OICR
Ontario Institute for Cancer Research

EMBL-EBI

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

REACTOME
A CURATED PATHWAY DATABASE

Linking biological pathways, networks and disease

Robin Haw

19th August 2014

COMBINE 2014 Meeting

www.reactome.org

International Cancer Genome Consortium

Ontario

Ministry of Research and Innovation

National Human Genome Research Institute

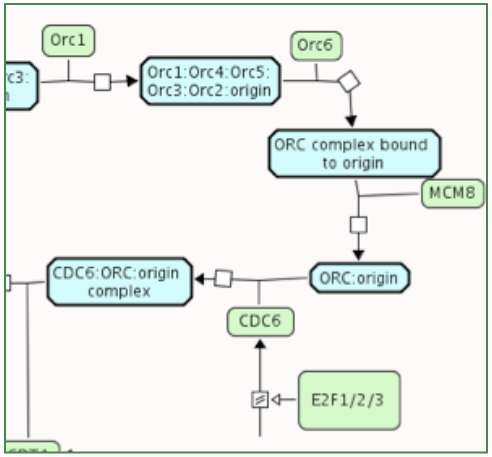
National Institutes of Health

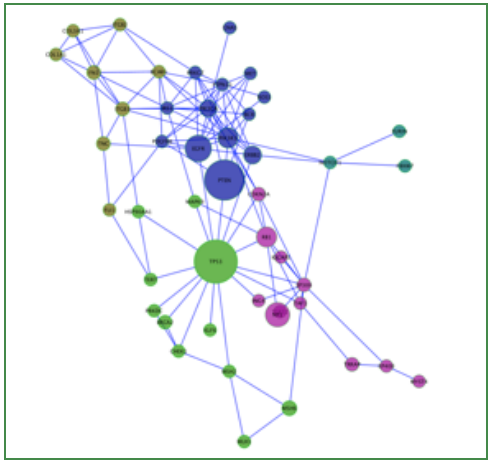
Overview

- Content
  - Controlled Vocabularies
  - Disease Pathway Curation
- Visualization and Analysis
  - New Pathway Browser
- Exchange
  - SBML, SBGN, BioPAX, PSIQUIC

What is Reactome?

- Open source and open access pathway database
- 1500+ pathway modules encompassing many areas of human biology.
  - V49 has annotations for 7684 human proteins, 7332 reactions, and 1462 small molecules, based on data from 16900 literature references.
- Provides tools and datasets for browsing and visualizing pathway data.





www.reactome.org

Controlled vocabularies for entities and events

- Created unique, unambiguous names for pathway events and molecular entities.
- Significantly improves consistency and readability of names.
- Benefits for searching and data mining within and between databases.
- Reduced curation burden.

Peptide CV Names

- Gene Symbol Core - HGNC approved gene symbols
- Peptide coordinates suffix
  - Refers to UniProt 'Chain' Feature
  - e.g. large and small subunits of CASP9 - **CASP9(1-315)** and **CASP9(316-416)**
- Post-translational modification (PTM) prefixes
  - Abbreviated from PSI-MOD
  - Includes coordinate of PTM (if known)
- Phosphorylation subtypes grouped as one class, ordered by coordinate.
  - DAPP1 phosphorylated on tyrosine-139 = **p-Y139-DAPP1**
  - WASF2 phosphorylated on tyrosine-150, serine-343 and threonine-346 = **p-Y150,S343,T346-WASF2**
  - GAB2 tyrosine-phosphorylated at unknown coordinate = **p-Y-GAB2**
  - GLI3 phosphorylated, unknown residue = **p-GLI3**

Complex and set CV names

- Concatenated string of component or set member names.
- Comma separator for sets, colon for complexes.
  - G protein-activated inward rectifier potassium channels = **KCNJ3,KCNJ5,KCNJ6,KCNJ9**
  - Complex of IL3, IL3RA, IL3RB, and JAK2 = **IL3:IL3RA:IL3RB:JAK2**
- Entity occurs more than once, name is preceded by *n*x
  - Complex of 2 molecules of PPOX and one of FAD = **2xPPOX:FAD**
  - *n*mer names e.g. dimer, hexamer allowed for homomers
- 'Candidates' (possible members) named in round brackets
  - Set of HRH2, HRH3, plus possible 'candidate' members HRH6 and HRH8 = **HRH2,HRH3,(HRH6,HRH8)**
- Precedence or hierarchical structure indicated with square brackets
  - Complex ABC1:ABC2 binds complex ABC3:ABC4 = **[ABC1:ABC2]:[ABC3:ABC4]**

Small molecule (chemical) CV names

- Standardised abbreviation for >1400 small molecules in Reactome
- Sourced from ChEBI, KEGG Compounds, PubChem, or literature
- Checked for precedence/ambiguity at http://www.allacronyms.com
- e.g. Calcium ion = **Ca2+**
- adenosine triphosphate = **ATP**
- Diacylglycerol = **DAG**
- 4-(4-(dimethylamino)styryl)-N-methylpyridinium = **4-Di-2-ASP**
- D-Glyceraldehyde 3-phosphate = **GA3P**

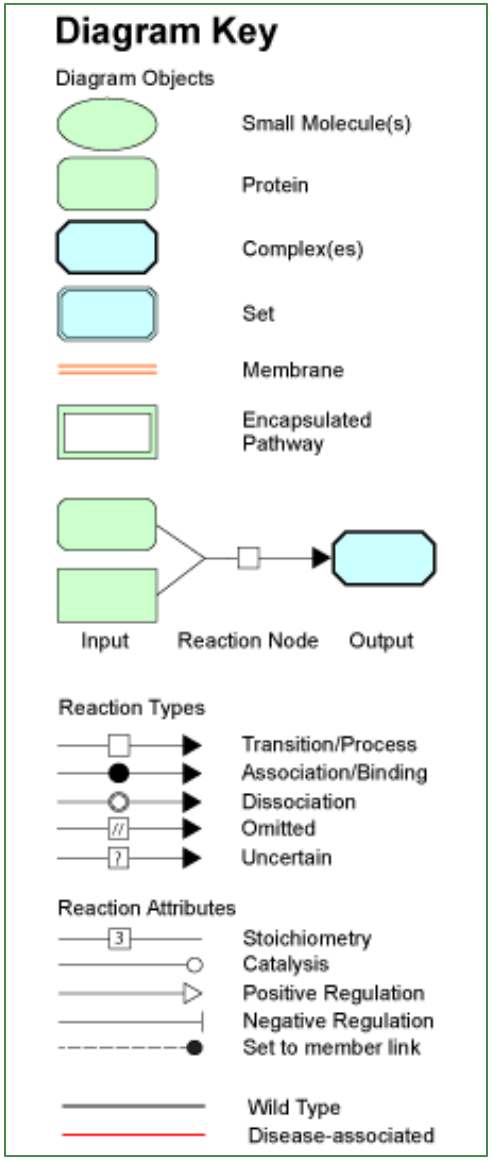
Pathway event (reaction) CV names

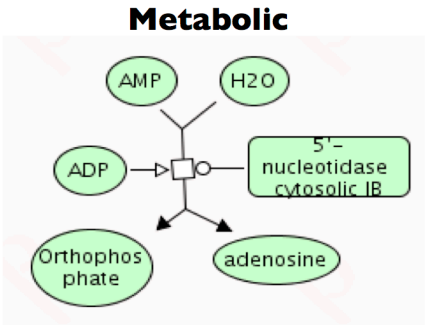
- A small set of CV terms - can be applied automatically using simple rules:

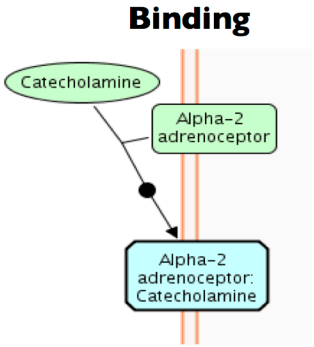
– Transformation (default)	a (,b and c) **TRANSFORMS TO** d (,e and f)
– Binding, Dissociation	a BINDS b forming c, c **DISSOCIATES TO** a **AND** b
– Polymerization	a (,b and c) **POLYMERIZE TO** x
– Transfer reactions	a **TRANSFERS** b **TO** c1 (**TO FORM** c2)
– Passive transport	a **TRANSLOCATES FROM** [x] **TO** [y]
– Active transport	a **TRANSPORTS** b (**FROM** x **TO** y)
– Antiporter	a **EXCHANGES** b **FOR** c (across x membrane)
– Cotransporter	a **COTRANSPORTS** b (,c) **WITH** d (,e)
– Activation (conformational)	a (,b) **IS (ARE) ACTIVATED**
– Catalysis* (default)	a **CATALYZES** b (,c and d) **TO** e (,f and g)

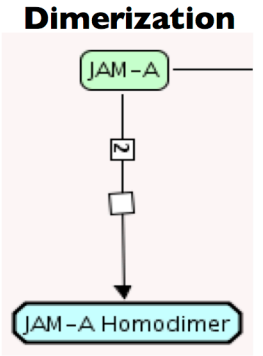
*Catalyst with defined GO molecular function

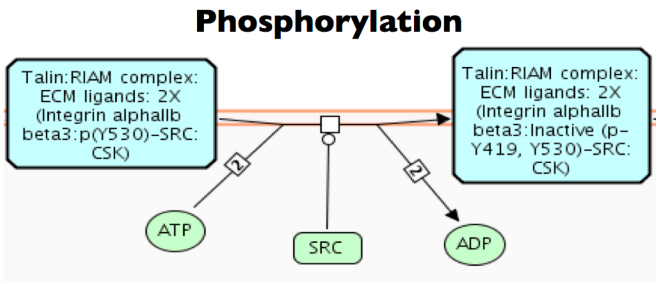
Reactome supports minimal SBGN PD glyphs





