A complex network diagram in the background, composed of numerous small blue dots connected by thin, curved lines, resembling a molecular or cellular structure.

The Reactome Pathway Knowledgebase

Robin Haw, Joel Weiser, Solomon Shorser
and Justin Cook

Ontario Institute for Cancer Research
BCC2020, July 17-18th 2020



reactome

Outline

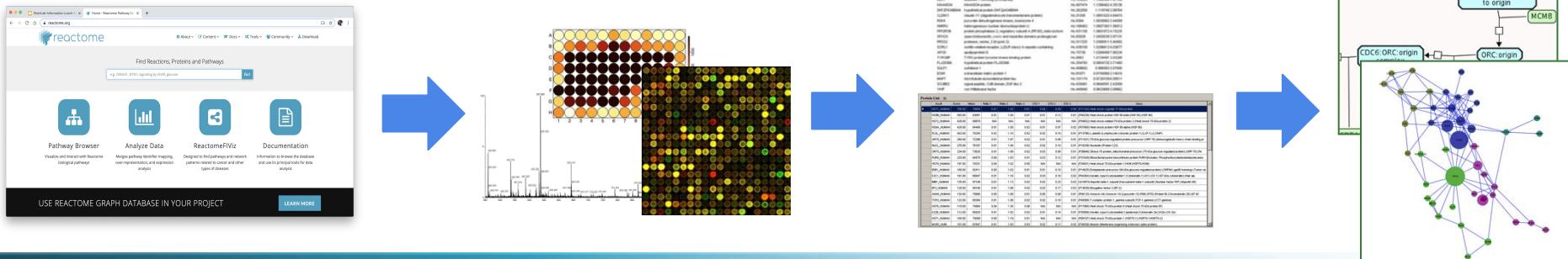
- Introduction to reactome.org
 - Robin Haw
- Content Service
 - Solomon Shorser
- Analysis Service
 - Justin Cook
- Graph Database
 - Joel Weiser
- Questions and Answers

Motivation for Pathway Databases and Analysis

- Intuitive to Scientists:
 - Intuitive display for biological data.
 - Visualize multiple data types.
 - Computational methods available to automate analysis.
- Pathway Analysis satisfy common “use cases” in biological research:
 - Identifying hidden patterns in gene lists.
 - Predicting the function of unannotated genes.
 - Establishing the framework for quantitative modeling.
 - Assisting in the development of molecular signatures.

What is Reactome?

- Open source and open access pathway knowledgebase
 - Dates to 2002 @ CSHL & EBI
 - Human pathways encompassing metabolism, signaling, gene regulation, and other biological processes
 - Every pathway is traceable to primary literature
 - Extensively cross-referenced to external bioinformatics databases
 - Provides software tools for browsing and visualizing pathway data and analyzing user data



 reactome

EMBL-EBI

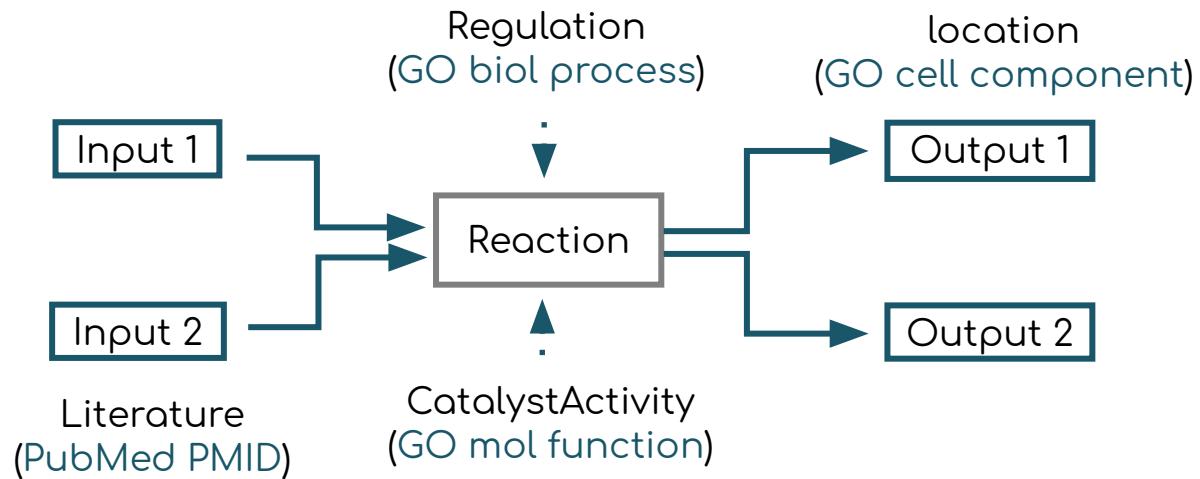
 NYU Langone
MEDICAL CENTER



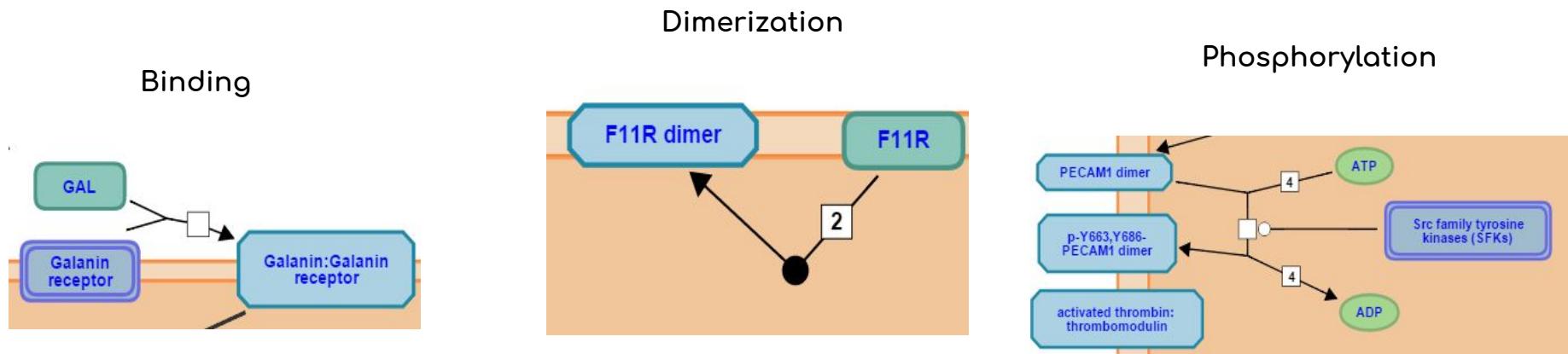
Data Model

- Reactome is a reaction-based database
- Explicitly describe biological processes as a series of biochemical reactions and events

