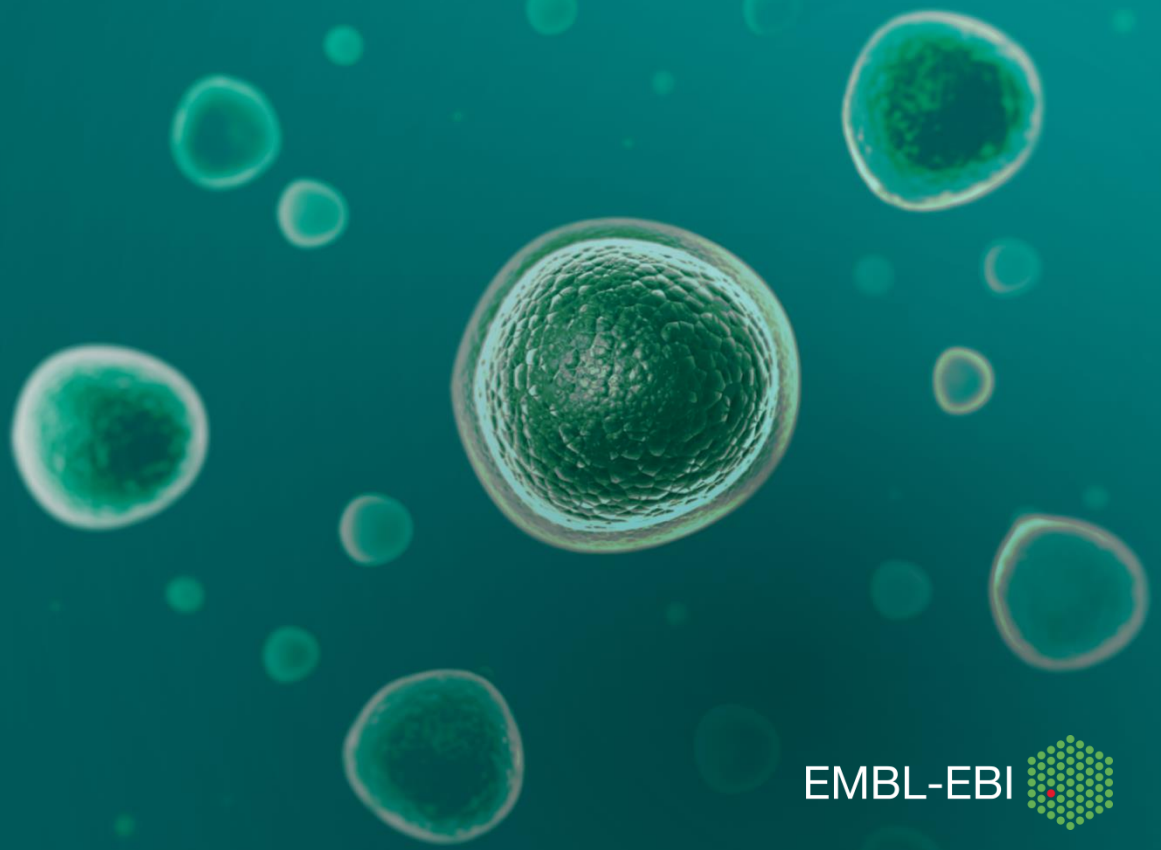
Exploring the Proteomics Services

Sandra Orchard



EMBL-EBI

Why study proteins

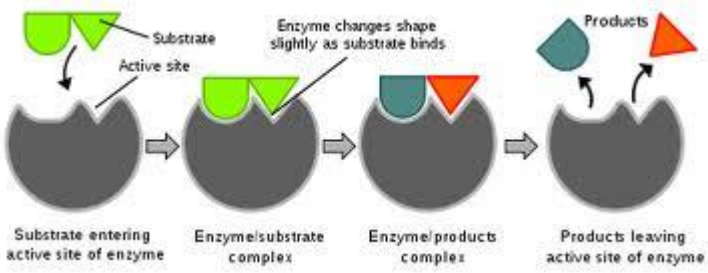
- Proteins are the work horses of the cell – and it is proteins which are largely the targets of therapeutic agents

- Knowing what proteins are being expressed in your healthy cells and how this alters in disease is critical to understanding the disease process

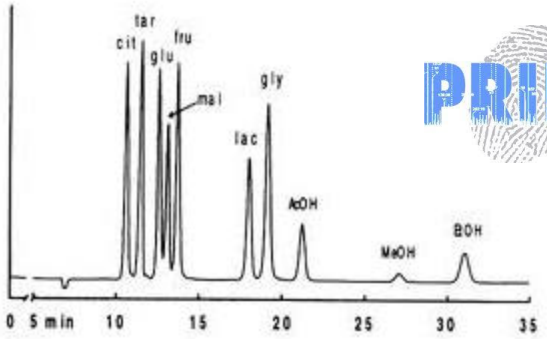
- Proteins cannot be studied in isolation – they need to be seen in the context of the network of interactions they make in the cell

MGCGCSSHPEDDWMENIDVCENCHYPIVPLDGKGTLLIR
NGSEVRDPLVTYEGSNPPASP LQDNLVIALHSYEPSHDG
DLGFEKGEQLRILEQSGEWWKAQSLTTGQEGFIPFNFVA
KAN SLEPEPWFFKNLSRKDAERQLLAPGNTHGSFLIRES