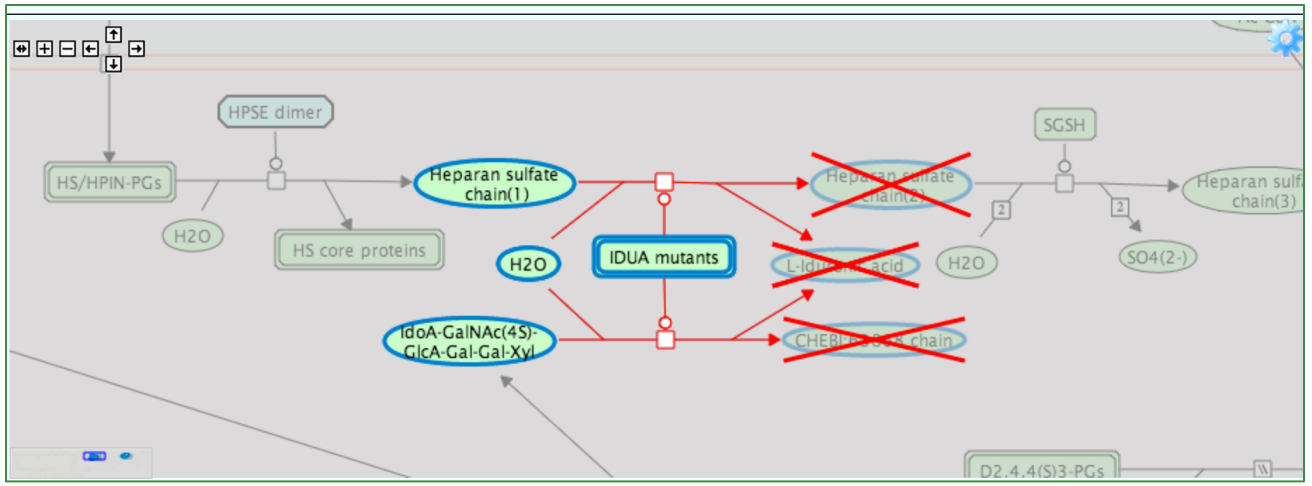




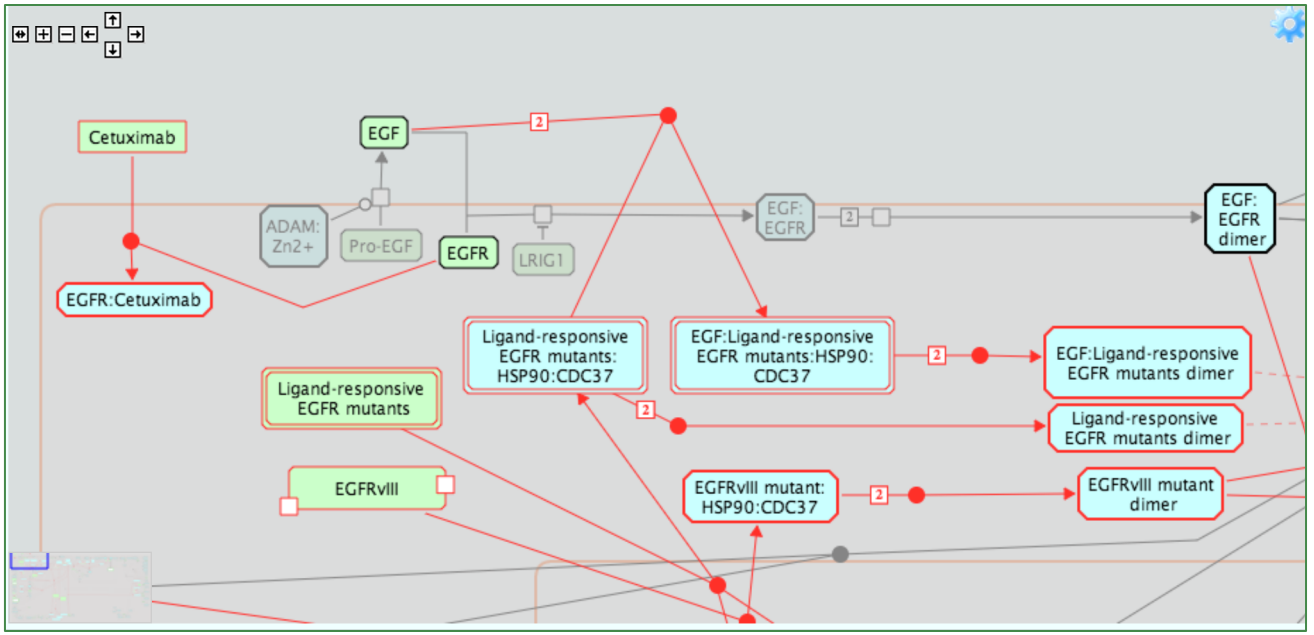


Disease Pathway Curation

Monofactorial genetic disease



Polygenic disease

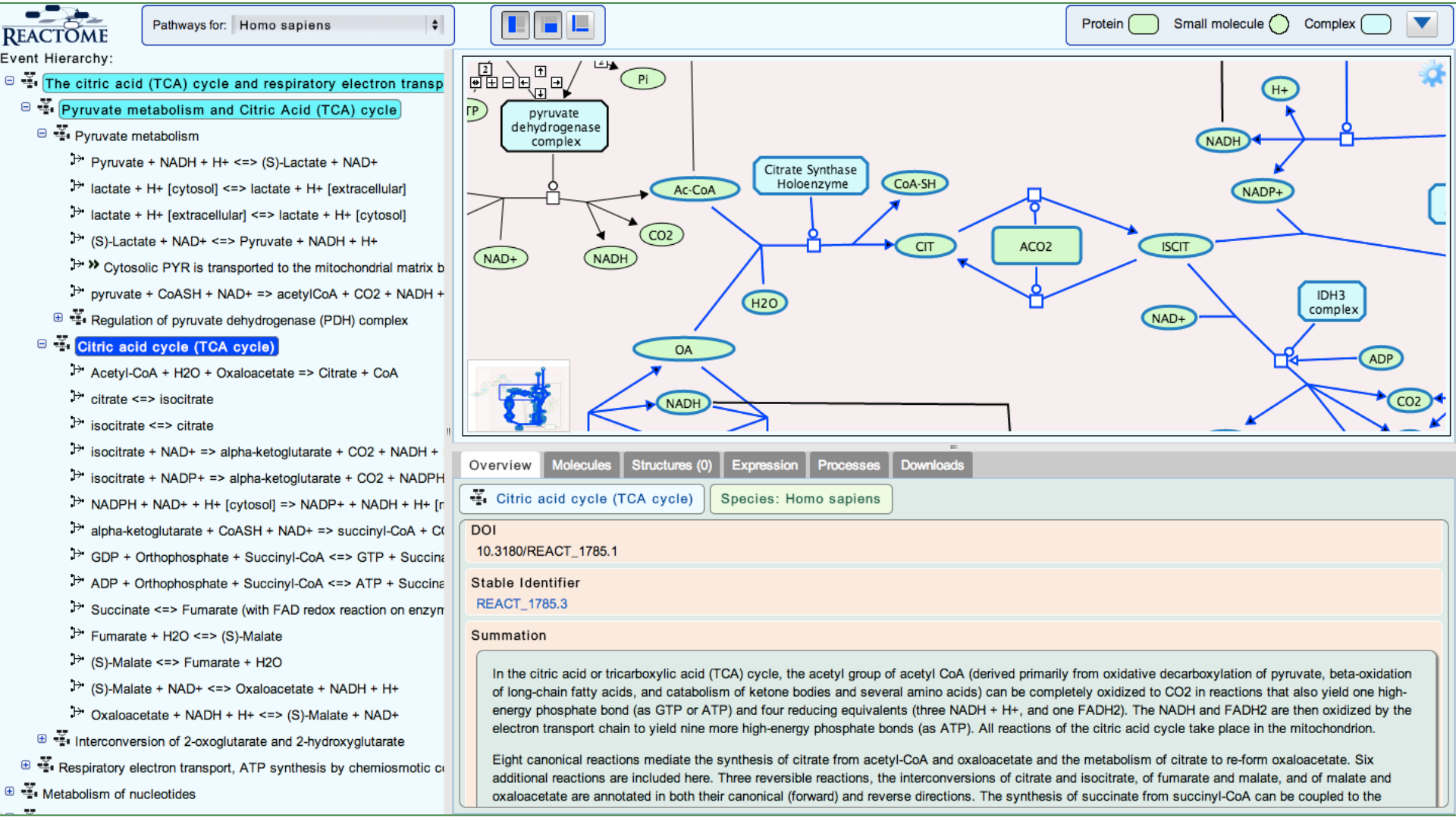


OICR Ontario Institute for Cancer Research

