Biological Knowledge Assembly and Interpretation

We have looked at interesting changes and patterns in microarray and RNA-Seq data:

1. PSON: Change in RNA-Seq expression levels between brca cell lines

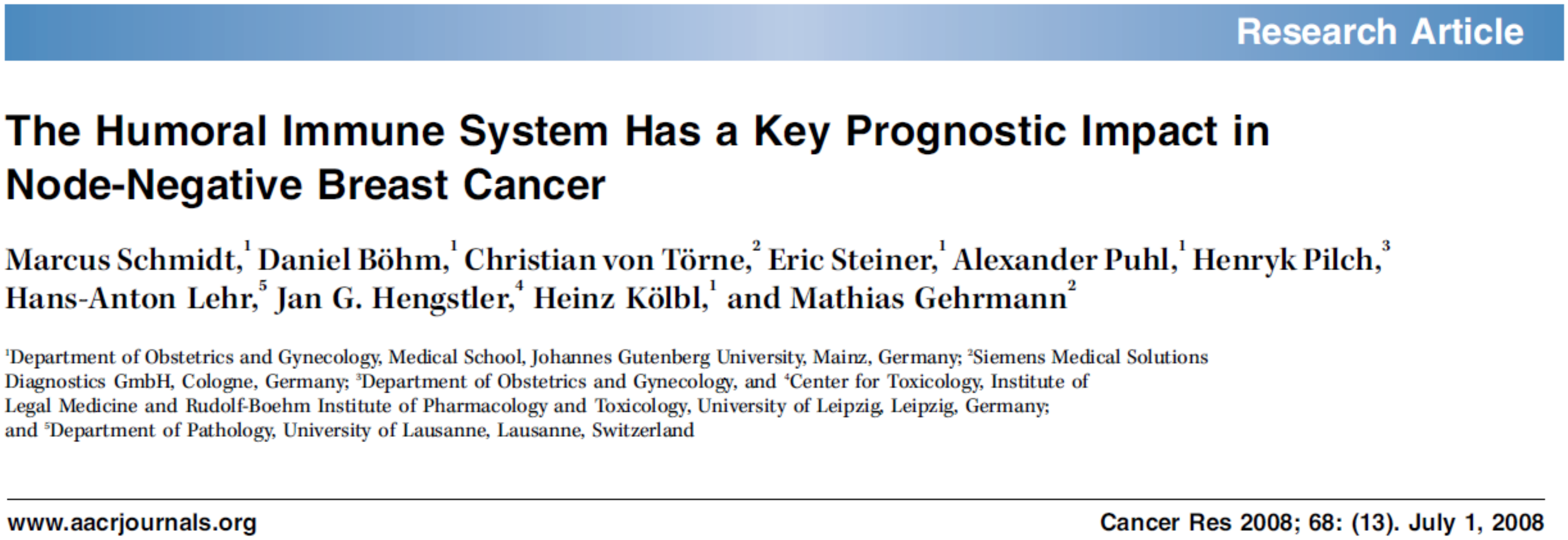
2. TCGA: 500 genes with highest variance across breast cancer samples

3. Hierarchical agglomerative clustering: ER status

4. Principal components analysis: Metagenes

Exploratory data analysis

Next goal: Extract biological meaning



Research Article

The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast Cancer

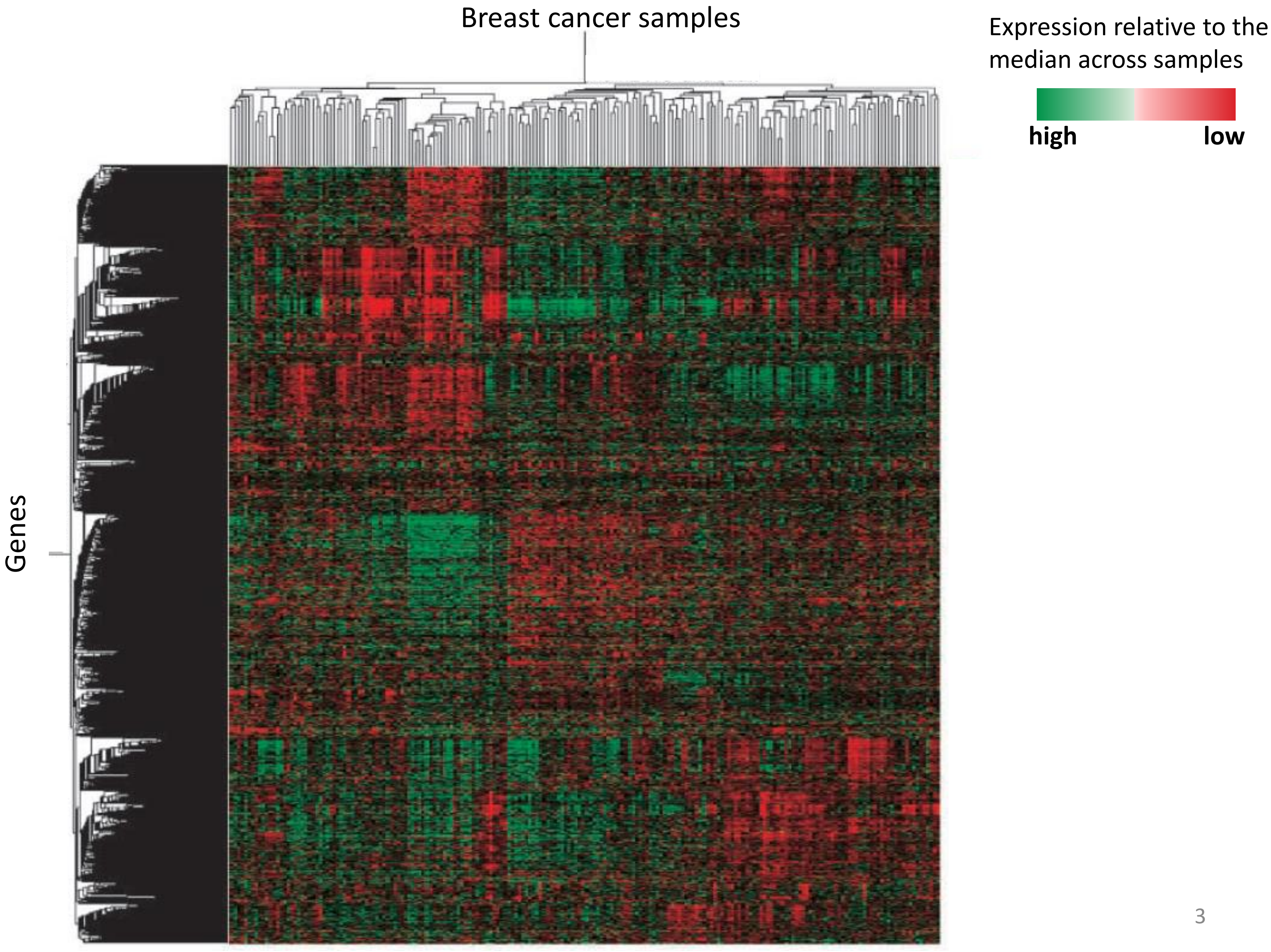
Marcus Schmidt,[1] Daniel Böhm,[1] Christian von Törne,[2] Eric Steiner,[1] Alexander Puhl,[1] Henryk Pilch,[3] Hans-Anton Lehr,[5] Jan G. Hengstler,[4] Heinz Kölbl,[1] and Mathias Gehrmann[2]

[1]Department of Obstetrics and Gynecology, Medical School, Johannes Gutenberg University, Mainz, Germany; [2]Siemens Medical Solutions Diagnostics GmbH, Cologne, Germany; [3]Department of Obstetrics and Gynecology, and [4]Center for Toxicology, Institute of Legal Medicine and Rudolf-Boehm Institute of Pharmacology and Toxicology, University of Leipzig, Leipzig, Germany; and [5]Department of Pathology, University of Lausanne, Lausanne, Switzerland

www.aacrjournals.org

Cancer Res 2008; 68: (13). July 1, 2008

Mainz data set from Schmidt et al.



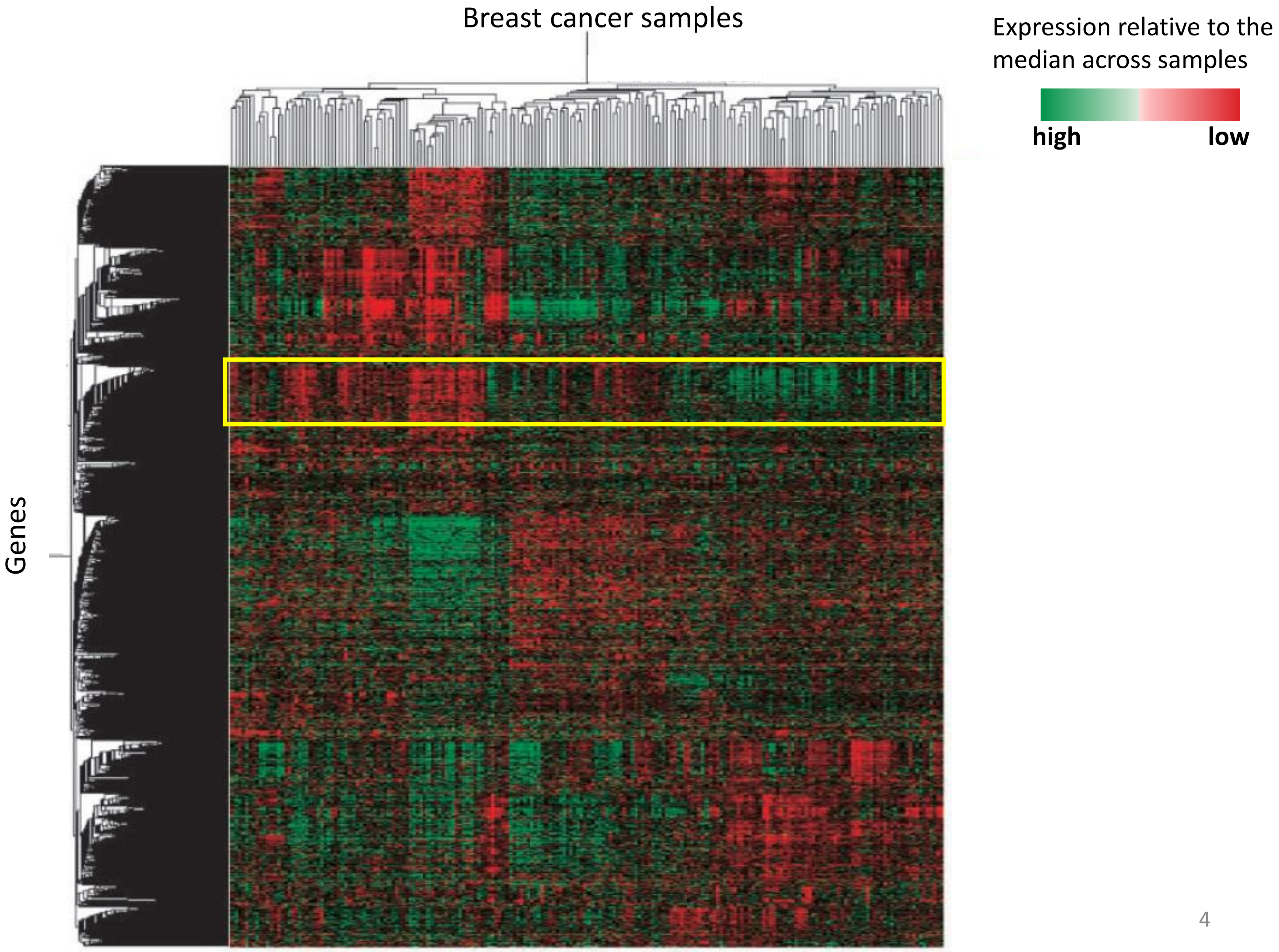
Breast cancer samples

Genes

Expression relative to the median across samples

high low

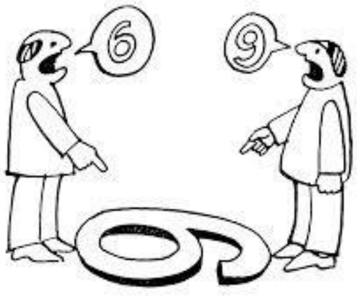
Mainz data set from Schmidt et al.



