

# UniProt

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

**UniProtKB**

UniProt Knowledgebase

**Swiss-Prot (550,552)**  
Manually annotated and reviewed.  
Records with information extracted from literature and curator-evaluated computational analysis.

**TrEMBL (60,971,489)**  
Automatically annotated and not reviewed.  
Records that await full manual annotation.

**UniRef**

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

**UniParc**

UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

**Proteomes**

A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

**Supporting data**

Literature citations  
Cross-ref. databases  
Taxonomy  
Diseases  
Subcellular locations  
Keywords

- A comprehensive, high-quality and freely accessible resource of protein sequence and functional information
- Provides protein sequences, domains and structural information, subcellular location, etc. for many species
- Also includes some alignment and mapping tools

# Reading protein information

## UniProtKB - P38398 (BRCA1\_HUMAN)

Basket ▾

### Display

BLAST Align Format Add to basket History

Help video Add a publication Feedback

Entry Protein Breast cancer type 1 susceptibility protein

Gene BRCA1

Organism Homo sapiens (Human)

Status ★ Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level<sup>i</sup>

None

### Function<sup>i</sup>

E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage. It is unclear whether it also mediates the formation of other types of polyubiquitin chains. The E3 ubiquitin-protein ligase activity is required for its tumor suppressor function. The BRCA1-BARD1 heterodimer coordinates a diverse range of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability. Regulates centrosomal microtubule nucleation. Required for normal cell cycle progression from G2 to mitosis. Required for appropriate cell cycle arrests after ionizing irradiation in both the S-phase and the G2 phase of the cell cycle. Involved in transcriptional regulation of P21 in response to DNA damage. Required for FANCD2 targeting to sites of DNA damage. May function as a transcriptional regulator. Inhibits lipid synthesis by binding to inactive phosphorylated ACACA and preventing its dephosphorylation. Contributes to homologous recombination repair (HRR) via its direct interaction with PALB2, fine-tunes recombinational repair partly through its modulatory role in the PALB2-dependent loading of BRCA2-RAD51 repair machinery at DNA breaks. Component of the BRCA1-RBBP8 complex which regulates CHEK1 activation and controls cell cycle G2/M checkpoints on DNA damage via BRCA1-mediated ubiquitination of RBBP8. Acts as a transcriptional activator (PubMed:20160719).

16 Publications ▾

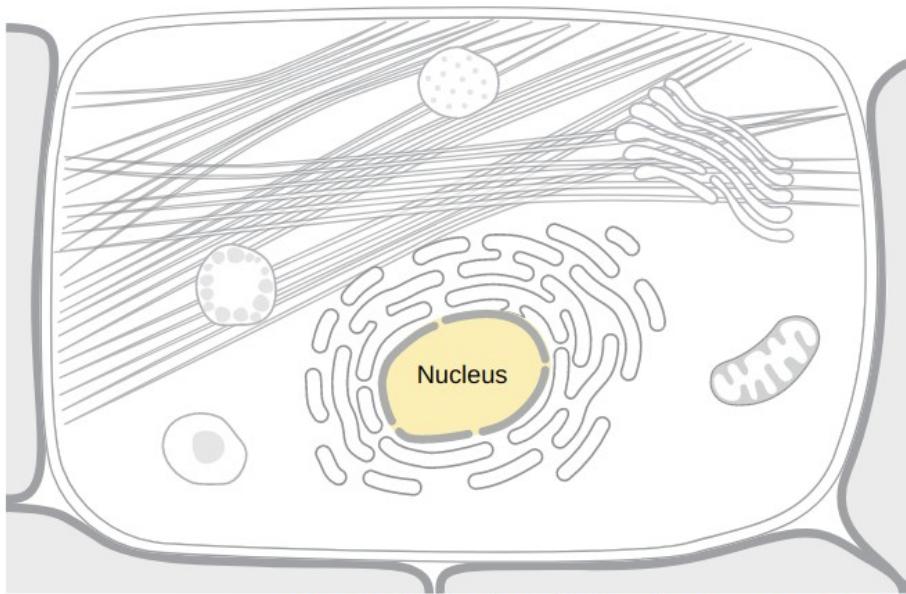
### Catalytic activity<sup>i</sup>

- S-ubiquitinyl-[E2 ubiquitin-conjugating enzyme]-L-cysteine + [acceptor protein]-L-lysine = [E2 ubiquitin-conjugating enzyme]-L-cysteine + N(6)-ubiquitinyl-[acceptor protein]-L-lysine.

6 Publications ▾ EC:2.3.2.27

# Reading protein information

## Subcellular location<sup>i</sup>



  Manual annotation     Automatic computational assertion

UniProt annotation   GO - Cellular component

Nucleus

Nucleus i 6 Publications ▾

Other locations

Chromosome i 3 Publications ▾

Cytoplasm i 1 Publication ▾

Note: Localizes at sites of DNA damage at double-strand breaks (DSBs); recruitment to DNA damage sites is mediated by ABRAXAS1 and the BRCA1-A complex (PubMed:26778126). Translocated to the cytoplasm during UV-induced apoptosis (PubMed:20160719). 2 Publications ▾

Isoform 3:

Other locations

Cytoplasm i

Isoform 5:

Other locations

Cytoplasm i 1 Publication ▾

# Reading protein information

## Structure<sup>i</sup>

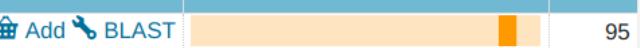


PDB Entry	Method	Resolution	Chain	Positions	Links
1JM7	NMR		A	1-110	PDBe RCSB PDB PDBj PDBsum
1JNX	X-ray	2.50 Å	X	1646-1859	PDBe RCSB PDB PDBj PDBsum
1N5O	X-ray	2.80 Å	X	1646-1859	PDBe RCSB PDB PDBj PDBsum
1OQA	NMR		A	1755-1863	PDBe RCSB PDB PDBj PDBsum
1T15	X-ray	1.85 Å	A	1646-1859	PDBe RCSB PDB PDBj PDBsum
1T29	X-ray	2.30 Å	A	1646-1859	PDBe RCSB PDB