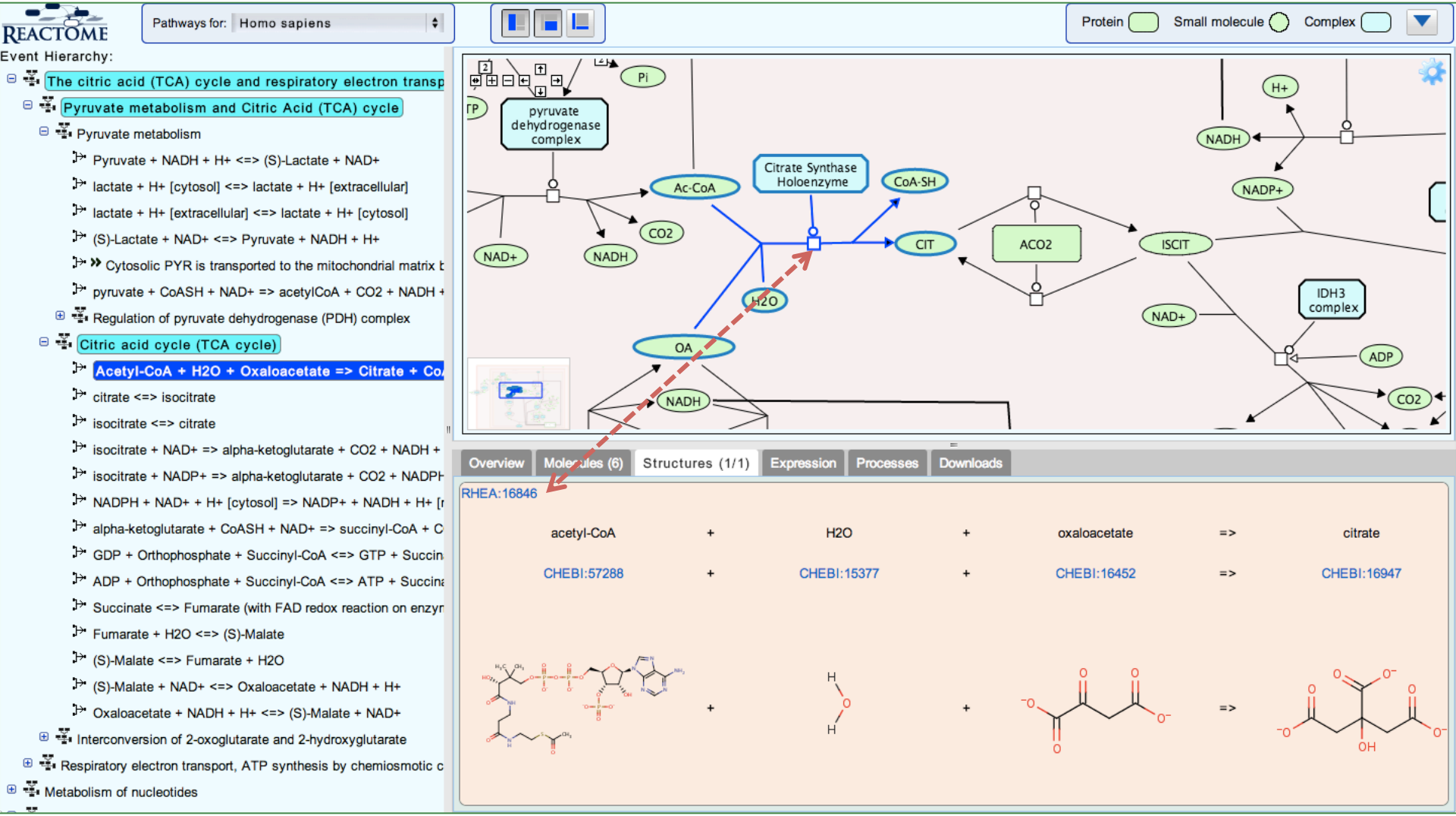
REACTOME A CURATED PATHWAY DATABASE

SBGN Pathway Browser

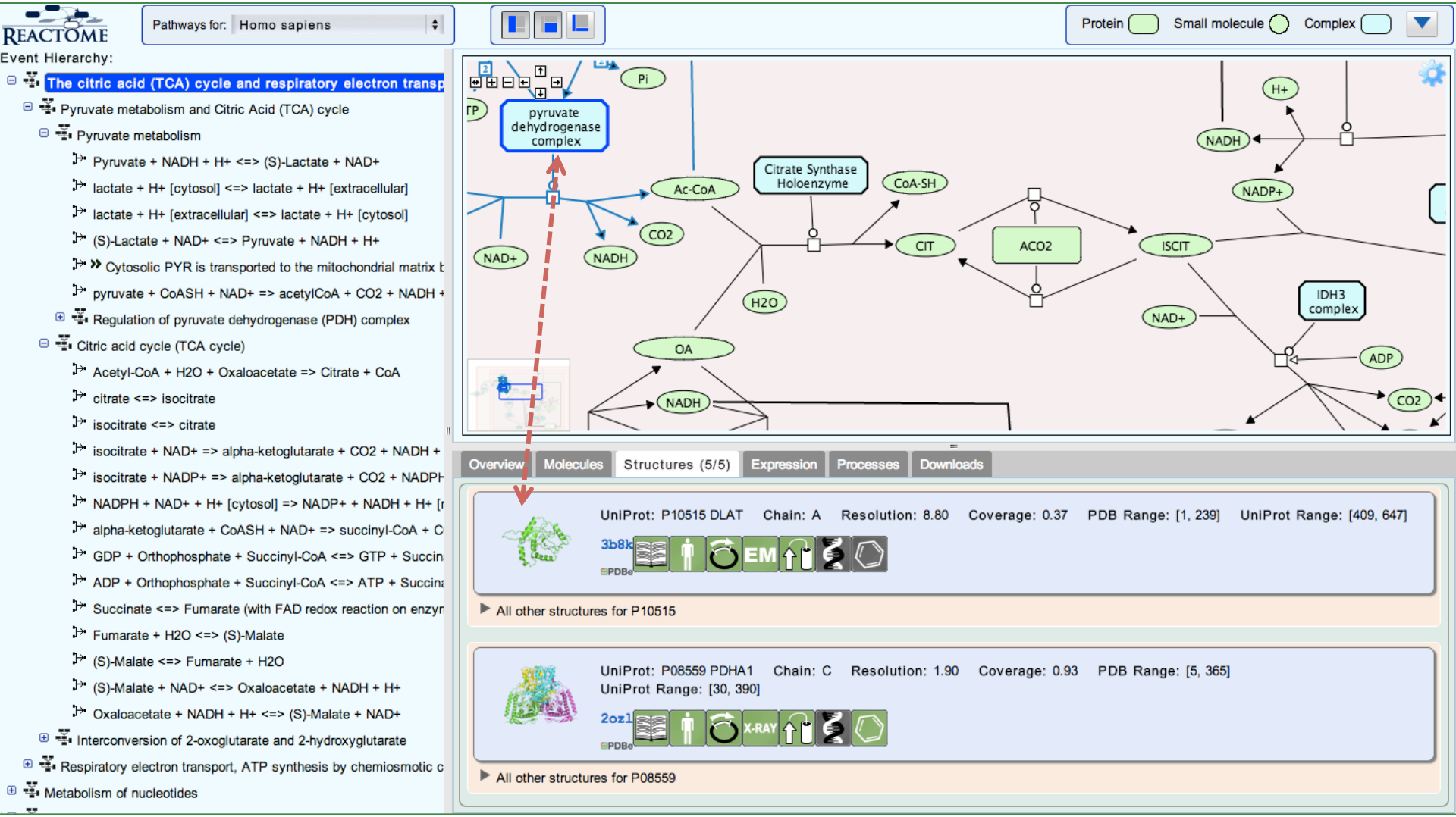
- Google-map style pathway diagrams



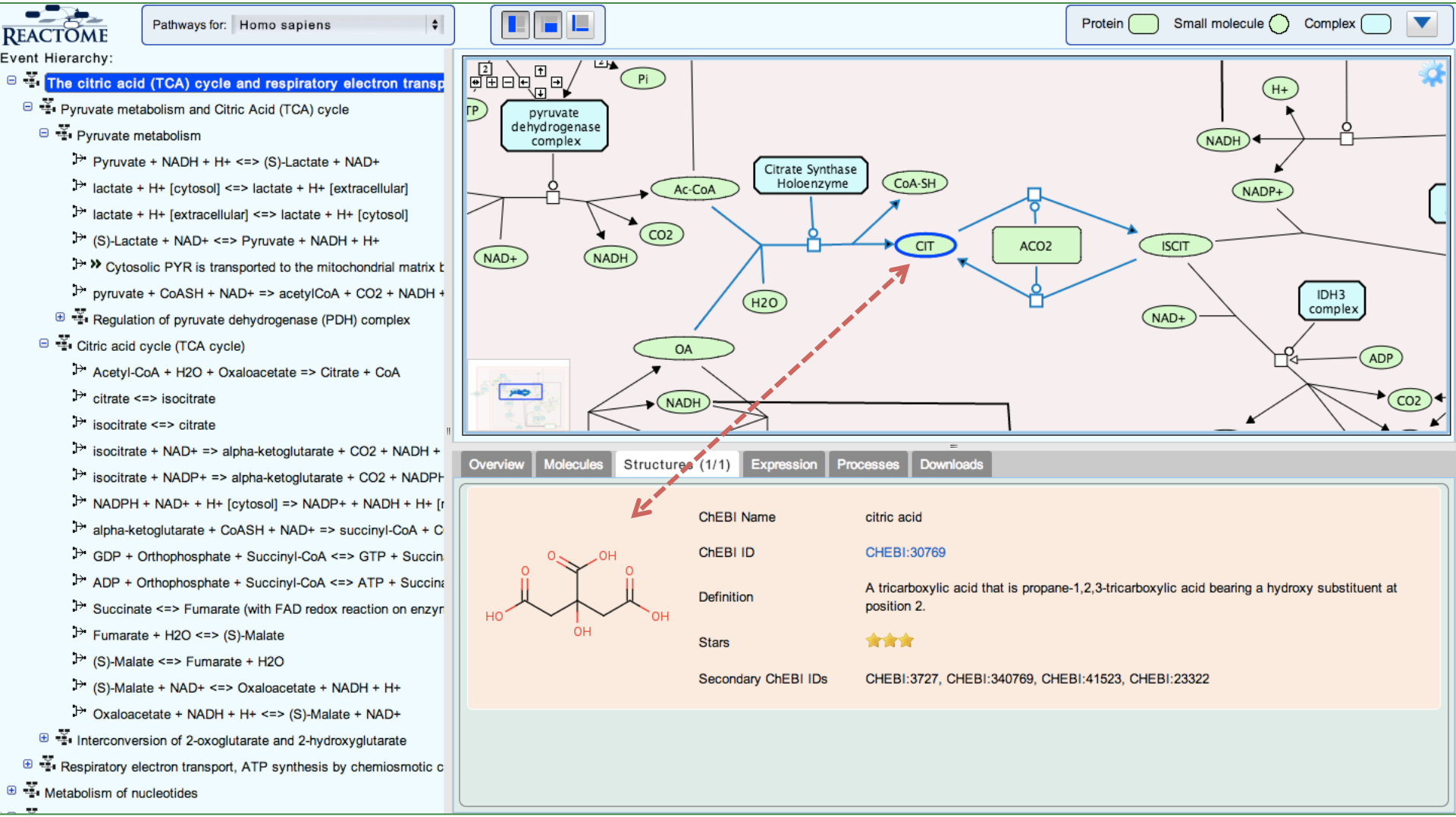