Protein Function

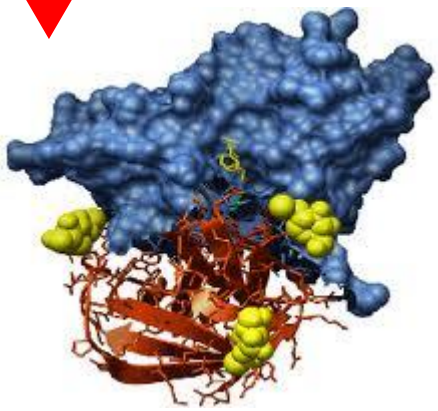


Protein identification



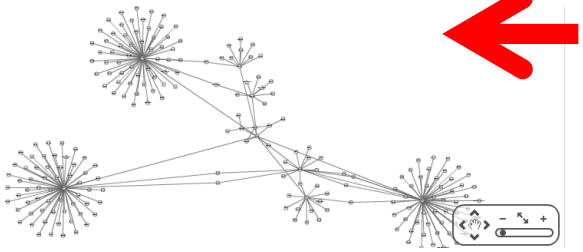
PRIDE

UniProt



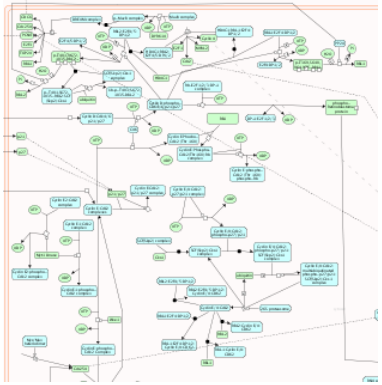
Protein Structure

PDBe
EMBL-EBI
PROTEIN DATA BANK EUROPE



IntAct

Network analysis

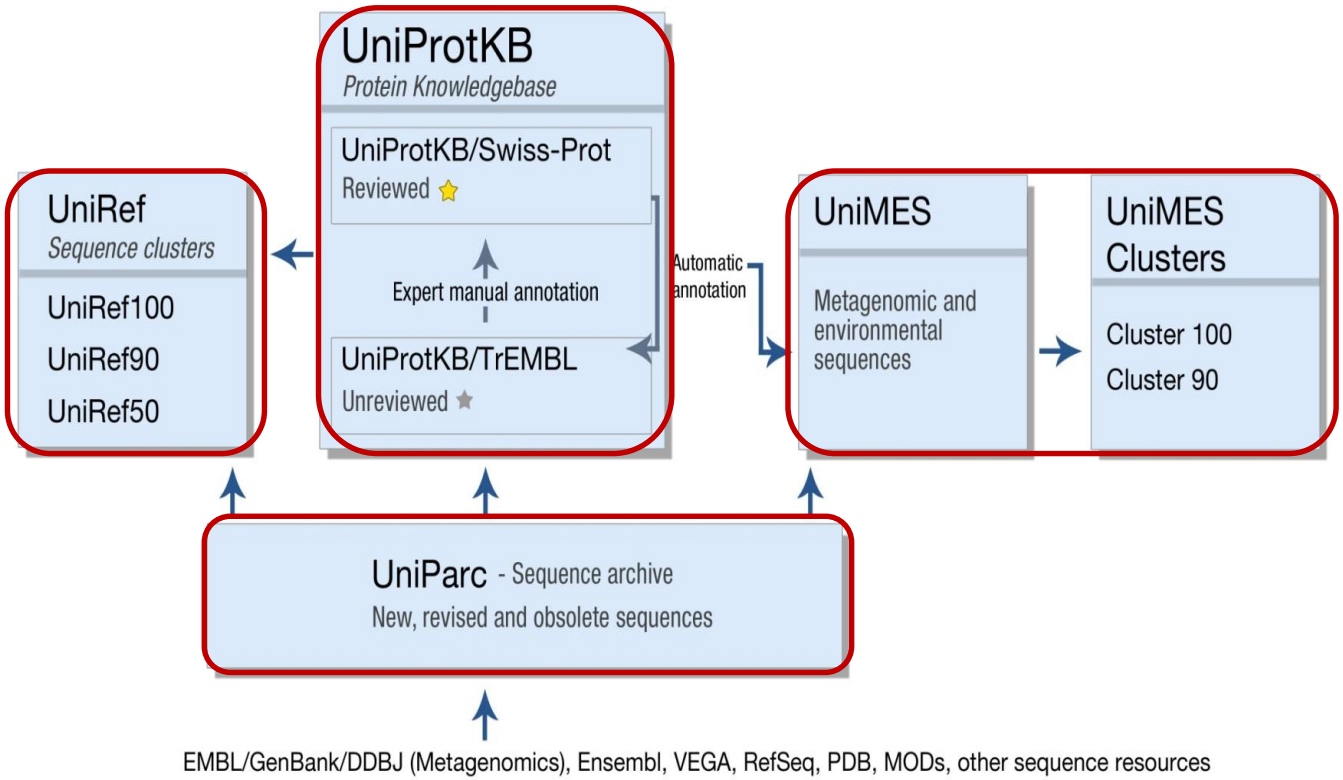


REACTOME

Pathway

UniProt mission

- Provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information



UniProt Knowledgebase (UniProtKB)

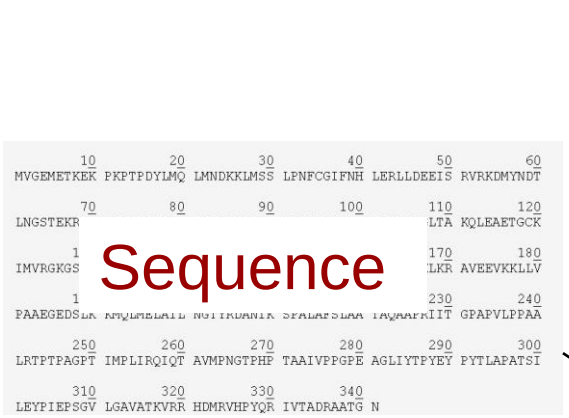
UniProtKB/Swiss-Prot

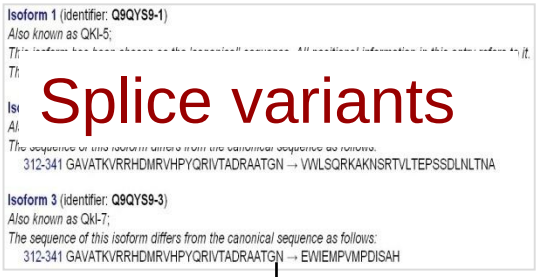
Non-redundant
High-quality
Manually curated
Reviewed

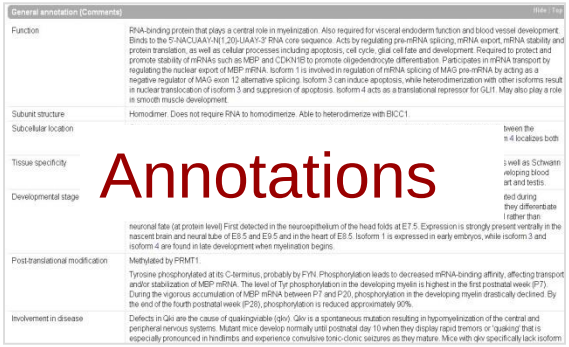
UniProtKB/TrEMBL

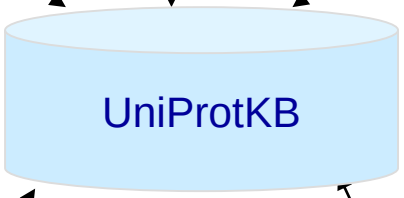
Redundant
Computationally generated
Automatically annotated
Unreviewed

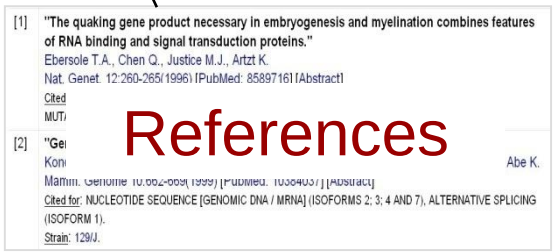
Manual annotation of UniProtKB/Swiss-Prot