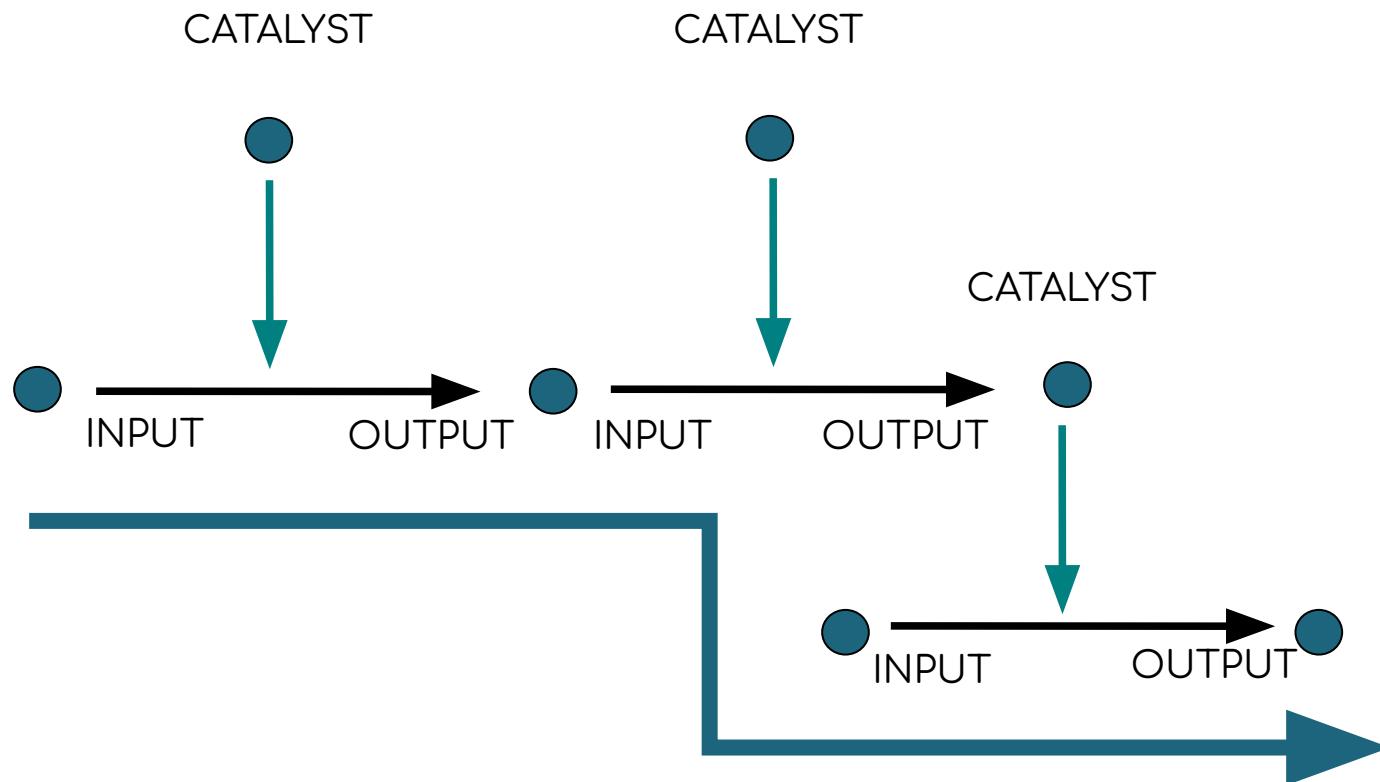
protein (UniProt)
molecule (ChEBI)
complex (GO/PRO)
ncRNA (miRBase)
disease variant (ClinGen, COSMIC)
drug (ChEBI, IUPHAR)