Balanced reactions from Rhea



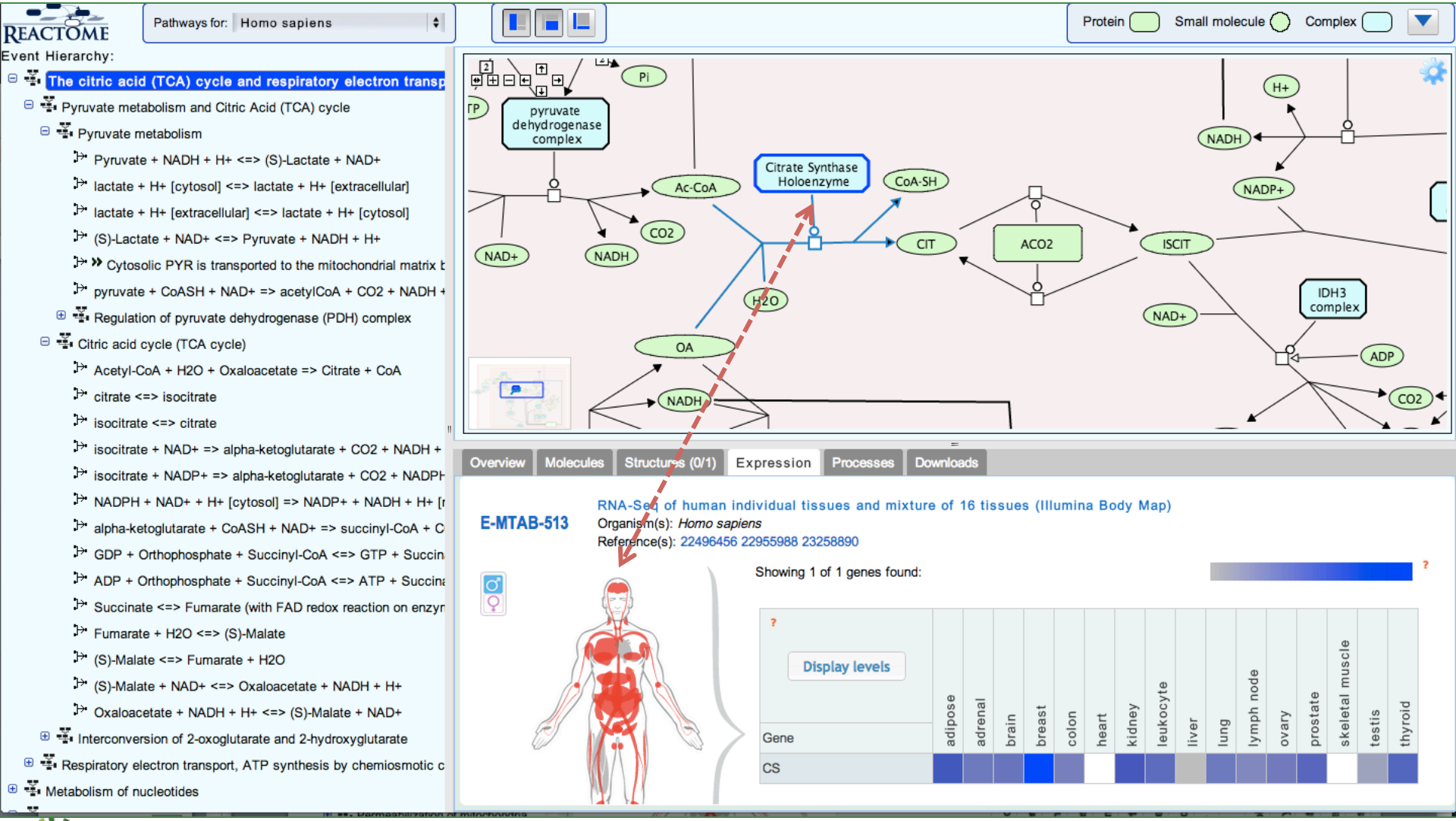
Protein structures from PDBe



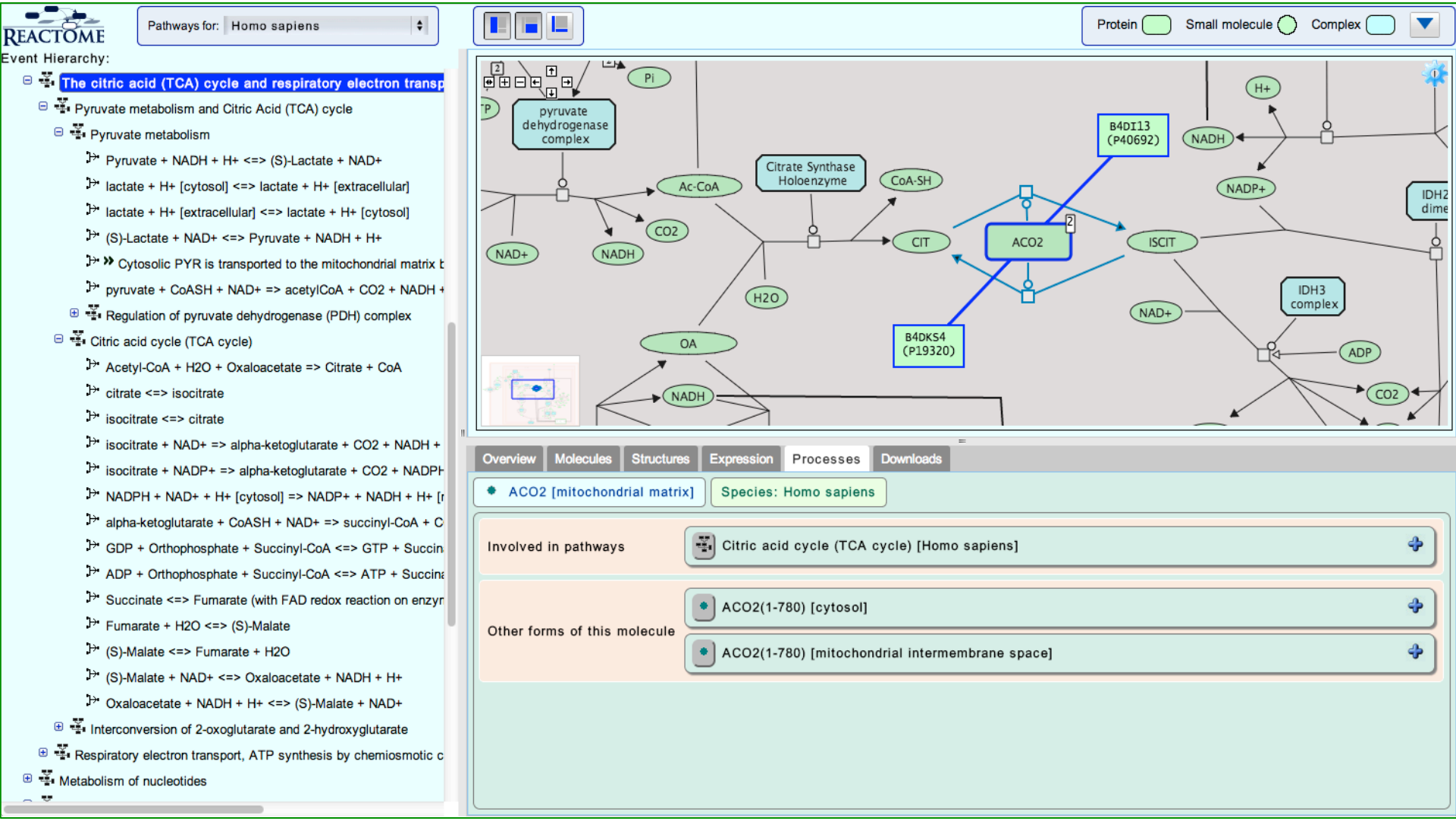
Chemical structures from ChEBI



Expression data from the Expression Atlas