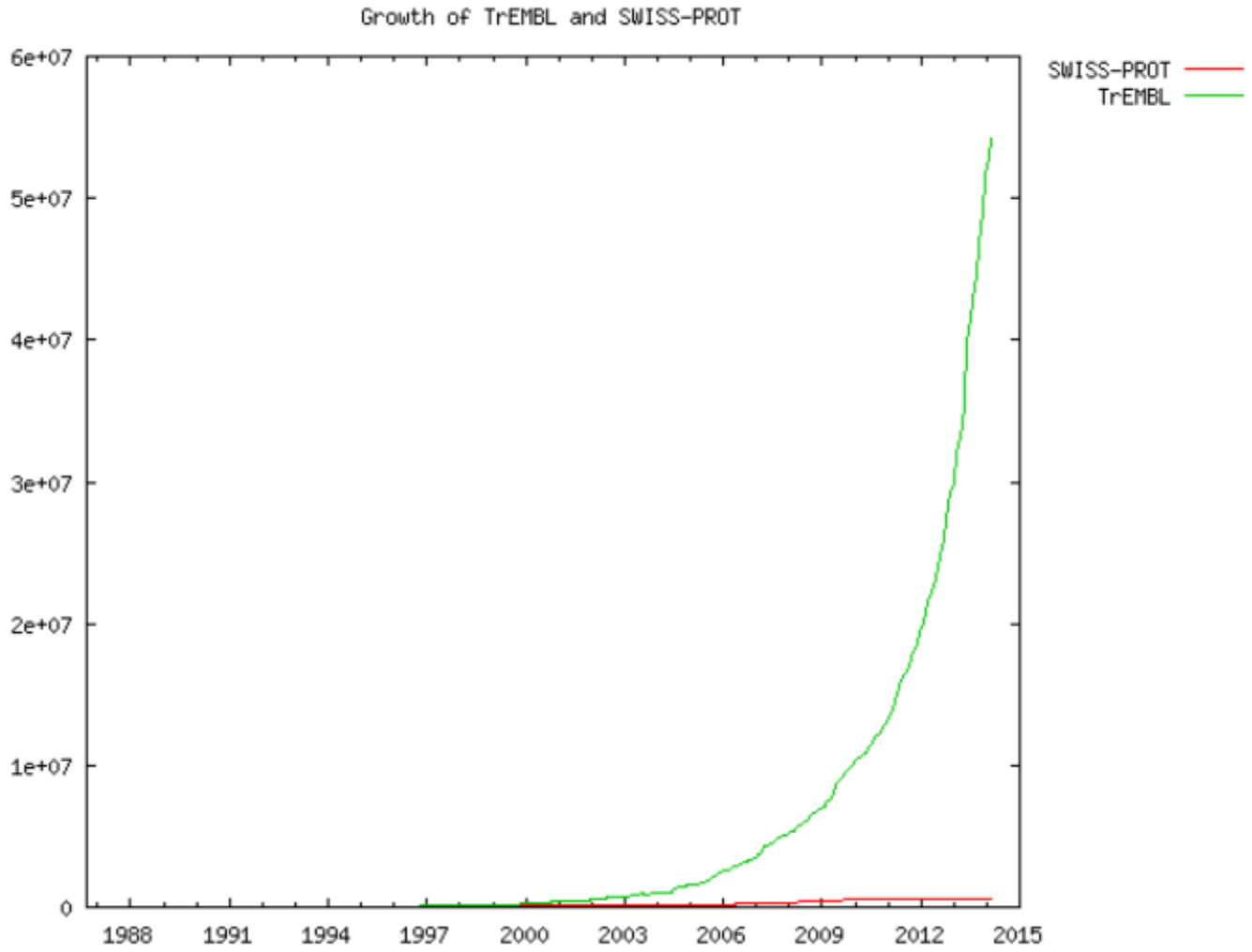








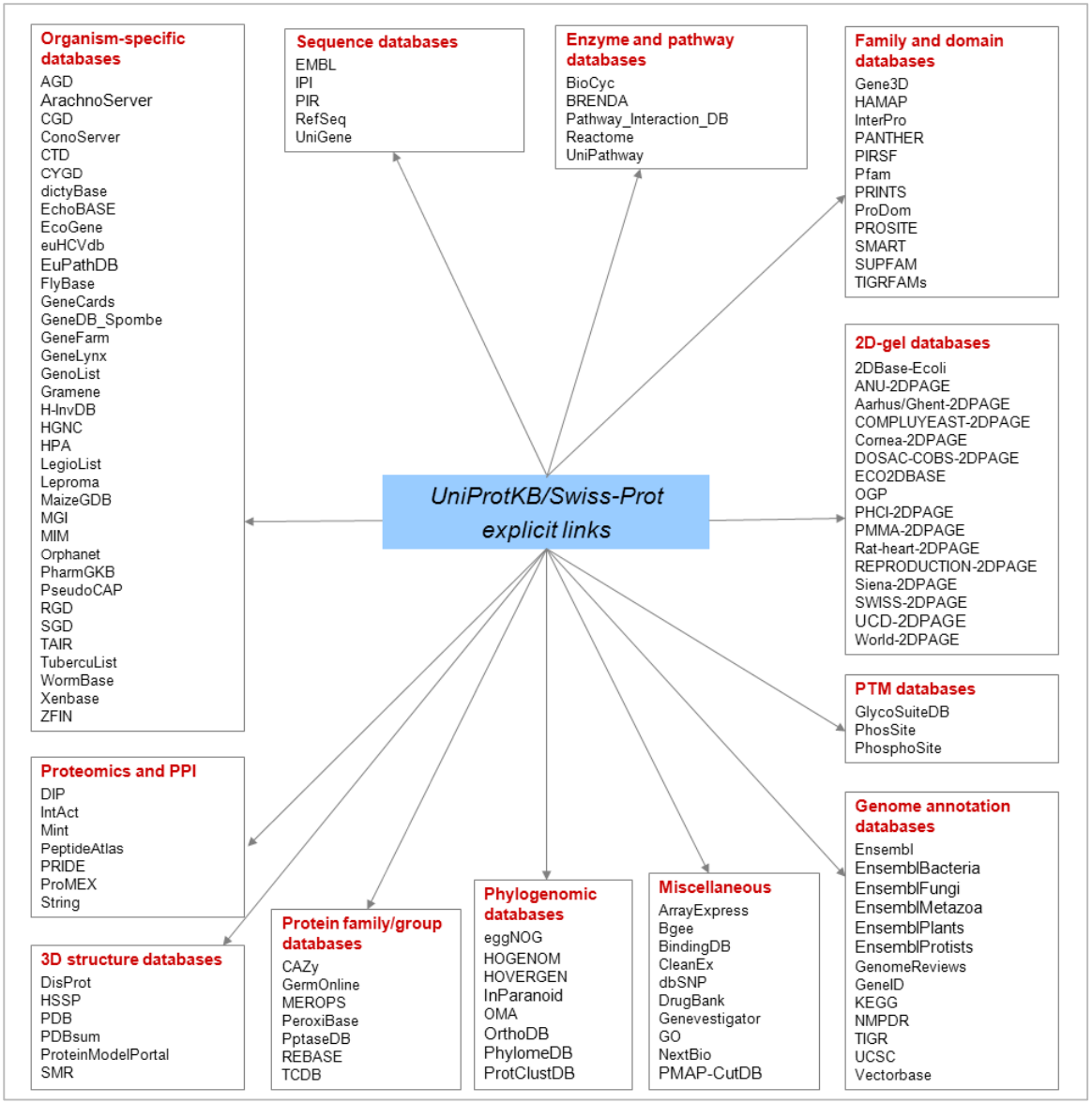
Swiss-Prot/TrEMBL



Automatic annotation

- Allows annotation of UniProtKB/TrEMBL in an efficient and scalable manner with a high degree of accuracy
- Based on annotation rules which are created, tested and validated against published experimental data in UniProtKB/Swiss-Prot
- Rules are linked to InterPro member database signatures and define annotations to be added and conditions which must be fulfilled
- Signatures identify family members in UniProtKB
- Common annotation in Swiss-Prot is transferred to related family members in TrEMBL if they fulfil rule conditions

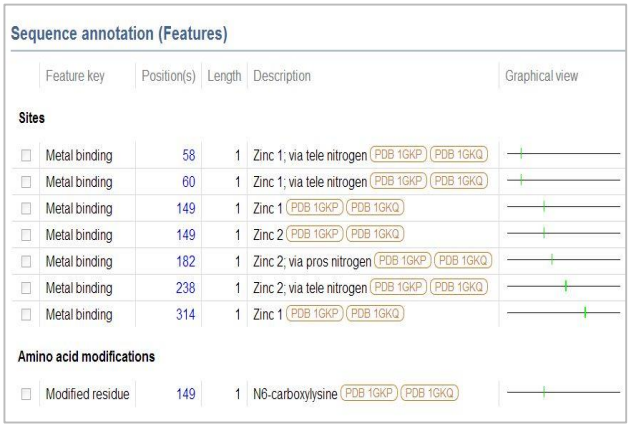
Extensively cross-referenced



Data import

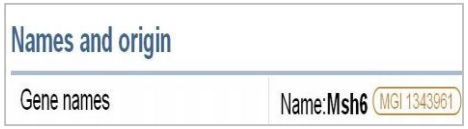
Gene names

Data from structures



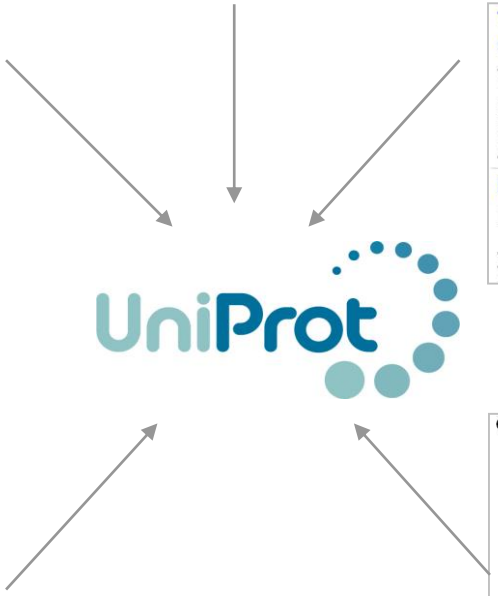
Sequence annotation (Features)

	Feature key	Position(s)	Length	Description	Graphical view
Sites					
	Metal binding	58	1	Zinc 1; via tele nitrogen (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	60	1	Zinc 1; via tele nitrogen (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	149	1	Zinc 1 (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	149	1	Zinc 2 (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	182	1	Zinc 2; via pros nitrogen (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	238	1	Zinc 2; via tele nitrogen (PDB 1GKP) (PDB 1GKQ)	
	Metal binding	314	1	Zinc 1 (PDB 1GKP) (PDB 1GKQ)	
Amino acid modifications					
	Modified residue	149	1	N6-carboxylysine (PDB 1GKP) (PDB 1GKQ)	



Names and origin

Gene names	Name:**Msh6** (MGI 1343961)



UniProt

Additional bibliography

The 4.1/ezrin/radixin/moesin domain of the DAL-1/Protein 4.1B tumour suppressor interacts with 14-3-3 proteins.

Yu T., Robb V.A., Singh V., Gutmann D.H., Newsham I.F.

Biochem. J. 365:783-789(2002) (GeneRIF 7529) · Mapped (22)
Results show that three 14-3-3 isoforms beta gamma and eta are DAL-1/Protein 4.1B-binding proteins. (GeneRIF 7529)

Regulation of TSC2 by 14-3-3 binding.

Li Y., Inoki K., Yeung R., Guan K.L.

J. Biol. Chem. 277:44593-44596(2002) (GeneRIF 7529) · Mapped (22)
TSC2 associates with 14-3-3 in vivo (GeneRIF 7529)

Interactions

Binary interactions

With	Entry	#Exp.	IntAct
itself		3	EBI-945751,EBI-945751
AK6	Q9Y3D8	5	EBI-945751,EBI-2896123
ATXN1	P54253	3	EBI-945751,EBI-930964
ATXN1	P54253-1	6	EBI-945751,EBI-930975
CEP70	Q8NHQ1	3	EBI-945751,EBI-739624
LNX1	Q8TBB1	2	EBI-945751,EBI-739832
PICK1	Q9NRD5	3	EBI-945751,EBI-79165
PSME3	P61289	3	EBI-945751,EBI-355546
VRK1	Q99986	9	EBI-945751,EBI-1769146
XRCC5	P13010	6	EBI-945751,EBI-357997
XRCC6	P12956	3	EBI-945751,EBI-353208

GO terms

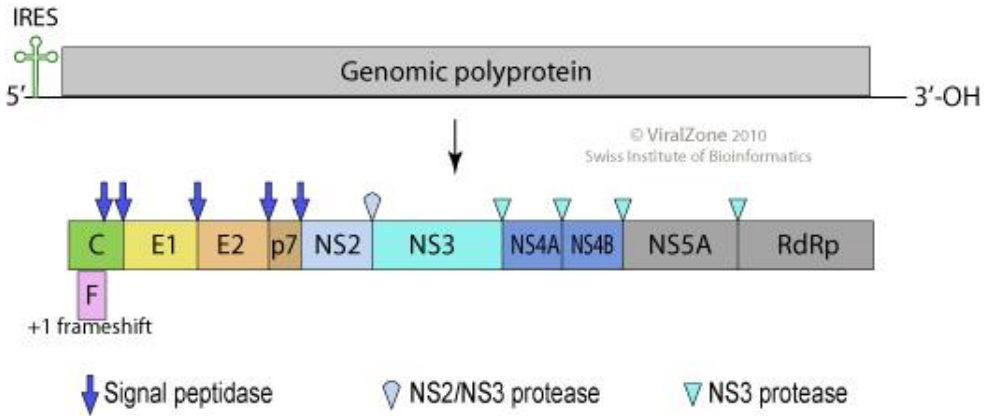
Gene Ontology (GO)

