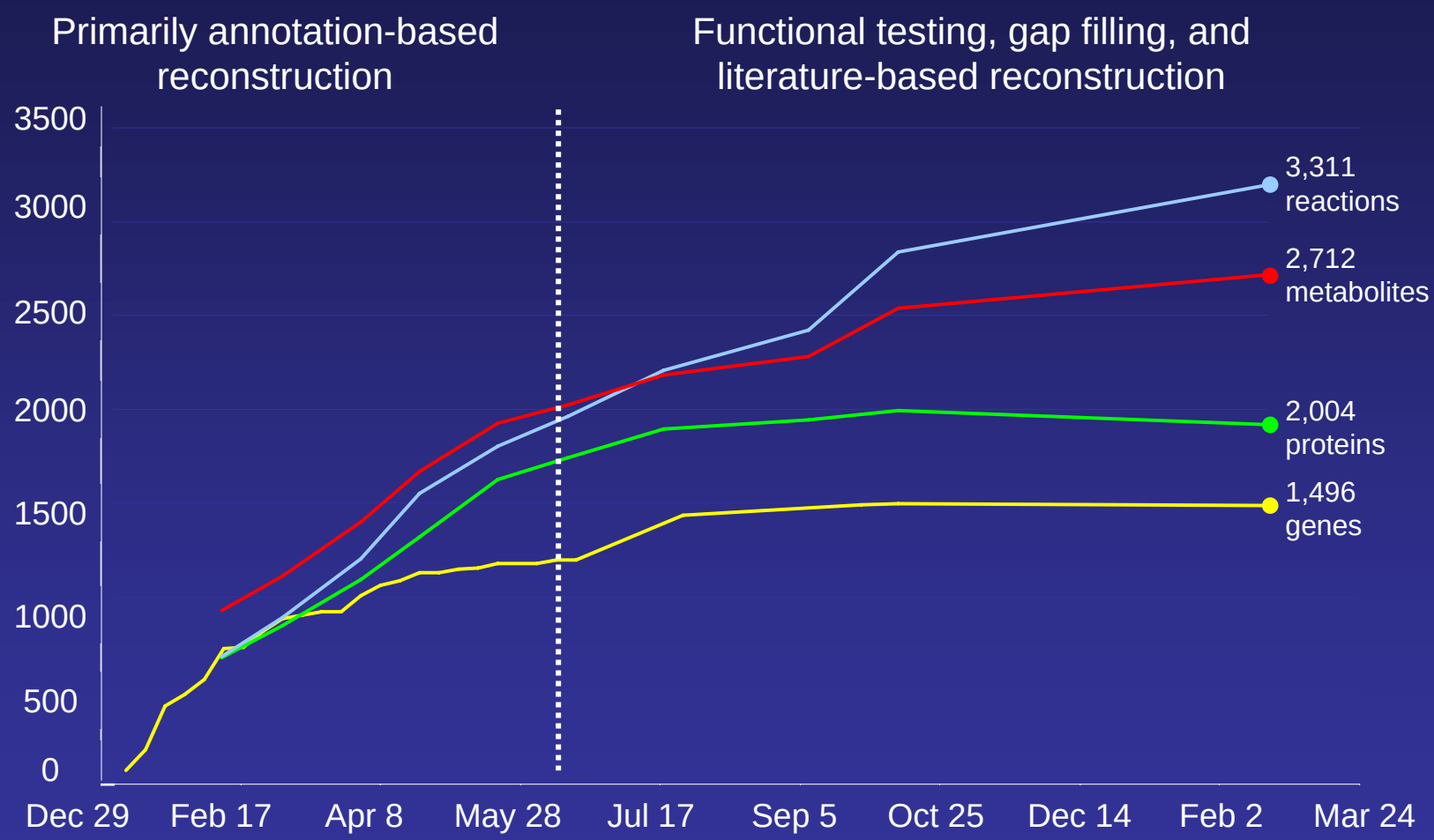
How do we integrate these data?

- The genome annotation is the starting point for manual, bottom-up reconstruction



1. Make an **initial parts list** from the genome annotation
2. Collect and record **biological evidence** from the literature
3. Fix contents and **test performance**
4. Identify **inconsistencies & gaps**
5. **Repeat** steps 2-4

Human Reconstruction: Process



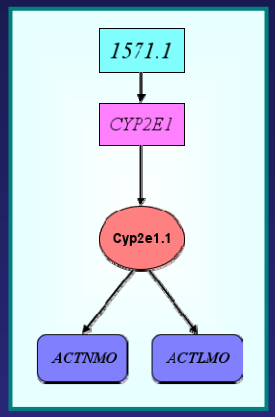
The global human metabolic map

Genes (1,496)

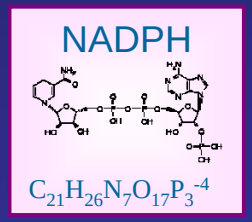
Transcripts (1,905)

Proteins (2,004)

Reactions (3,311)

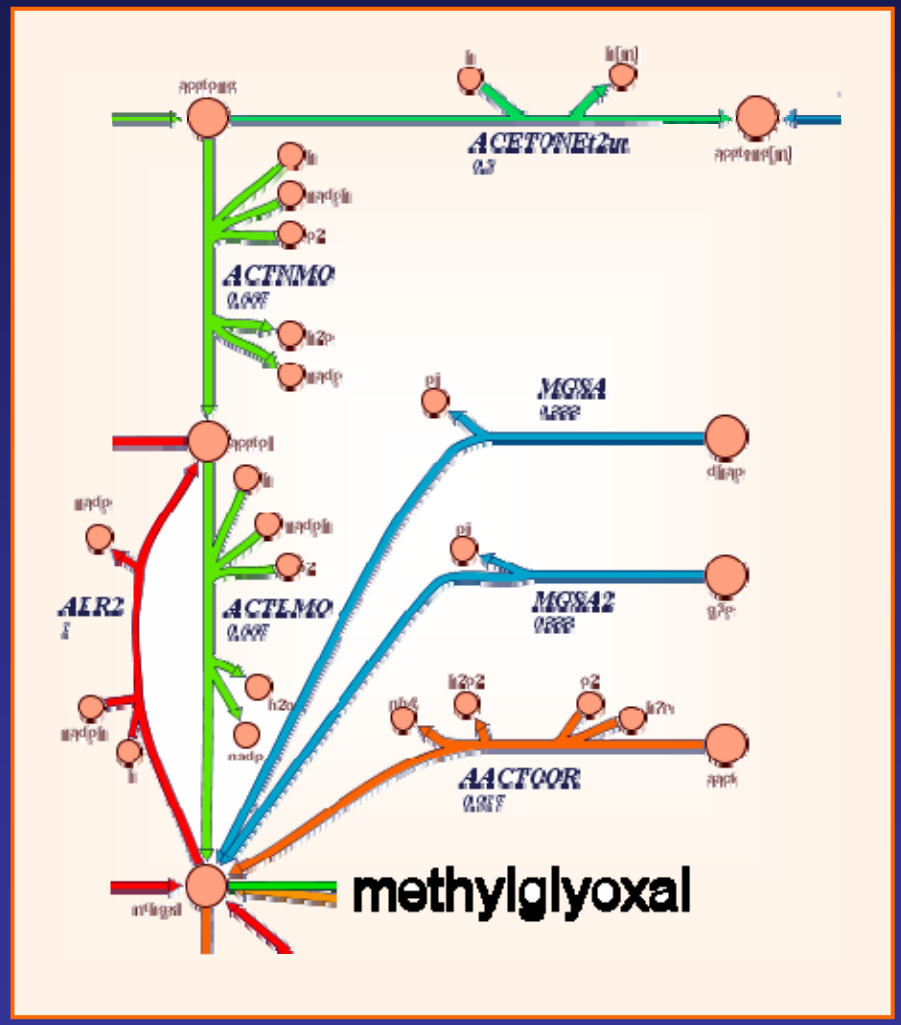


Compounds (2,712)