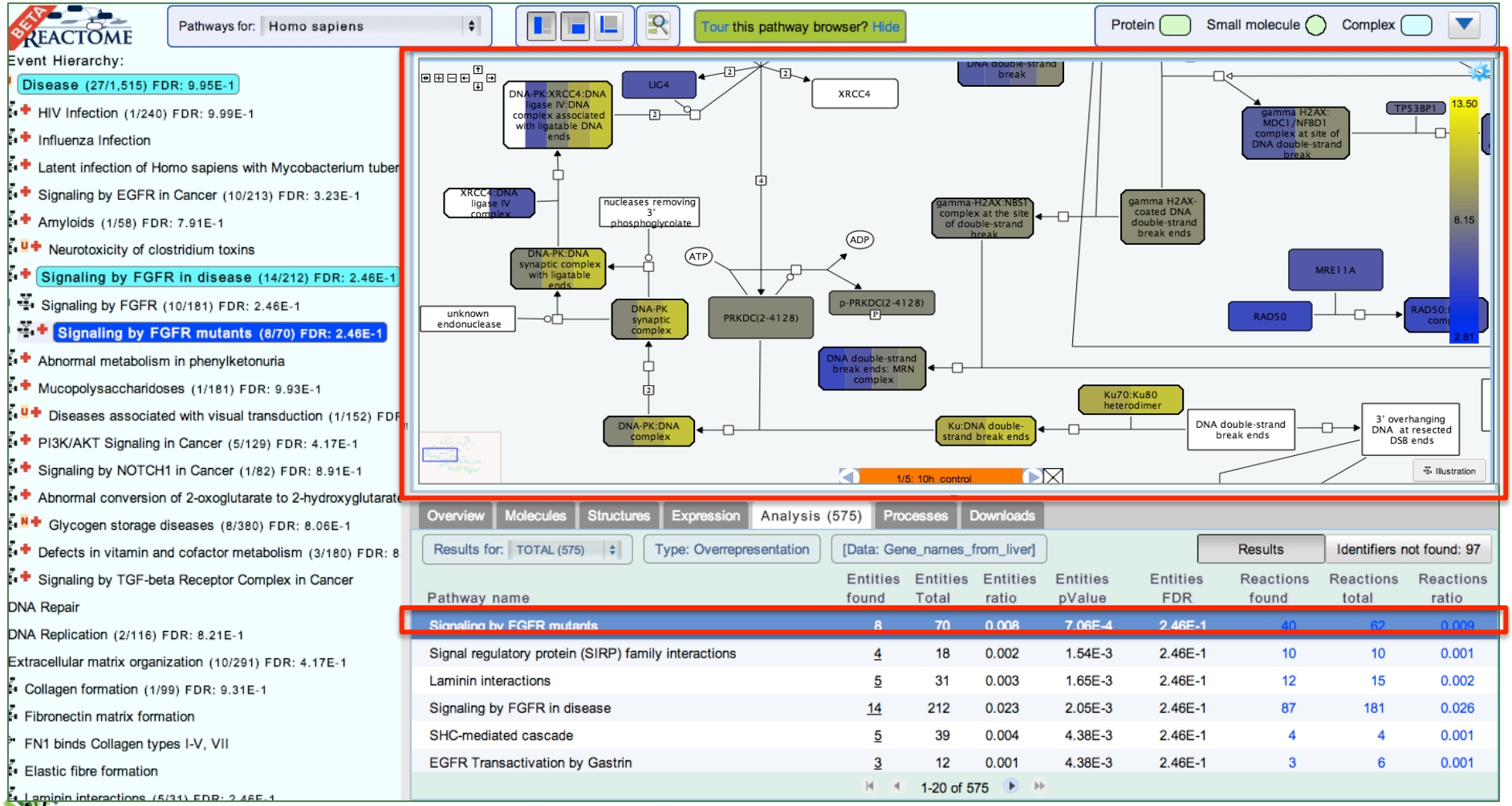
PSICQUIC - Interaction Data from STRING



Pathway Diagrams support Reactome Tools

- Pathway Mapping and Enrichment Analysis
- Expression Overlay onto Pathways
- Compare Pathways between Species