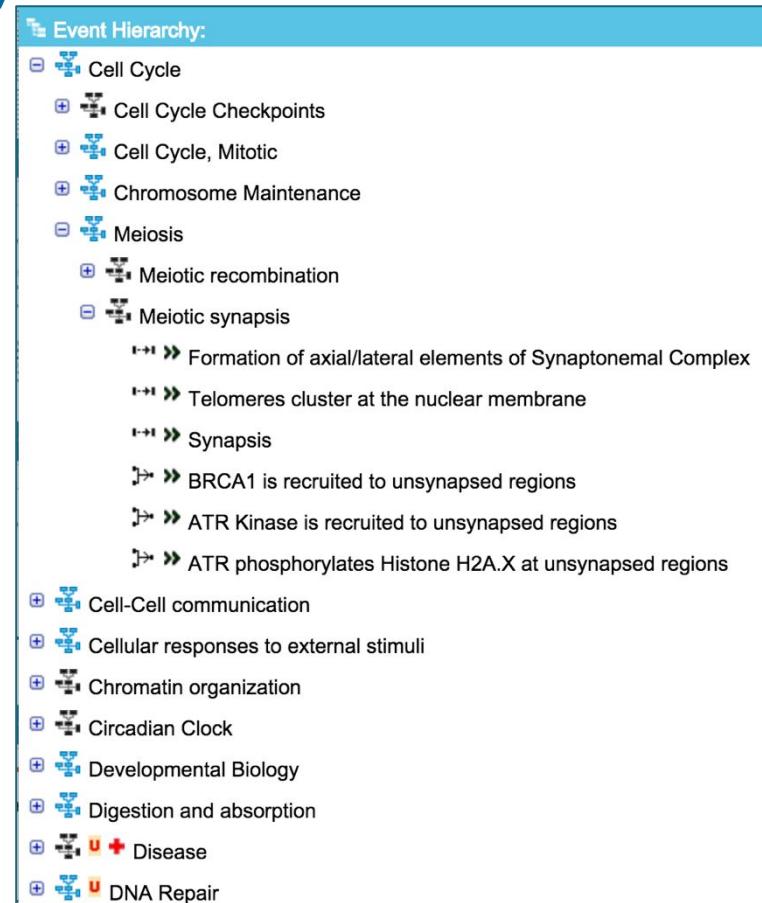
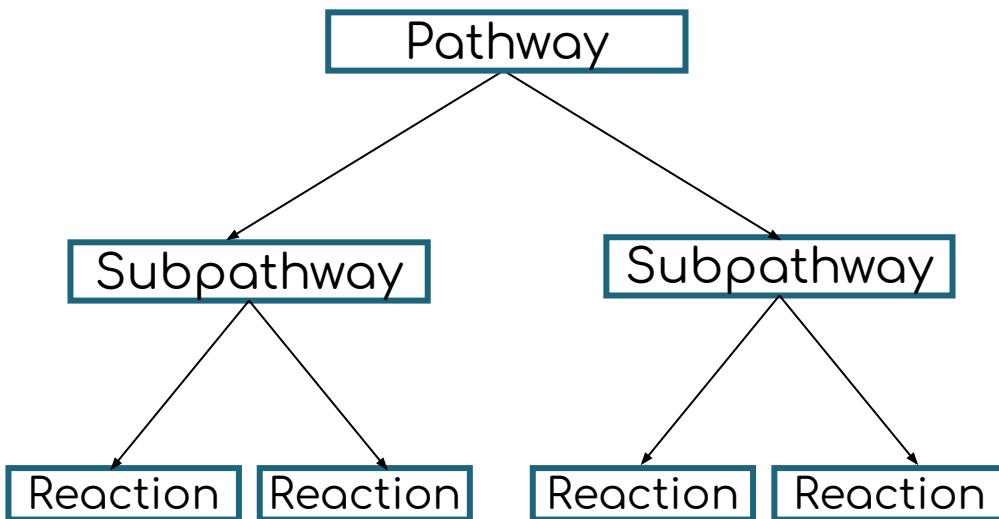
Reactome supports multiple types of reaction