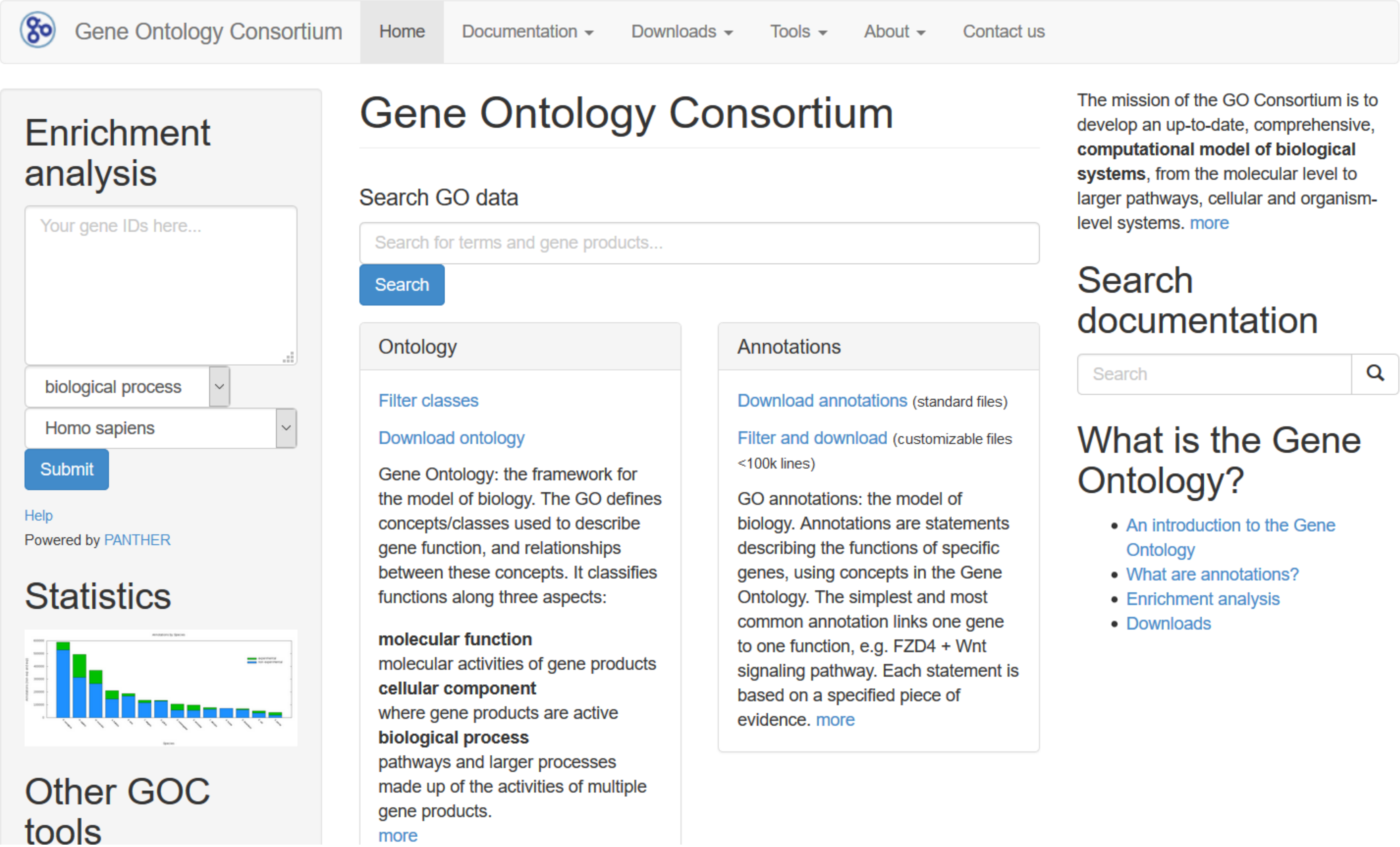
Find biological associations “by hand”?

Time-consuming

Extremely subjective and not systematic



http://www.geneontology.org



Enrichment analysis

Help

Powered by PANTHER

Statistics

Other GOC tools

Gene Ontology Consortium

Search GO data

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products
cellular component
where gene products are active
biological process
pathways and larger processes made up of the activities of multiple gene products.
more

Annotations

Download annotations (standard files)

Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. more

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. more

Search documentation

What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Enrichment analysis
- Downloads

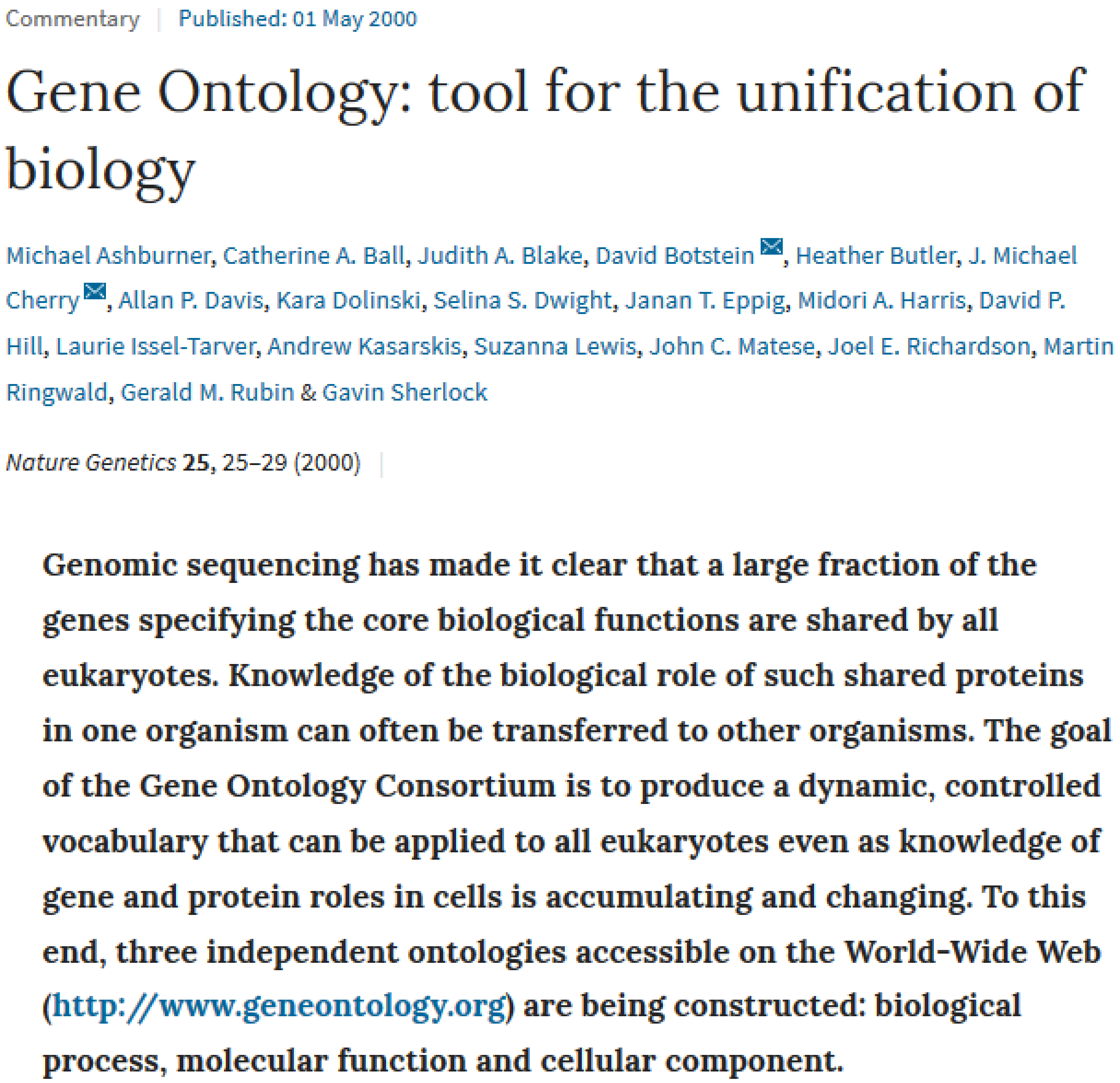
Complexity of biological systems and datasets of increasing size.

→ We depend on knowledge in stored computable form to analyze biomedical research.

The Gene Ontology (GO) project is the most comprehensive resource for computable knowledge regarding the functions of genes and gene products.

Two primary components:

1) The **Gene Ontology (GO)** provides the logical structure of biological functions (‘terms’) and their relationships to one another.
2) The **GO annotations** are evidence-based statements relating a gene product to a specific ontology term



Commentary | Published: 01 May 2000

Gene Ontology: tool for the unification of biology

Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein ✉, Heather Butler, J. Michael Cherry ✉, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin & Gavin Sherlock

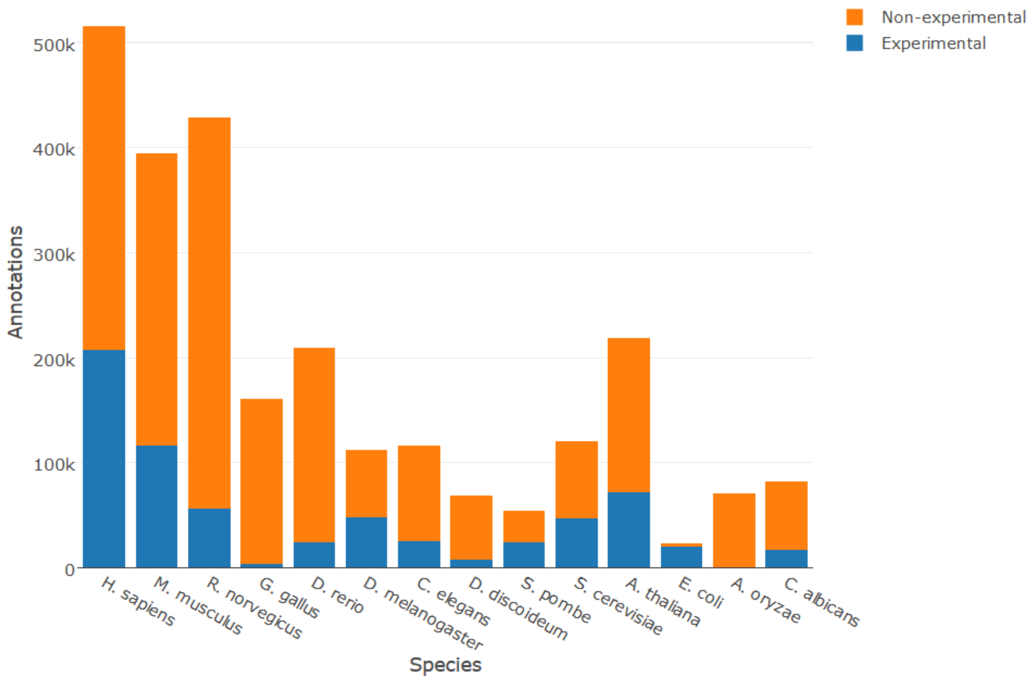
Nature Genetics **25**, 25–29 (2000)

Genomic sequencing has made it clear that a large fraction of the genes specifying the core biological functions are shared by all eukaryotes. Knowledge of the biological role of such shared proteins in one organism can often be transferred to other organisms. The goal of the Gene Ontology Consortium is to produce a dynamic, controlled vocabulary that can be applied to all eukaryotes even as knowledge of gene and protein roles in cells is accumulating and changing. To this end, three independent ontologies accessible on the World-Wide Web (http://www.geneontology.org) are being constructed: biological process, molecular function and cellular component.

Evolution of GO

- Original GO created in 2000
- Three databases involved:
  - FlyBase (*Drosophila*)
  - MGI (Mouse)
  - SGD (*S. cerevisae*)

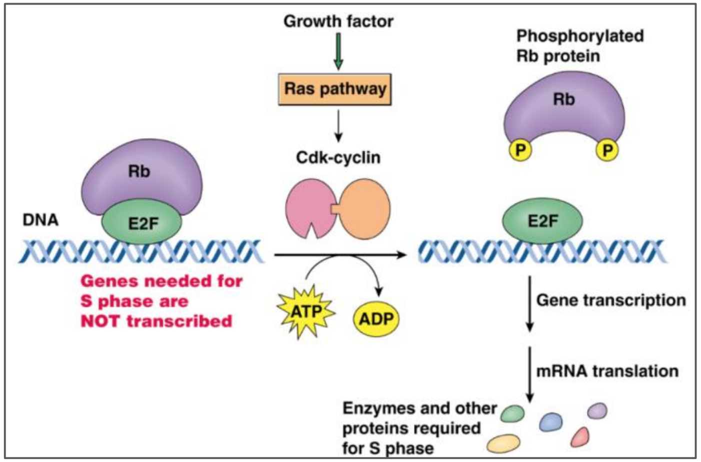
Experimental annotations by species



Three "aspects" of GO

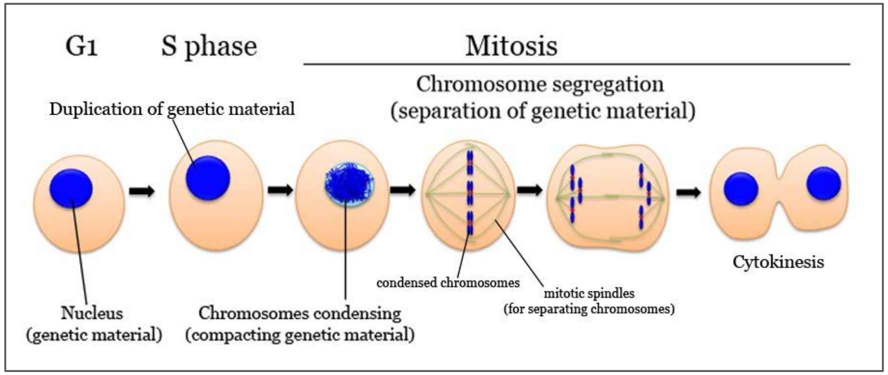
1. Molecular Function (MF)

   An elemental activity



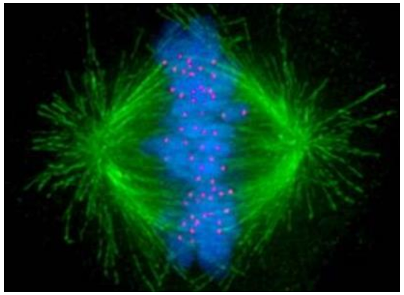
2. Biological Process (BP)

