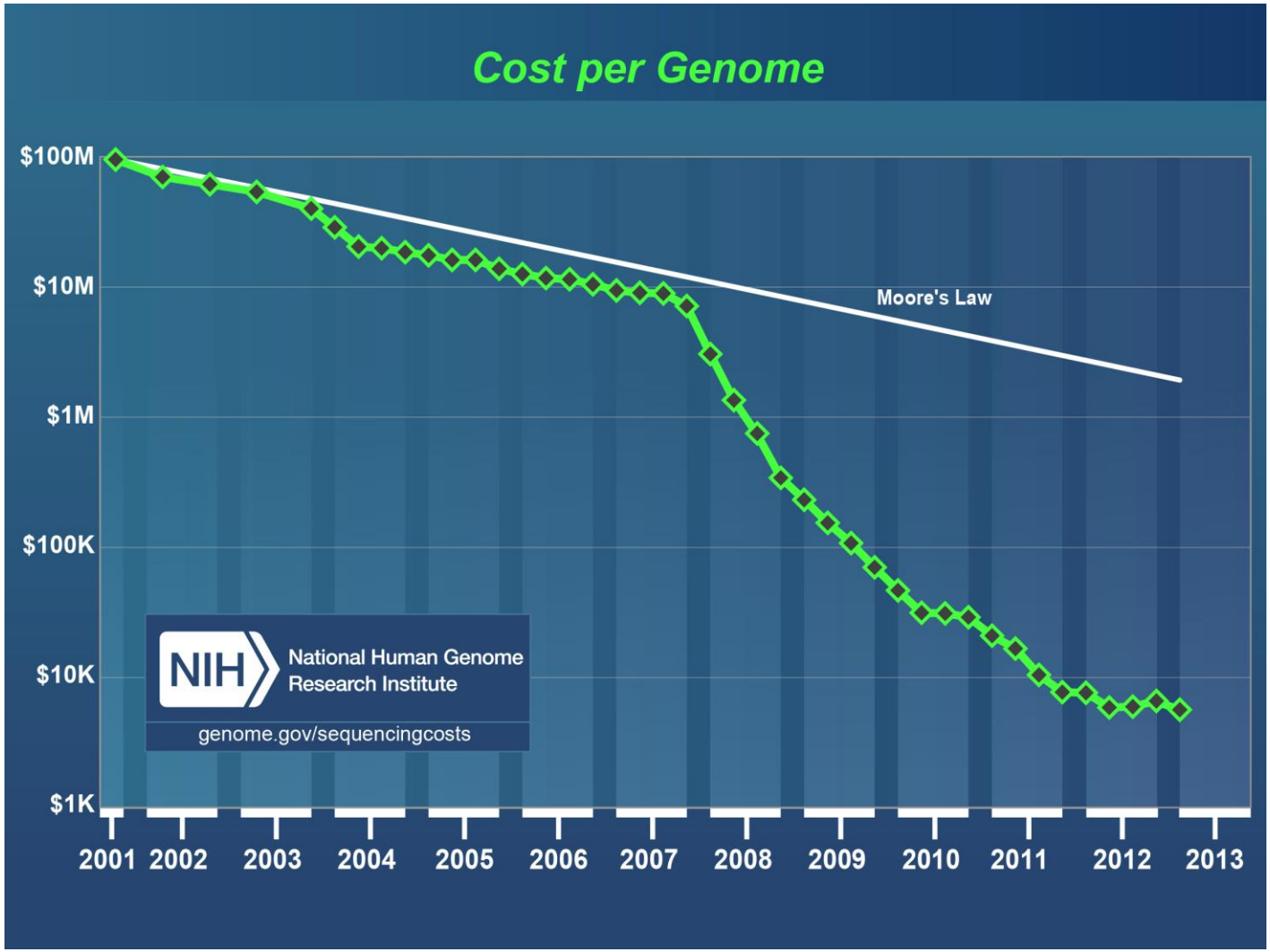
Bioinformatics and Computational Pharmacology

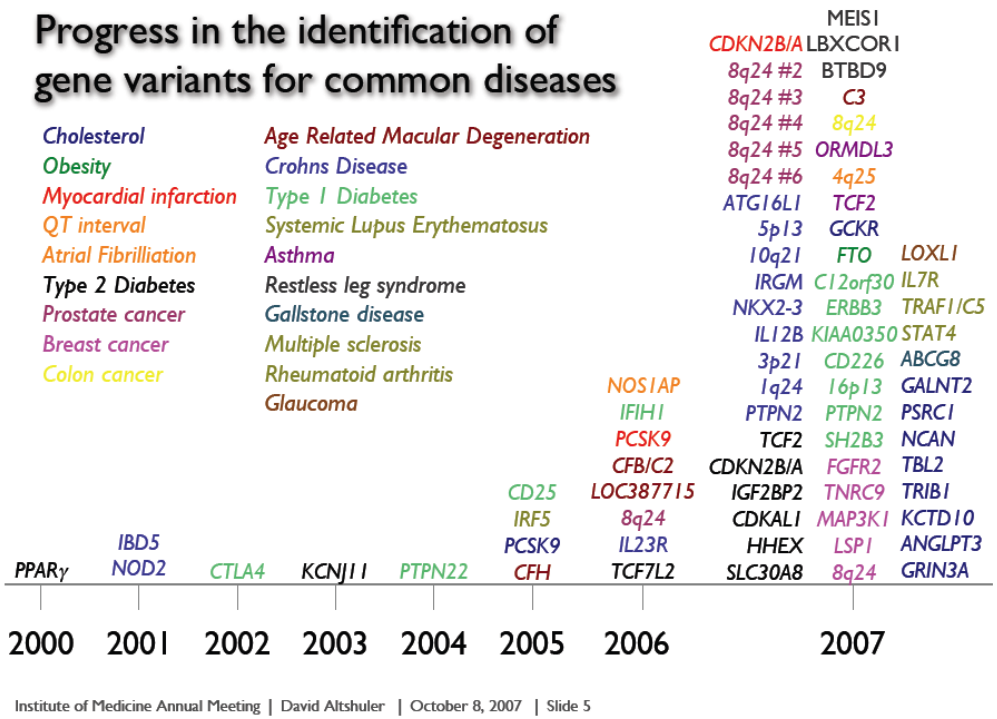
Lawrence Hunter, Ph.D.
Professor and Director
Computational Bioscience Program
University of Colorado School of Medicine