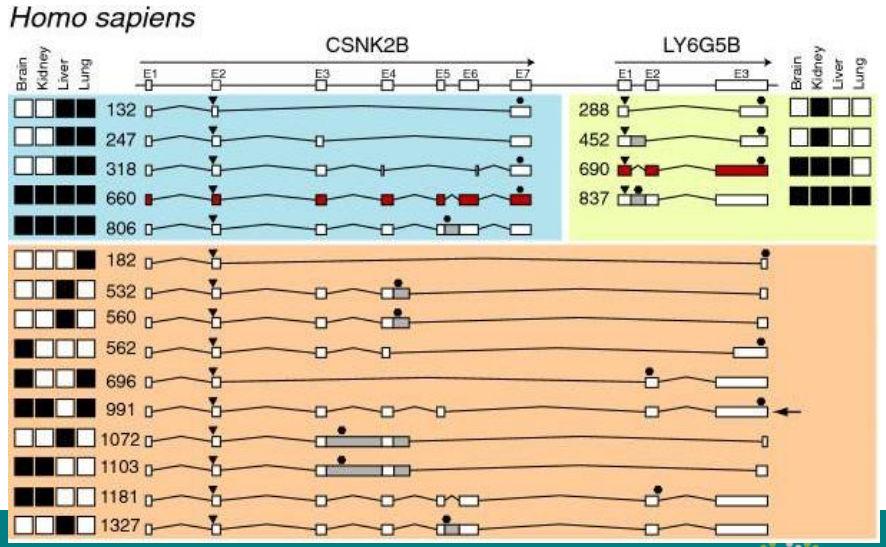
Biological process	Ras protein signal transduction Inferred from Experiment. Source: Reactome activation of pro-apoptotic gene products Inferred from Experiment. Source: Reactome cytoplasmic sequestering of protein Inferred from direct assay. Source: BHF-UCL negative regulation of protein amino acid dephosphorylation Inferred from direct assay. Source: BHF-UCL positive regulation of catalytic activity Inferred from direct assay. Source: BHF-UCL
Cellular component	cytosol Inferred from Experiment. Source: Reactome melanosome Inferred from electronic annotation. Source: UniProtKB-SubCell perinuclear region of cytoplasm Inferred from direct assay. Source: UniProtKB
Molecular function	histone deacetylase binding Inferred from physical interaction. Source: BHF-UCL phosphoserine binding Inferred from physical interaction. Source: BHF-UCL protein domain specific binding Inferred from physical interaction. Source: UniProtKB

Sequence curation

Improve sequence curation and representation, focus on capturing complexity of protein products –

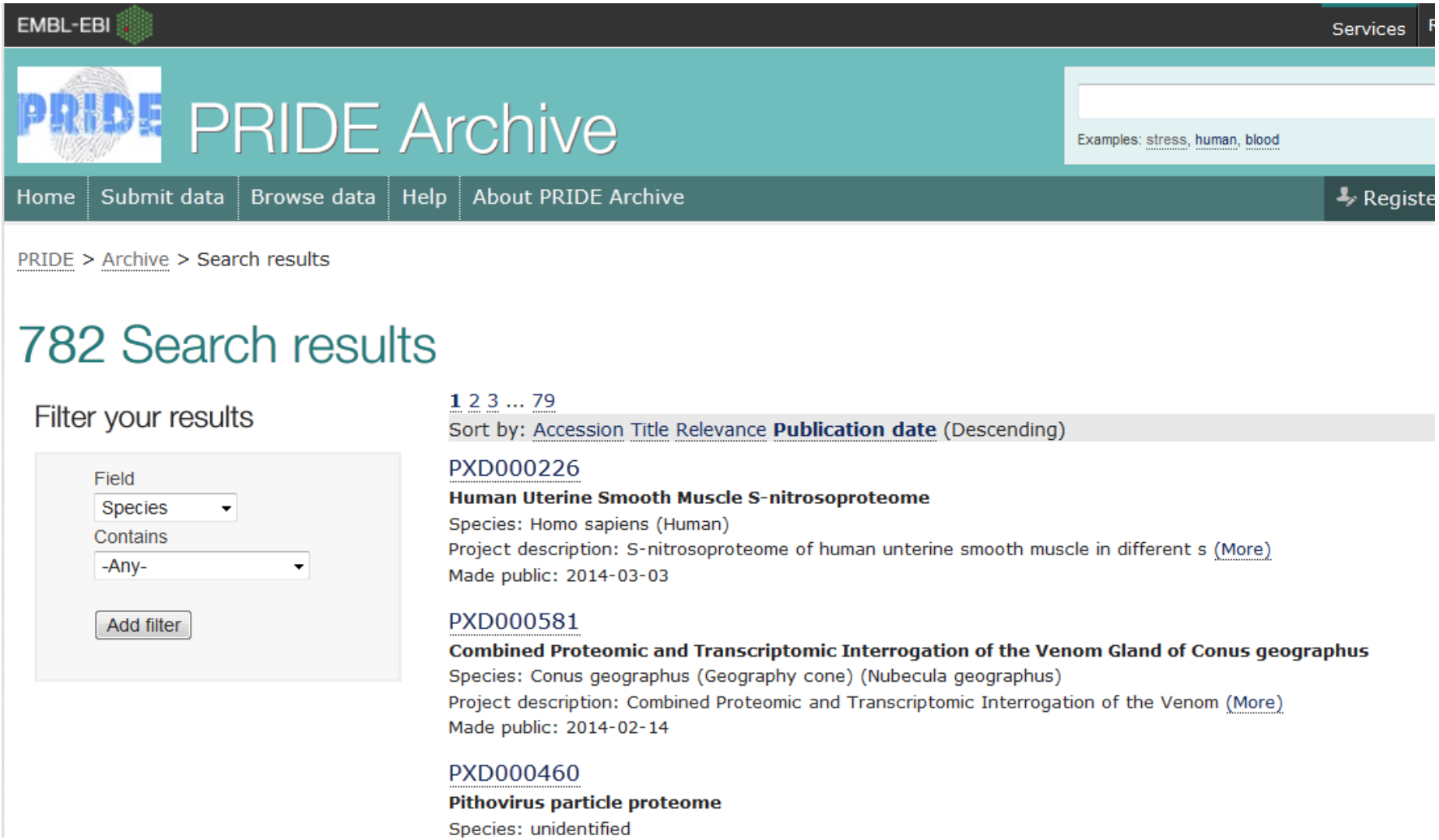
splicing, protein cleavage, transcriptional read-through, uORFS