Reactions connect to one another

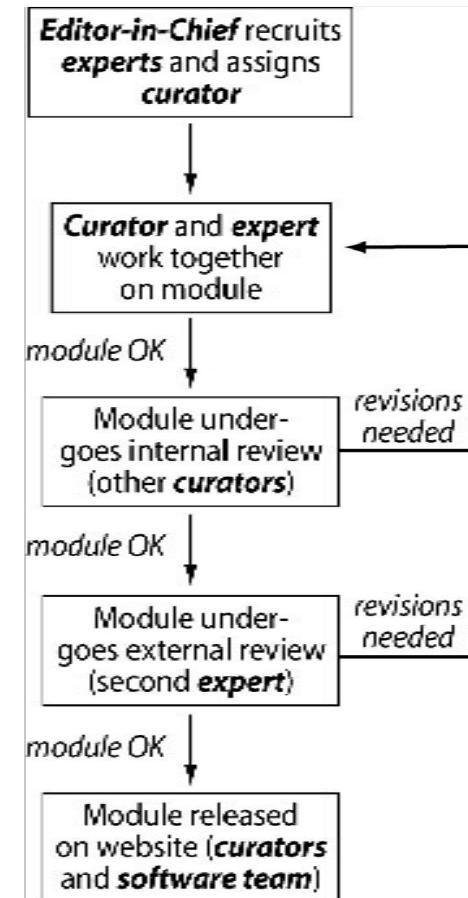


Hierarchical arrangements for pathways

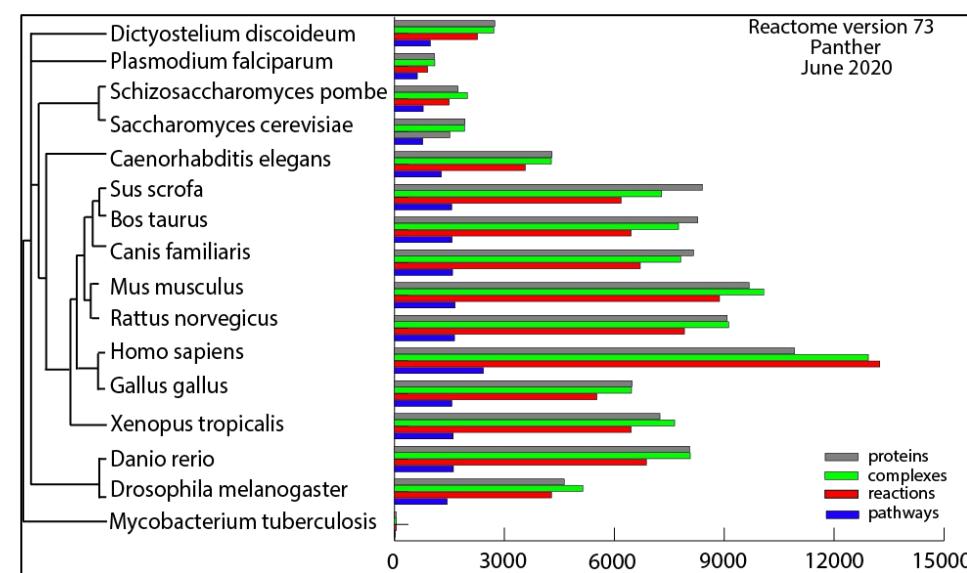
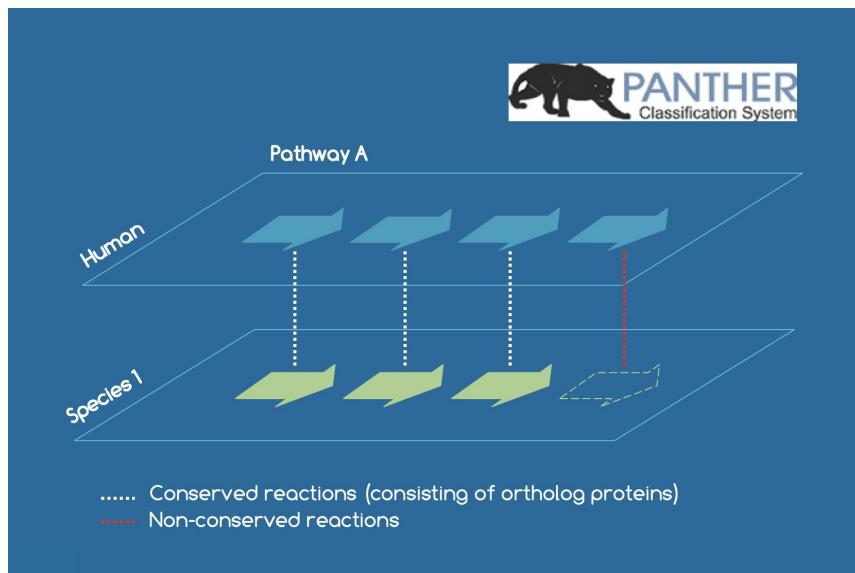


Where does the data comes from?

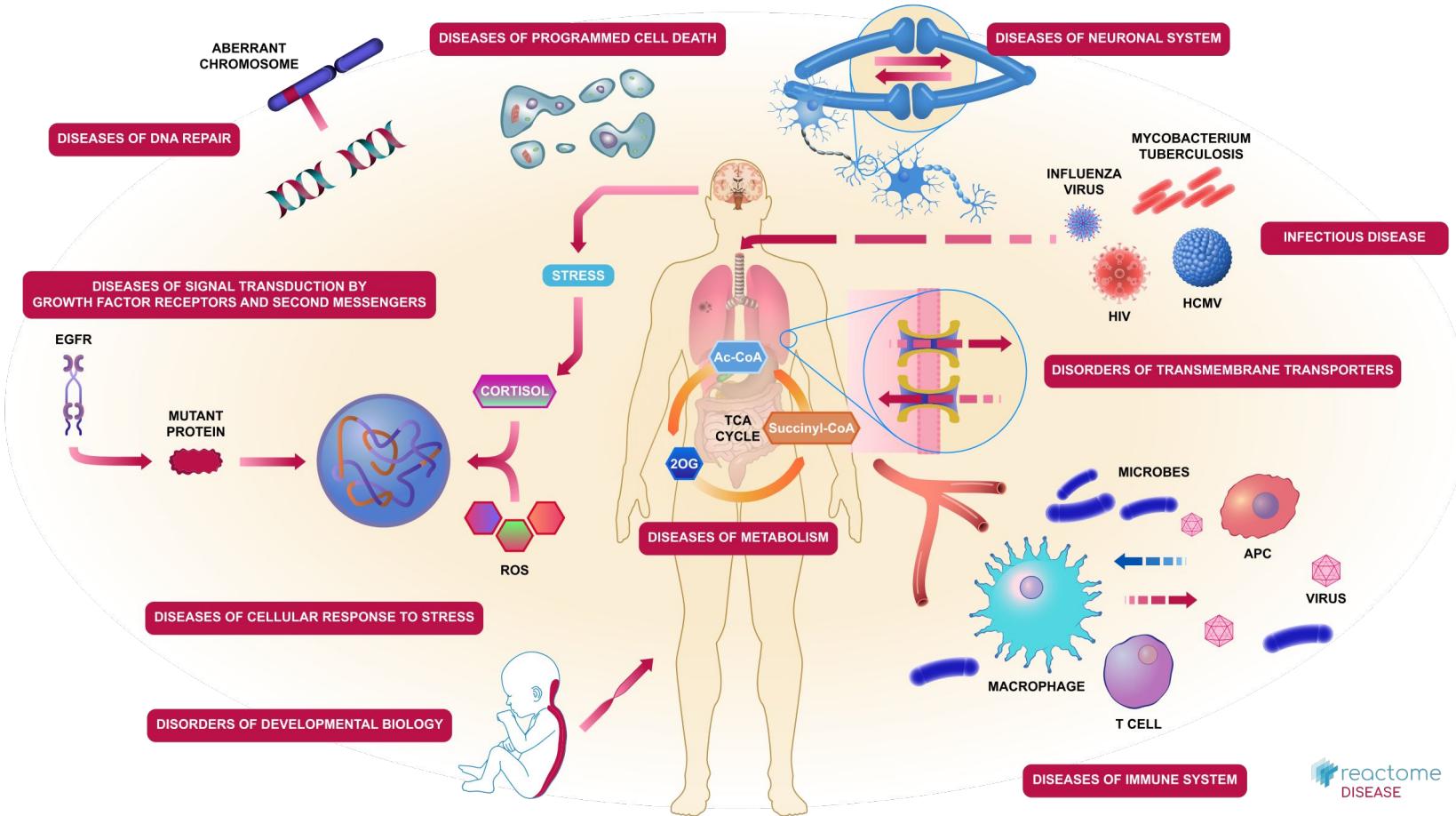
- Recruit bench scientists to write modules.
- Authors describe their pathway(s):
 - All molecules must be identified.
 - All assertions backed up by literature references.
- Curators work with authors to ensure consistency and completeness.
- Module checked by peer review and software before publications.
- Public Release of Curated data every 3 months.
- Rolling Review every 2 years.
- Regular Pathway updates.