   A commonly recognized series of events



3. Cellular component

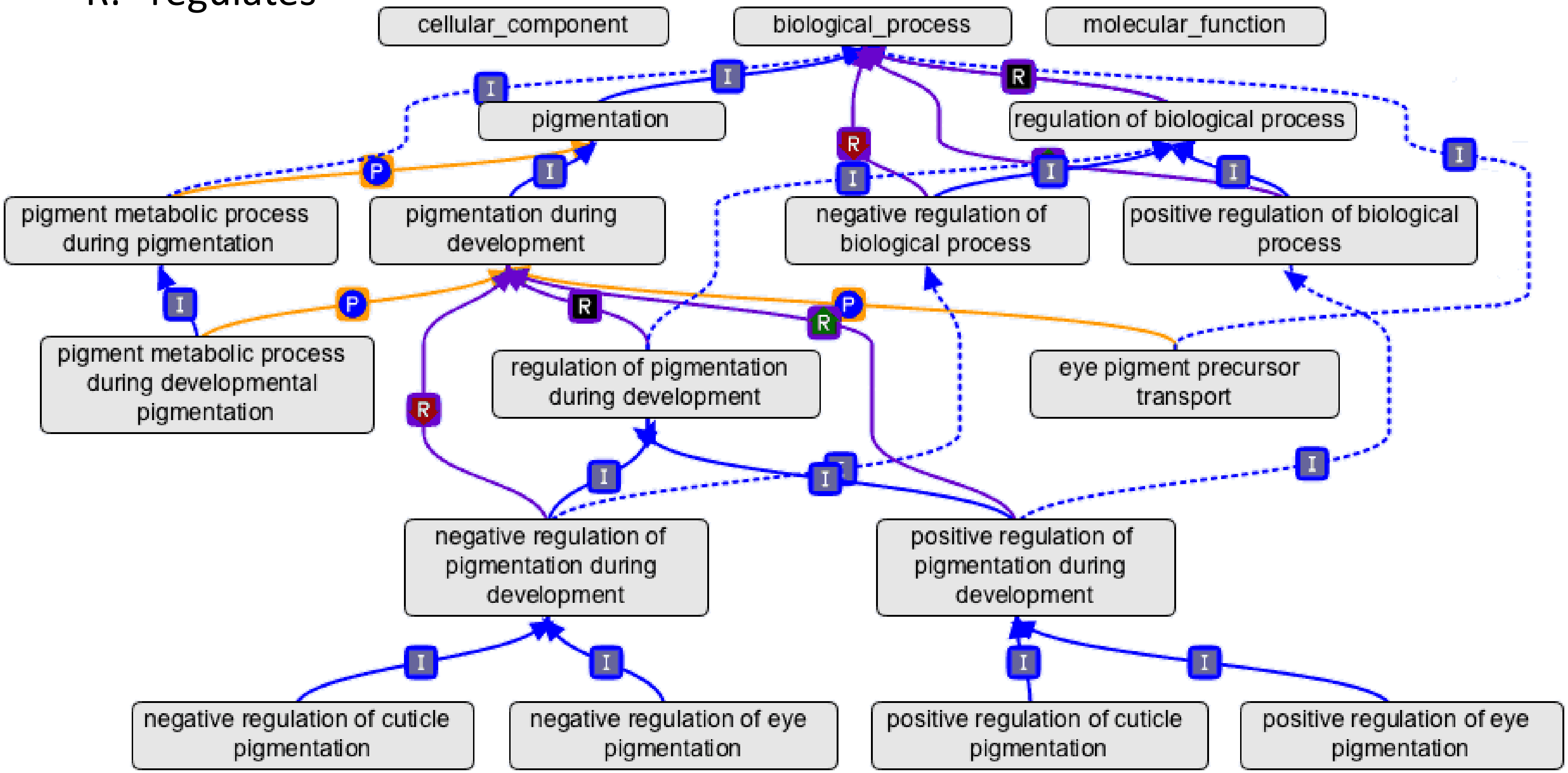
   Where a gene product is located



Spindle
Microtubule cytoskeleton

Hierarchical structure and relationships in GO

I: “is a”
P: “is part of”
R: “regulates”


cellular_component
biological_process
molecular_function
pigmentation
regulation of biological process
pigment metabolic process during pigmentation
pigmentation during development
negative regulation of biological process
positive regulation of biological process
pigment metabolic process during developmental pigmentation
regulation of pigmentation during development
eye pigment precursor transport
negative regulation of pigmentation during development
positive regulation of pigmentation during development
negative regulation of cuticle pigmentation
negative regulation of eye pigmentation
positive regulation of cuticle pigmentation
positive regulation of eye pigmentation
I
P
R

Less specific concepts

More specific concepts

GENEONTOLOGY
Unifying Biology

A major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases.

Three main goals:

1. Maintain and further develop its vocabulary
2. Annotate genes and gene products, disseminate annotation data
3. Provide tools to facilitate access to data

GENEONTOLOGY
Unifying Biology

The most common use of the Gene Ontology annotations is for interpretation of large-scale molecular biology experiments.

Given a set of genes that are up-regulated under certain conditions, **Gene Ontology (GO) enrichment analysis** will find which GO terms are over- or under-represented using the annotations for the set of genes.

We want to interpret the underlying molecular differences between:

A cancer cell and a normal cell,
Two different cells lines,
Across tumor samples, etc.

GO enrichment analysis identifies relevant groups of genes that function together.

Reduces thousands of molecular changes to a much smaller number of biological functions.

GO Annotation Tools

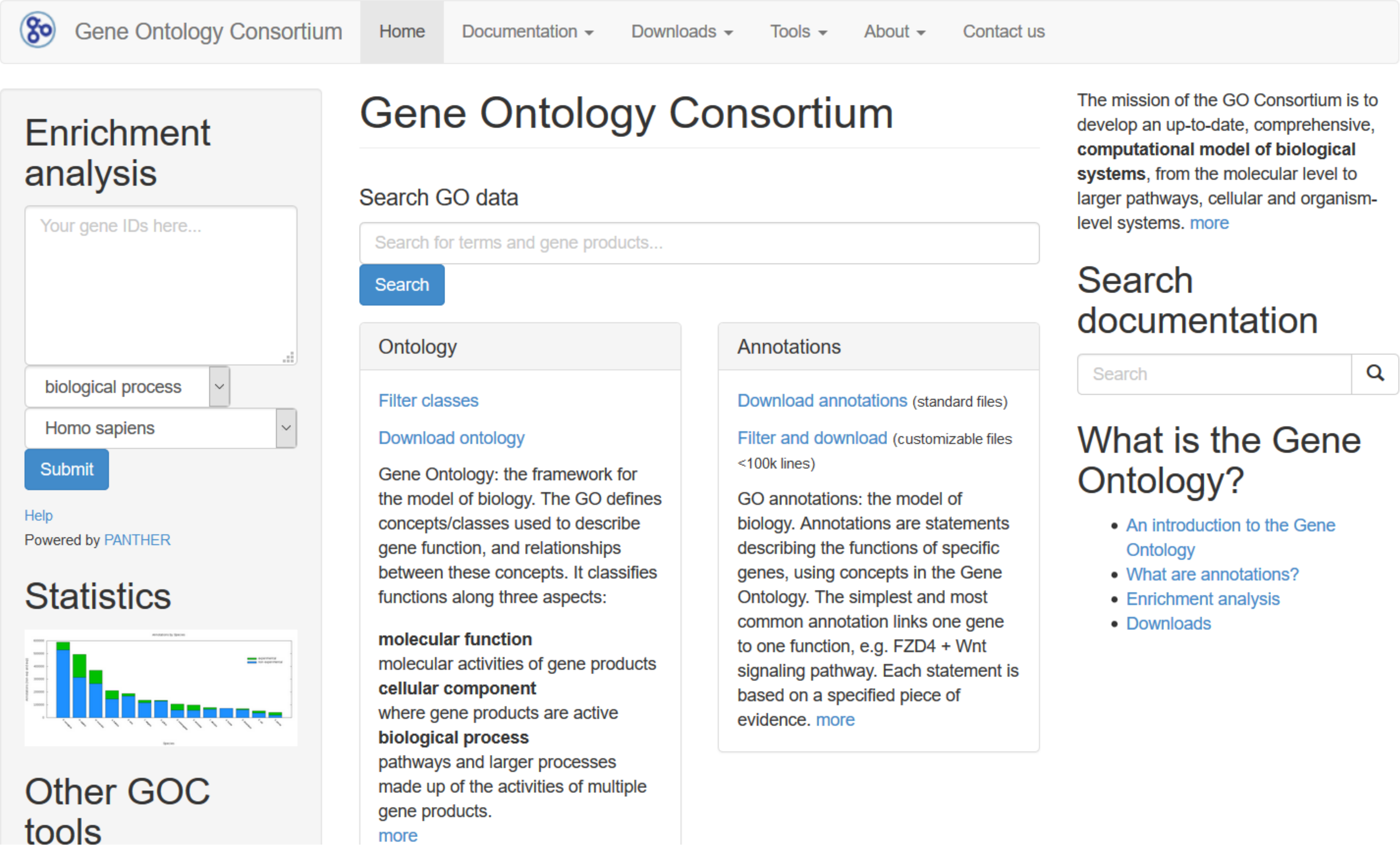
Most of these tools work in a similar way:

> input a gene list and a subset of ‘interesting’ genes

> tool shows which GO categories have most interesting genes associated with them i.e. which categories are ‘enriched’ for interesting genes

> tool provides a statistical measure to determine whether enrichment is significant

http://www.geneontology.org



Enrichment analysis

Powered by PANTHER

Statistics

Other GOC tools

Gene Ontology Consortium

Search GO data

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products
cellular component
where gene products are active
biological process
pathways and larger processes made up of the activities of multiple gene products.
more

Annotations

Download annotations (standard files)

Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. more

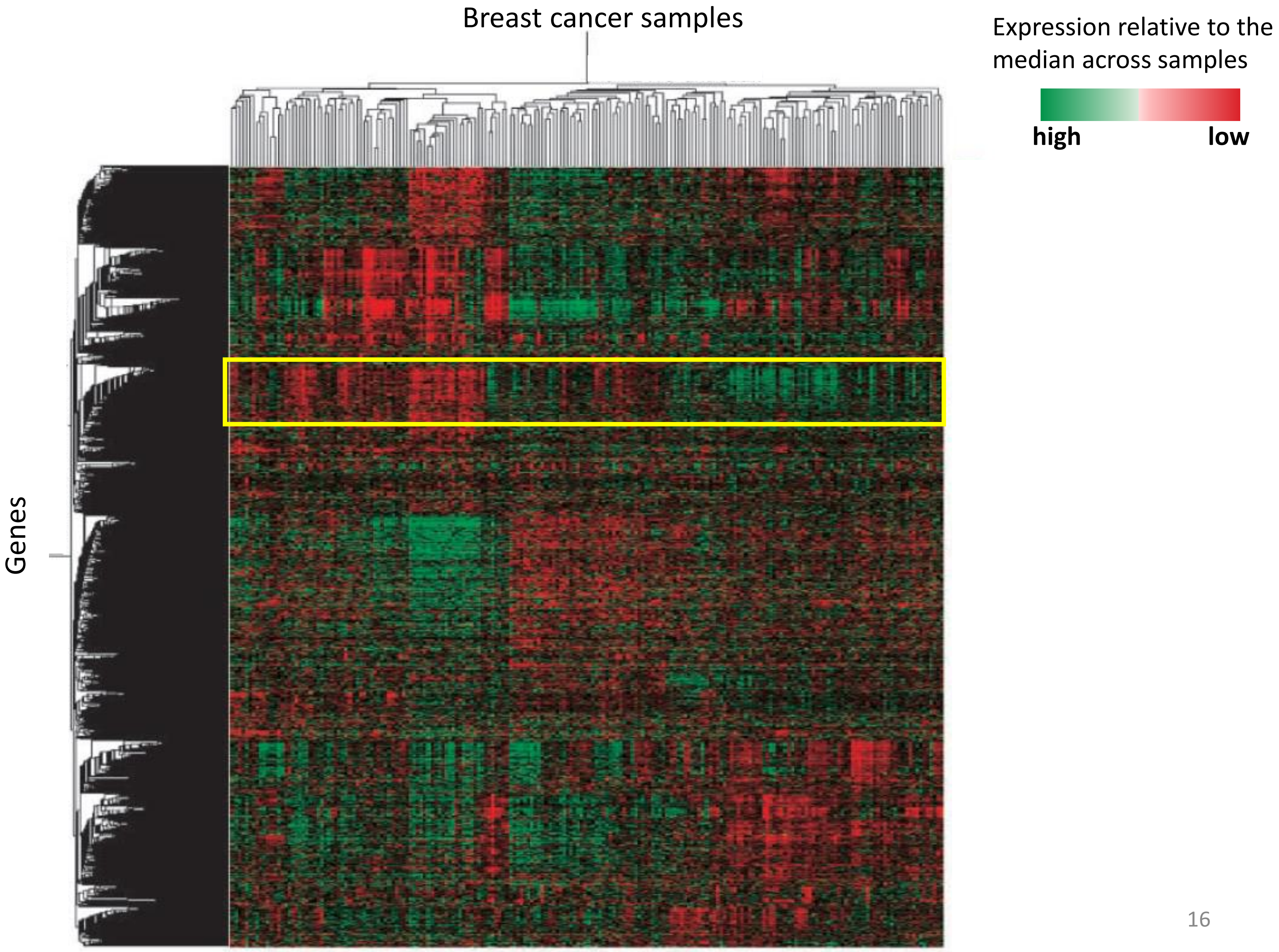
The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. more