Compartments (7)

Metabolic Pathways (98)



UCSD

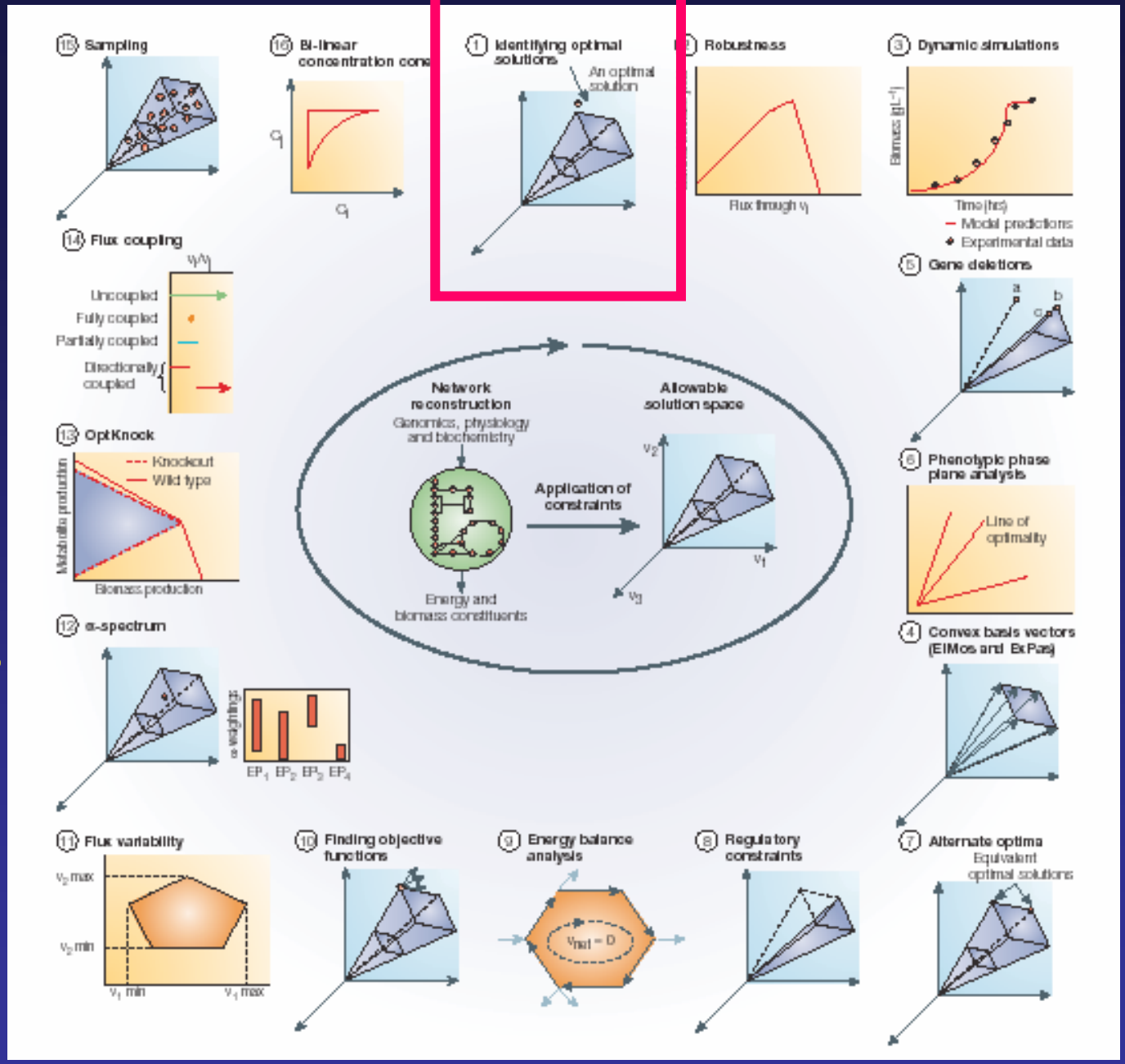
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Use #1: defining our knowledge gaps:

validation against 288 physiological functions,

Genome-scale constraint-based modeling:

a rapidly growing field

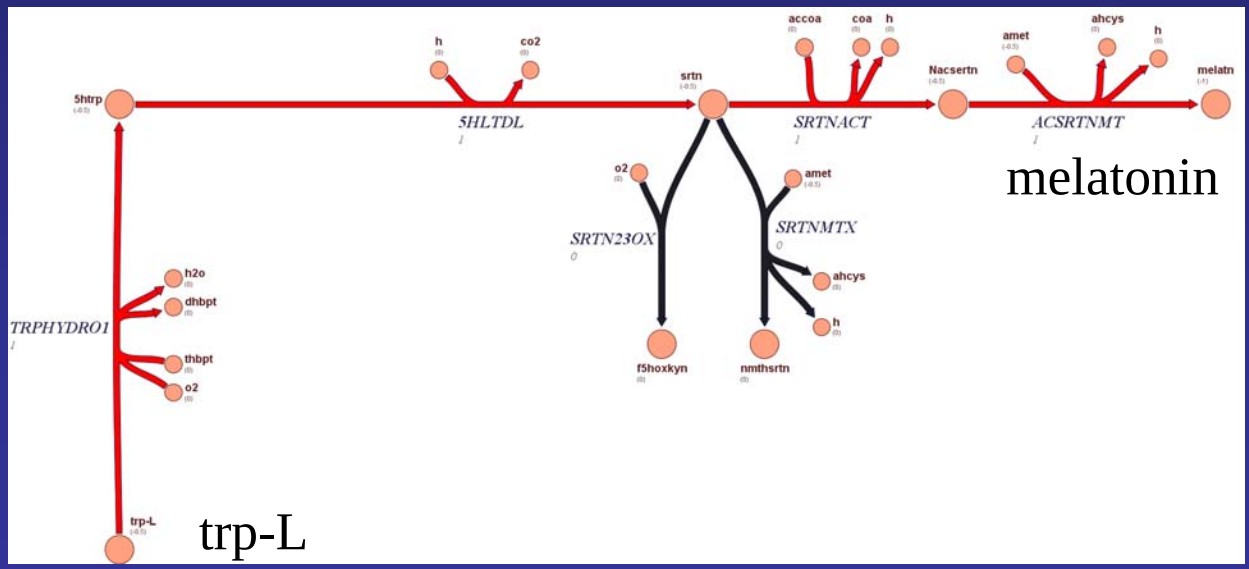


Price, et al Nat Rev Microbiol. Nov 2004

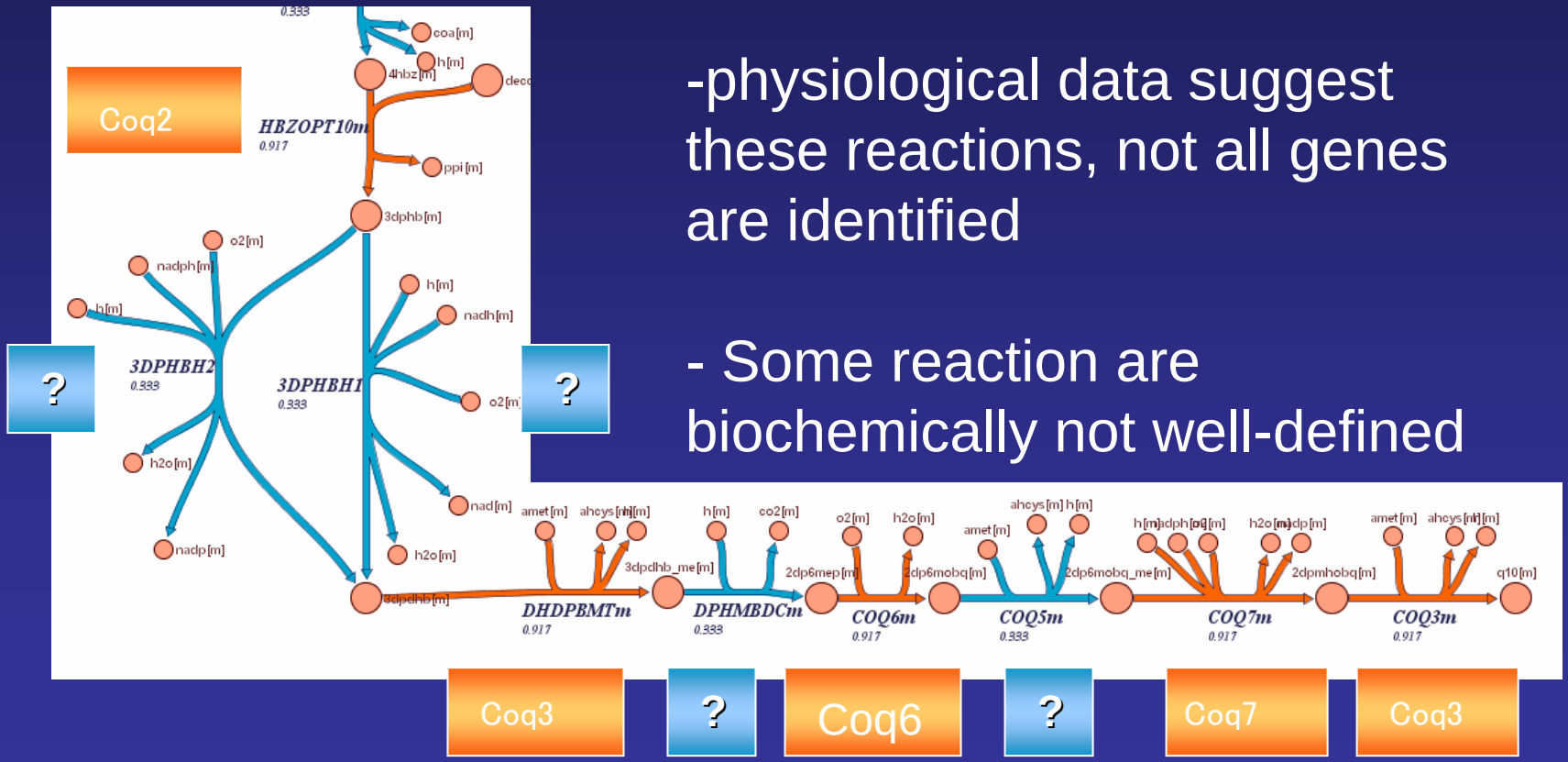
Network Validation

- Over 288 distinct metabolic functions tested with FBA
  - Many are simple:

Desired Functionality
Degrade Trp to accoa (and many intermediates)
Synthesize serotonin
Synthesize melatonin
Synthesize L-Formylkynurenine
Synthesize L-Kynurenine
Synthesize N-Formylanthranilate



Knowledge gaps
Ubiquinone 10 Biosynthesis



-physiological data suggest these reactions, not all genes are identified

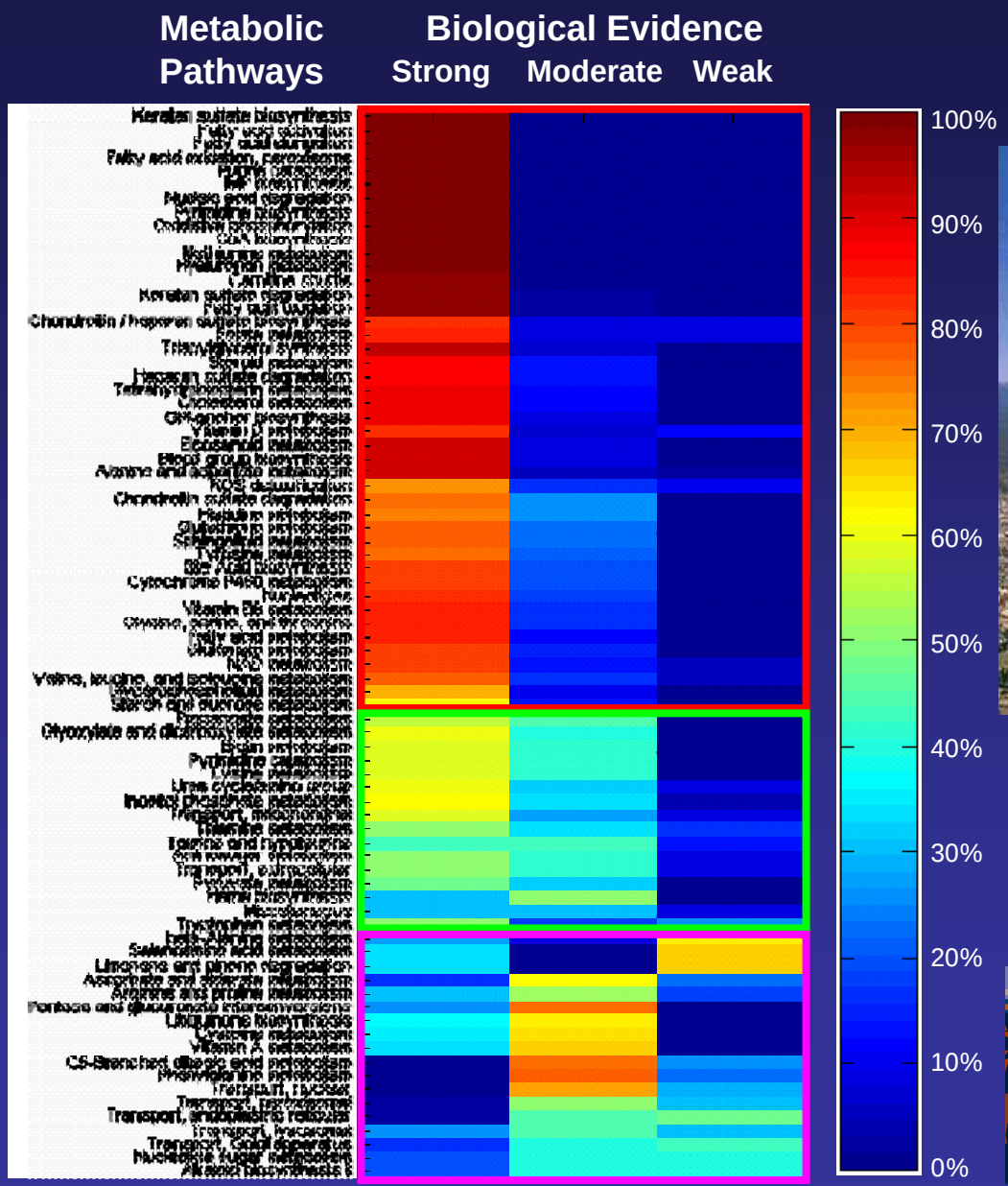
- Some reaction are biochemically not well-defined