PANTHER used to identify model organism orthologs



Disease annotation in Reactome: variant proteins



Community Curation

Enabling the research community to contribute pathway annotations directly to Reactome

CREDIT



Get the recognition
for sharing your
expertise

DISCOVERY



Make your
contributions
searchable and
discoverable

OPEN SCIENCE



Resources and data
are available to all

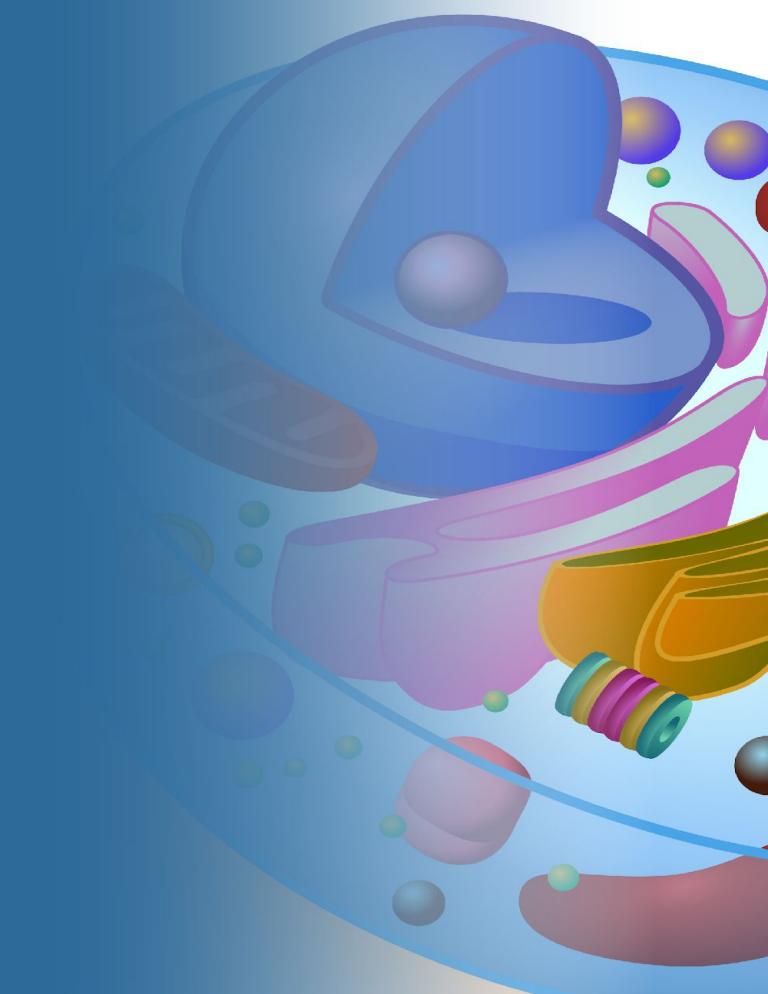
REUSE



Data curation and
standards enable
data reuse

Contact Us: help@reactome.org

<https://reactome.org/community/collaboration>



Credit

- Get the recognition for sharing your expertise
- Credit attribution through DOI and ORCID iD



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

Go! Expand All

Robin Haw

ORCID https://orcid.org/0000-0002-2013-7835 Are you Robin Haw? Register or Connect your ORCID

Affiliation Ontario Institute for Cancer Research

Reviewed Pathways (15/26)

Date	Identifier	Pathway	Reference
2004-12-15	R-HSA-9013694	Signaling by NOTCH4	BibTeX
2018-05-01	R-HSA-9013695	NOTCH4 Intracellular Domain Regulates Transcription	BibTeX
2018-05-01	R-HSA-9013700	NOTCH4 Activation and Transmission of Signal to the Nucleus	BibTeX
2018-05-01	R-HSA-9604323	Negative regulation of NOTCH4 signaling	BibTeX
2012-02-06	R-HSA-1912408	Pre-NOTCH Transcription and Translation	BibTeX
2004-12-15	R-HSA-9012852	Signaling by NOTCH3	BibTeX
2017-10-30	R-HSA-9013508	NOTCH3 Intracellular Domain Regulates Transcription	BibTeX
2017-10-30	R-HSA-9017802	Noncanonical activation of NOTCH3	BibTeX
2017-10-30	R-HSA-9013507	NOTCH3 Activation	BibTeX
2013-02-10	R-HSA-2644603	Signaling by NOTC	BibTeX
2013-02-10	R-HSA-2691230	Signaling by NOTC	BibTeX
2013-02-10	R-HSA-2660826	Constitutive Signaling	BibTeX
2013-02-10	R-HSA-2660825	Signaling by NOTC	BibTeX
2013-02-10	R-HSA-2644606	Constitutive Signaling	BibTeX
2013-02-10	R-HSA-2644602	Signaling by NOTC	BibTeX

Show all reviewed pathways...

Robin Haw

ORCID ID https://orcid.org/0000-0002-2013-7835

Also known as R. Haw, R.A. Haw, Robin A. Haw

Websites & Social Links Ontario Institute for Cancer Research Reactome, a curated pathway database