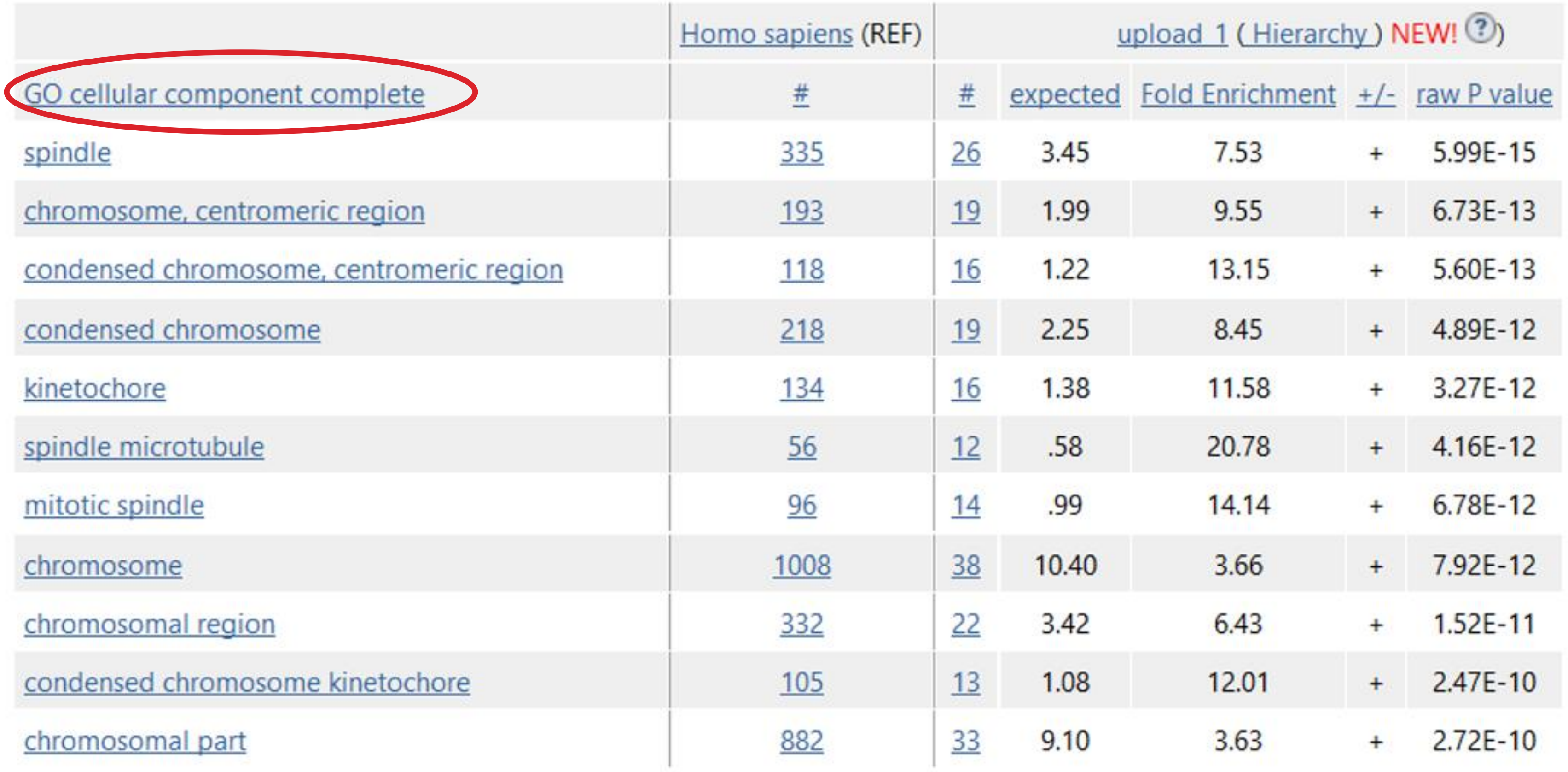
Search documentation

What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Enrichment analysis
- Downloads

Mainz data set from Schmidt et al.

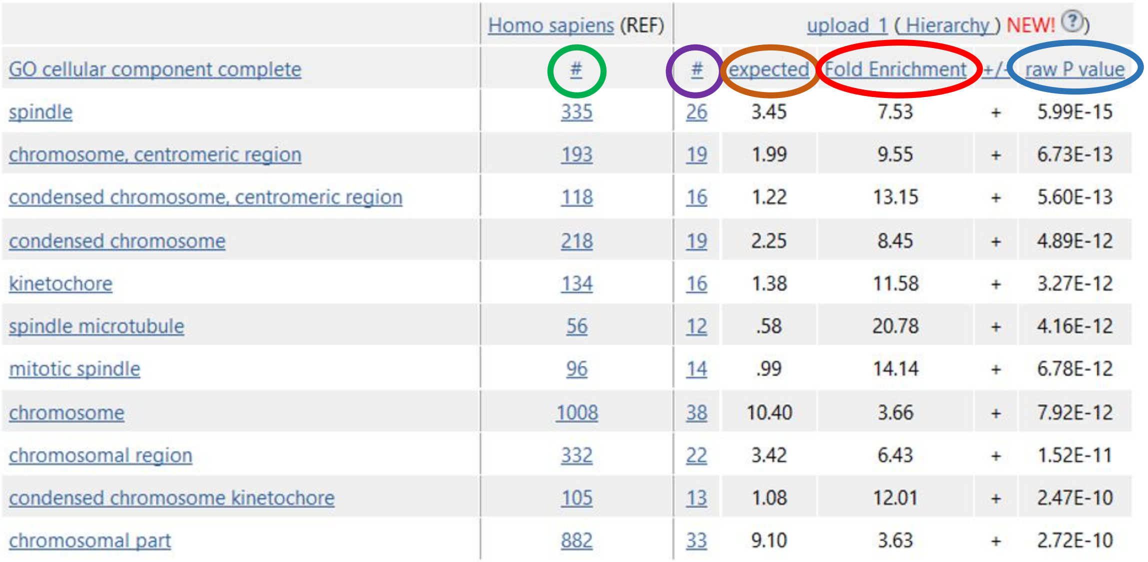




	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?)				
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value
spindle	335	26	3.45	7.53	+	5.99E-15
chromosome, centromeric region	193	19	1.99	9.55	+	6.73E-13
condensed chromosome, centromeric region	118	16	1.22	13.15	+	5.60E-13
condensed chromosome	218	19	2.25	8.45	+	4.89E-12
kinetochore	134	16	1.38	11.58	+	3.27E-12
spindle microtubule	56	12	.58	20.78	+	4.16E-12
mitotic spindle	96	14	.99	14.14	+	6.78E-12
chromosome	1008	38	10.40	3.66	+	7.92E-12
chromosomal region	332	22	3.42	6.43	+	1.52E-11
condensed chromosome kinetochore	105	13	1.08	12.01	+	2.47E-10
chromosomal part	882	33	9.10	3.63	+	2.72E-10

Interpreting the Results Table

List of **significant shared GO terms** used to describe the set of genes



	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?				
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value
spindle	335	26	3.45	7.53	+	5.99E-15
chromosome, centromeric region	193	19	1.99	9.55	+	6.73E-13
condensed chromosome, centromeric region	118	16	1.22	13.15	+	5.60E-13
condensed chromosome	218	19	2.25	8.45	+	4.89E-12
kinetochore	134	16	1.38	11.58	+	3.27E-12
spindle microtubule	56	12	.58	20.78	+	4.16E-12
mitotic spindle	96	14	.99	14.14	+	6.78E-12
chromosome	1008	38	10.40	3.66	+	7.92E-12
chromosomal region	332	22	3.42	6.43	+	1.52E-11
condensed chromosome kinetochore	105	13	1.08	12.01	+	2.47E-10
chromosomal part	882	33	9.10	3.63	+	2.72E-10

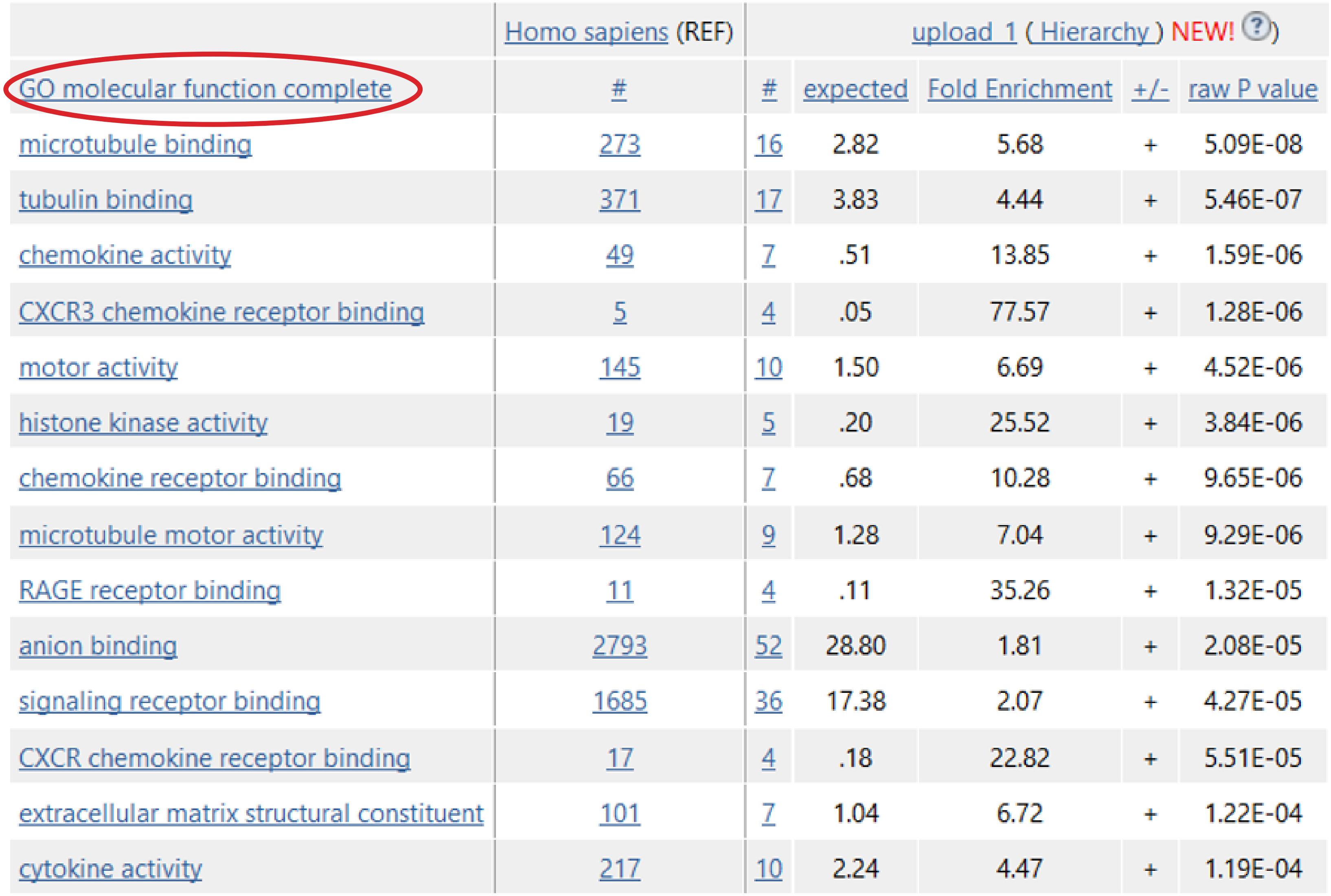
The number of genes annotated to a GO term in the entire background set,

The number of genes annotated to that GO term in the input list.

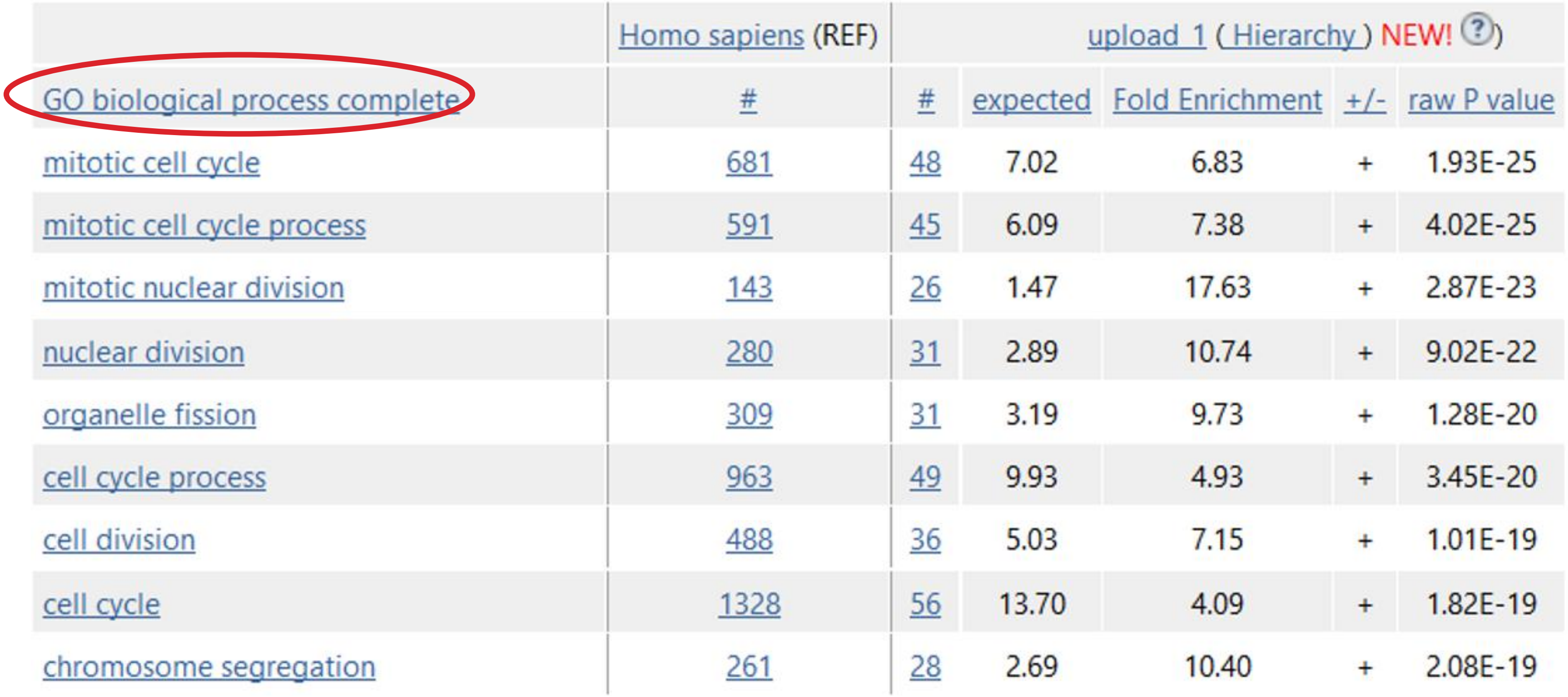
The number of genes expected in the input list for this category, based on the reference list.