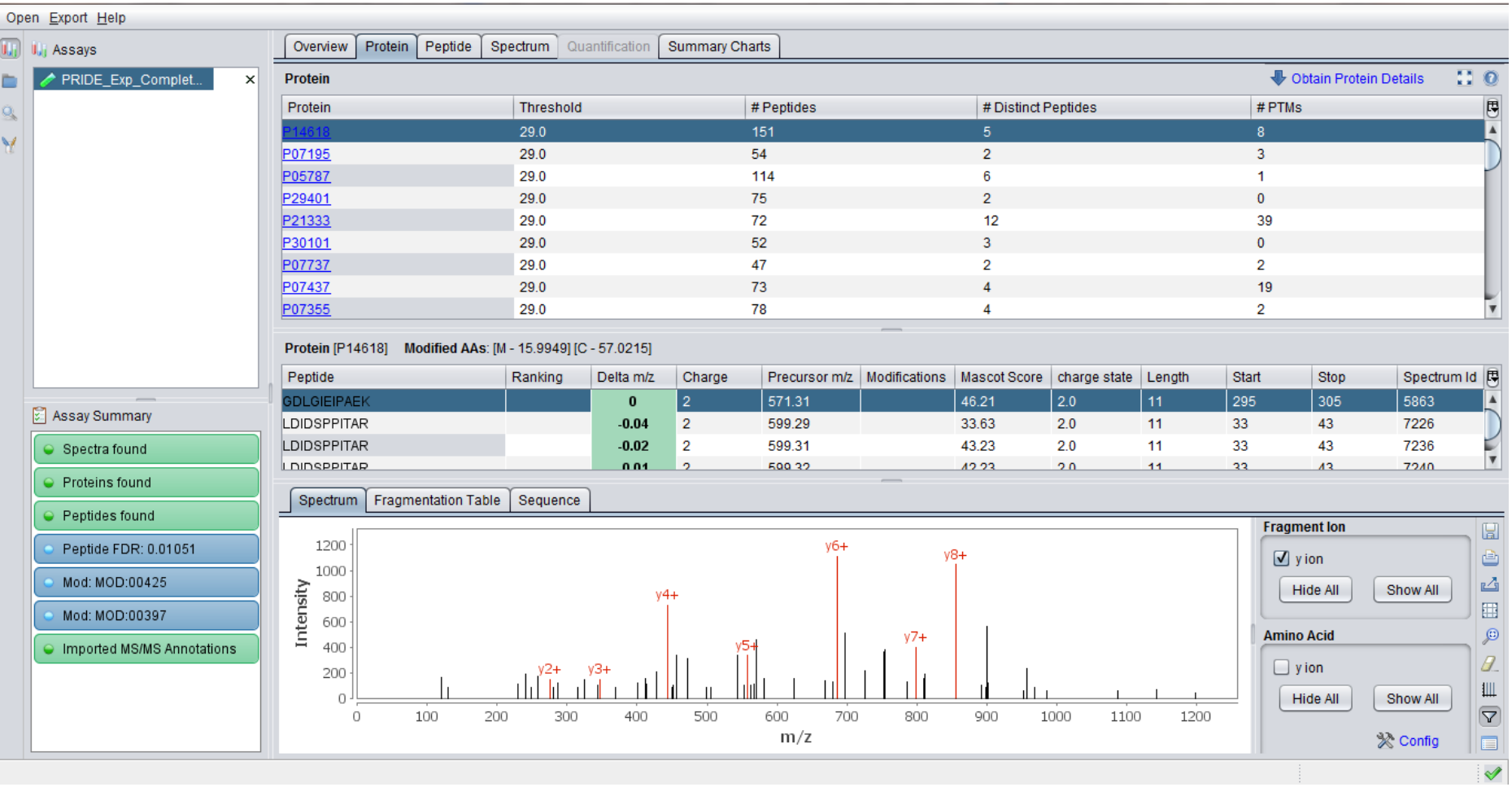


PRIDE – MS identifications

- PRIDE-Archive - stores submitted MS identifications plus meta-data describing sample
- PRIDE-Proteomes will hold a filtered subset- high quality identifications
- Will then build an expression atlas to give organism/cellular/subcellular map at the protein level
- Sample annotation enables comparison of normal/cell line/diseased cell protein profile
- Protein identification provides proof of protein existence and will also be a source of PTM data → UniProtKB.



www.ebi.ac.uk/pride



Protein	Threshold	# Peptides	# Distinct Peptides	# PTMs
P14618	29.0	151	5	8
P07195	29.0	54	2	3
P05787	29.0	114	6	1
P29401	29.0	75	2	0
P21333	29.0	72	12	39
P30101	29.0	52	3	0
P07737	29.0	47	2	2
P07437	29.0	73	4	19
P07355	29.0	78	4	2

Peptide	Ranking	Delta m/z	Charge	Precursor m/z	Modifications	Mascot Score	charge state	Length	Start	Stop	Spectrum Id
GDLGIEIPAEK		0	2	571.31		46.21	2.0	11	295	305	5863
LDIDSPPITAR		-0.04	2	599.29		33.63	2.0	11	33	43	7226
LDIDSPPITAR		-0.02	2	599.31		43.23	2.0	11	33	43	7236

PRIDE Inspector – suite of analysis tools

‘Omics analysis

- The end result of such experiments is usually a list of identifiers
- Biological context needs to be derived from this list
- Tools available for this, using EBI data
  - Network analysis
  - Pathway analysis
  - GO representation analysis

Network analysis

- Protein interaction networks aim to represent the complex web of molecular interactions within a cell

- Overlay of expression data on those networks can help identify clusters of interacting (i.e. probably functionally related) proteins which are up(down)-regulated in your sample

- Graph analysis tools (e.g as supplied in R) can enable a deeper understanding of the data

IntAct

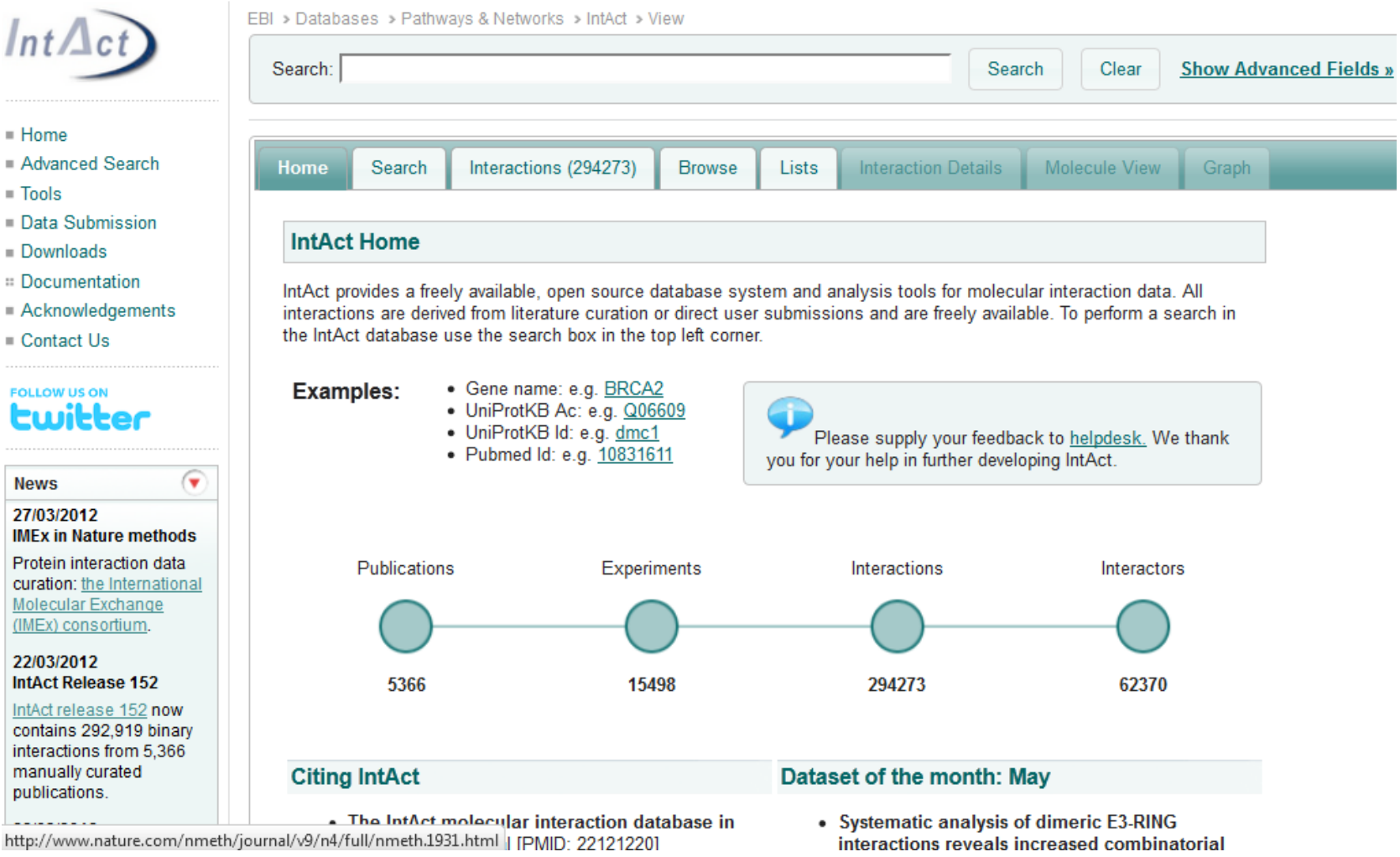
- Publicly available repository of molecular interactions (mainly PPIs) - ~445K binary interactions taken from >12,500 publications (February 2014)

- Data is standards-compliant and available via our website, for download at our ftp site or via PSICQUIC

http://www.ebi.ac.uk/intact
ftp://ftp.ebi.ac.uk/pub/databases/intact
www.ebi.ac.uk/Tools/webservices/psicquic/view/main.xhtml

- Provide open-access versions of the software to allow installation of local IntAct nodes.