Score: 0.30 0.90

Model-Driven Discovery

- Optimal stoichiometric identification
  - Herrgard, M.J., et al , *PLoS Computational Biology,* in press (2006)
- Phenotyping of KOs ('Smiley')
  - Reed, JR, et al *PNAS,* in review
- Double perturbation experiments
  - Covert M, et al *Nature*, **429**: 92-96 (2004)
  - Barrett, C.L., et al *BMC Bioinformatics* **7**:132 (2006)
  - Cho, B.K., et al, *Microbiology*, accepted (2006)

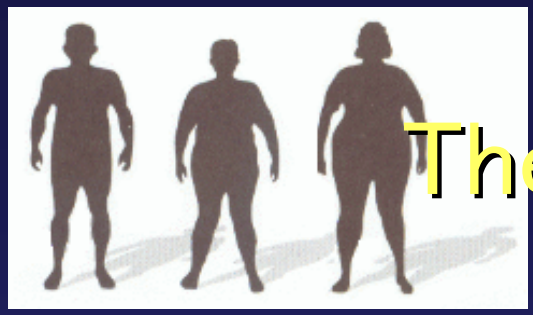
Metabolic knowledge landscapes



- Strong evidence
  - Genetic
  - Biochemical
- Moderate evidence
  - Physiological
  - Homological
- Weak evidence
  - Inferred from modeling

Use #2:
context for content

a basis for multi-omic integration & analysis



The Obesity Epidemic

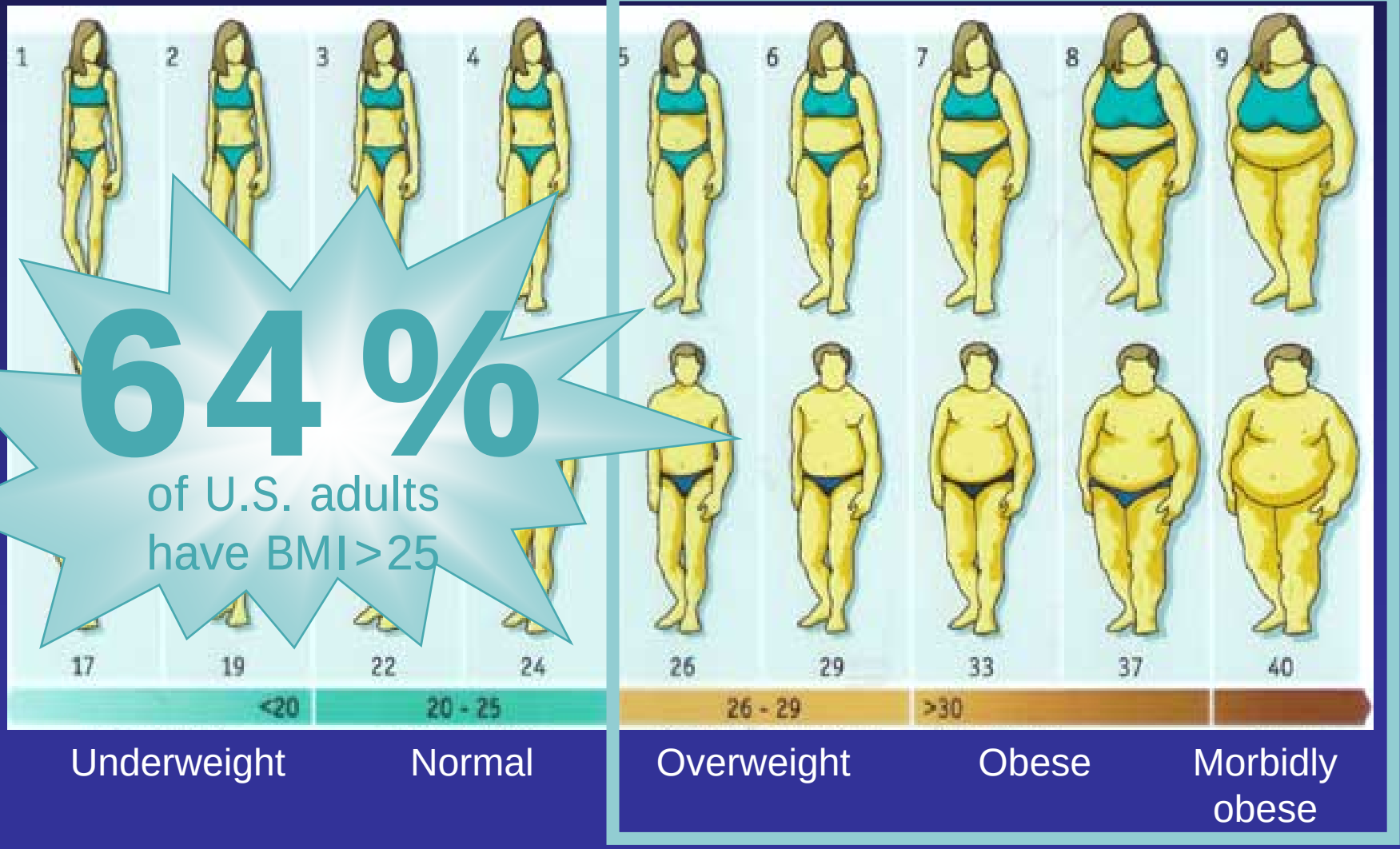
- Biggest public health problem of the century
- **>1.5 billion adults** worldwide and **10% of children** are overweight or obese
- Significantly increases risk of:
  - Diabetes mellitus
  - Hypertension
  - Dyslipidemia
  - Osteoarthritis

International Obesity Conference (2006)