# Reading protein information

## Family & Domains<sup>i</sup>

### Domains and Repeats

Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain <sup>i</sup>	1642 – 1736	BRCT 1 PROSITE-ProRule annotation	Add BLAST		95
Domain <sup>i</sup>	1756 – 1855	BRCT 2 PROSITE-ProRule annotation	Add BLAST		100

### Region

Feature key	Position(s)	Description	Actions	Graphical view	Length
Region <sup>i</sup>	1397 – 1424	Interaction with PALB2 1 Publication	Add BLAST		28

### Compositional bias

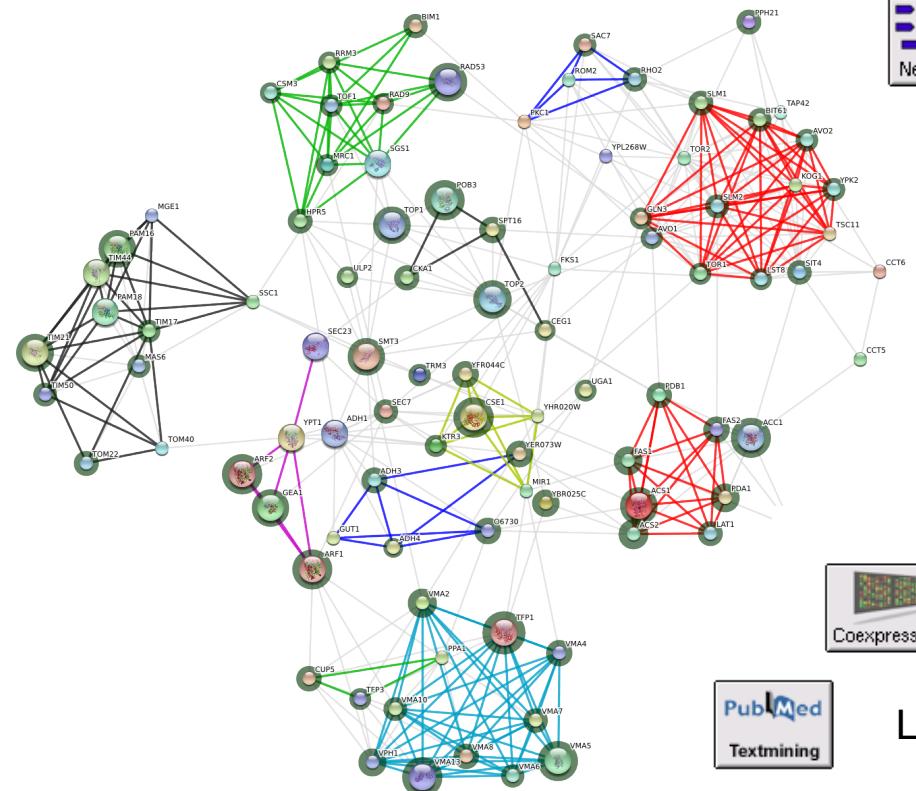
Feature key	Position(s)	Description	Actions	Graphical view	Length
Compositional bias <sup>i</sup>	651 – 654	Poly-Lys			4

# Practical example

- Look for the ELAVL1 gene. How many human proteins there are ? How many are reviewed?
- Which type of domains are in the protein that is the first result of our query ?
- How many isoforms of the protein are present ? Which is their mass ?
- How many protein-protein interactions does ELAVL1 make ?

Home · Download · Help/Info

STRING



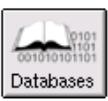
Genomic Neighborhood



Genes/Species Co-occurrence



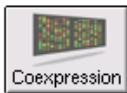
Gene Fusions



Database Imports



Exp. Interaction Data



Co-expression



Literature co-occurrence

- viewers for all types of evidence
- focus on usability and speed
- integrated scoring scheme
- *information transfer between species*

# Other Interaction Sources

## Interaction Databases



## Pathway Databases



Reactome



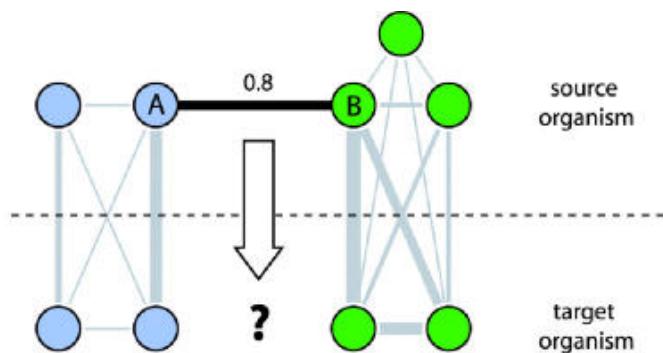
*the Gene Ontology*

Pathway**Interaction**Database

## Automated Textmining



## Interolog Transfer

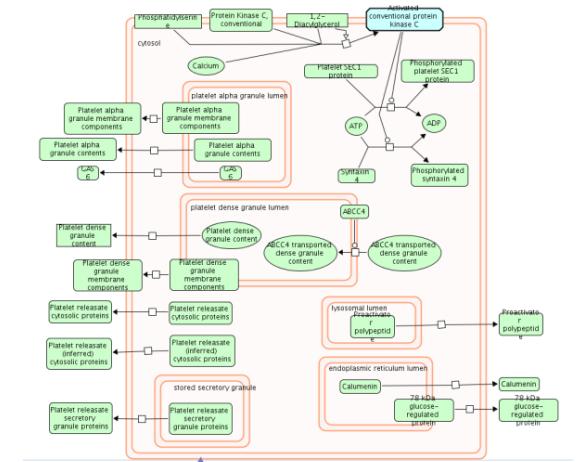
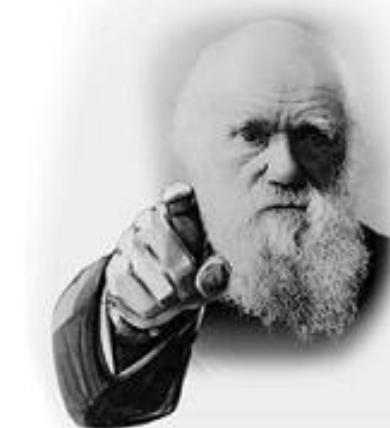