IntAct – Home Page

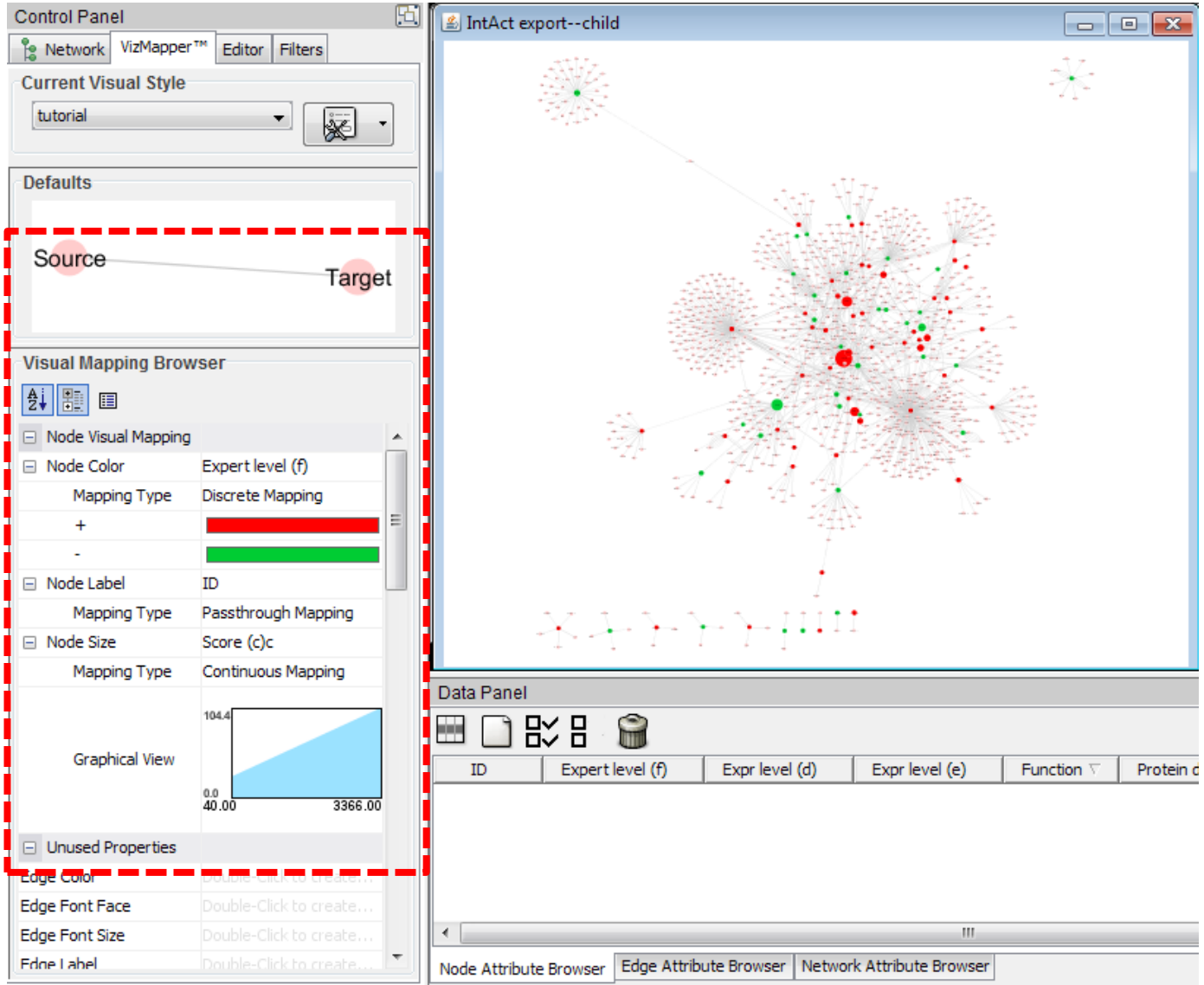


http://www.ebi.ac.uk/intact

Use IntAct to….

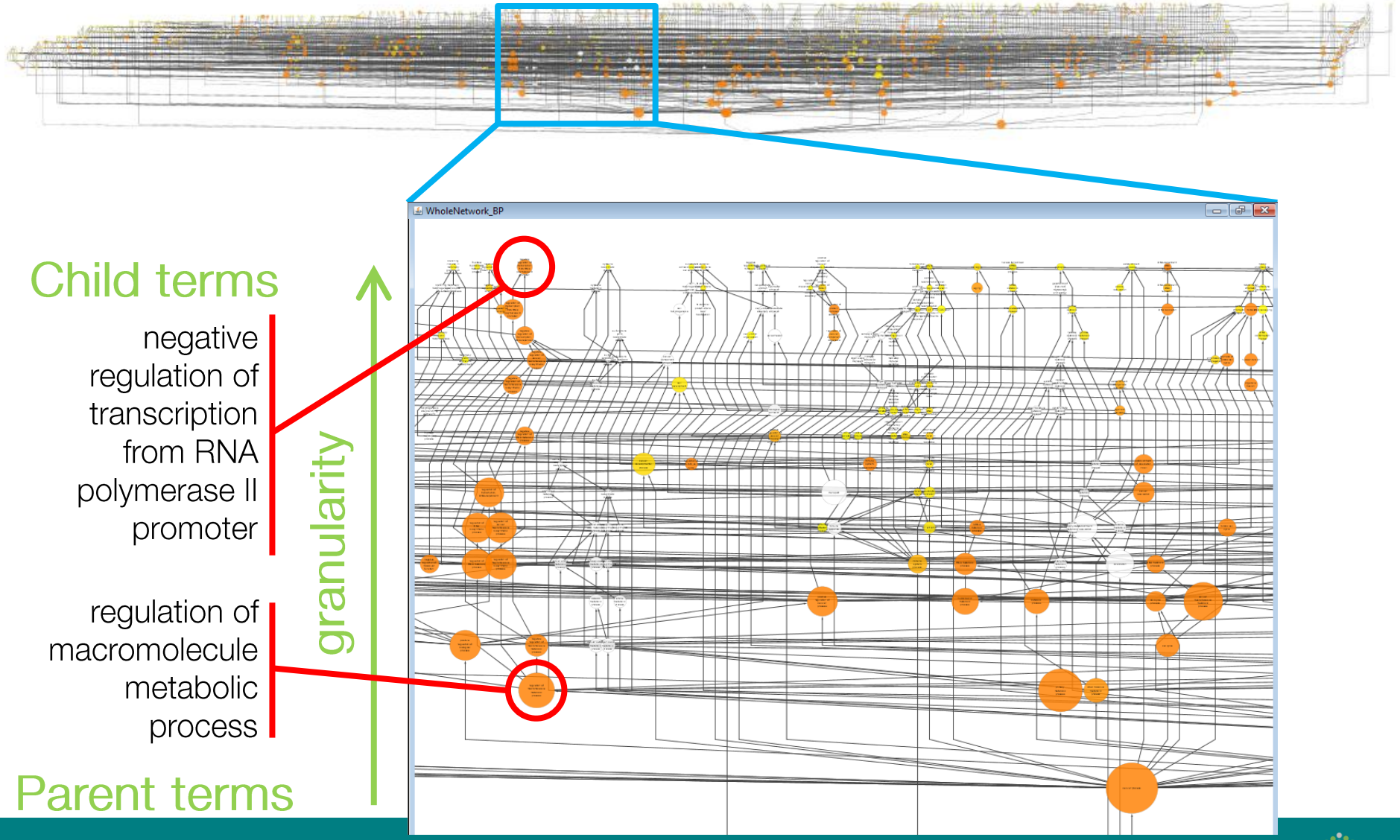
- Search for the interactors of one, or a set of, proteins
- Perform an ontology-driven search (e.g. find all the interaction of human peroxisomal proteins – GO:0005777 peroxisome)
- Download a large dataset (e.g. the human interactome) in standard formats (XML, tab-delineated, XGMML, RDF)