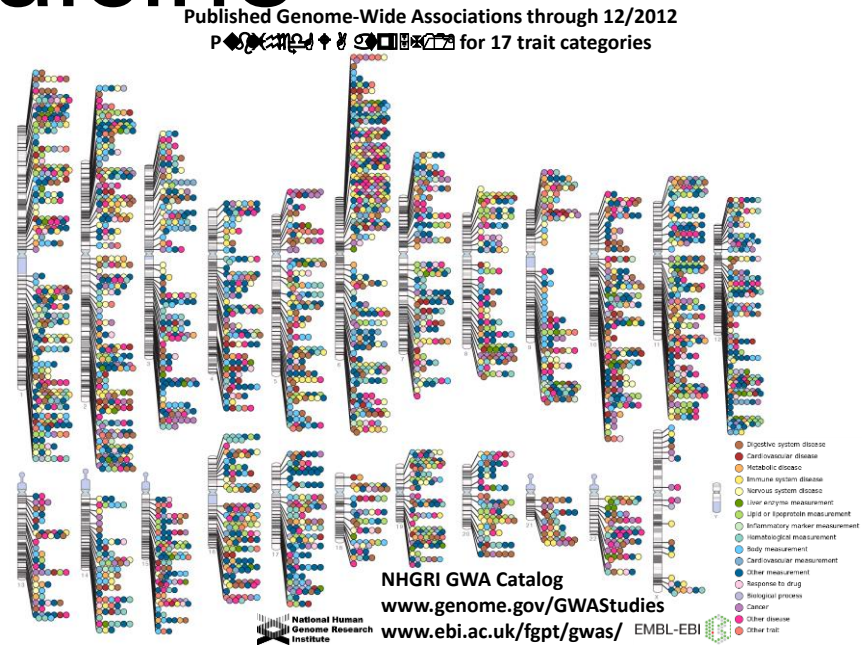
Larry.Hunter@uchsc.edu
http://compbio.uchsc.edu/Hunter

The revolution in genomics



We are at an inflection point in the history of life and medicine





Large UK GWAS Underscores Impact of CYP2C9, VKORC1, New SNP in Warfarin-Dosing Variability

March 25, 2009

Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy of Clopidogrel Therapy

JAMA®
The Journal of the American Medical Association

Vol. 302 No. 8, August 26, 2009

Analysis is the hard part

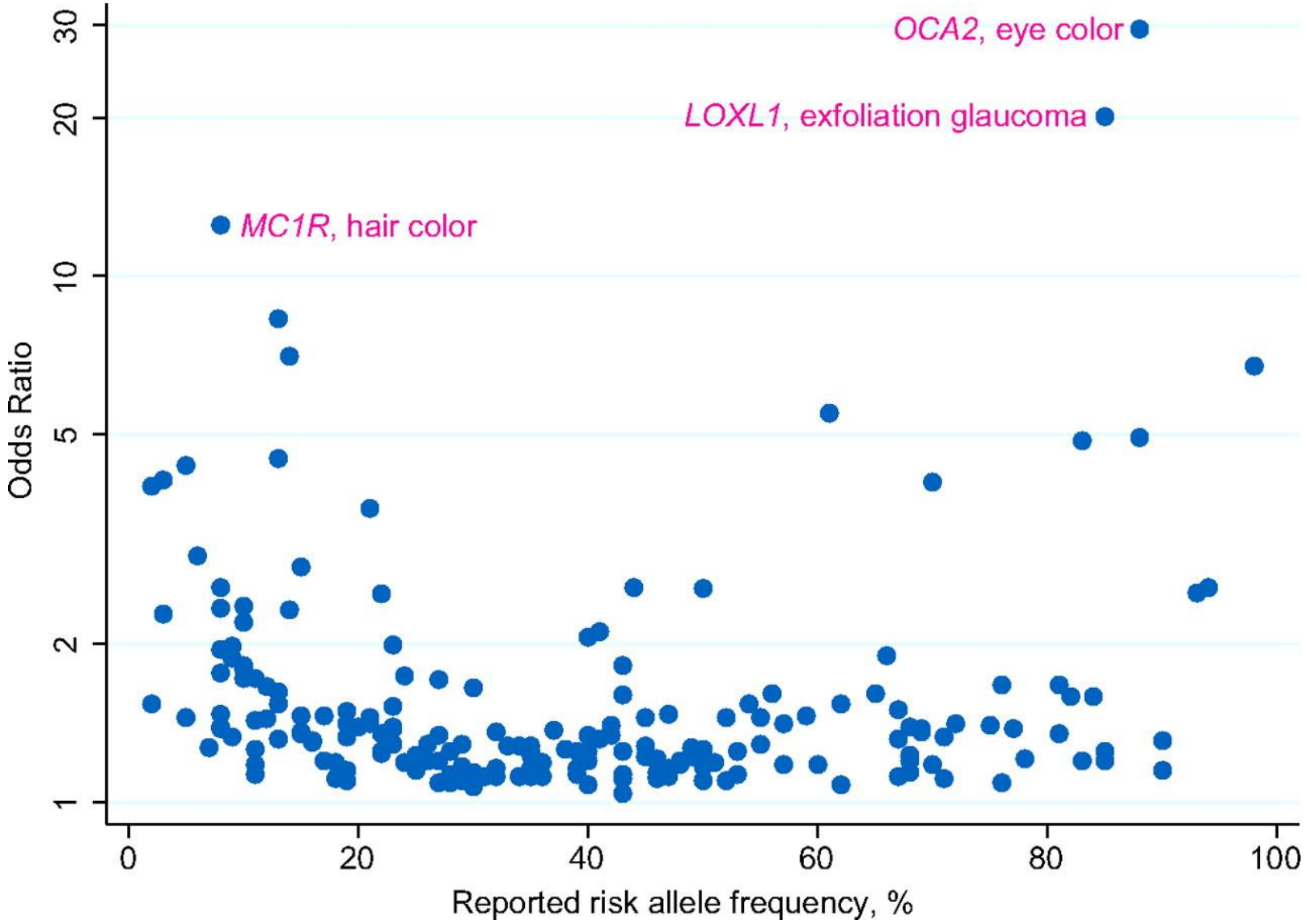
- “We are close to having a $1,000 genome sequence, but this may be accompanied by a $1,000,000 interpretation.”