# Reactome is...

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

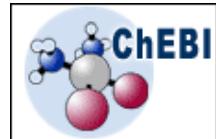
Authored by expert biologists,  
maintained by Reactome editorial  
staff (curators)

Mapped to cellular compartment



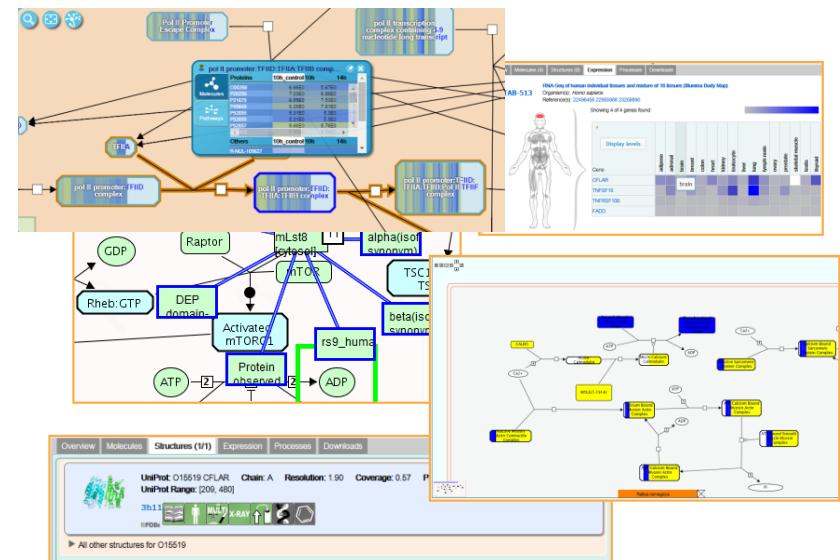
# Reactome is...

Extensively cross-referenced



Tools for data analysis –  
Pathway Analysis,  
Expression Overlay, Species  
Comparison

Used to infer orthologous  
events in 17 other species

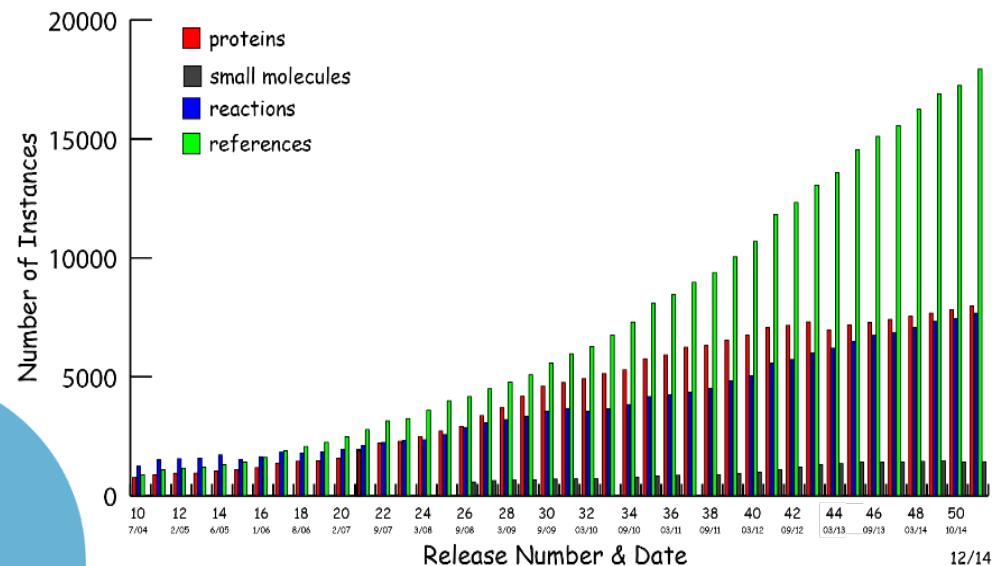
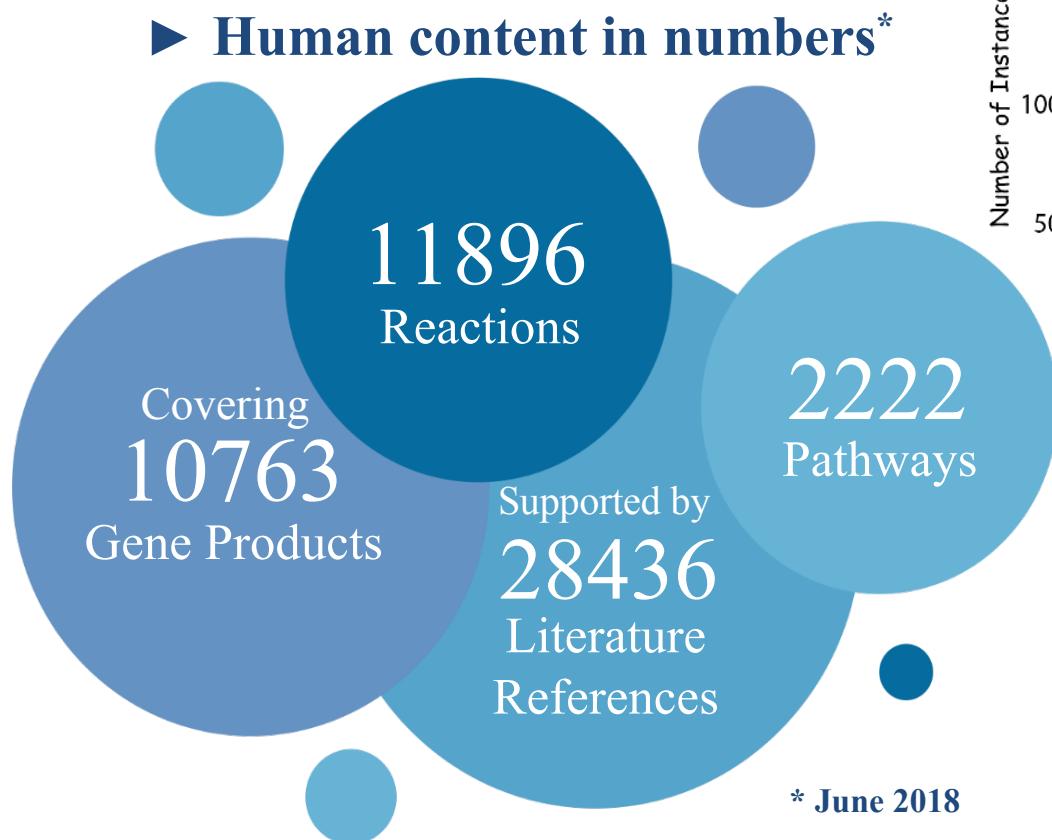