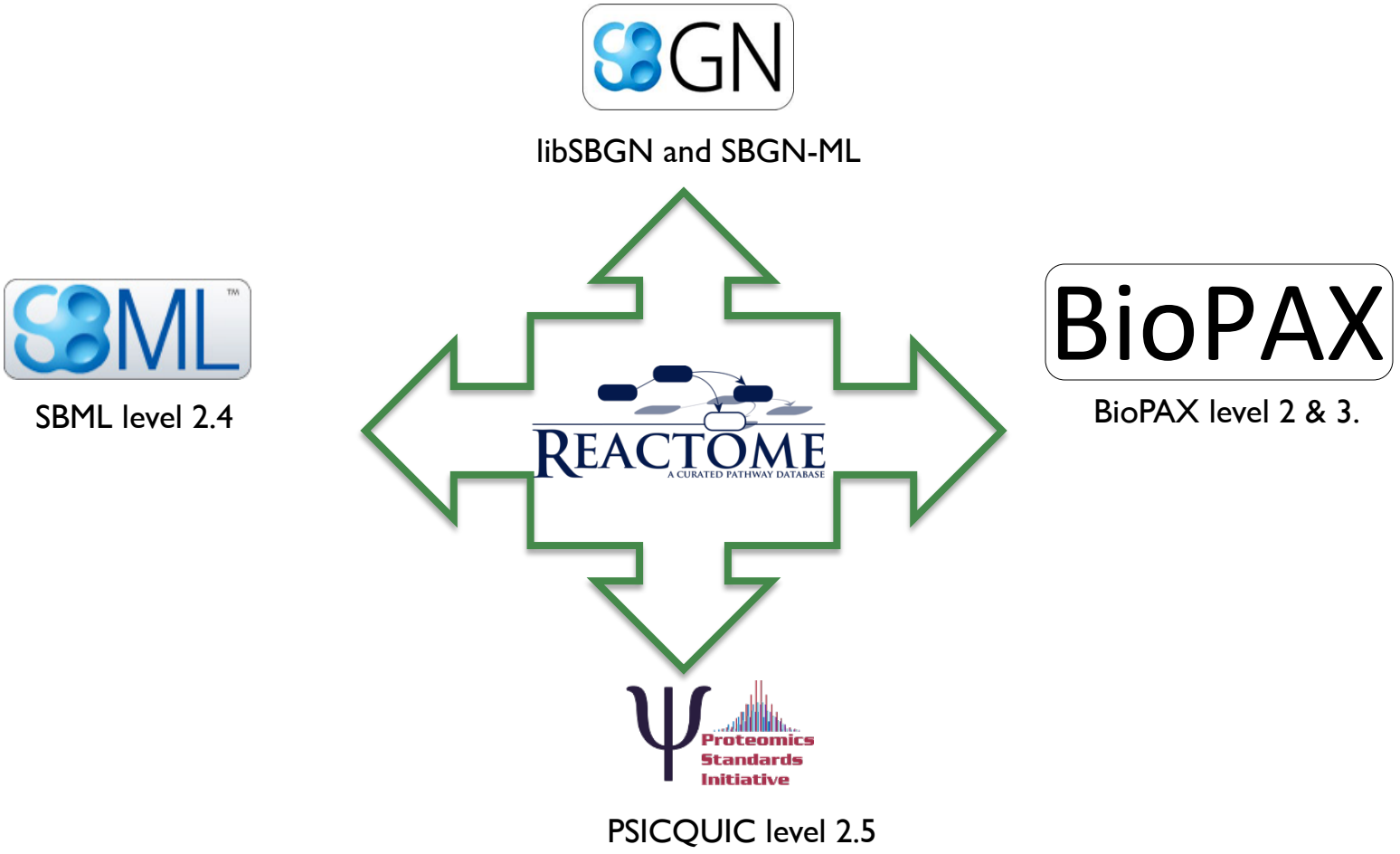
In 2 months reached 6,622 analysis (15/09/14)



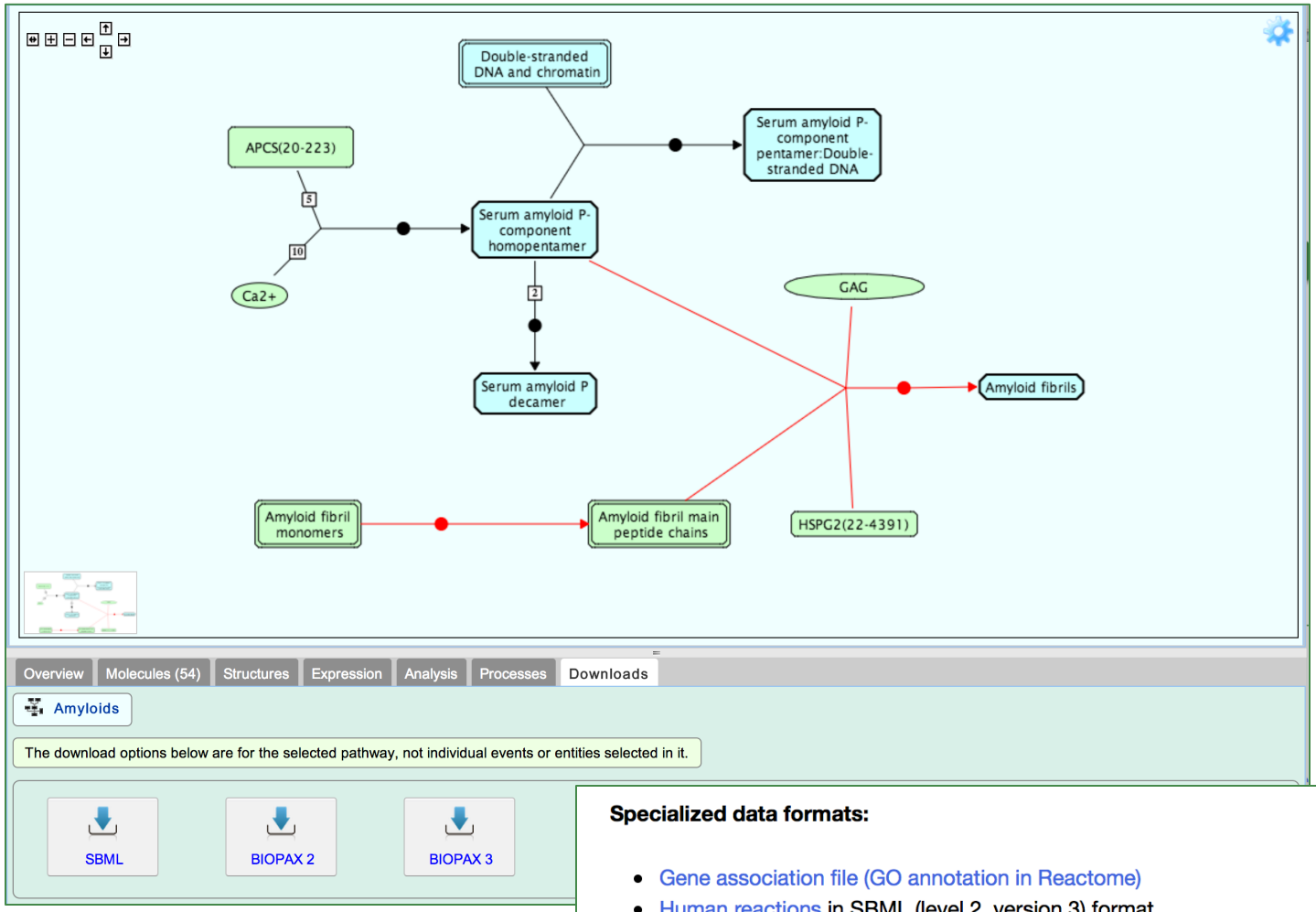
OICR Ontario Institute for Cancer Research

REACTOME A CURATED PATHWAY DATABASE

Reactome supports Open Data Standards



Single vs Bulk Pathway Downloads



Specialized data formats:

- Gene association file (GO annotation in Reactome)
- Human reactions in SBML (level 2, version 3) format.
- Events in the BioPAX level 2 format.
- Events in the BioPAX level 3 format.
- Reactome Pathways Gene Set.
- Protégé 2.0 ontology for Reactome data model: pins file, pont file, pprj file.
- Human pathway diagrams: PDF, or PNG