Country Canada

Keywords Bioinformatics curation cancer genetics genomics proteomics yeast

Other IDs ResearcherID: D-1393-2009

Notch4 binds Smad3
2018-04-11 | dataset
OTHER-ID: R-MMU-9605444

Source: Reactome

Items per page: 50 1 - 50 of 313 < >

Source: Reactome

Notch4 binds Smad3
2018-04-11 | dataset
OTHER-ID: R-MMU-9605444

Source: Reactome

ARTICLE doi:10.1038/nature10659

New gene functions in megakaryopoiesis and platelet formation

A full list of authors and their affiliations appears at the end of the article

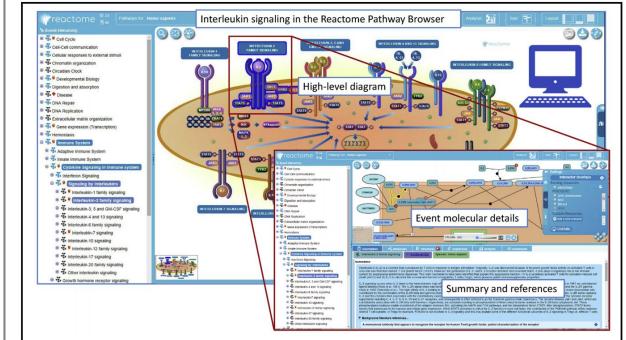
Translational and clinical immunology

Interleukins and their signaling pathways in the Reactome biological pathway database



Steve Jupe, PhD,^a Keith Ray, PhD,^b Corina Duenas Roca, MSc,^b Thawfeek Varusal, PhD,^a Veronica Shamovsky, MSc,^c Lincoln Stein, MD, PhD,^a Peter D'Eustachio, PhD,^a and Henning Hermjakob, MSc^{a,*} Hinxton and Cambridge, United Kingdom; New York, NY; Toronto, Ontario, Canada; and Beijing, China

GRAPHICAL ABSTRACT



Reuse

- Data curation and standards enable data reuse

 Open Targets For biomedical researchers who need to identify a biological target for a new therapy, Open Targets is a public-private initiative to generate evidence on the validity of therapeutic targets <input type="checkbox"/> Analysis Service <input checked="" type="checkbox"/> Widgets <input type="checkbox"/> Graph Database Open Targets	 THE HUMAN PROTEIN ATLAS Contains information for a large majority of all human protein-coding genes regarding the expression and localization of the corresponding proteins based on both RNA and protein data. <input checked="" type="checkbox"/> Analysis Service <input type="checkbox"/> Widgets <input type="checkbox"/> Graph Database The Human Protein Atlas	 Ensembl Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes. <input type="checkbox"/> Analysis Service <input checked="" type="checkbox"/> Widgets <input type="checkbox"/> Graph Database Ensembl	 BLUEPRINT The BLUEPRINT consortium has been formed with the aim to further the understanding of how genes are activated or repressed in both healthy and diseased human cells. <input checked="" type="checkbox"/> Analysis Service <input type="checkbox"/> Widgets <input type="checkbox"/> Graph Database BLUEPRINT
---	--	---	---



BioPAX

PSICQUIC



Open Science

- Resources and data are available to all
- Third parties can integrate Reactome components:
 - <https://reactome.org/dev>
 - <https://reactome.org/download-data>

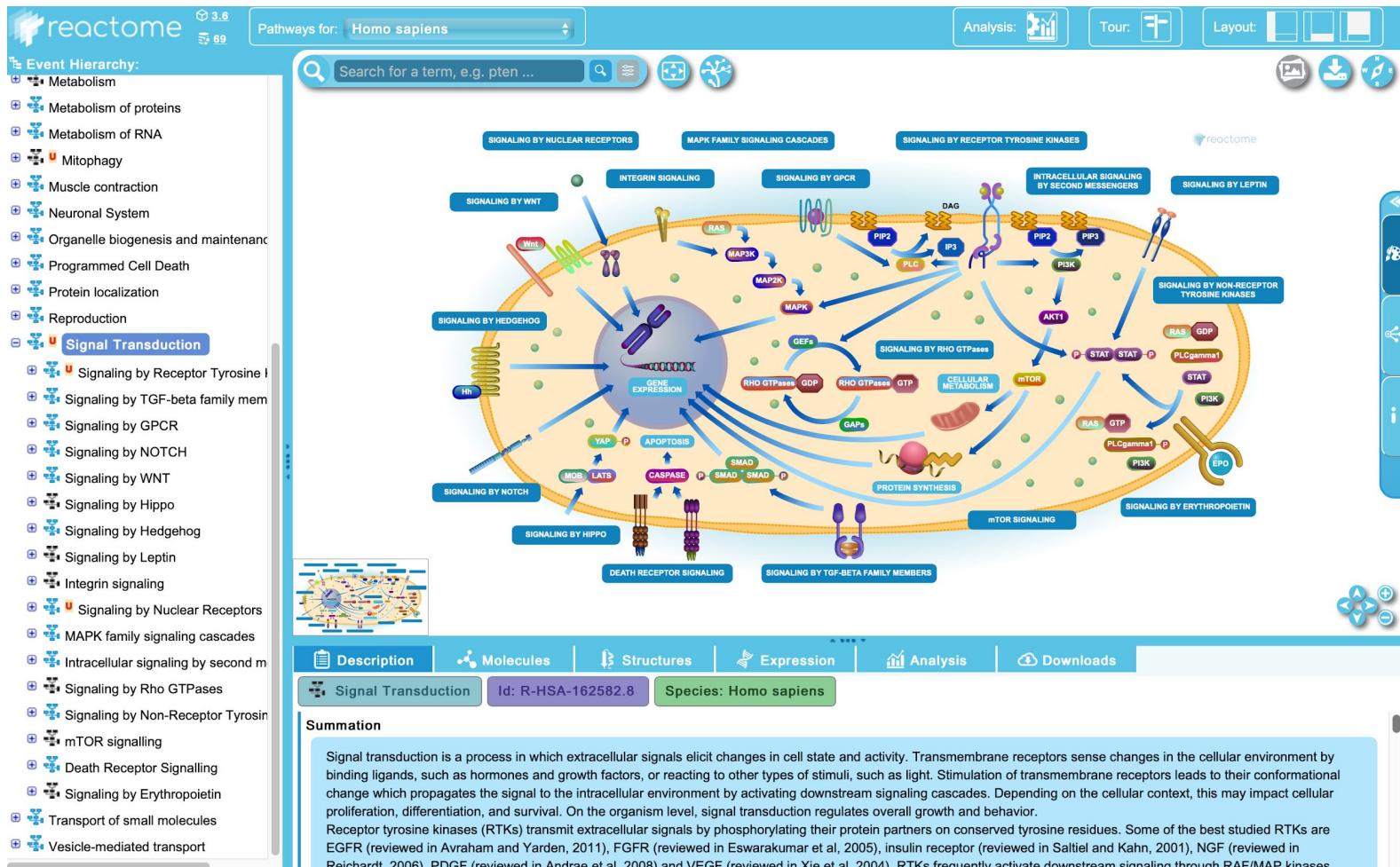
Developer's Zone

Explore our tools and web services and learn how to include them in your applications