The number of category genes observed in the uploaded list over the expected number.
> 1, category is overrepresented in your experiment.
< 1 the category is underrepresented.

This is the probability that the number of genes you observed in this category occurred by chance (randomly), as determined by your reference list.



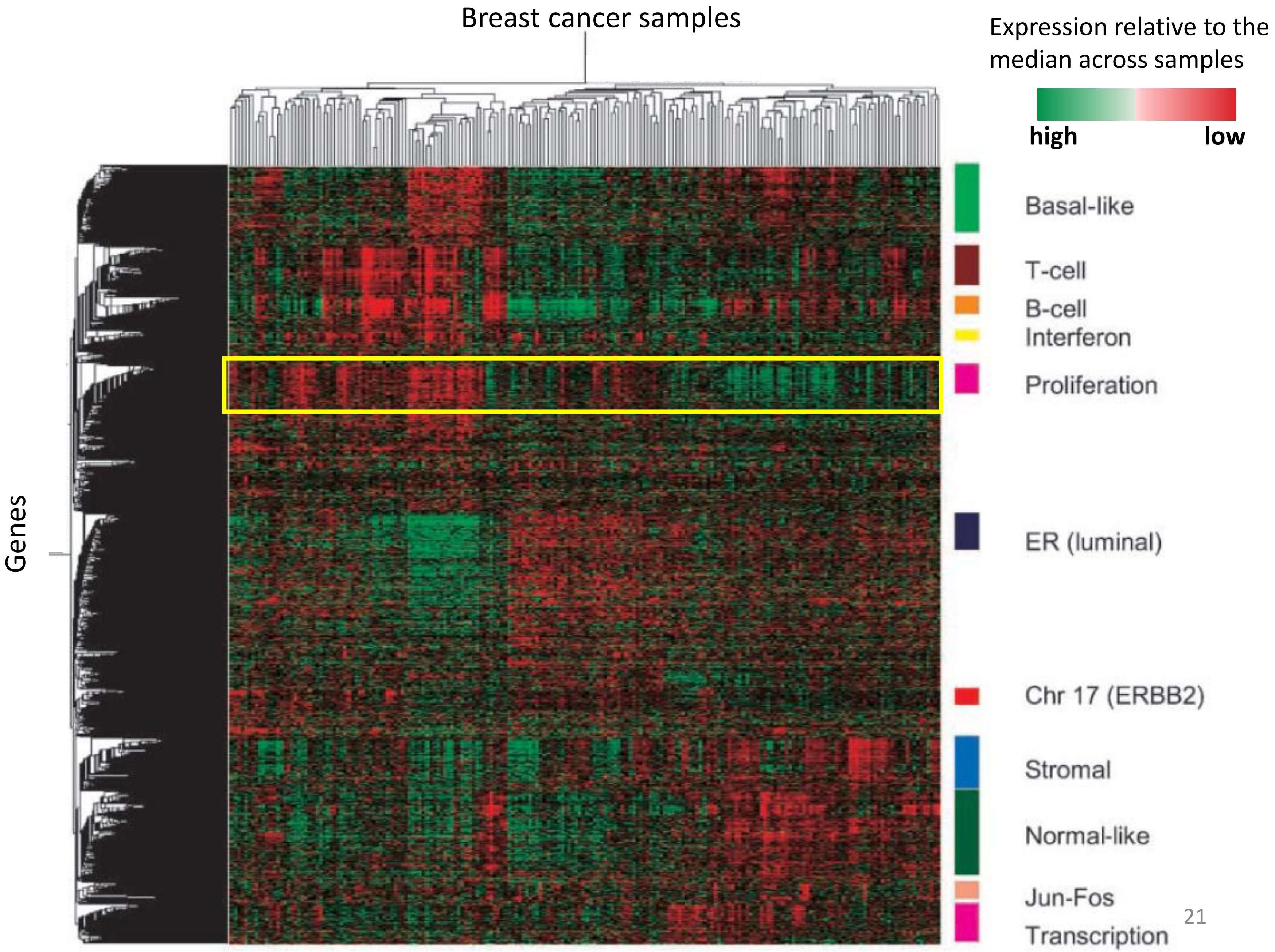
	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?)				
GO molecular function complete	#	#	expected	Fold Enrichment	+/-	raw P value
microtubule binding	273	16	2.82	5.68	+	5.09E-08
tubulin binding	371	17	3.83	4.44	+	5.46E-07
chemokine activity	49	7	.51	13.85	+	1.59E-06
CXCR3 chemokine receptor binding	5	4	.05	77.57	+	1.28E-06
motor activity	145	10	1.50	6.69	+	4.52E-06
histone kinase activity	19	5	.20	25.52	+	3.84E-06
chemokine receptor binding	66	7	.68	10.28	+	9.65E-06
microtubule motor activity	124	9	1.28	7.04	+	9.29E-06
RAGE receptor binding	11	4	.11	35.26	+	1.32E-05
anion binding	2793	52	28.80	1.81	+	2.08E-05
signaling receptor binding	1685	36	17.38	2.07	+	4.27E-05
CXCR chemokine receptor binding	17	4	.18	22.82	+	5.51E-05
extracellular matrix structural constituent	101	7	1.04	6.72	+	1.22E-04
cytokine activity	217	10	2.24	4.47	+	1.19E-04



	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?)				
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value
mitotic cell cycle	681	48	7.02	6.83	+	1.93E-25
mitotic cell cycle process	591	45	6.09	7.38	+	4.02E-25
mitotic nuclear division	143	26	1.47	17.63	+	2.87E-23
nuclear division	280	31	2.89	10.74	+	9.02E-22
organelle fission	309	31	3.19	9.73	+	1.28E-20
cell cycle process	963	49	9.93	4.93	+	3.45E-20
cell division	488	36	5.03	7.15	+	1.01E-19
cell cycle	1328	56	13.70	4.09	+	1.82E-19
chromosome segregation	261	28	2.69	10.40	+	2.08E-19

Knowledge about the molecular mechanisms involved in the processes of estrogen-dependent tumor growth and proliferative activity has led to the successful development of therapeutic concepts.

Mainz data set from Schmidt et al.



What do we need?

- A shared, consistent functional vocabulary
- Systematic linkage between genes and functions
- A way to determine which genes are relevant to the study condition
- Statistical analysis
- A way to identify a set of “related” genes we want to functionally annotate

What do we need?

- A shared, consistent functional vocabulary

  GO: Gene Ontology

- Systematic linkage between genes and functions

  GO annotation

- A way to determine which genes are relevant to the study condition

  Fold change, ranking

- Statistical analysis

  Enrichment analysis

- A way to identify a set of “related” genes we want to functionally annotate

  Exploratory data analysis

Function annotation of proteins

GENEONTOLOGY
Unifying Biology

STRING

REACTOME
A CURATED PATHWAY DATABASE

Pfam

UniProt

HumanNet
Probabilistic Functional Gene Network of *Homo sapiens*

KEGG
Kyoto Encyclopedia of Genes and Genomes

IntAct

SMART

RCSB PDB
PROTEIN DATA BANK

InterPro

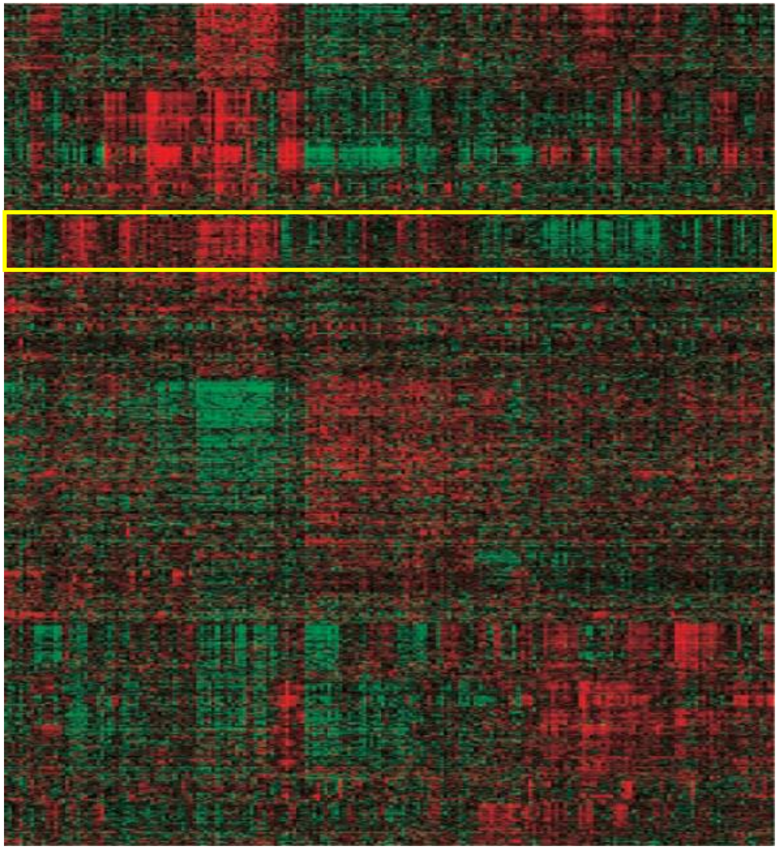
> **“The nice thing about standards is that there are so many to choose from”**
>
> Andrew S. Tanenbaum

Picking relevant genes

Significant differential expression

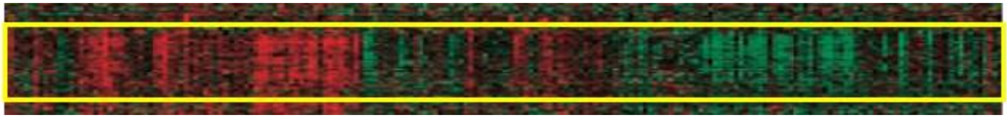
Fold change cutoff (e.g., > two fold change)

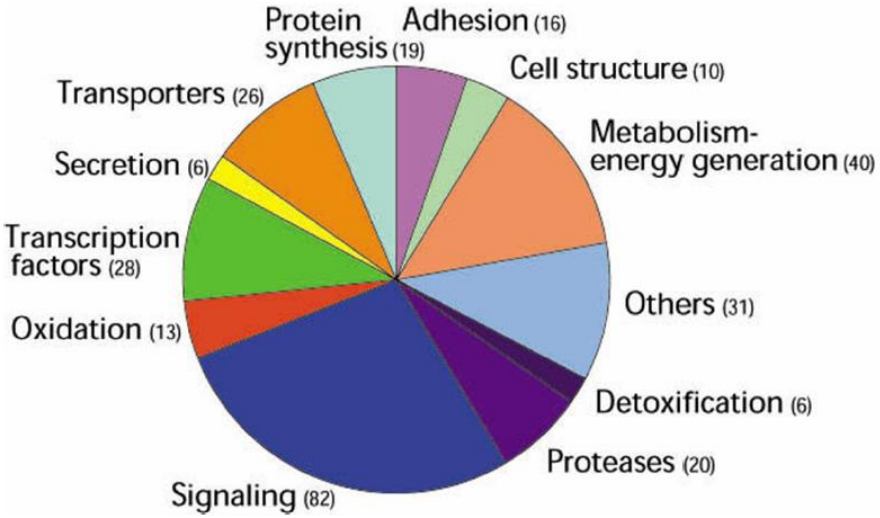
Fold change rank (e.g., top 10%)



Functional enrichment analysis

Study set of genes





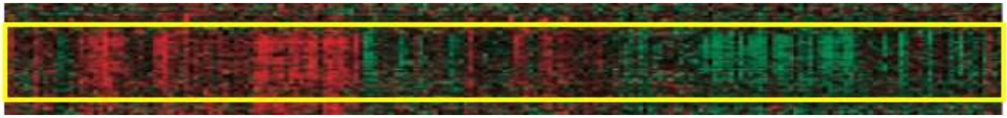
Functional category	# genes in study set	% in study set
Signaling	82	28%
Metabolism	40	14%
Other	31	10%
Trans factors	28	9%
Transporters	26	9%
Proteases	20	7%
Protein synthesis	19	7%
Adhesion	16	5%
Oxidation	13	4%
Cell structure	10	3%
Secretion	6	2%
Detoxification	6	2%

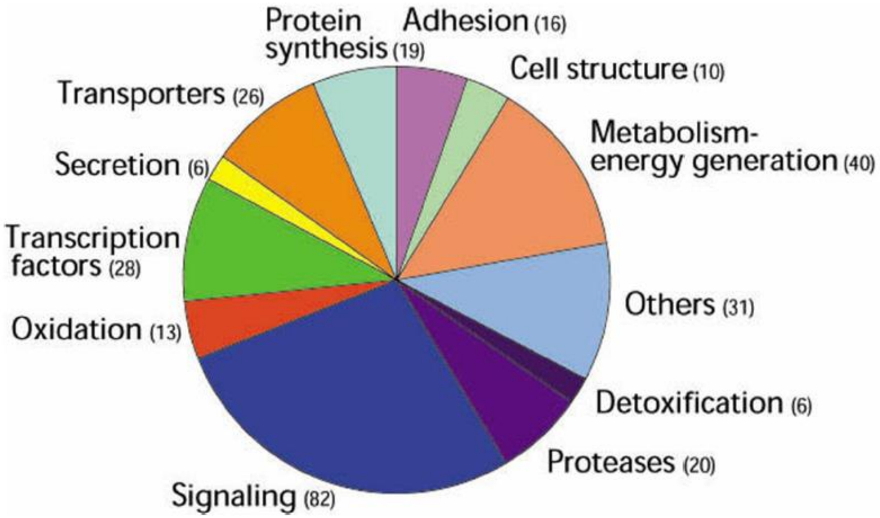
Largest category is Signaling: contains 27.6% of all genes in the study set

Conclude: Signaling may be important in the condition under study.