# Primary external sources

- Gene Ontology
  - Molecular Function
  - Compartment
  - Biological Process
- ChEBI – small molecules
- UniProt – proteins
- Ensembl – genes and transcripts
- PubMed – literature evidence for events

# Curation

- 52.6% of the 20,296 predicted human protein-coding genes



# Reactome Tools

- Interactive Pathway Browser
- Analysis
  - Over-representation
  - Pathway topology
  - Expression overlay
- Molecular Interaction overlay
- Species Comparison

# Front Page

<http://www.reactome.org>

The screenshot shows the Reactome website's homepage. At the top, there is a navigation bar with links for About, Content, Docs, Tools, Community, and Download. Below the navigation is a search bar with the placeholder "Find Reactions, Proteins and Pathways" and a "Go!" button. The main content area features four large blue icons representing different tools: "Pathway Browser" (with a 3D bar chart icon), "Analyze Data" (with a bar chart icon), "ReactomeFIViz" (with a network graph icon), and "Documentation" (with a document icon). Each tool has a brief description below it. A black banner at the bottom of the page reads "USE REACTOME GRAPH DATABASE IN YOUR PROJECT" and includes a "LEARN MORE" button. The background of the page features a network graph design.

reactome

About Content Docs Tools Community Download

Find Reactions, Proteins and Pathways

e.g. O95631, NTN1, signaling by EGFR, glucose

Pathway Browser

Analyze Data

ReactomeFIViz

Documentation

Visualize and interact with Reactome biological pathways

Merges pathway identifier mapping, over-representation, and expression analysis

Designed to find pathways and network patterns related to cancer and other types of diseases

Information to browse the database and use its principal tools for data analysis.

USE REACTOME GRAPH DATABASE IN YOUR PROJECT

LEARN MORE

Why Reactome

Reactome is a free, open-source, curated and peer reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. The current version (v62) of Reactome was released on September 27, 2017.

If you use Reactome in Asia, we suggest using our Chinese mirror site at [reactome.ncpsb.org](http://reactome.ncpsb.org).

EMBL-EBI NYU Langone Health OICR

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751 and 1U54GM114833-01), Ontario Research Fund, and the European Molecular Biology Laboratory.

Tweets

reactome @reactome

Exciting times for us! Our brand new reactome.org website features a #clean and #userfriendly interface that adjusts to browsers on your desktop and all your devices #webdesign #uxdesign #ResponsiveDesign #newlogo

reactome

# The Pathway Browser

The screenshot shows the Reactome Pathway Browser interface. At the top, there are navigation links: Home, Species, Analyse Data, Video Tour, and Layout. On the left, the Hierarchy Panel displays a tree structure of biological pathways. The central area features the Pathway Panel, which contains a complex network diagram of pathways like Immune System, Signal Transduction, and Metabolism. Below the diagram is the Detail Panel, which provides a description of selected items. The bottom navigation bar includes buttons for Description, Molecules, Structures, Expression, Analysis, and Downloads. On the right side, there are additional controls: Key, Export, Settings Sidebar, and Zoom/Move. A large orange box labeled "Thumbnail" points to a small preview image at the bottom left of the main panel.

Home

Species

Analyse Data

Video Tour

Layout

Event Hierarchy:

- Cell Cycle
- Cell-Cell communication
- Cellular responses to external stimuli
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Digestion and absorption
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Gene
- Hemostasis
- Immune
- Metab
- Metabol
- Metabol
- Mitophagy
- Muscle contraction
- Neuronal System
- Organelle biogenesis and maintenance
- Programmed Cell Death
- Reproduction
- Signal Transduction
- Transport of small molecules
- Vesicle-mediated transport

Pathways for: Homo sapiens

Analysis: Tour: Layout:

Key

Export

Settings Sidebar

Zoom/Move

Open Diagram

Search Diagram

Fit to Page

Illustrations

Thumbnail

Description

Molecules

Structures

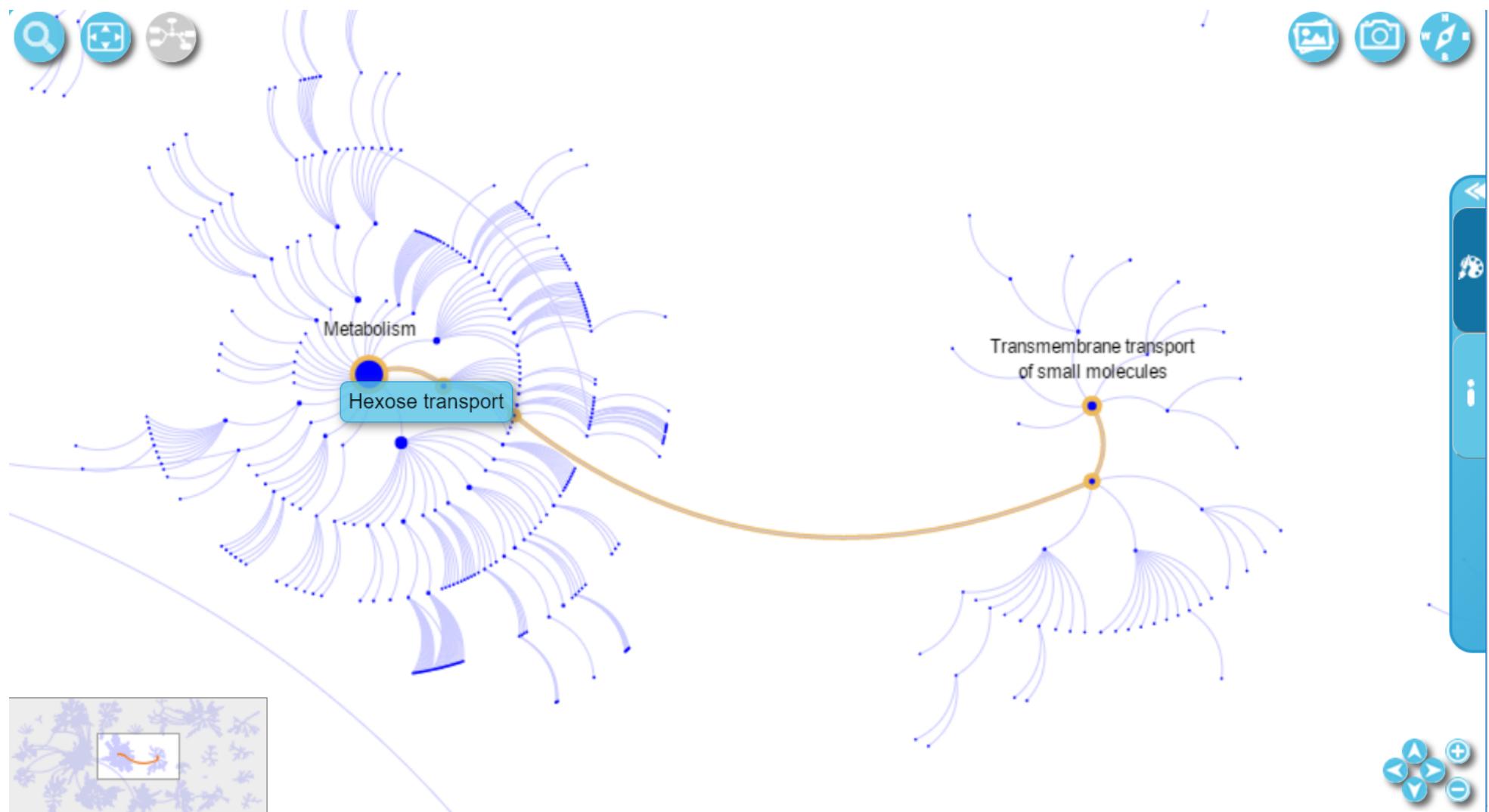
Expression

Analysis

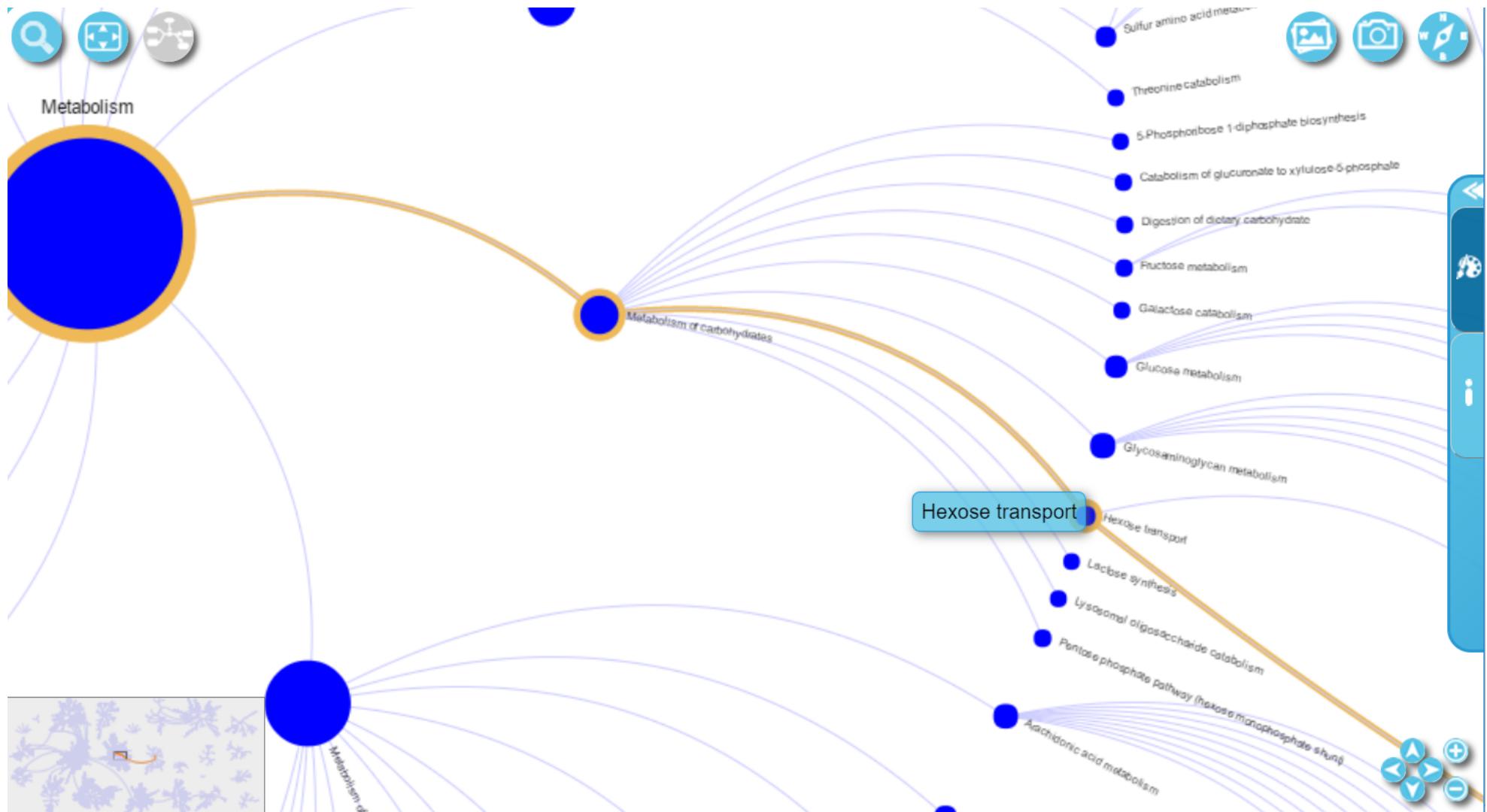
Downloads

Detail Panel

# Edges = shared pathways



# Zoom in for pathway names



# Hierarchy Panel

REACTOME 3.0    54 Pathways for: **Homo sapiens**

**Event Hierarchy:**

- ⊕  Cell Cycle
- ⊕  Cell-Cell communication
- ⊕  Cell junction organization
- ⊕  Signal regulatory protein (SIRP) family interactions
- ⊕  DSCAM interactions
- ⊕  Nephrin interactions
- ⊕  Cellular responses to stress
- ⊕  Chromatin organization
- ⊕  Circadian Clock