Reactome Web Services

```
<?xml version='1.0' encoding='UTF-8' standalone='no'?>
<sbml xmlns="http://www.sbml.org/sbml/level2/version4" level="2" version="4">
    <model id="pathway_977225" name="Amyloids" metaid="metaid_0">
        <notes><p xmlns="http://www.w3.org/1999/xhtml">Amyloid is a term used to describe typically extracellular deposits of aggregated
            proteins, sometimes known as plaques. Abnormal accumulation of amyloid is amyloidosis, a term associated with diseased organs
            and tissues, particularly neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntingdon's. Amyloid deposits consist
            predominantly of amyloid fibrils, rigid, nonbranching structures that form ordered assemblies, characteristically with a cross
            betasheet structure where the sheets run parallel to the direction of the fibril (Sawaya et al. 2007). Often the fibril has a
            lefthanded twist (Nelson    Eisenberg 2006). At least 27 human proteins form amyloid fibrils (Sipe et al. 2010). Many of these
            proteins have nonpathological functions; the trigger that leads to abnormal aggregations differs between proteins and is not
            well understood but in many cases the peptides are abnormal fragments or mutant forms arising from polymorphisms, suggesting
            that the initial event may be aggregation of misfolded or unfolded peptides. Early studies of AmyloidBeta assembly led to a
            widely accepted model that assembly was a nucleationdependent polymerization reaction (Teplow 1998) but it is now understood to
            be more complex, with multiple 'offpathway' events leading to a variety of oligomeric structures in addition to fibrils
            (Roychaudhuri et al. 2008). An increasing body of evidence suggests that these oligomeric forms are primarily responsible for
            the neurotoxic effects of Amyloidbeta (Roychaudhuri et al. 2008), alphasynuclein (Winner et al. 2011) and tau (Dance    Strobel
            2009, MerazRios et al. 2010). Amyloid oligomers are believed to have a common structural motif that is independent of the
            protein involved and not present in fibrils (Kayed et al. 2003). Conformation dependent, aggregation specific antibodies suggest
            that there are 3 general classes of amyloid oligomer structures (Glabe 2009) including annular structures which may be
            responsible for the widely reported membrane permeabilization effect of amyloid oligomers. Toxicity of amyloid oligomers
            preceeds the appearance of plaques in mouse models (Ferretti et al. 2011). </p><p xmlns="http://www.w3.org/1999/xhtml">Fibrils
            are often associated with other molecules, notably heparan sulfate proteoglycans and Serum Amyloid Pcomponent, which are
            universally associated and seem to stabilize fibrils, possibly by protecting them from degradation.</p></notes>
        <annotation>
            <p xmlns="http://www.w3.org/1999/xhtml">SBML engine: Jsbml</p><sbgn xmlns="http://sbgn.org/libsbgn/0.2"><map language="process
                description"><glyph class="macromolecule" id="entityVertex_7193679"><label text="APCS_20_223"/><bbox h="36.0" w="168.0"
                x="273.0" y="267.0"/><glyph class="unit of information" id="entityVertex_7193679_mt"><label text="mt:prot"/><bbox h="24.0"
                w="56.0" x="289.0" y="255.0"/></glyph></glyph><glyph class="simple chemical" id="entityVertex_7193680"><label text="Ca2"/><
                bbox h="36.0" w="90.0" x="282.0" y="675.0"/></glyph><glyph class="complex" id="entityVertex_7193681"><label text="Serum"/><
                bbox h="45.0" w="168.0" x="930.0" y="468.0"/><glyph class="annotation" id="entityVertex_7193681_annotation"><label text=
                "APCS(20-223), Ca2+"/><callout target="entityVertex_7193681"><point x="1098.0" y="513.0"/></callout><bbox h="24.0" w="144.0"
                x="1114.0" y="537.0"/></glyph></glyph><glyph class="unspecified entity" id="entityVertex_7193682"><label text=
                "Amyloid_fibril"/><bbox h="36.0" w="133.5" x="354.0" y="1281.0"/></glyph><glyph class="unspecified entity" id=
                "entityVertex_7193683"><label text="Amyloid_fibril_main"/><bbox h="36.0" w="181.5" x="1251.0" y="1281.0"/></glyph><glyph
                class="simple chemical" id="entityVertex_7193684"><label text="GAG"/><bbox h="36.0" w="
                glyph><glyph class="complex" id="entityVertex_7193685"><label text="Amyloid_fibrils"/><
                y="933.0"/><glyph class="annotation" id="entityVertex_7193685_annotation"><label text="
                Serum amyloid P-component homopentamer, HSPG2(22-4391), GAG"/><callout target="entityVe
                y="969.0"/></callout><bbox h="24.0" w="760.0" x="2450.5" y="993.0"/></glyph></glyph><gl
                "entityVertex_7193686"><label text="HSPG2_22_4391"/><bbox h="36.0" w="166.5" x="1791.0"
                information" id="entityVertex_7193686_mt"><label text="mt:prot"/><bbox h="24.0" w="56.0
                glyph><glyph class="complex multimer" id="entityVertex_7193687"><label text="Serum"/><b
                y="930.0"/><glyph class="annotation" id="entityVertex_7193687_annotation"><label text="
                homopentamer"/><callout target="entityVertex_7193687"><point x="1099.5" y="966.0"/></ca
                x="1115.5" y="990.0"/></glyph></glyph><glyph class="unspecified entity" id="entityVerte
                "Double_stranded"/><bbox h="36.0" w="183.0" x="915.0" y="42.0"/></glyph><glyph class="c
                label text="Serum"/><bbox h="57.6" w="175.5" x="1671.0" y="219.0"/><glyph class="annota
                "entityVertex_7193689_annotation"><label text="Serum amyloid P-component homopentamer,
                callout target="entityVertex_7193689"><point x="1846.5" y="276.6"/></callout><bbox h="2
                ></glyph></glyph><glyph class="process" id="reactionVertex_7193690"><bbox h="36.0" w="3
```

Reactome Web services

Our new RESTful API provides outside users with direct access to pathway data in Reactome. The Reactome pathway analysis tools are also available for integration into third party websites.

- RESTful API is available to access the Reactome data (Note: The RESTful API has been moved to reactomews.oicr.on.ca for better performance.). For details about this API, please see this document: Reactome RESTful API.
- SOAP based Web Services API is available to access the Reactome data. For details about this API, please follow the following links:
  - Simple Description for the Reactome Web Services API.
  - Training Materials for the Reactome Web Services API
    - Reactome SOAP WS User's Guide in PDF (1M).
    - Reactome SOAP WS Tutorial in Power Point Slides (2M).
    - Reactome SOAP WS Tutorial in Flash Movie (640 x 480) (11M).
    - Reactome SOAP WS Tutorial in Flash Movie (800 x 600) (12M).
    - XML Schema for the data model.
    - WSDL file for the Reactome Web Servies API.

http://reactomews.oicr.on.ca:8080/ReactomeRESTfulAPI/RESTfulWS/sbmlExporter/977225

OICR Ontario Institute for Cancer Research

REACTOME A CURATED PATHWAY DATABASE

Current and Future Work

- ORCID iD as a key mechanism for credit attribution for Reactome contributors.
- Supplement normal pathways with variant reactions representing disease states.
- Improve annotation consistency and enrich the data model.
- Work with the biocuration community for improved ontology support.
- Continued support for SBGN, SBML, BioPAX and PSI data exchange.
  - SBGN Single pathway and bulk pathways downloads.
  - Custom exports for SBML, BioPAX, SBGN via RESTful API.
- Improving the web site and resources to meet the needs of a growing and diverse user community.

Acknowledgements

- Michael Caudy
- David Croft
- Eric Dawson
- Adrian Duong
- Phani Garapati
- Marc Gillespie
- Bijay Jassal
- Steve Jupe
- Irina Kalatskaya
- Maulik Kamdar
- Bruce May
- Sheldon MacKay
- Lisa Matthews
- Antonio Fabregat Mundo
- Marija Orlic-Milacic
- Karen Rothfels
- Veronica Shamovsky
- Heeyeon Song
- Joel Weiser
- Mark Williams
- Guanming Wu
- Christina Yung

- **Henning Hermjakob**
- **Peter D'Eustachio**
- **Lincoln Stein**