Functional enrichment analysis: The Wrong Way

Study set of genes





Functional category	# genes in study set	% in study set
Signaling	82	28%
Metabolism	40	14%
Other	31	10%
Trans factors	28	9%
Transporters	26	9%
Proteases	20	7%
Protein synthesis	19	7%
Adhesion	16	5%
Oxidation	13	4%
Cell structure	10	3%
Secretion	6	2%
Detoxification	6	2%

Largest category is Signaling: contains 27.6% of all genes in the study set

Conclude: Signaling may be important in the condition under study.

Functional enrichment analysis: A better way

What if ~27% of the genes on the array are involved in signaling?

What is the number of signaling genes in the set is what expected by chance?

Functional category	# genes in study set	% in study set	% on array
Signaling	82	28%	26%
Metabolism	40	14%	15%
Other	31	10%	11%
Trans factors	28	9%	10%
Transporters	26	9%	2%
Proteases	20	7%	7%
Protein synthesis	19	7%	7%
Adhesion	16	5%	6%
Oxidation	13	4%	4%
Cell structure	10	3%	8%
Secretion	6	2%	2%
Detoxification	6	2%	2%

Which categories are **enriched and over-represented?**

We need to consider not only the number of genes in the set for each category, but also the total number on the array.

Functional enrichment analysis: A better way

Categories that **depleted and under-represented** are important, too.

Functional category	# genes in study set	% in study set	% on array
Signaling	82	28%	26%
Metabolism	40	14%	15%
Other	31	10%	11%
Trans factors	28	9%	10%
Transporters	26	9%	2%
Proteases	20	7%	7%
Protein synthesis	19	7%	7%
Adhesion	16	5%	6%
Oxidation	13	4%	4%
Cell structure	10	3%	8%
Secretion	6	2%	2%
Detoxification	6	2%	2%

Suggests that maintenance of normal cell structure is not necessary or impaired.

Functional enrichment analysis: An even better way

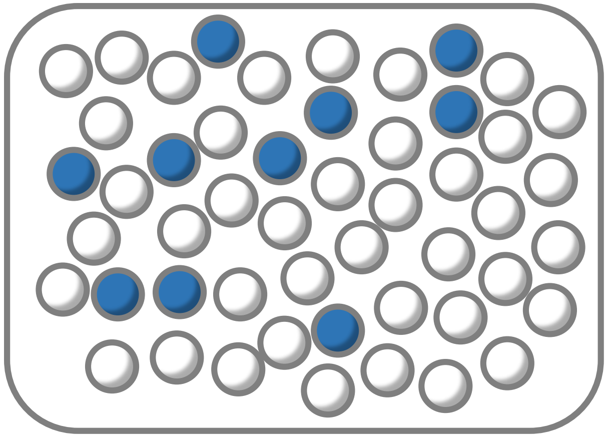
Assume the study set has nothing to do with the specific function at hand and was selected randomly, *would we be surprised to see a certain number of genes annotated with this function?*

Need a statistical test based on a null model

The “urn” version: You pick a set of 8 balls from an urn that contains 50 white and blue balls. How surprised will you be to find that 4 of the balls you picked are blue?

Functional enrichment analysis: An even better way

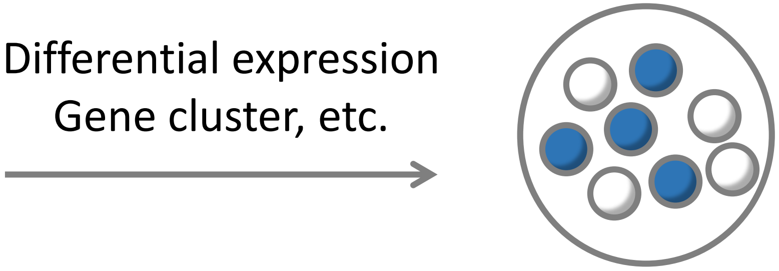
Genes (balls) in our experiment



10 out of 50

Differential expression
Gene cluster, etc.

Study set

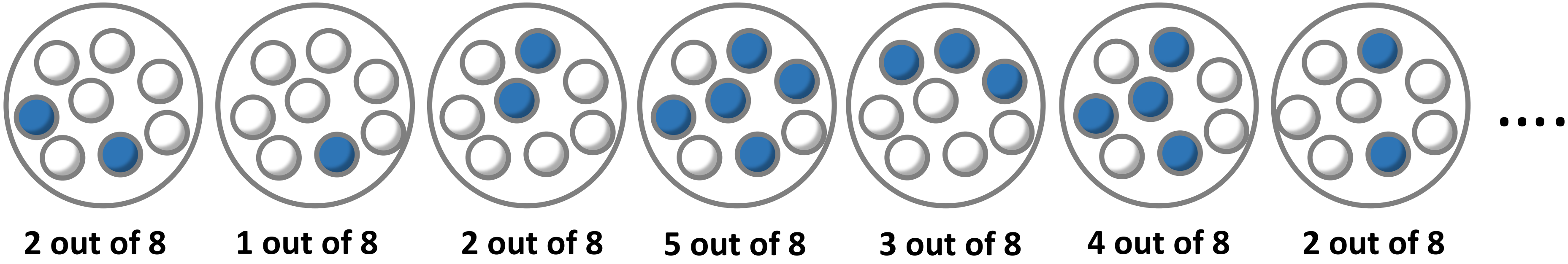


4 out of 8

Do I have a surprisingly high number of blue genes, e.g. annotated as “signaling”?

Null model: The 8 genes (balls) are selected randomly

50 balls, 10 of which are blue. Pick 8 at random; what is the probability that *k* are blue?

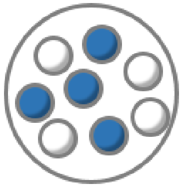


Modified Fisher’s exact test

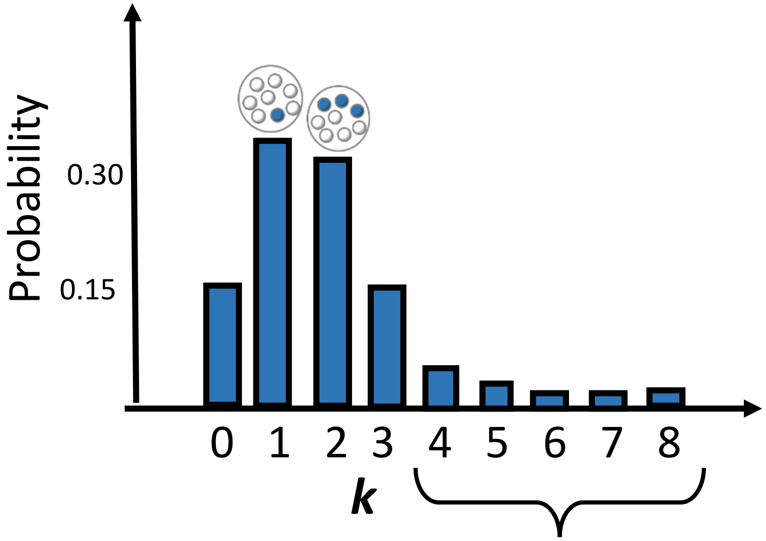
50 balls in experiment, 10 of which are blue.

8 balls in study set, 4 of which are blue.

Pick 8 at random; what is the probability that *k* balls in the study set are blue?



Do I have a surprisingly high number of blue genes, e.g. annotated as “signaling”?



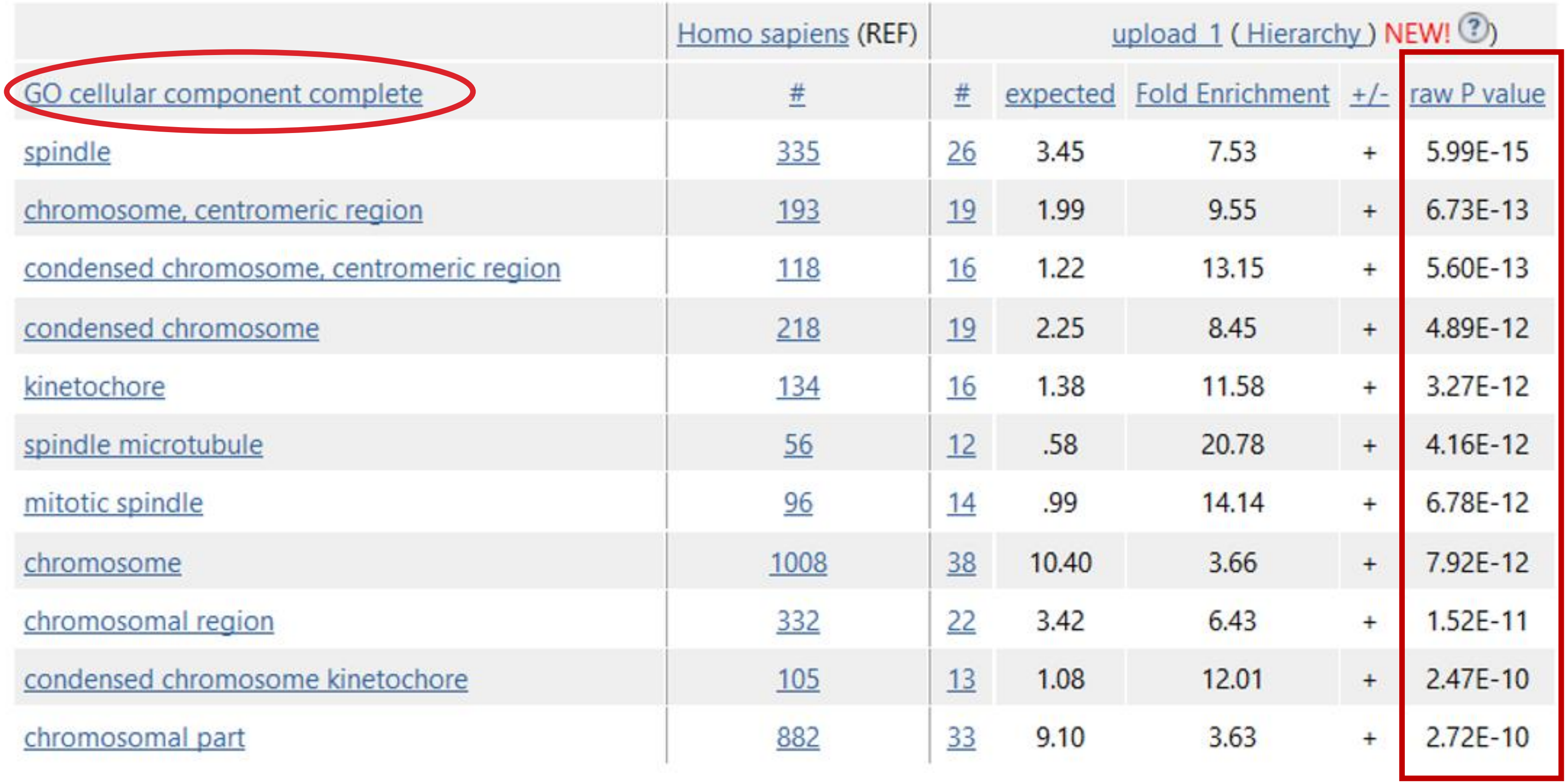
What is the probability of getting at least 4 blue genes in the null model?