1999-2000 National Health and Nutrition Examination Survey

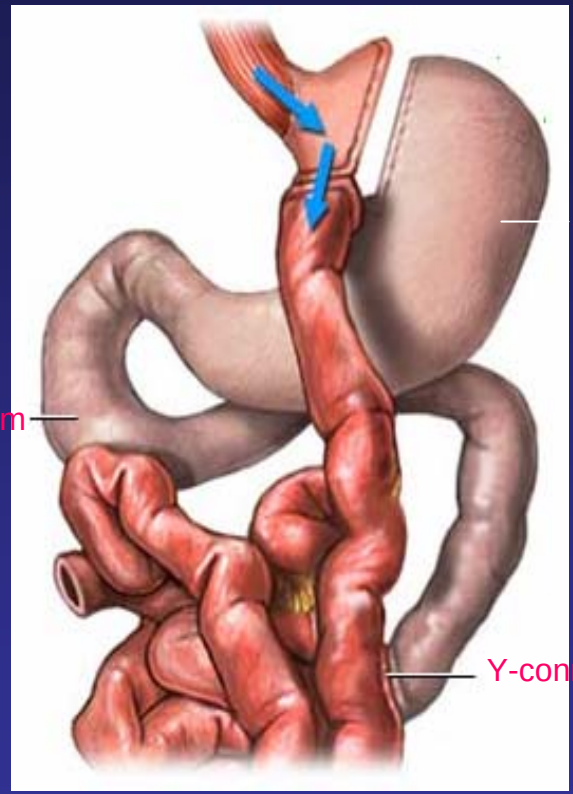
Body Mass Index (BMI)

$$\mathrm{BMI} = 703 \frac{weight\ (\mathrm{lb})}{height \times height\ (\mathrm{in} \times \mathrm{in})}$$



Gastric bypass surgery

- Established treatment for severely obese
  - 103,000 U.S. procedures in 2003
- Short-term loss of 40-80% excess body weight
- Significant improvement in comorbidities



Roux-en-Y
gastric bypass

Genome-scale data analysis

The data set:

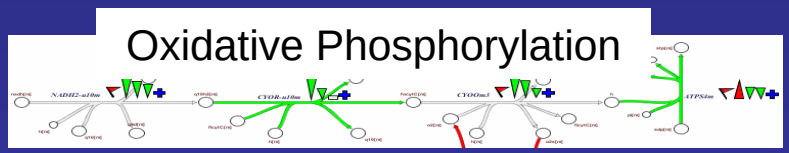
- 3 morbidly obese patients (BMI>40) underwent gastric bypass surgery
- Skeletal muscle was expression profiled pre-surgery and 1 year afterwards

359 lbs
BMI 61.6
Morbidly obese

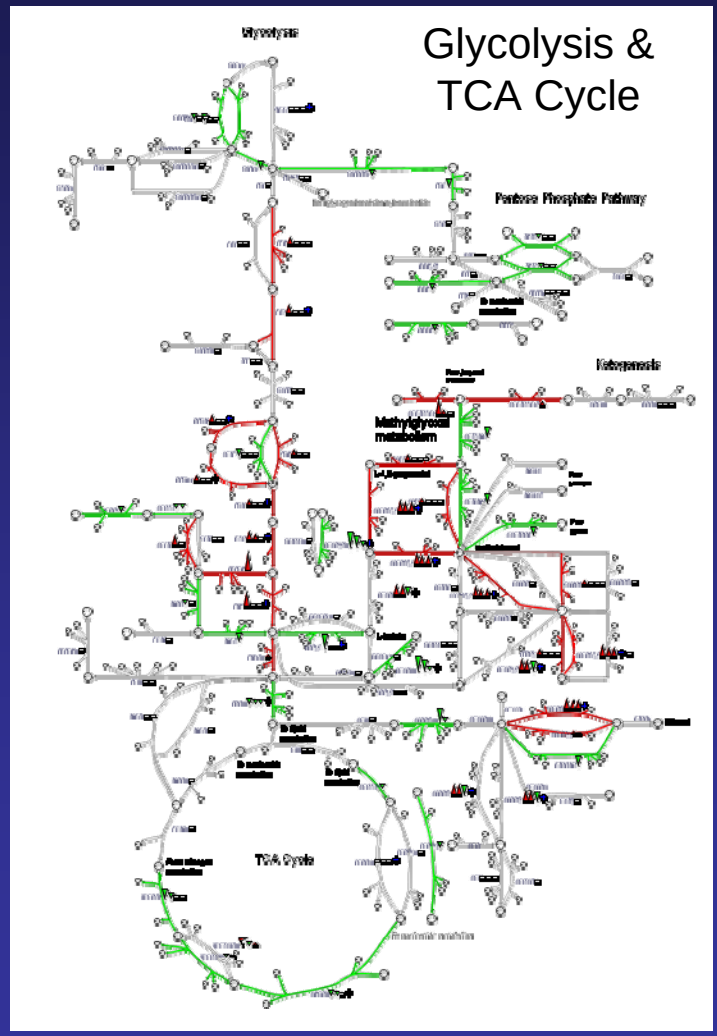
181 lbs
BMI 31.1
Obese

Recon 1 as a context for content

- Recon 1 was used to visualize gene expression changes pre- and post-surgery
- Oxidative metabolism decreased after surgery and glycolytic metabolism increased