  - Bruce Korf, president American College of Medical Genetics

- Not only is the cost of sequencing essentially free, but big computers and big storage are cheap, too. What will keep us busy for the next 50 years is understanding the data”

  - Russ Altman, chair of Biomedical Engineering at

The gene? No, a gene.

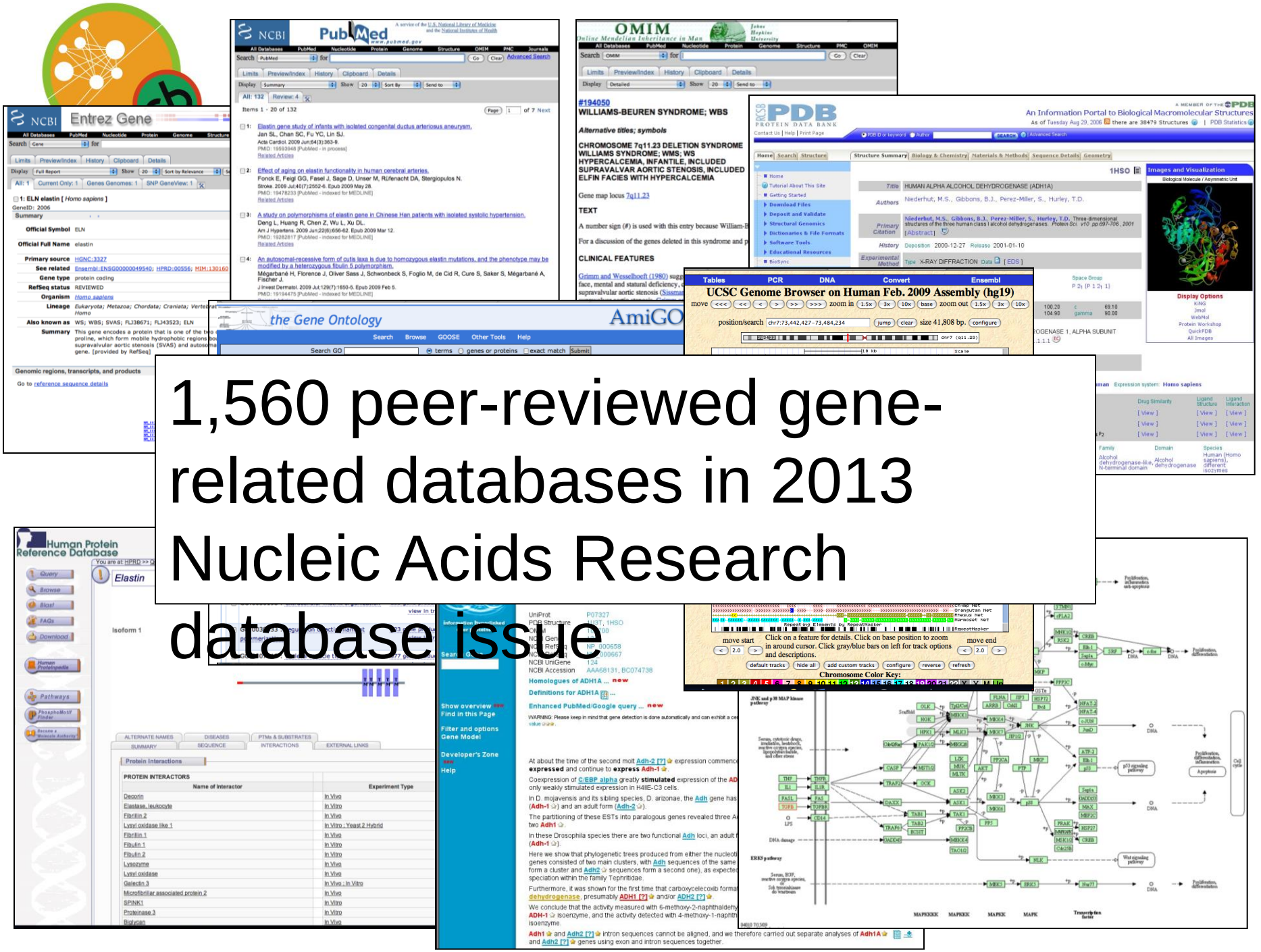


The basics

- What sorts of explanations play a role in understanding life at the molecular level?
- Where do you find knowledge about genes that is relevant to understanding their effects?
- How do you put together knowledge and data about multiple genetic differences into a coherent story about mechanism?

Information about genes

- Integrated summary resources:
  - Entrez Gene http://www.ncbi.nlm.nih.gov/gene
  - EBI Gene Summaries http://www.ebi.ac.uk
  - UCSC Genome Browser http://genome.ucsc.edu/
- Multigene interactions & groups
  - Protein interactions http://www.ebi.ac.uk/intact/
  - Metabolic pathways http://www.reactome.org/
- Gene / drug interactions


1,560 peer-reviewed gene-related databases in 2013
Nucleic Acids Research database issue