$P(\sigma_t >= 4)$

Hypergeometric distribution

$$\boxed{\mathbb{P}(\sigma_t = k) = \frac{\binom{m_t}{k}\binom{m-m_t}{n-k}}{\binom{m}{n}}}$$

$m=50, m_t=10, n=8, \sigma_t = 4$



	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?)				
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value
spindle	335	26	3.45	7.53	+	5.99E-15
chromosome, centromeric region	193	19	1.99	9.55	+	6.73E-13
condensed chromosome, centromeric region	118	16	1.22	13.15	+	5.60E-13
condensed chromosome	218	19	2.25	8.45	+	4.89E-12
kinetochore	134	16	1.38	11.58	+	3.27E-12
spindle microtubule	56	12	.58	20.78	+	4.16E-12
mitotic spindle	96	14	.99	14.14	+	6.78E-12
chromosome	1008	38	10.40	3.66	+	7.92E-12
chromosomal region	332	22	3.42	6.43	+	1.52E-11
condensed chromosome kinetochore	105	13	1.08	12.01	+	2.47E-10
chromosomal part	882	33	9.10	3.63	+	2.72E-10

KEGG PATHWAY database: Kyoto Encyclopedia of Genes and Genomes

http://www.kegg.com

Consolidated set of databases that cover

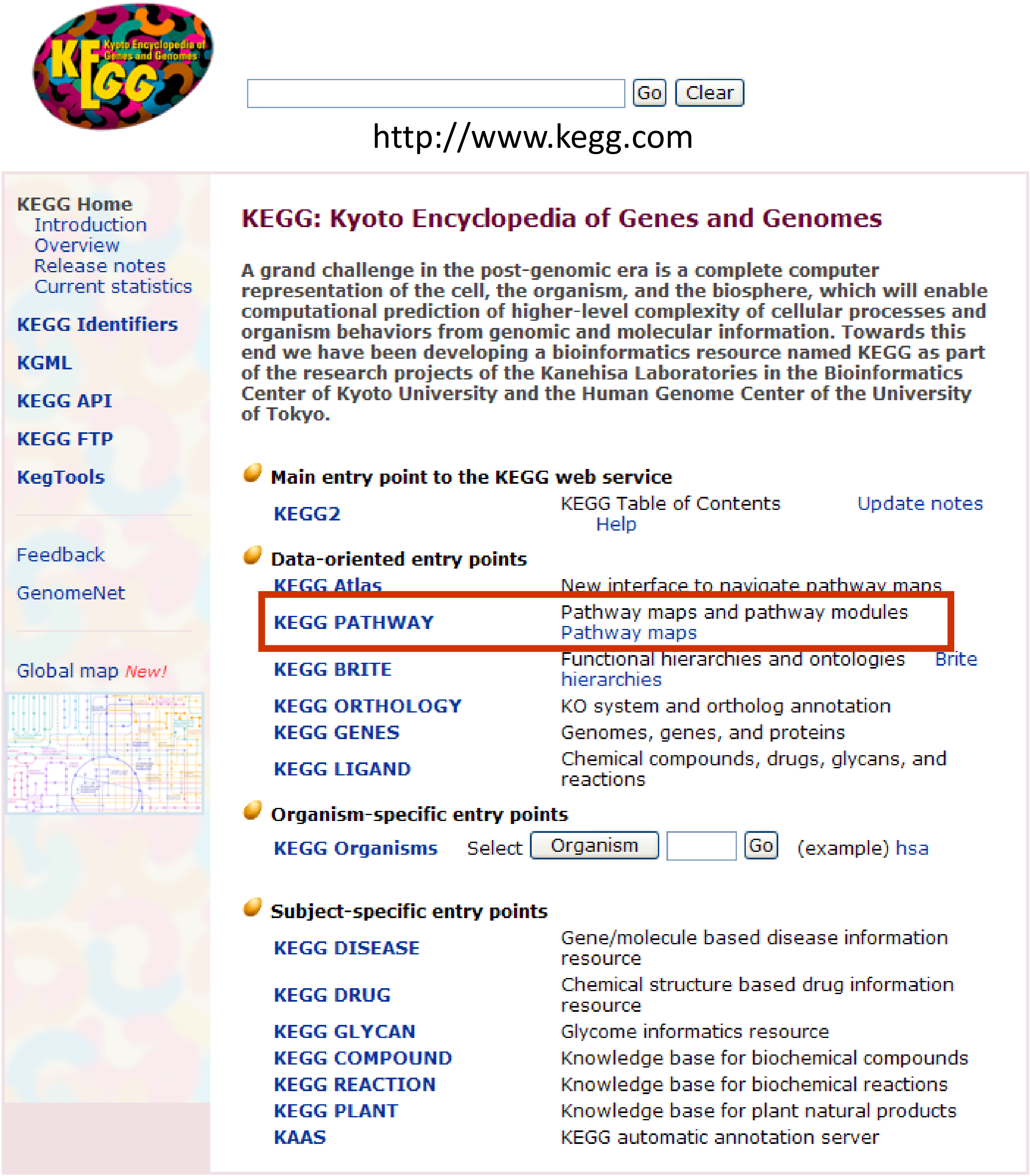
genomics (GENE),

chemical compounds (LIGAND), and

reaction networks (PATHWAY)

Broad focus on metabolics, signal transduction, disease, *etc*.

Species-specific views available



http://www.kegg.com

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents Help	Update notes

Data-oriented entry points

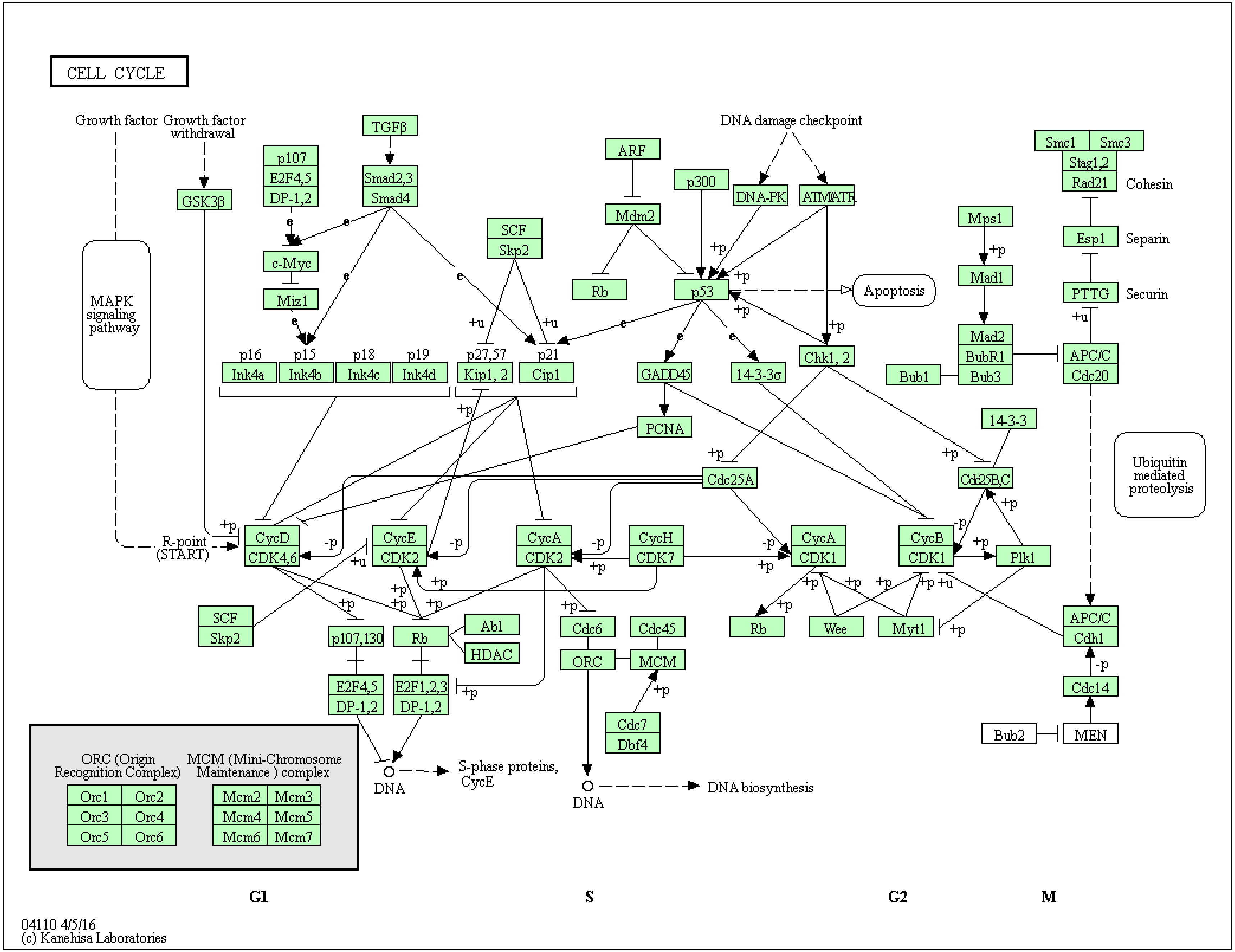
KEGG Atlas	New interface to navigate pathway maps
KEGG PATHWAY	Pathway maps and pathway modules Pathway maps
KEGG BRITE	Functional hierarchies and ontologies Brite hierarchies
KEGG ORTHOLOGY	KO system and ortholog annotation
KEGG GENES	Genomes, genes, and proteins
KEGG LIGAND	Chemical compounds, drugs, glycans, and reactions

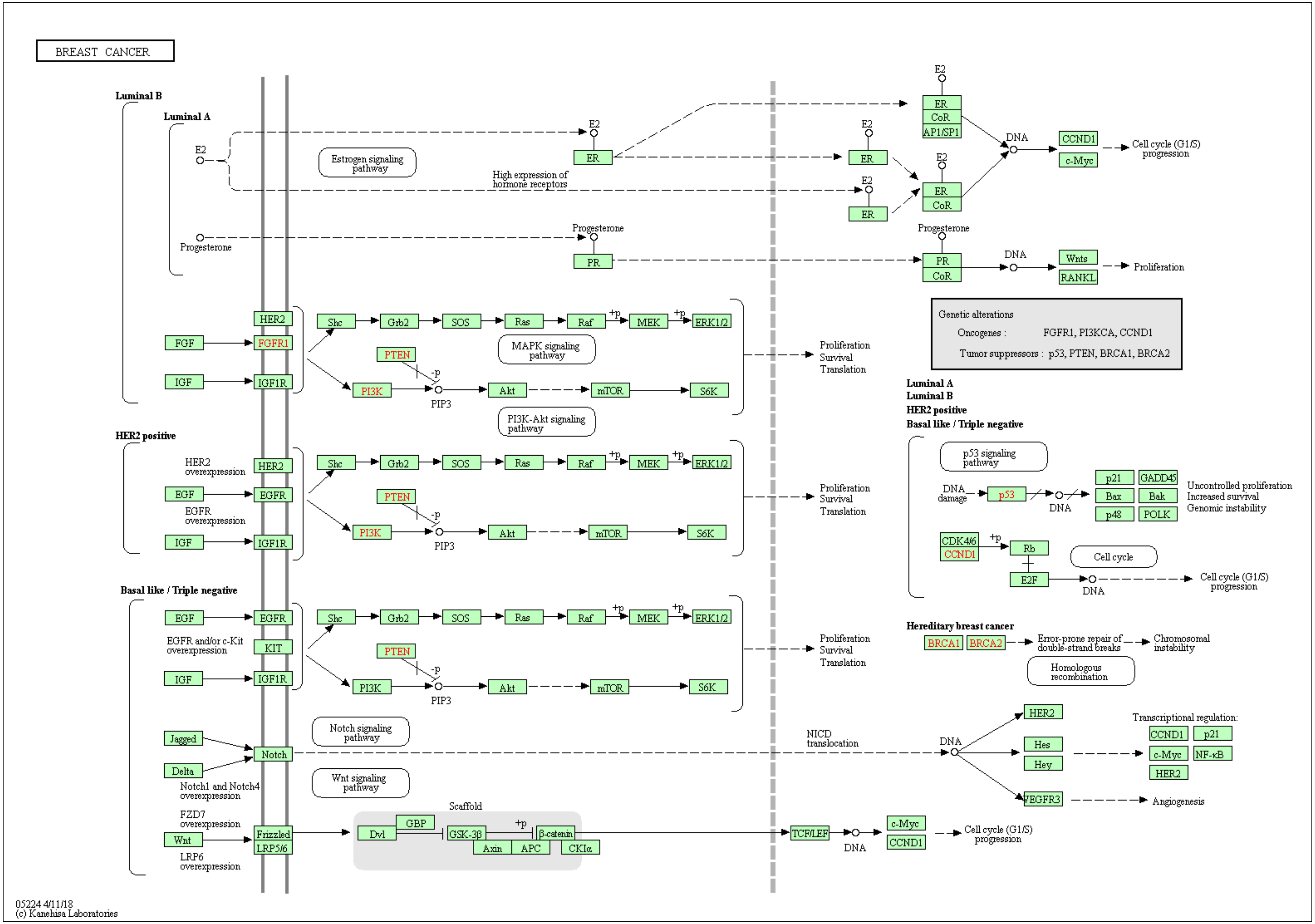
Organism-specific entry points

KEGG Organisms Select Organism Go (example) hsa

Subject-specific entry points

KEGG DISEASE	Gene/molecule based disease information resource
KEGG DRUG	Chemical structure based drug information resource
KEGG GLYCAN	Glycome informatics resource
KEGG COMPOUND	Knowledge base for biochemical compounds
KEGG REACTION	Knowledge base for biochemical reactions
KEGG PLANT	Knowledge base for plant natural products
KAAS	KEGG automatic annotation server

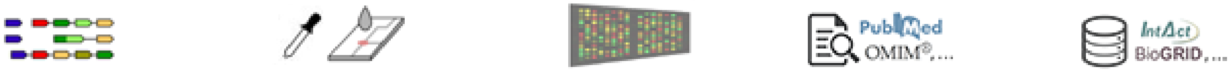

CELL CYCLE
Growth factor
Growth factor withdrawal
GSK3β
MAPK signaling pathway
p107
E2F4,5
DP-1,2
TGFβ
Smad2,3
Smad4
c-Myc
Miz1
SCF
Skp2
ARF
Mdm2
Rb
p300
DNA damage checkpoint
DNA-PK
ATM/ATR
p53
Apoptosis
Mps1
Mad1
Mad2
BubR1
Bub3
Bub1
Smc1
Smc3
Stag1,2
Rad21
Cohesin
Esp1
Separin
PTTG
Securin
APC/C
Cdc20
p16
Ink4a
p15
Ink4b
p18
Ink4c
p19
Ink4d
p27,57
Kip1, 2
p21
Cip1
GADD45
14-3-3σ
Chk1, 2
PCNA
14-3-3
Cdc25A
Cdc25B,C
Ubiquitin mediated proteolysis
R-point (START)
CycD
CDK4,6
CycE
CDK2
CycA
CDK2
CycH
CDK7
CycA
CDK1
CycB
CDK1
Plk1
SCF
Skp2
p107,130
Rb
Abl
HDAC
Cdc6
Cdc45
ORC
MCM
Rb
Wee
Myt1
APC/C
Cdh1
E2F4,5
DP-1,2
E2F1,2,3
DP-1,2
Cdc7
Dbf4
Cdc14
Bub2
MEN
ORC (Origin Recognition Complex)
MCM (Mini-Chromosome Maintenance) complex
Orc1
Orc2
Orc3
Orc4
Orc5
Orc6
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7
DNA
S-phase proteins, CycE
DNA
DNA biosynthesis
+p
-p
+u
e
G1
S
G2
M
04110 4/5/16
(c) Kanehisa Laboratories