Network analysis – adding quantitative expression data to an interaction network



Gene ontology enrichment analysis

Choose Layouts → Cytoscape Layouts → Hierarchical layout



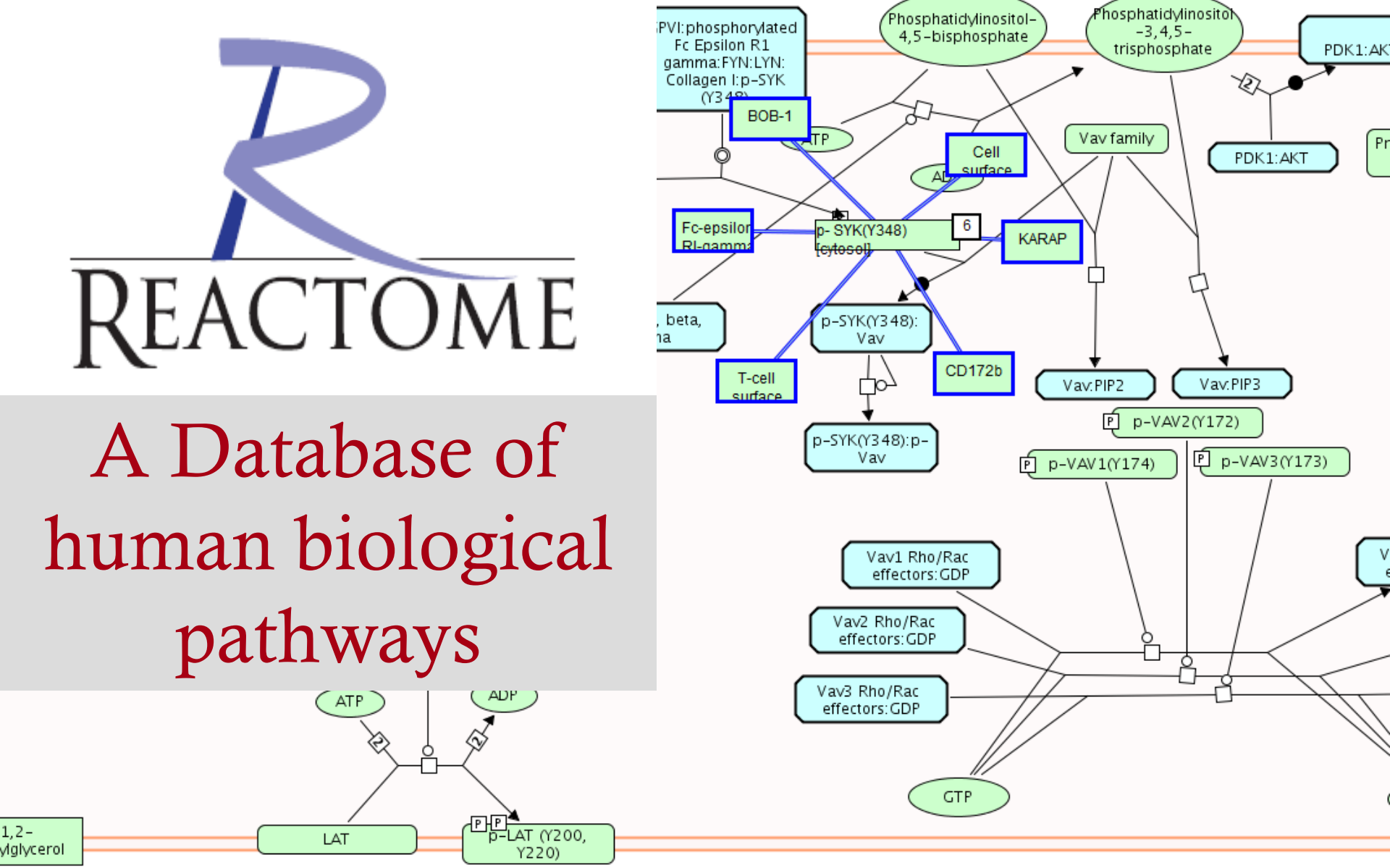
Pathway analysis of large datasets

- Network analysis highlights interacting clusters of molecules (often indicates protein complexes stable/transient) – may have similar function and subcellular location

- To group molecules by Process, can also subject identifier link to pathway analysis

REACTOME

A Database of human biological pathways

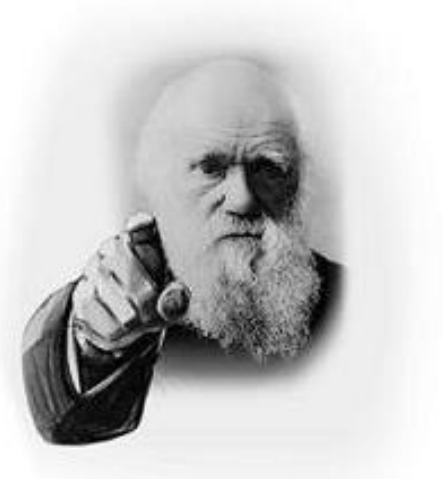


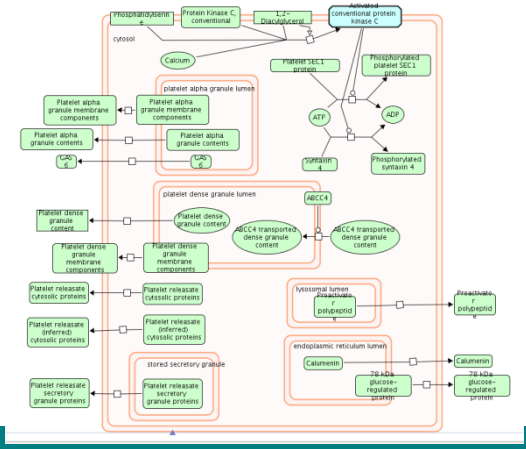
Reactome is…

Free, online, open-source curated database of pathways and reactions in human biology

Authored by expert biologists, maintained by Reactome editorial staff (curators)

Mapped to cellular compartment – agnostic of cell/tissue type/state



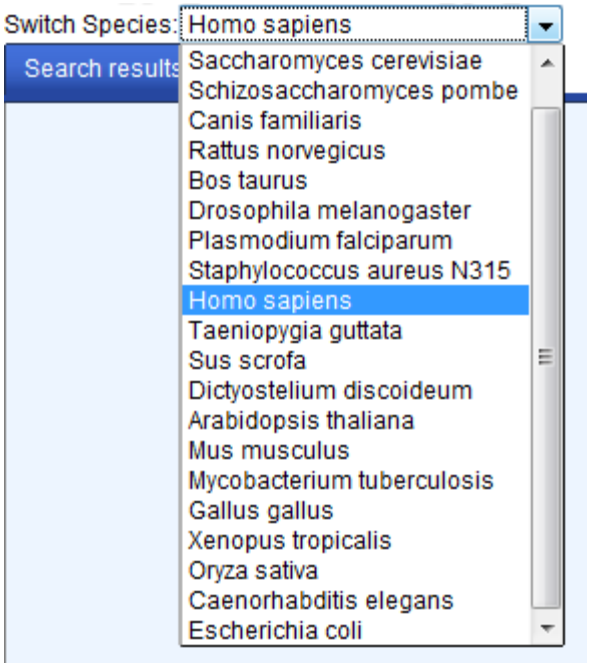


Data Expansion - Link-outs From Reactome

- GO
  - Molecular Function
  - Compartment
  - Biological process

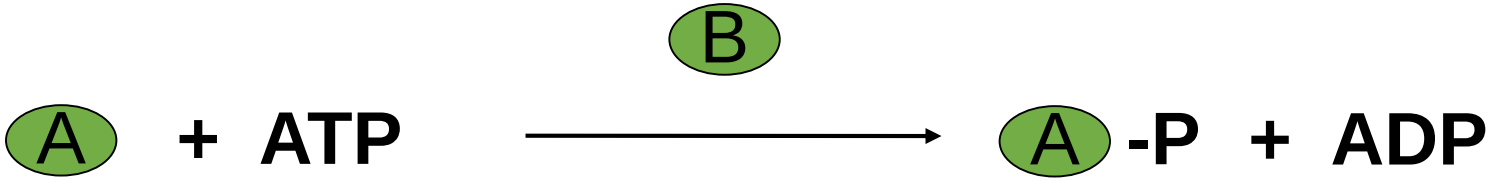
- KEGG, ChEBI – small molecules
- UniProt – proteins
- Sequence dbs – Ensembl, OMIM, Entrez Gene, RefSeq, HapMap, UCSC, KEGG Gene
- PubMed references – literature evidence for events