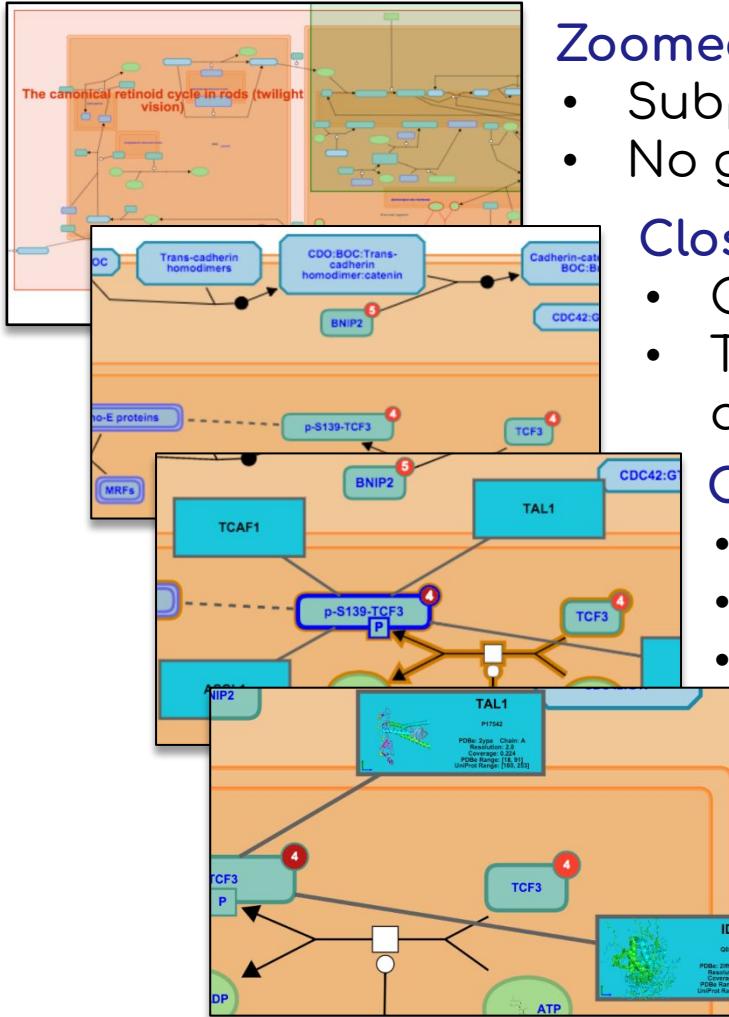
 Analysis Service	 Content Service	 Graph Database
Use the Analysis Service to analyse your data against Reactome's content	Use the Content Service to access all our knowledgebase content from your client	Access to the Reactome knowledgebase content as an interconnected graph database
		
Pathways Overview	Pathway Diagrams	Reactome Partners
Use this widget to include our pathways overview in your web application	Use this widget to include our pathway diagrams in your web application	Check out who is currently using Reactome web services and widgets