Gene functions

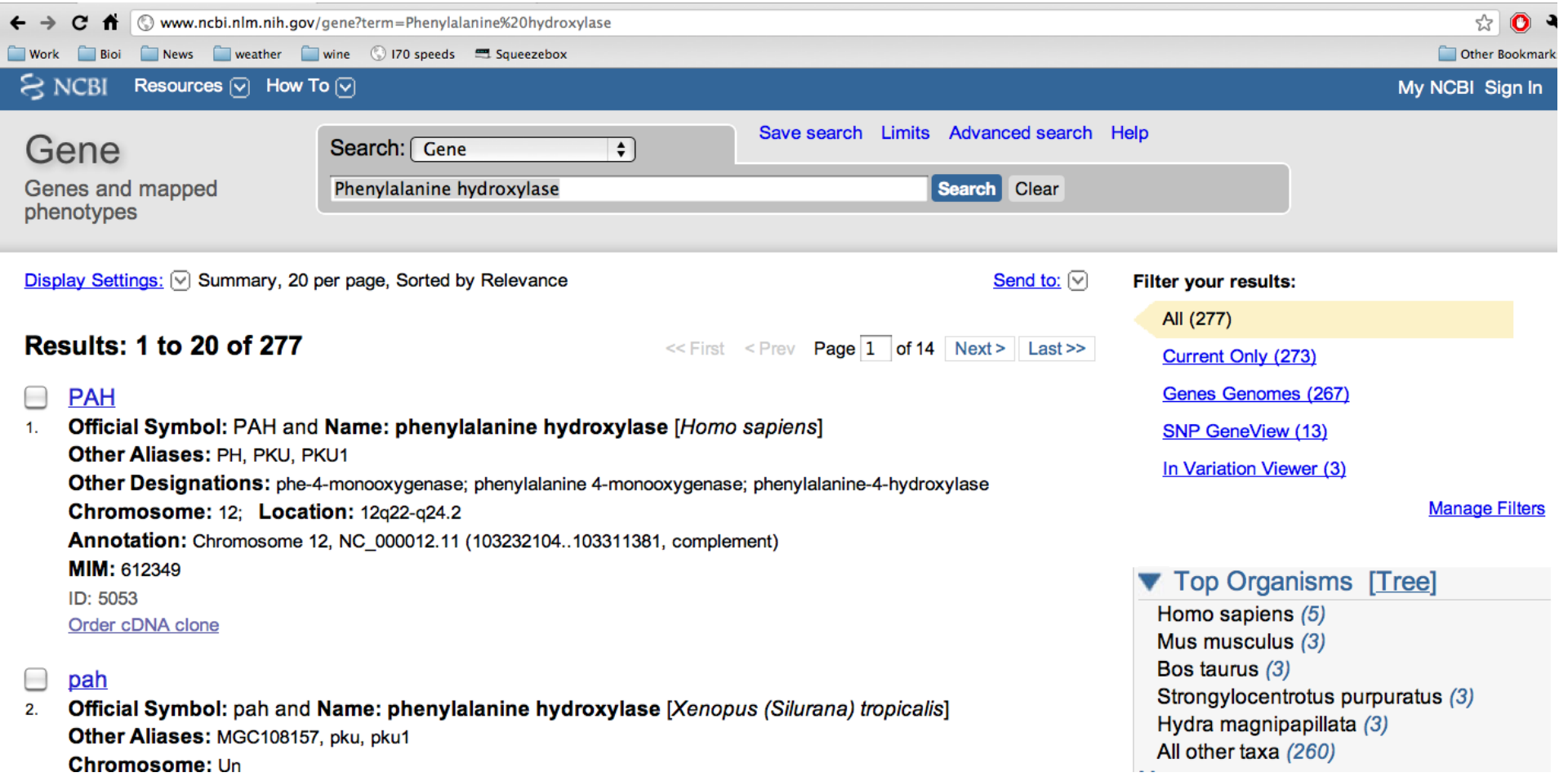
- All organisms must:
  - Grow: turn available substances into living parts
  - Reproduce
- To do this, their genes must:
  - Manage energy
  - Synthesize biochemicals (DNA, protein, lipids)
  - Sense the state of the environment
  - Maintain homeostasis
- These are the functions that explain

A simple example

- *Phenylalanine hydroxylase* (PAH) is a gene whose product catalyzes transformation the amino acid *phenylalanine* to *tyrosine*.
- If there isn't enough PAH, then phenylalanine accumulates, and can be toxic. This happens in a disease called *phenylketonuria.*
- We test newborns for this gene defect, since avoiding phenylalanine in the diet