- ⊕  Developmental Biology
- ⊕  Disease
- ⊕  DNA Repair
- ⊕  DNA Replication
- ⊕  Extracellular matrix organization
- ⊕  Gene Expression
- ⊕  Hemostasis
- ⊕  Immune System
- ⊕  Mitophagy
- ⊕  Metabolism
- ⊕  Metabolism of proteins
- ⊕  Muscle contraction
- ⊕  Neuronal System
- ⊕  Organelle biogenesis and maintenance
- ⊕  Programmed Cell Death
- ⊕  Reproduction
- ⊕  Signal Transduction
- ⊕  Transmembrane transport of small molecules
- ⊕  Vesicle-mediated transport

-  Pathway
-  Reaction
-  Black-box
-  Inferred from
-  New
-  Updated
-  Disease

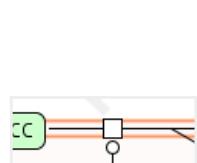
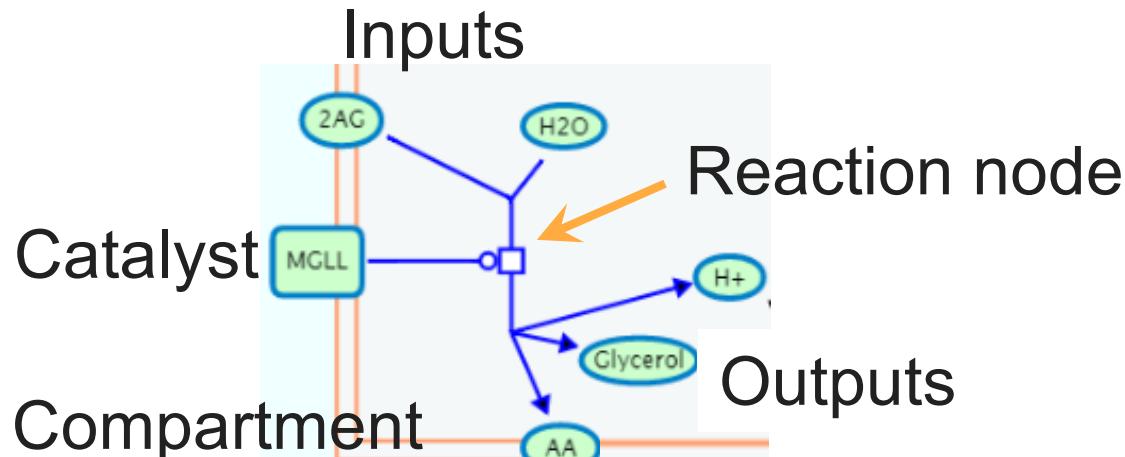
# The Pathway Browser - Pathway Diagrams

Ovals are small molecules (or sets of)