Species Selection

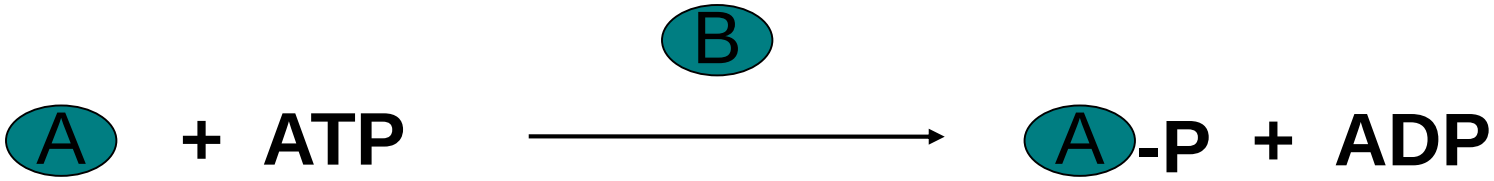


Data Expansion – Projecting to Other Species

Human



Mouse



Drosophila

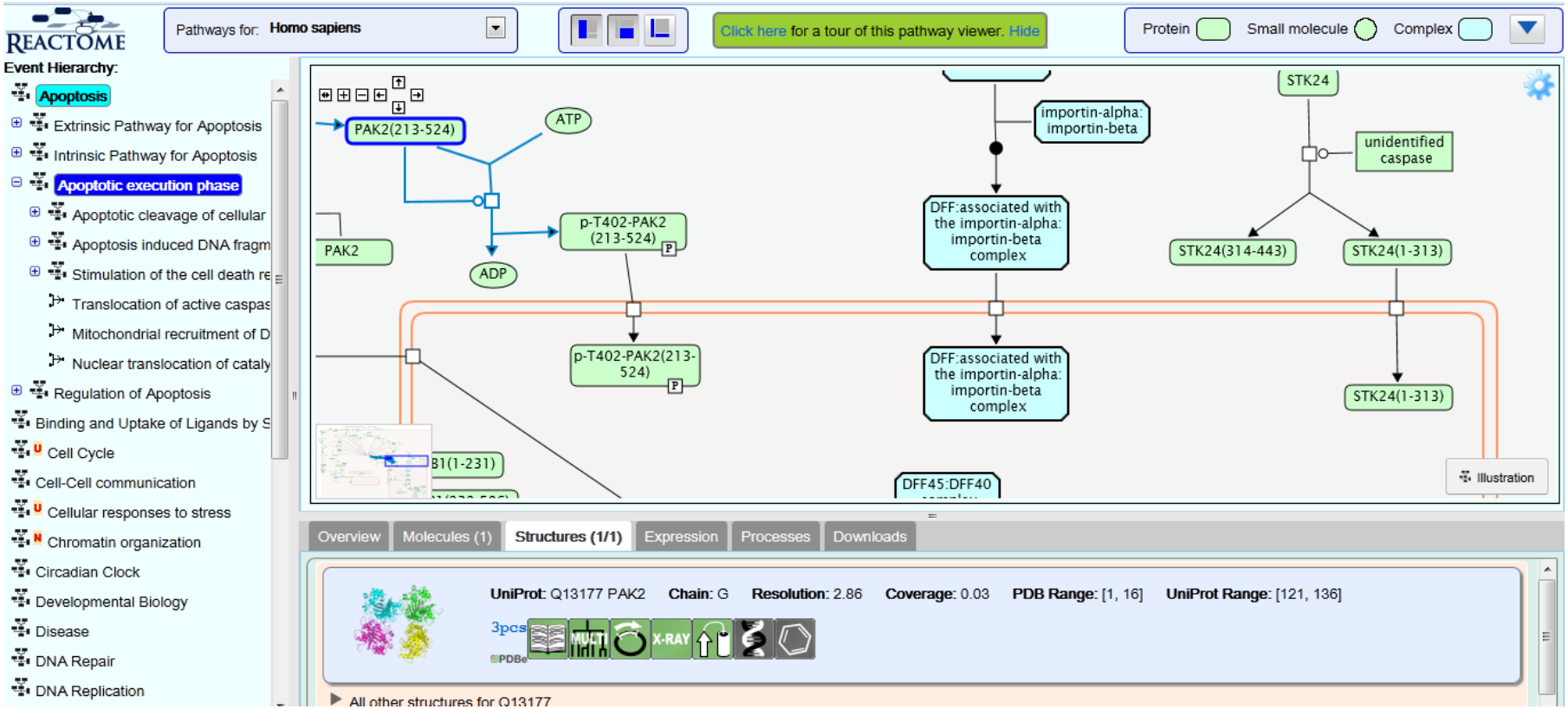
Reaction not inferred

No orthologue - Protein not inferred

Reactome Tools

- Interactive Pathway Browser
- Pathway Mapping and Over-representation
- Expression overlay onto pathways
- Molecular Interaction overlay
- Biomart

Pathway Browser



Expression analysis

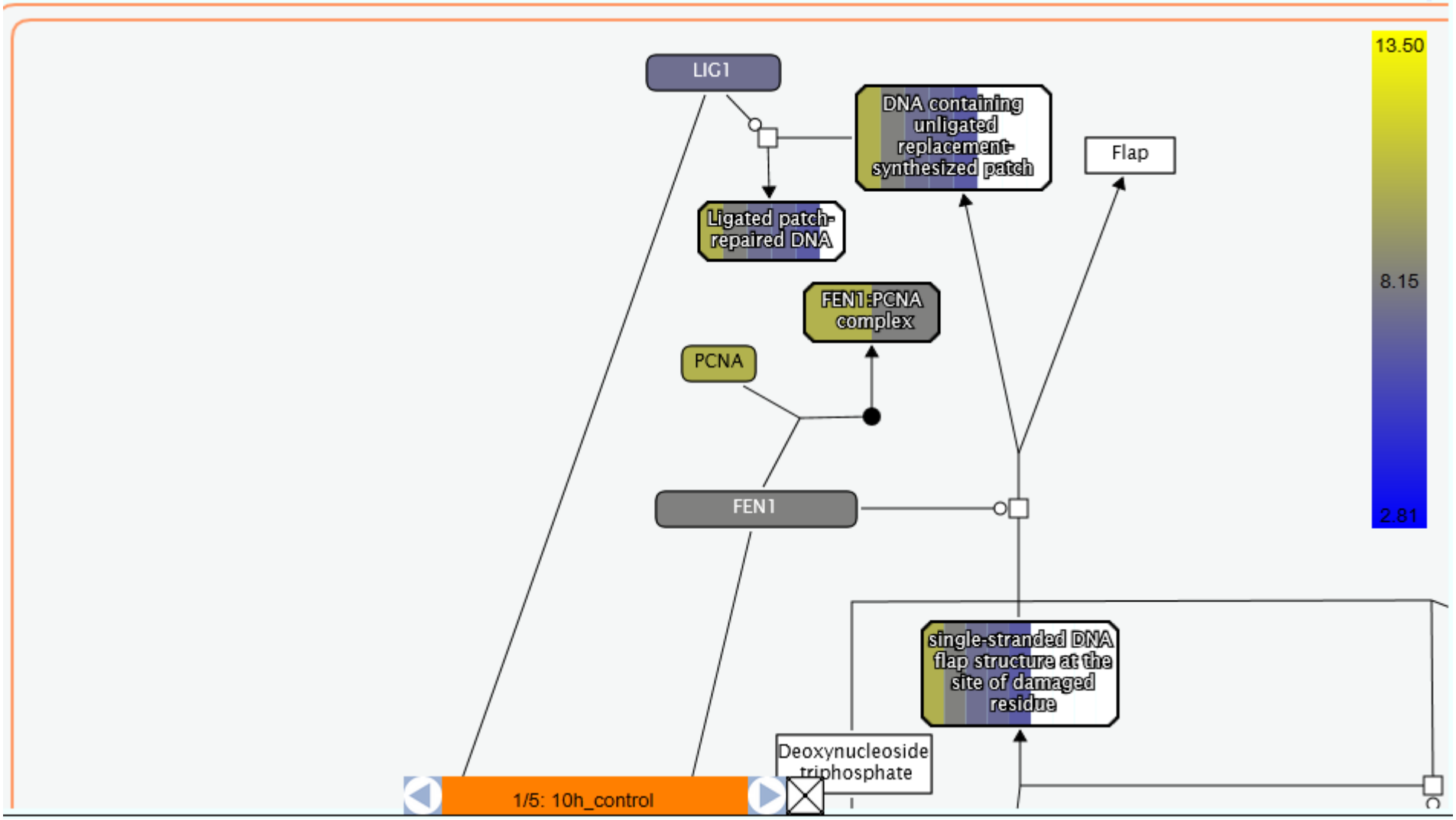
Submit analysis data | **Analysis results** | Identifier results | Summary results

Analysis results, per pathway

This table provides an overview of your expression data in a pathway context. For each Reactome pathway, the total number of proteins is shown, plus the number of genes/proteins in your dataset that match. By clicking on a pathway name, you will be taken to an interactive graphical representation of the pathway, where your expression levels are represented as coloration of proteins.

Select format to download this table: microsoft xcel (tsv) Download

Pathway	Species	IDs in pathway (%)	10h_control	10h	14h	18h	24h
Not assigned	*Not known*	21 (0%)	7.3	6.8	7.0	6.9	6.5
Apoptosis	Homo sapiens	74 (48%)	7.9	7.6	7.7	7.7	7.4
Binding and Uptake of Ligands by Scaveng	Homo sapiens	6 (3%)	4.8	5.2	5.2	5.0	5.4
Cell Cycle	Homo sapiens	115 (25%)	8.4	7.8	8.0	8.0	7.5
Cell-Cell communication	Homo sapiens	4 (3%)	7.8	7.8	7.9	7.7	7.8
Cellular responses to stress	Homo sapiens	24 (12%)	7.6	7.1	7.2	7.1	6.9
Chromatin organization	Homo sapiens	4 (4%)	8.1	7.6	7.5	7.8	7.2
Circadian Clock	Homo sapiens	1 (2%)	10.3	10.1	10.1	10.2	9.3
Developmental Biology	Homo sapiens	19 (5%)	6.5	6.5	6.6	6.5	6.6
Disease	Homo sapiens	180 (17%)	8.9	8.7	8.8	8.8	8.5
DNA Repair	Homo sapiens	74 (69%)	7.6	7.1	7.2	7.3	6.9
DNA Replication	Homo sapiens	77 (77%)	8.8	8.1	8.3	8.3	7.7
Extracellular matrix organization	Homo sapiens	22 (8%)	5.7	5.9	5.8	5.8	6.0
Gene Expression	Homo sapiens	272 (38%)	9.0	8.6	8.8	8.8	8.4
Hemostasis	Homo sapiens	76 (17%)	6.0	6.0	6.0	6.0	6.1
Immune System	Homo sapiens	95 (8%)	8.0	7.7	7.8	7.8	7.5


LIG1
DNA containing unligated replacement-synthesized patch
Flap
Ligated patch-repaired DNA
FEN1:PCNA complex
PCNA
FEN1
single-stranded DNA flap structure at the site of damaged residue
Deoxynucleoside triphosphate
1/5: 10h_control
13.50
8.15
2.81

Exploring proteomics

- UniProt can act as a central hub to explore other resources
- All resources designed for large-scale querires – web-services, APIs etc
- All data and most software is open-source and can be locally installed
- Our design is driven by your needs – please feedback