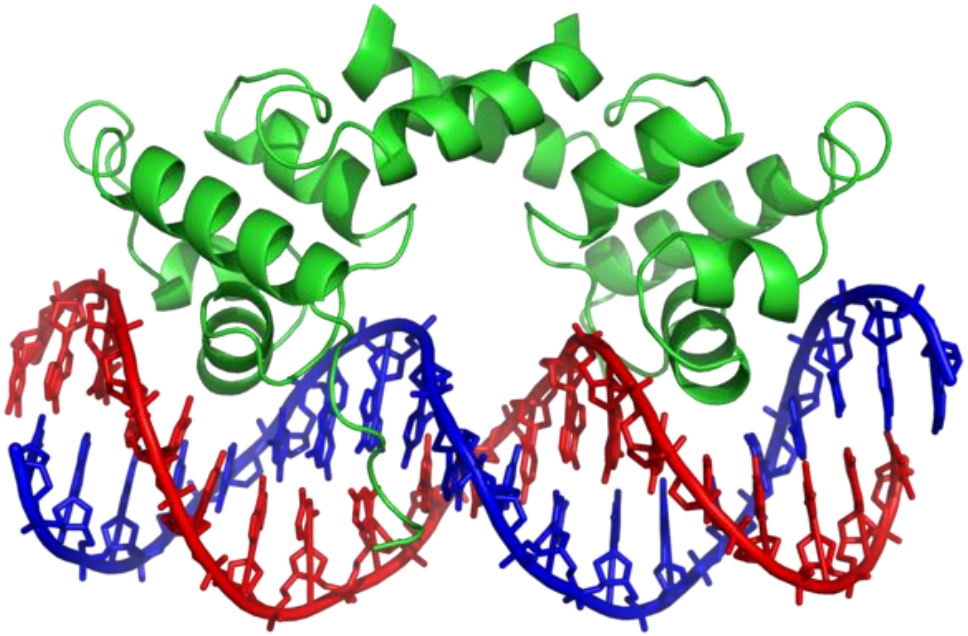
How to find out? Entrez Gene

http://www.ncbi.nlm.nih.gov/gene



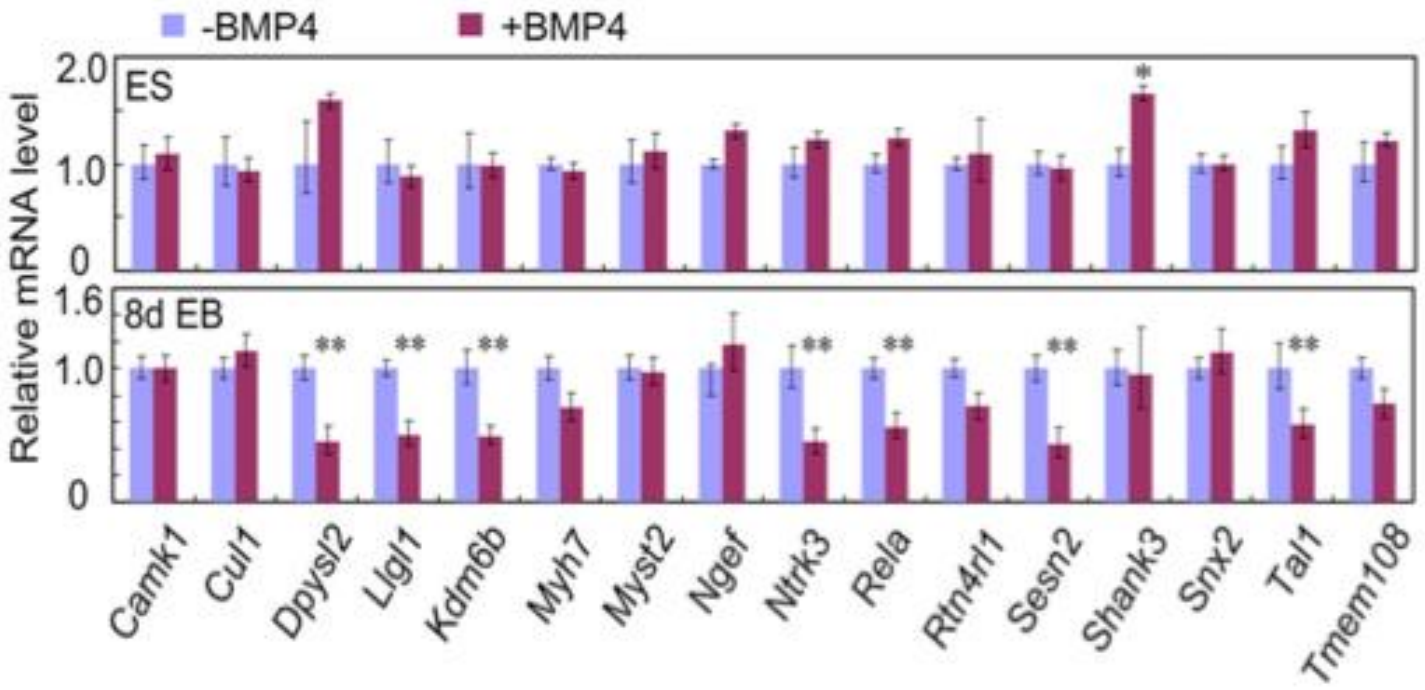
Transcriptional Regulation

- Transcription factors (TFs)
  - DNA-binding proteins
  - Control the expression of other genes
  - Changes in TFs have very widespread effects