Results are consistent with **nutrient-deprivation**



UCSD

Systems Biology Research Group
http://systemsbiology.ucsd.edu

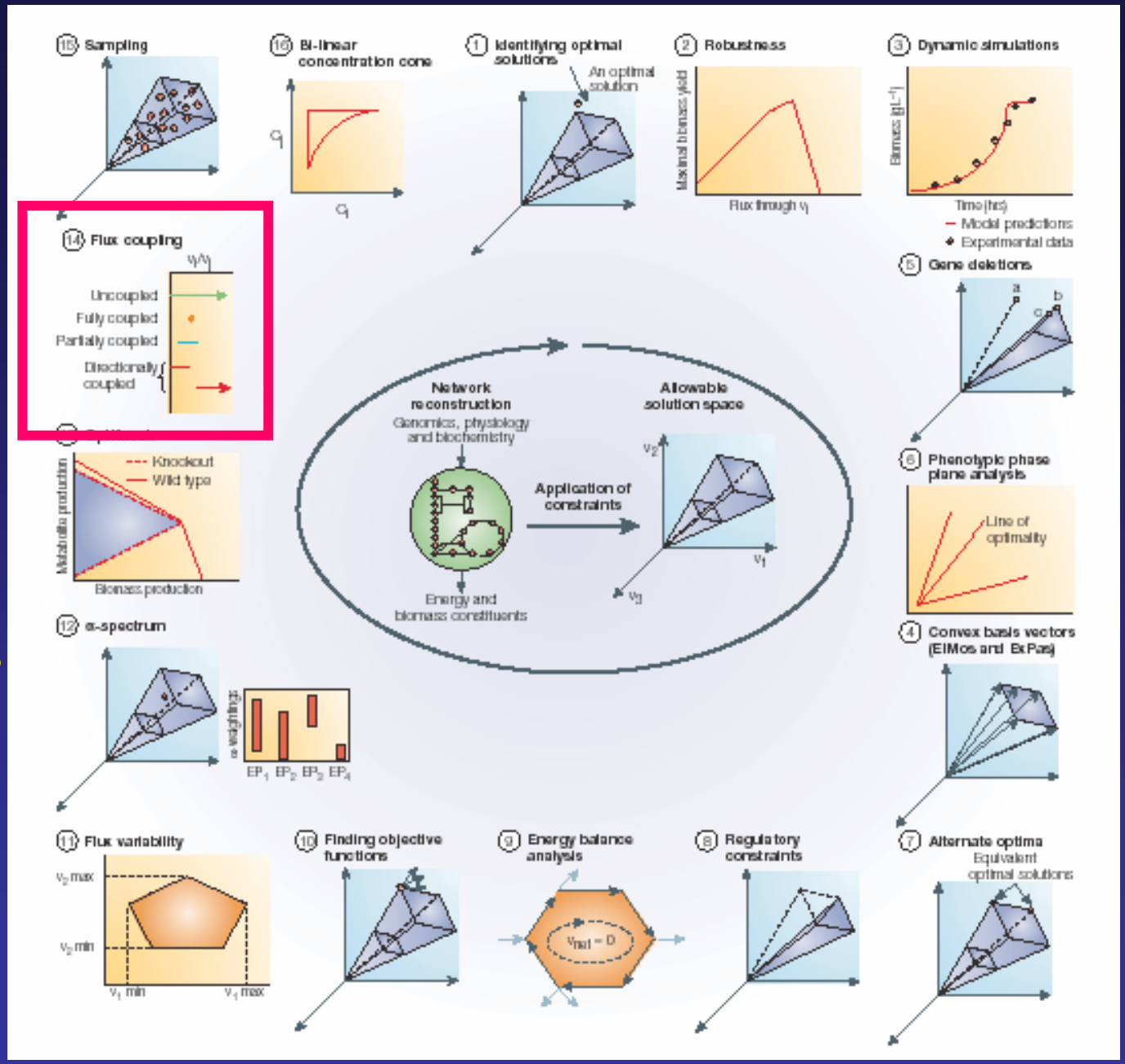
Use #3: Computational interrogation

tailoring to specific cases, and systems analysis of SNPs

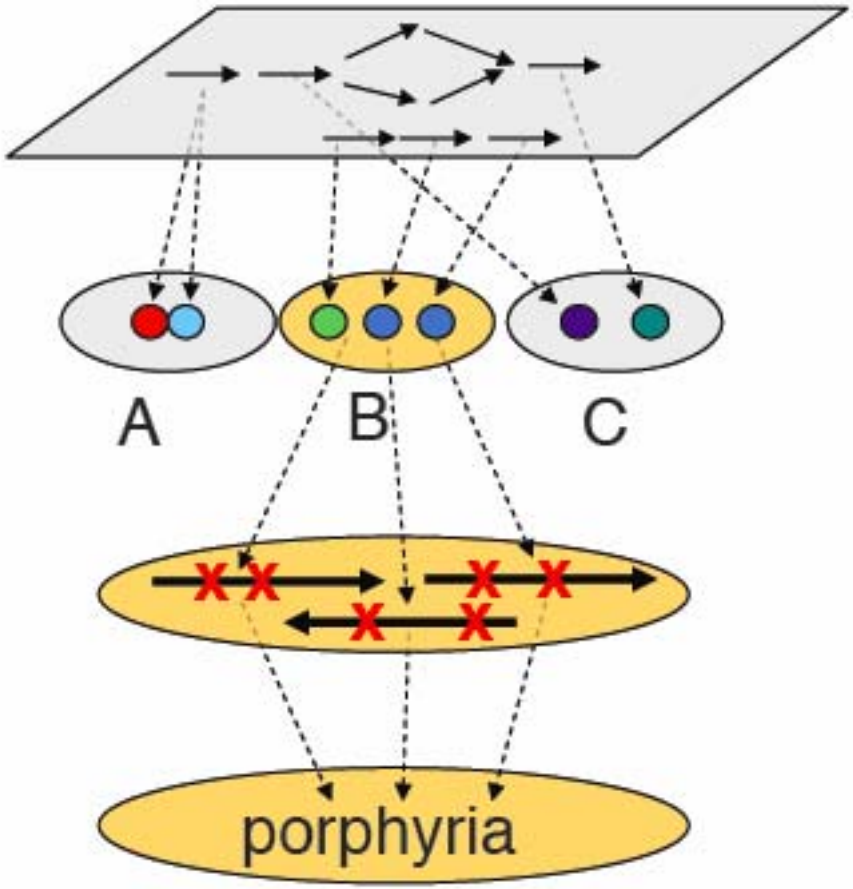
Ex #2:Using map properties to analyze causative SNP and their correlation

Genome-scale constraint-based modeling:

a rapidly growing field



Price, et al Nat Rev Microbiol. Nov 2004



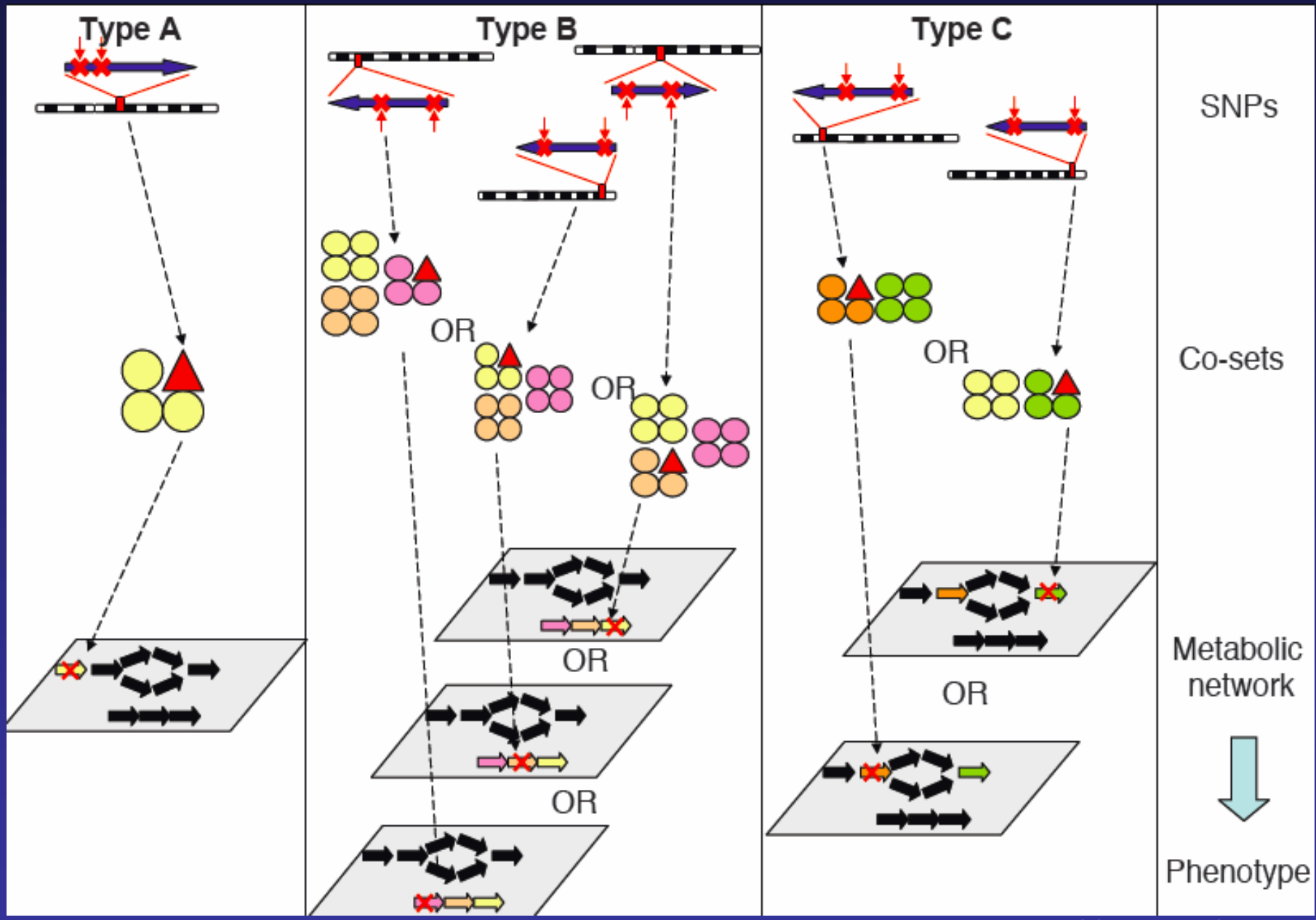
Metabolic network

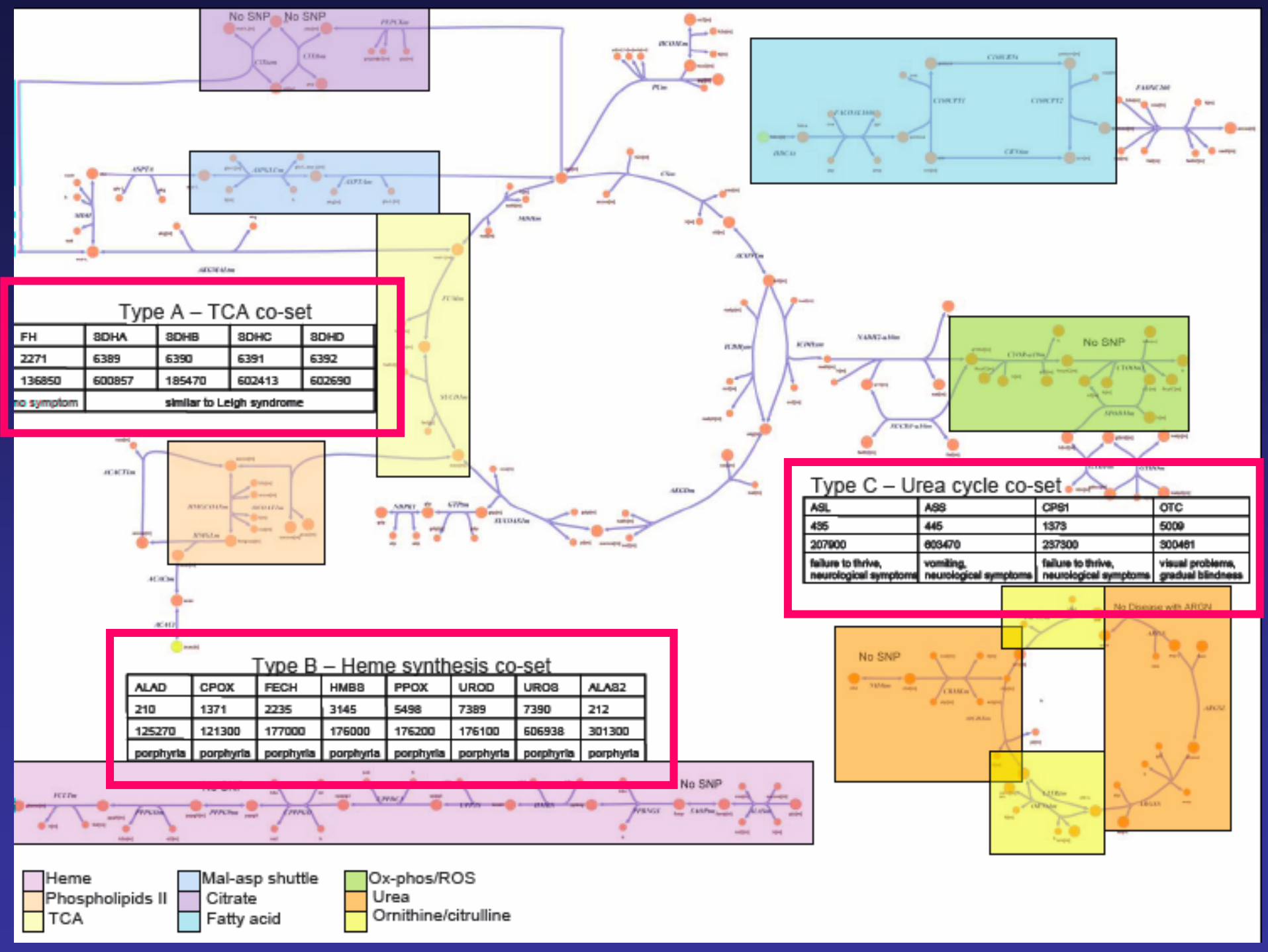
Co-sets

SNPs

Common Phenotype

Jamshidi, N., Palsson, B., "Systems biology of SNPs," *Molecular Systems Biology*, in press (2006)


Type A
Type B
Type C
SNPs
OR
OR
OR
OR
OR
Co-sets
Metabolic network
Phenotype



Type A – TCA co-set

FH	SDHA	SDHB	SDHC	SDHD
2271	6389	6390	6391	6392
136850	600857	185470	602413	602690
no symptom	similar to Leigh syndrome			

Type B – Heme synthesis co-set

ALAD	CPOX	FECH	HMBS	PPOX	UROD	UROS	ALAS2
210	1371	2235	3145	5498	7389	7390	212
125270	121300	177000	176000	176200	176100	606938	301300
porphyria	porphyria	porphyria	porphyria	porphyria	porphyria	porphyria	porphyria

Type C – Urea cycle co-set

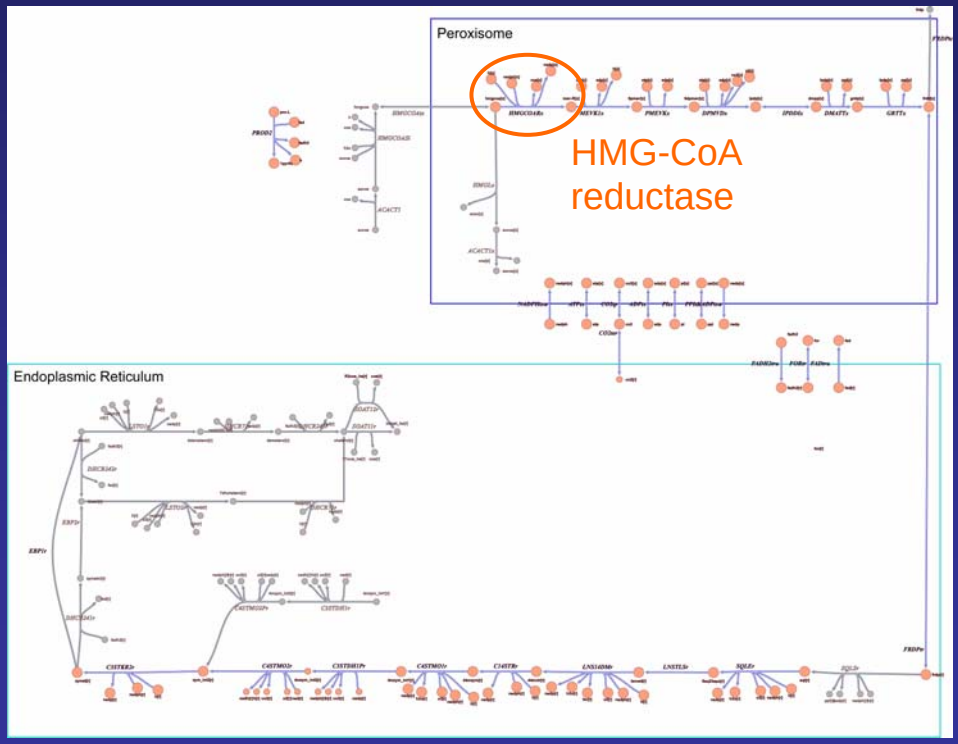
ASL	ASS	CPS1	OTC
435	445	1373	5009
207900	603470	237300	300461
failure to thrive, neurological symptoms	vomiting, neurological symptoms	failure to thrive, neurological symptoms	visual problems, gradual blindness