BREAST CANCER
Luminal B
Luminal A
E2
Estrogen signaling pathway
E2
ER
High expression of hormone receptors
E2
ER
E2
ER
E2
ER
CoR
AP1/SP1
E2
ER
CoR
DNA
CCND1
c-Myc
Cell cycle (G1/S) progression
Progesterone
Progesterone
PR
Progesterone
PR
CoR
DNA
Wnts
RANKL
Proliferation
HER2
FGF
FGFR1
IGF
IGF1R
Shc
Grb2
SOS
Ras
Raf
+p
MEK
+p
ERK1/2
PTEN
-p
MAPK signaling pathway
PI3K
PIP3
Akt
mTOR
S6K
PI3K-Akt signaling pathway
Proliferation
Survival
Translation
Genetic alterations
Oncogenes : FGFR1, PI3KCA, CCND1
Tumor suppressors : p53, PTEN, BRCA1, BRCA2
HER2 positive
HER2 overexpression
HER2
EGF
EGFR
EGFR overexpression
IGF
IGF1R
Shc
Grb2
SOS
Ras
Raf
+p
MEK
+p
ERK1/2
PTEN
-p
PI3K
PIP3
Akt
mTOR
S6K
Proliferation
Survival
Translation
Luminal A
Luminal B
HER2 positive
Basal like / Triple negative
p53 signaling pathway
DNA damage
p53
DNA
p21
GADD45
Bax
Bak
p48
POLK
Uncontrolled proliferation
Increased survival
Genomic instability
CDK4/6
CCND1
+p
Rb
E2F
Cell cycle
DNA
Cell cycle (G1/S) progression
Basal like / Triple negative
EGF
EGFR
EGFR and/or c-Kit overexpression
KIT
IGF
IGF1R
Shc
Grb2
SOS
Ras
Raf
+p
MEK
+p
ERK1/2
PTEN
-p
PI3K
PIP3
Akt
mTOR
S6K
Proliferation
Survival
Translation
Hereditary breast cancer
BRCA1
BRCA2
Error-prone repair of double-strand breaks
Chromosomal instability
Homologous recombination
Notch signaling pathway
Jagged
Notch
Delta
Notch1 and Notch4 overexpression
NICD translocation
DNA
HER2
Hes
Hey
VEGFR3
Transcriptional regulation:
CCND1
p21
c-Myc
NF-κB
HER2
Angiogenesis
Wnt signaling pathway
FZD7 overexpression
Wnt
Frizzled
LRP5/6
LRP6 overexpression
Scaffold
GBP
Dvl
GSK-3β
+p
β-catenin
Axin
APC
CKIα
TCF/LEF
DNA
c-Myc
CCND1
Cell cycle (G1/S) progression
05224 4/11/18
(c) Kanehisa Laboratories

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:



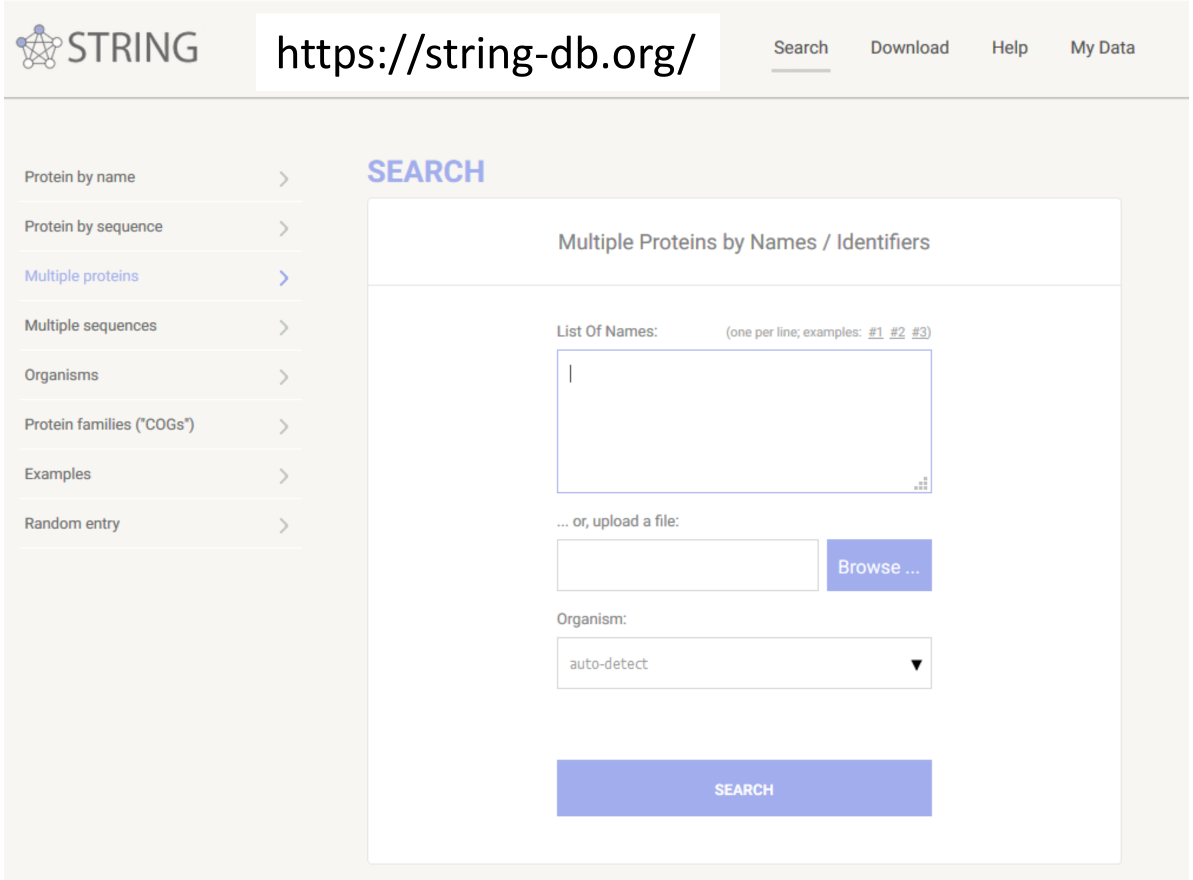
Genomic Context Predictions

High-throughput Lab Experiments

(Conserved) Co-Expression

Automated Textmining

Previous Knowledge in Databases



https://string-db.org/

SEARCH

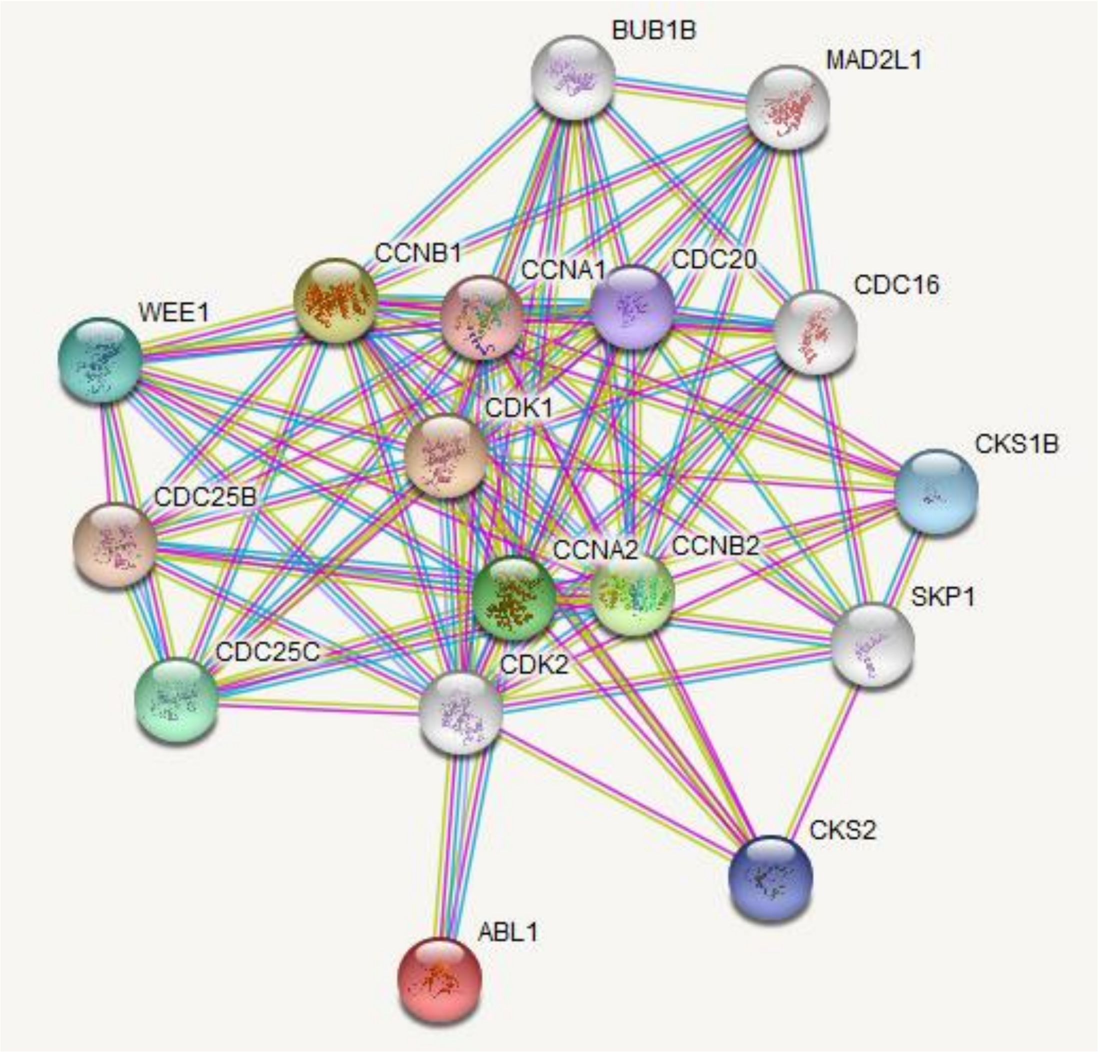
Multiple Proteins by Names / Identifiers

List Of Names: (one per line; examples: #1 #2 #3)

... or, upload a file:

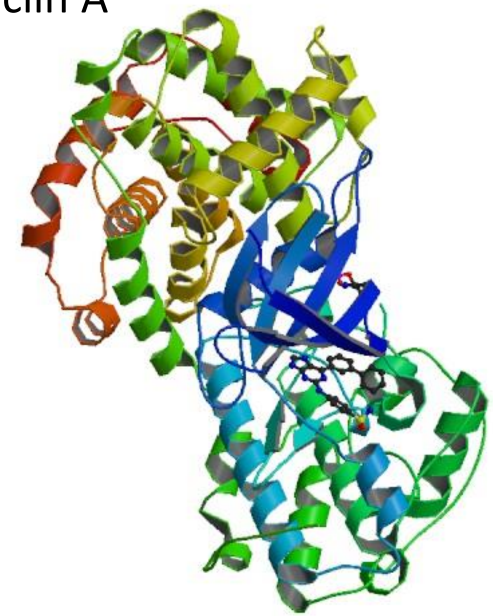
Organism: auto-detect

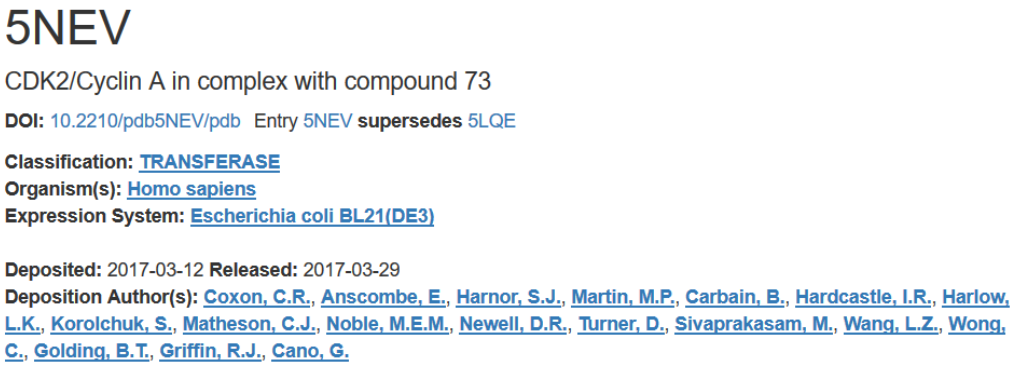
STRING network of functional protein interactions for ABL1 and CDK2



RCSB PDB PROTEIN DATA BANK 141842 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

http://www.rcsb.org/structure/5NEV





5NEV

CDK2/Cyclin A in complex with compound 73

DOI: 10.2210/pdb5NEV/pdb Entry 5NEV **supersedes** 5LQE

Classification: TRANSFERASE
Organism(s): Homo sapiens
Expression System: Escherichia coli BL21(DE3)

Deposited: 2017-03-12 **Released:** 2017-03-29
Deposition Author(s): Coxon, C.R., Anscombe, E., Harnor, S.J., Martin, M.P., Carbain, B., Hardcastle, I.R., Harlow, L.K., Korolchuk, S., Matheson, C.J., Noble, M.E.M., Newell, D.R., Turner, D., Sivaprakasam, M., Wang, L.Z., Wong, C., Golding, B.T., Griffin, R.J., Cano, G.

Thursday:

1) PSON cell line data: Expression and motility

2) TCGA brca data

Classification

Over/under-representation analysis

Statistical tests.

Gene set enrichment analysis