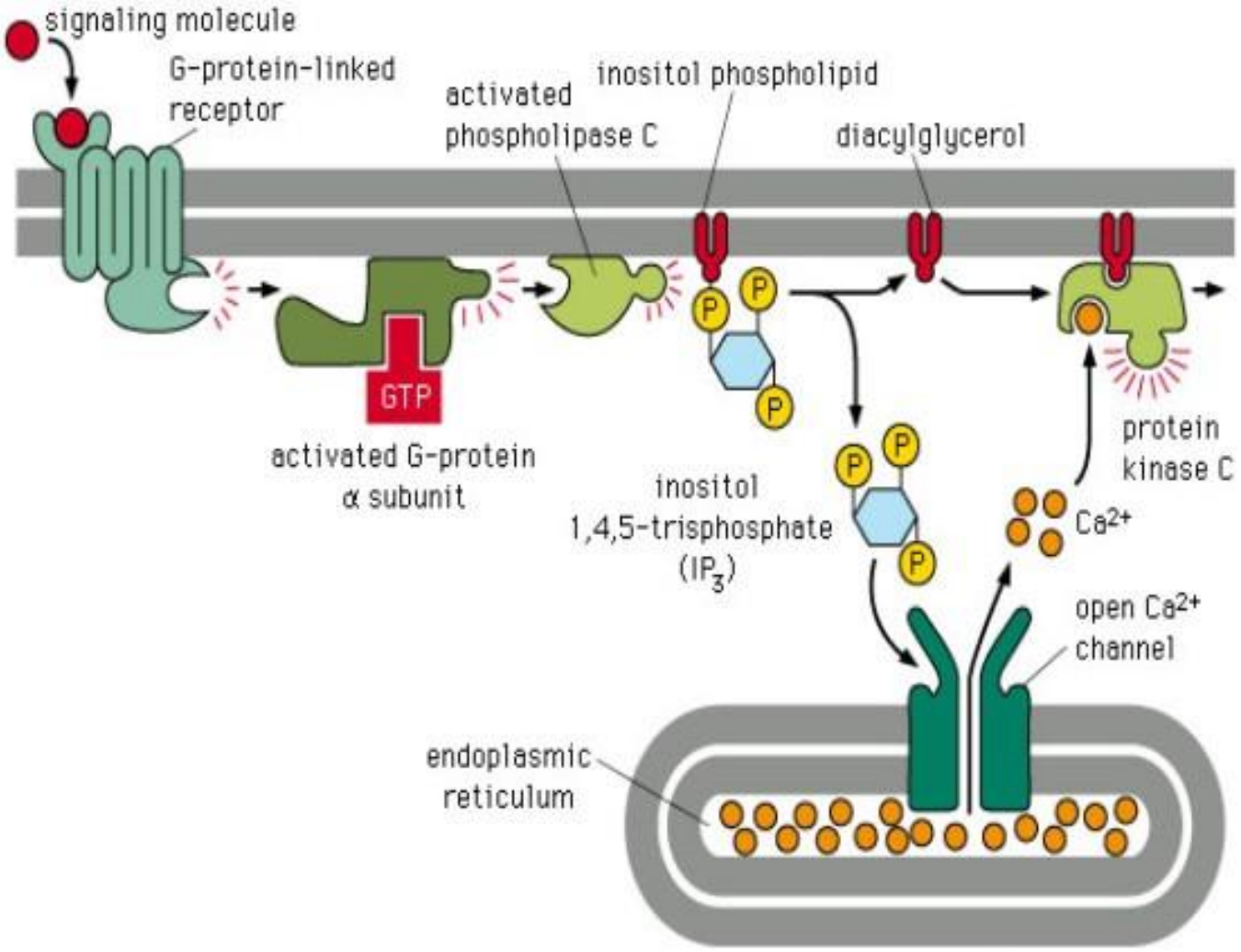
Downstream consequences

Transcription factors tend to influence many “downstream” gene
BMP4 plays a key role in bone development



Fei T et al. Genome Res. 2010;20:36-44

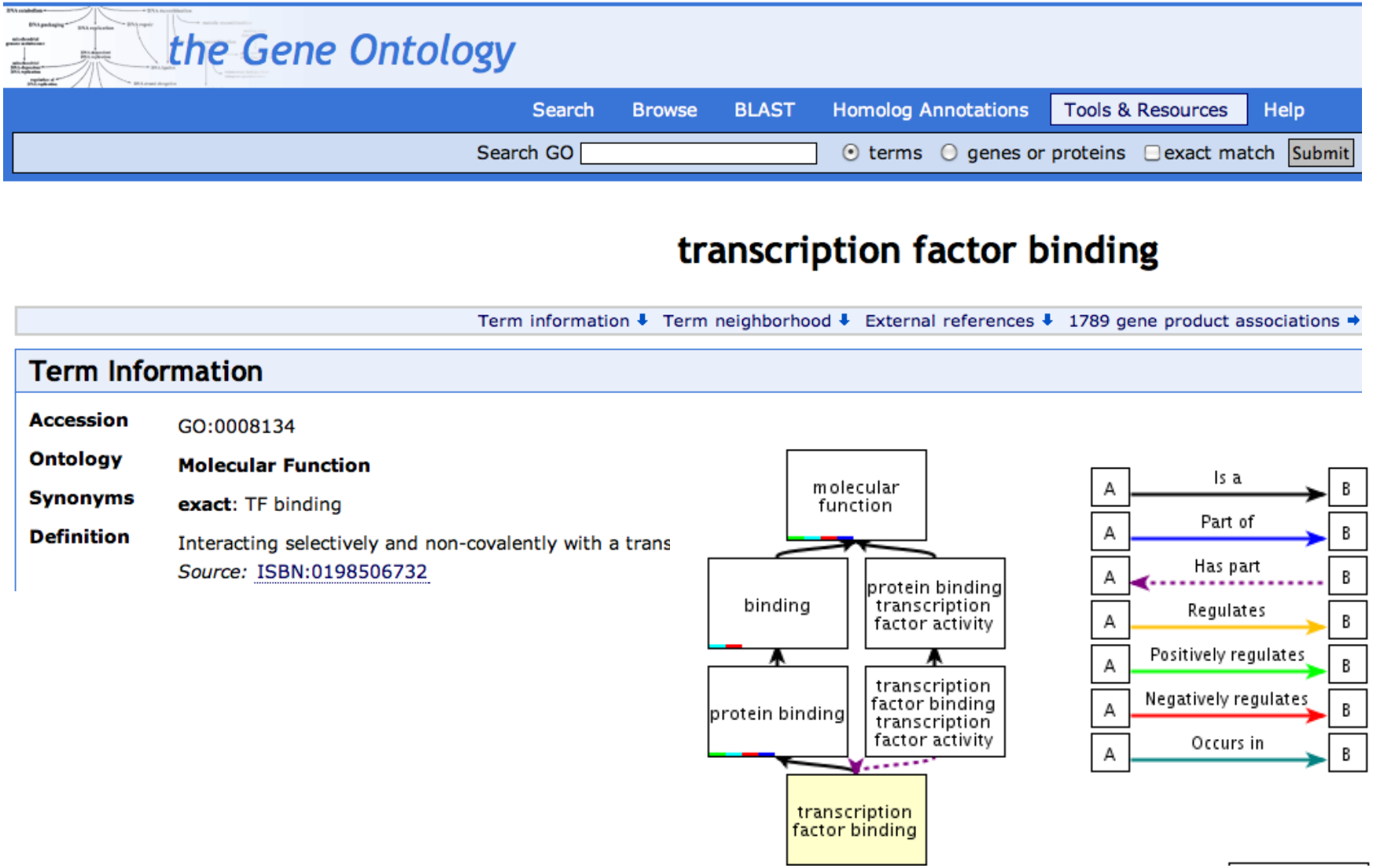
Signal Transduction



Describing functions

- Functions
  - As biochemical effect
  - As process participation
  - As interaction with other proteins
- How to name them?
  - Hierarchy (function is manifold)
  - Ontology (cross organisms)