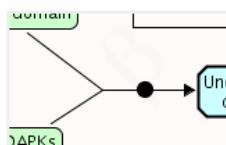
Green boxes are proteins,

Blue are complexes,

Blue with double-boundary are sets



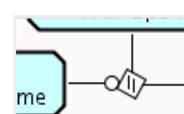
Transition



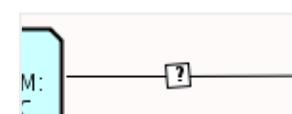
Binding



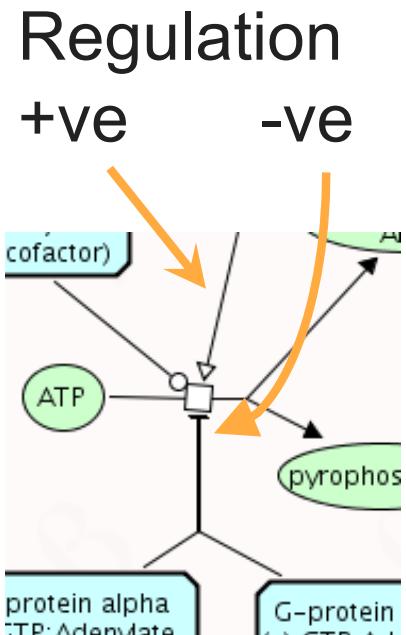
Dissociation



Omitted

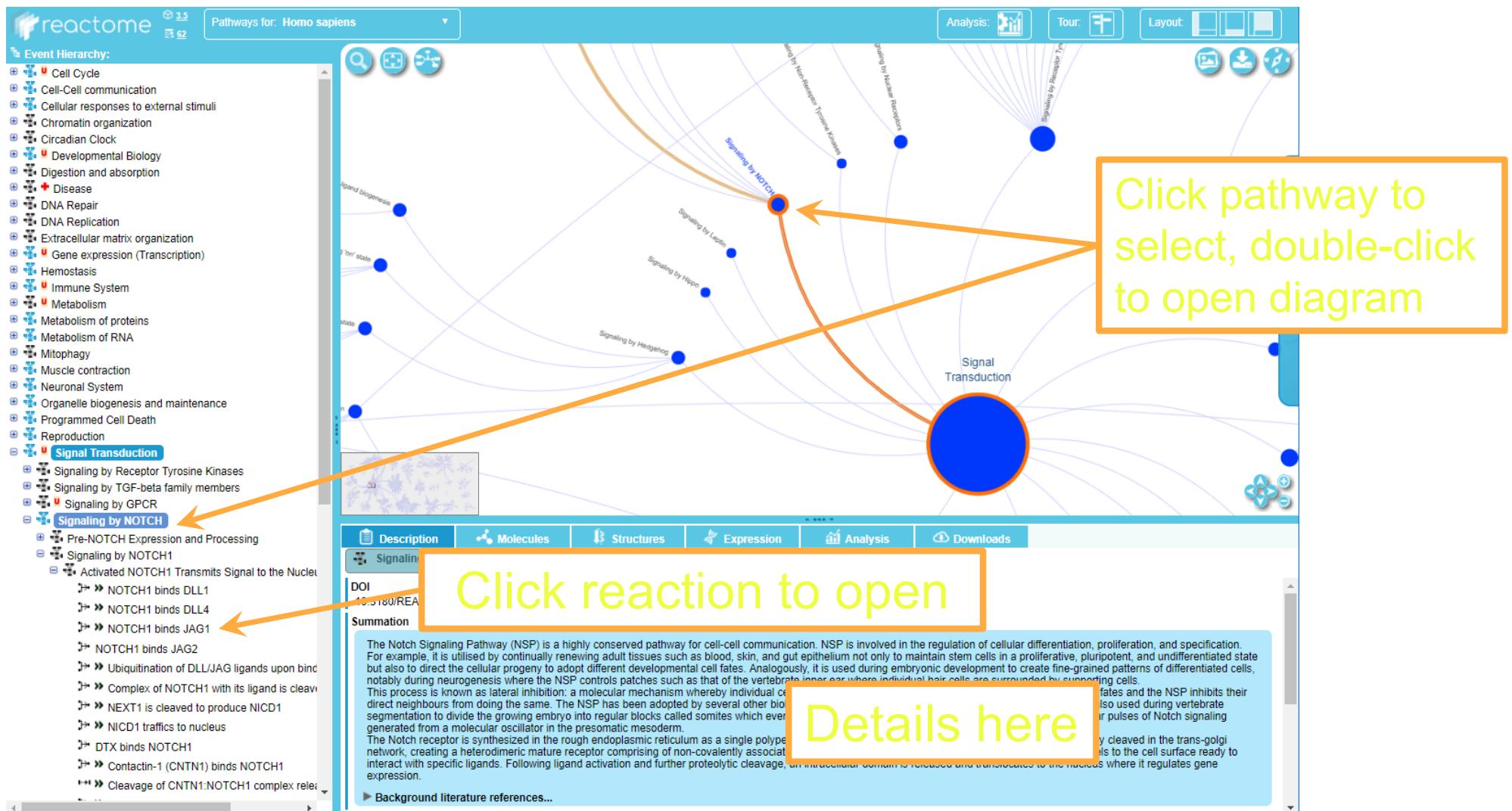


Uncertain

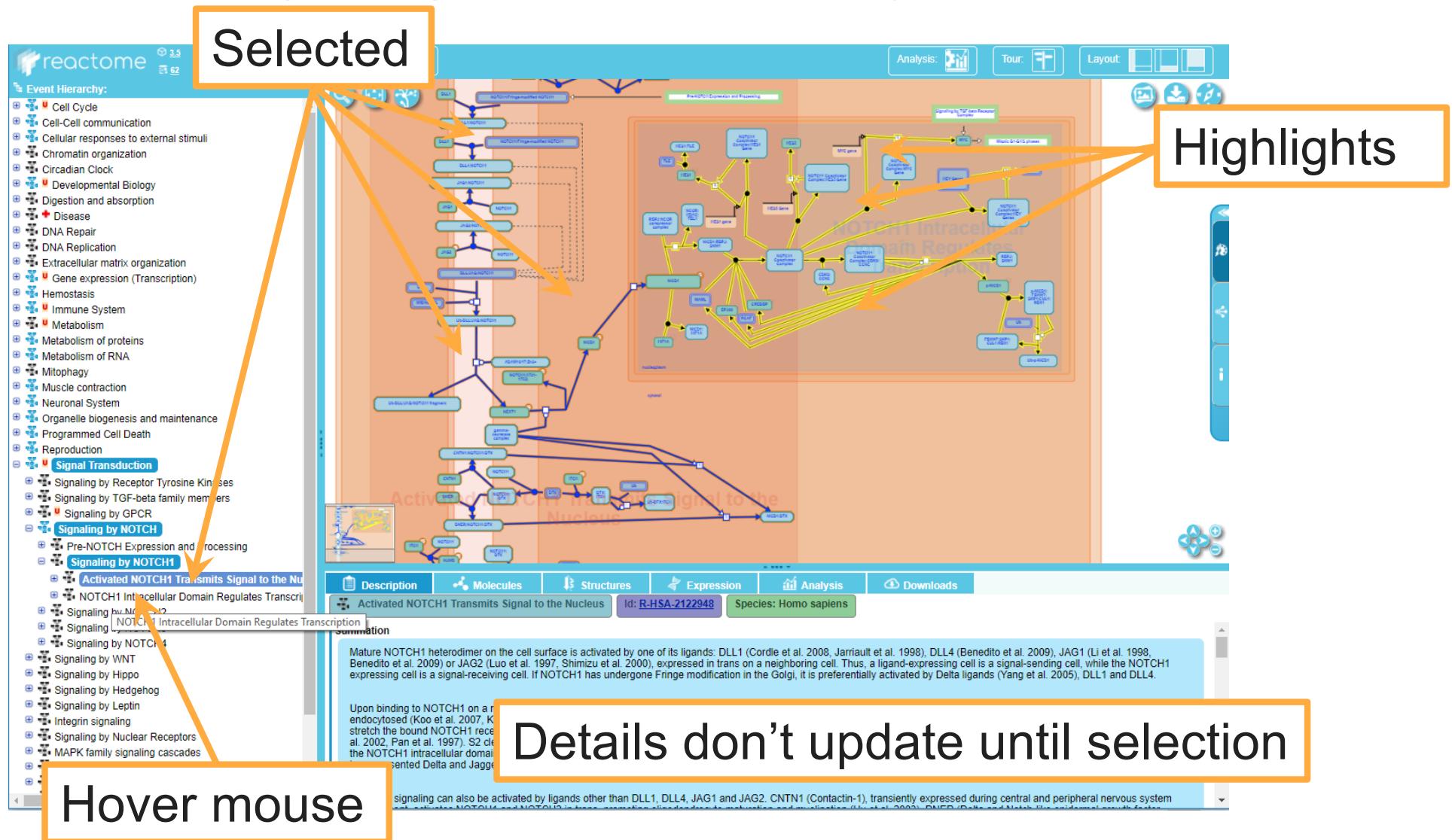


# Navigating in the Pathway Browser

# Home button



# Navigating in the Pathway Browser



# Show Illustration

REACTOME 3.2 Pathways for: Homo sapiens Analysis: Tour: Layout:

Event Hierarchy:

- Effects of PIP2 hydrolysis
- Response to elevated platelet cytosolic Ca<sup>2+</sup>
- Formation of Fibrin Clot (Clotting Cascade)
- Dissolution of Fibrin Clot
- Cell surface interactions at the vascular wall
- Factors involved in megakaryocyte development
- Immune System
- Mitophagy
- Metabolism
- Metabolism of proteins
- Muscle contraction
- Neuronal System
- Organelle biogenesis and maintenance**
- Mitochondrial biogenesis
- Mitochondrial translation
- Assembly of the primary cilium**
- Anchoring of the basal body to the plasma membrane
- Cargo trafficking to the periciliary membrane
- Intraflagellar transport
- ATAT acetylates microtubules
- HDAC6 deacetylates microtubules
- Programmed Cell Death
- Apoptosis
- Regulated Necrosis
- Reproduction
- Signal Transduction

**CILIOGENESIS**

**ASSEMBLY OF THE PRIMARY CILIJUM INtraflagellar Transport and Cargo Trafficking**

ANTEROGRADE TRAFFIC      RETROGRADE TRAFFIC

Description Molecules Structures Expression Analysis Downloads

Analysis results are shown here when an analysis has been run. To start an analysis, click on the Analyse Data button in the top bar.

# The Details Panel - Overview

The screenshot shows the 'Details' panel for a biological interaction. At the top, there are several tabs: Description, Molecules, Structures, Expression, Analysis, and Downloads. The 'Description' tab is active, showing the summation of the interaction: 'Collagen type I binds integrin alpha1beta1, alpha2beta1, alpha10beta1'. Below this, the ID is listed as 'R-HSA-114563' and the species as 'Homo sapiens'. The main content area is divided into sections: Summation, Input, Output, Cellular compartment, Inferred from another species, Authored, Reviewed, and Revised. Each section contains a list of items with a '+' icon to expand them. Orange callout boxes with arrows point to specific parts: 'Background' points to the 'Summation' section; 'Select' points to the 'Input' section; 'Reveal Details' points to the '+' icon in the 'Output' section; 'Orthologues' points to the 'Inferred from another species' section; and 'Key literature' points to the 'Reviewed' section.

Description

Molecules

Structures

Expression

Analysis

Downloads

Collagen type I binds integrin alpha1beta1, alpha2beta1, alpha10beta1

Id: R-HSA-114563

Species: Homo sapiens

**Summation**

Integrin alpha1beta1 binds to collagen type IV and VI with higher affinity than to types I-III, whereas alpha2beta1 has a higher affinity for collagen types I-III than for type IV. Integrin alpha10beta1 binds collagen types I, IV, and VI with similar affinities (Tulla et al. 2001). Integrin alpha11beta1 binds preferentially to the fibril-forming collagen types I and II, binding to type III is weaker and collagens IV and VI are poor ligands (Zhang et al. 2003).

Binding to collagen type I occurs at sites corresponding to the six-residue sequence G(F/L)GER (Knight et al. 1998, 2000, Xu et al. 2000).

Integrin alpha2beta1 is the major platelet collagen receptor (Kunicki et al. 1988). It requires Mg<sup>2+</sup> to interact with collagen and may require initiation mediated by the activation of Integrin alphaiibBeta3 (van de Walle 2007).

► Background literature references...

**Input**

- Collagen type I fibril [extracellular region]
- Mg<sup>2+</sup> [extracellular region]
- Integrin alpha1beta1, alpha2beta1, alpha10beta1 [plasma membrane]

**Output**

- Integrins alpha1beta1, alpha2beta1:Collagen type I fibril:Mg<sup>2+</sup> [plasma membrane]

**Cellular compartment**

- extracellular region
- plasma membrane

**Inferred from another species**

- Collagen type I binds integrin alpha1beta1, alpha2beta1, alpha10beta1 [Homo sapiens, Rattus norvegicus]

**Authored**

- Geiger, B, Horwitz, R, 2008-05-07 08:30:32

**Reviewed**

- Yamada, K, Humphries, MJ, Hynes, R, 2008-05-07 08:53:37
- Ricard-Blum, Sylvie, 2013-08-13

**Revised**

- Jupe, S, 2013-08-13

Background

Select

Reveal Details

Orthologues

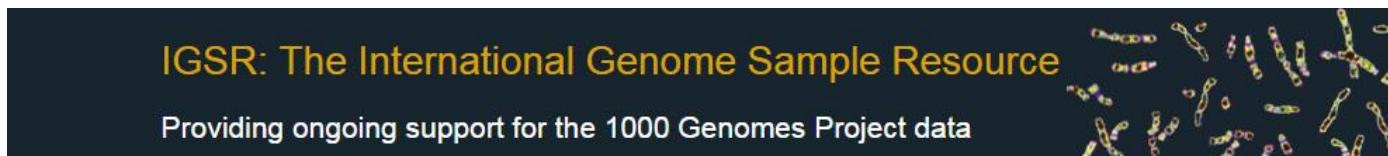
Key literature

# Genotype data catalogue

- HapMap : <http://hapmap.ncbi.nlm.nih.gov/>



- 1000 Genomes Project : [www.1000genome.org](http://www.1000genome.org)



- dbSNP : <http://www.ncbi.nlm.nih.gov/SNP/>



- Variation Viewer : <http://www.ncbi.nlm.nih.gov/variation/view/>



# Single Nucleotide Polymorphisms

- Although any two unrelated people are the same at about 99.9% of their DNA sequences, 0.1% is important because it contains the genetic variants that influence how people differ in their risk of disease or their response to drugs.
- Sites in the genome where the DNA sequences of many individuals differ by a single base are called **single nucleotide polymorphisms** (SNPs).
- 10 million SNPs exist in human populations (rarer SNP allele frequency is at least 1%).
- Alleles of SNPs that are close together tend to be inherited together (associated SNPs).
- A set of associated SNP alleles in a region of a chromosome is called a **haplotype**.
- A chromosome region may contain many SNPs, but only a few **tag SNPs** can provide most of the information on the pattern of genetic variation in the region.