Correlated reaction sets

Flux coupling results under glucose aerobic conditions

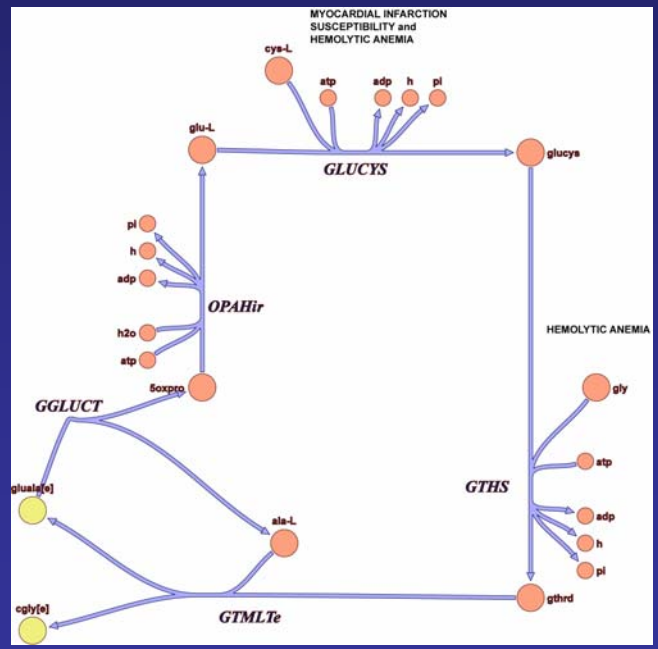
Cholesterol biosynthesis

Large, multi-compartment co-set



Glutathione metabolism

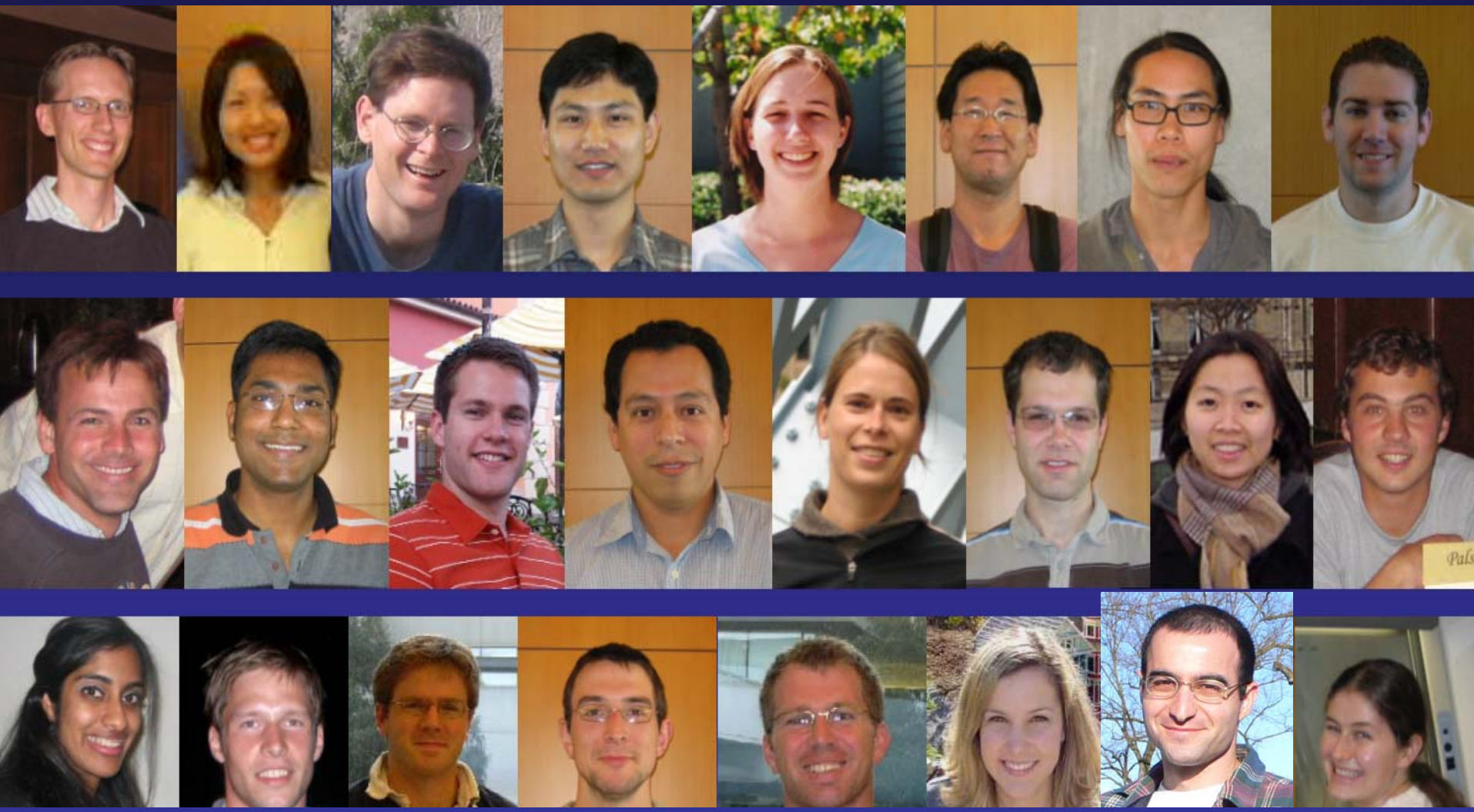
Genes have similar disease annotations



Conclusions

- We have reconstructed the genome-scale human metabolic map
- This global network was successfully reconstructed by using genomic and bibliomic data
- It can be computationally represented and can compute phenotypic states (288 examined)
- The reconstruction:
  1. Allows identification of knowledge gaps
  2. Provides a context for multi-omic integration
  3. Leads to computational interrogation of network properties

Systems Biology Research Group



UCSD Visit us at http://systemsbiology.ucsd.edu

Papers can be accessed from..

http://systemsbiology.ucsd.edu

Funding: NIGMS, NSF

Disclosure: Co-founder of Genomatica, a UCSD spin-off

UCSD