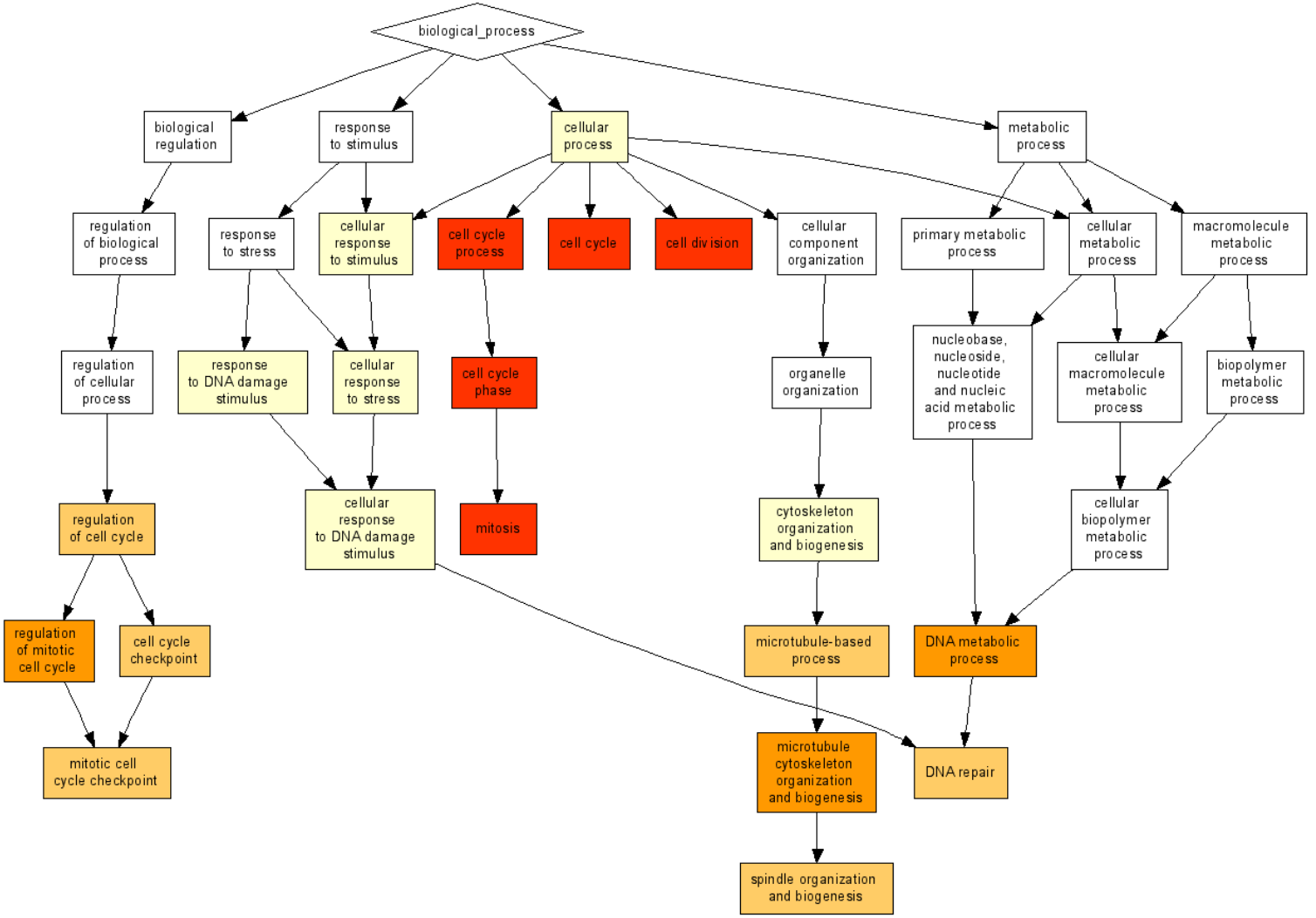
Gene Ontology



Multiplicity

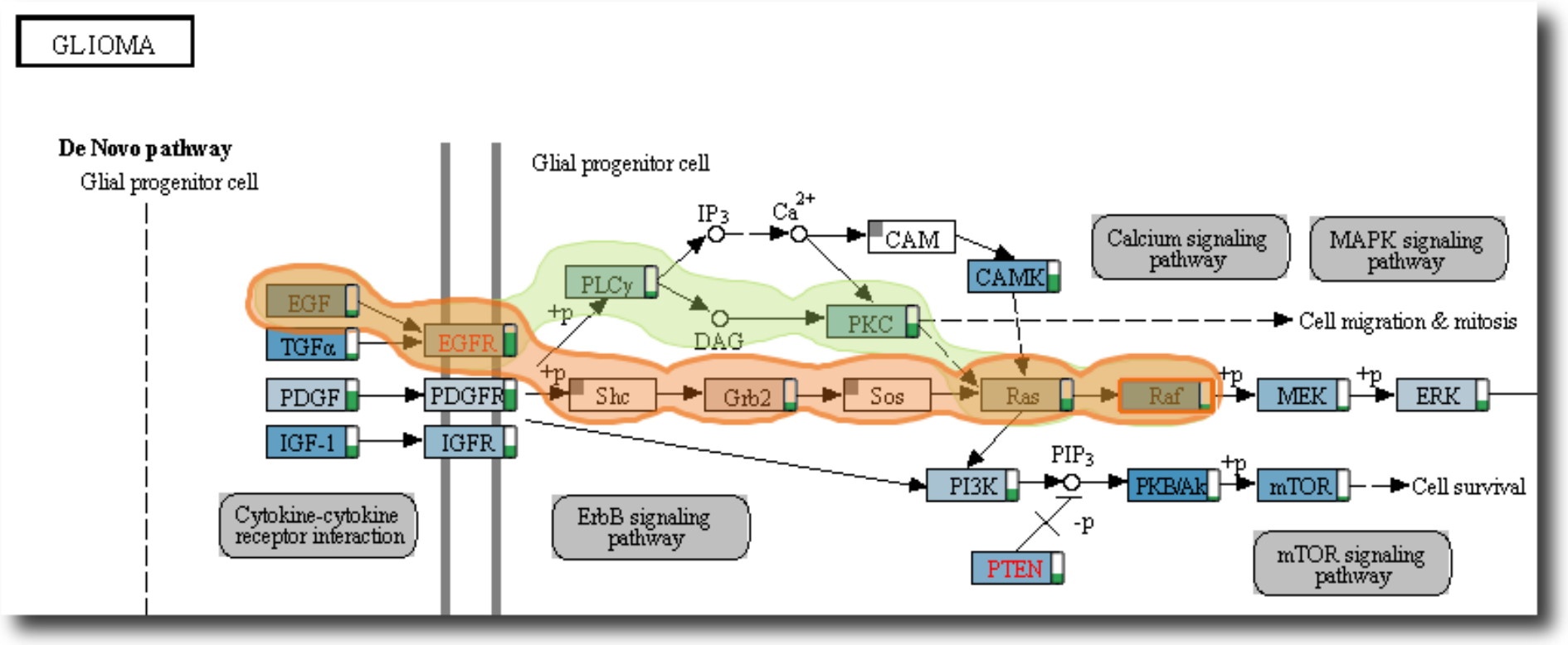
- Gene products have multiple functions
- Multiple genes work together
  - Modern experimental results generate lists of genes relevant to a phenomenon
- How do we understand the integrated functioning of many genes together?

Function enrichment

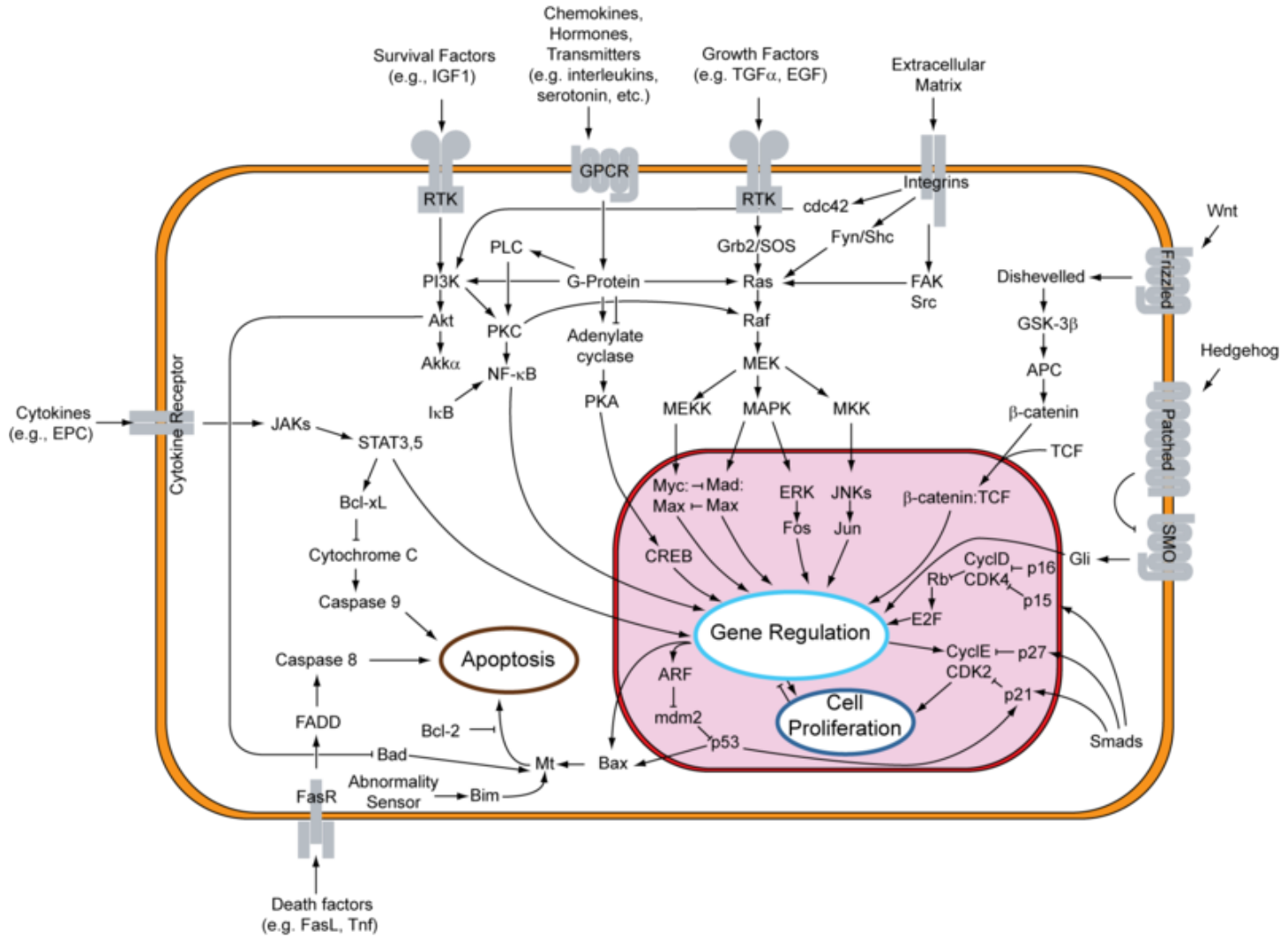


Mapping Genes to Pathways

- enRoute, 2012 BioVis best paper winner
  http://www.icg.tugraz.at/project/caleydo/projects-1/enroute

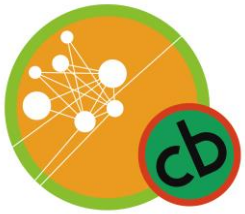


Combining functions into a story



But that's the ideal…

- I've shown you some of the best known, best studied pathways.
  - Most pathways have lots of missing pieces
- Many functions cannot be assembled easily into a reasonable story
  - Some protein functions remain unknown, too
- Most assay methods don't identify all the relevant genes anyway…



Many challenges

- Diversity: More types of data & knowledge than most application areas (across many orders of magnitude)
- Verification: biologists don't trust summaries; need easy access to experimental justifications
- Logical complexity: e.g. absence can be interesting
- Connectivity: Linkages among data can be more interesting than the data itself
- Focus: vast sea of data and documents; what's interesting to a user depends on context
- Timeliness: >2,500 new journal articles per day;