Pathway Visualization



Pathway Diagram Viewer



Zoomed-out view

- Subpathway boxes
- No glyph labels, No trivial molecules

Close view

- Glyph labels
- Trivial molecules and interactor summary appear

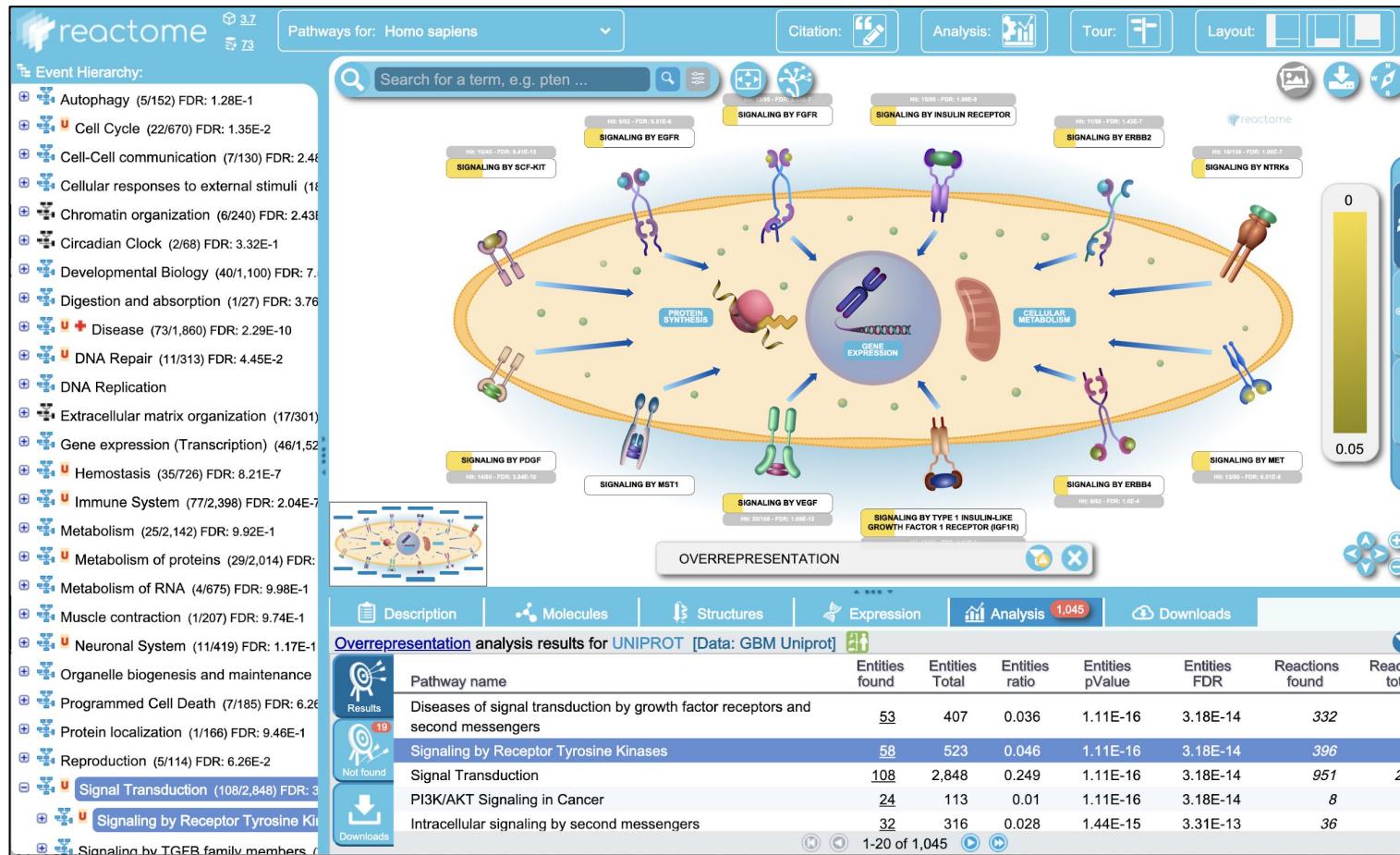
Closer view

- Stoichiometry shown
- Node attachments appear
- Interactors show gene or chemical name

Zoomed-in view

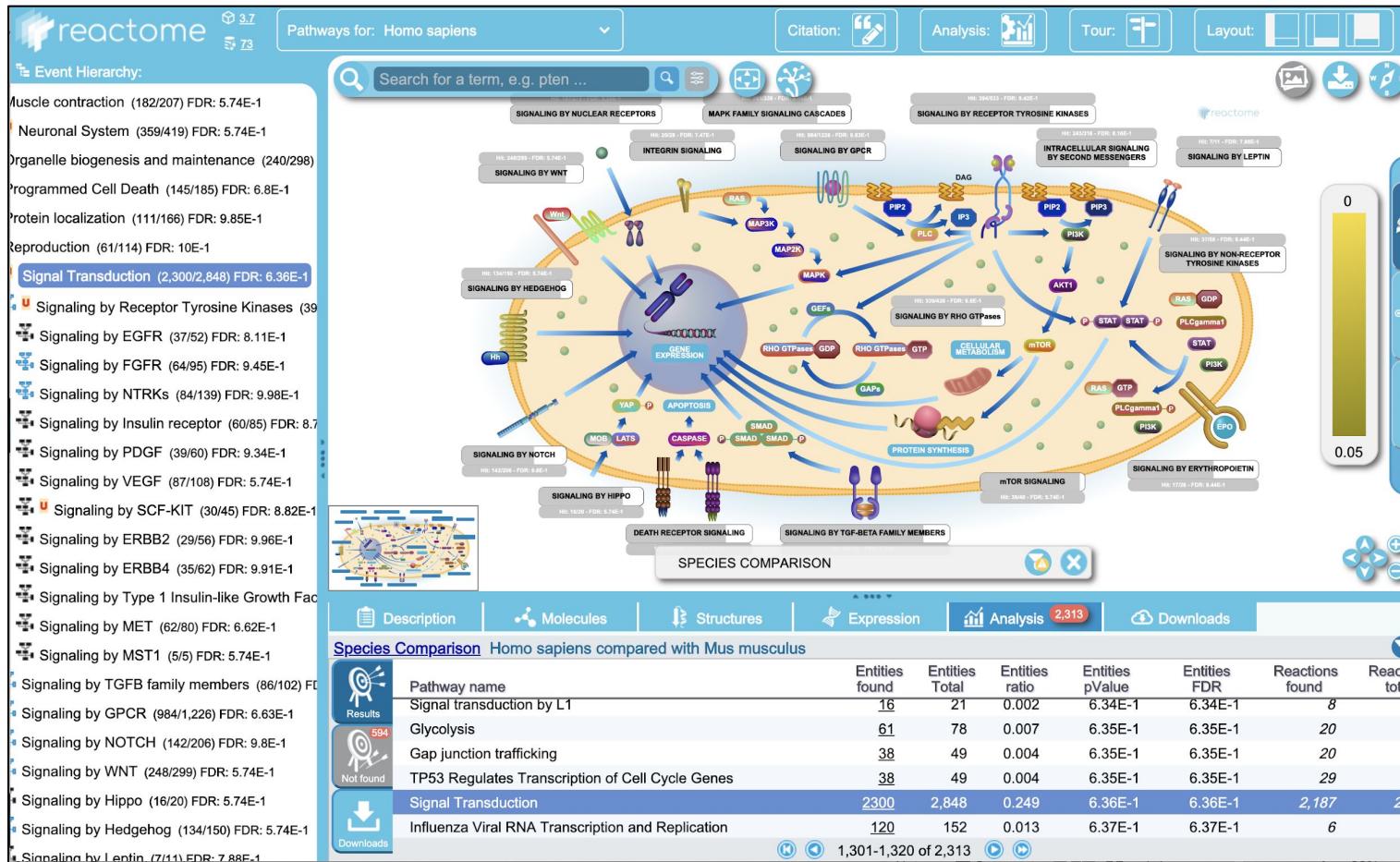
- Interactors show structures and details
- Diagram proteins, chemicals extra info

Pathway Enrichment Analysis



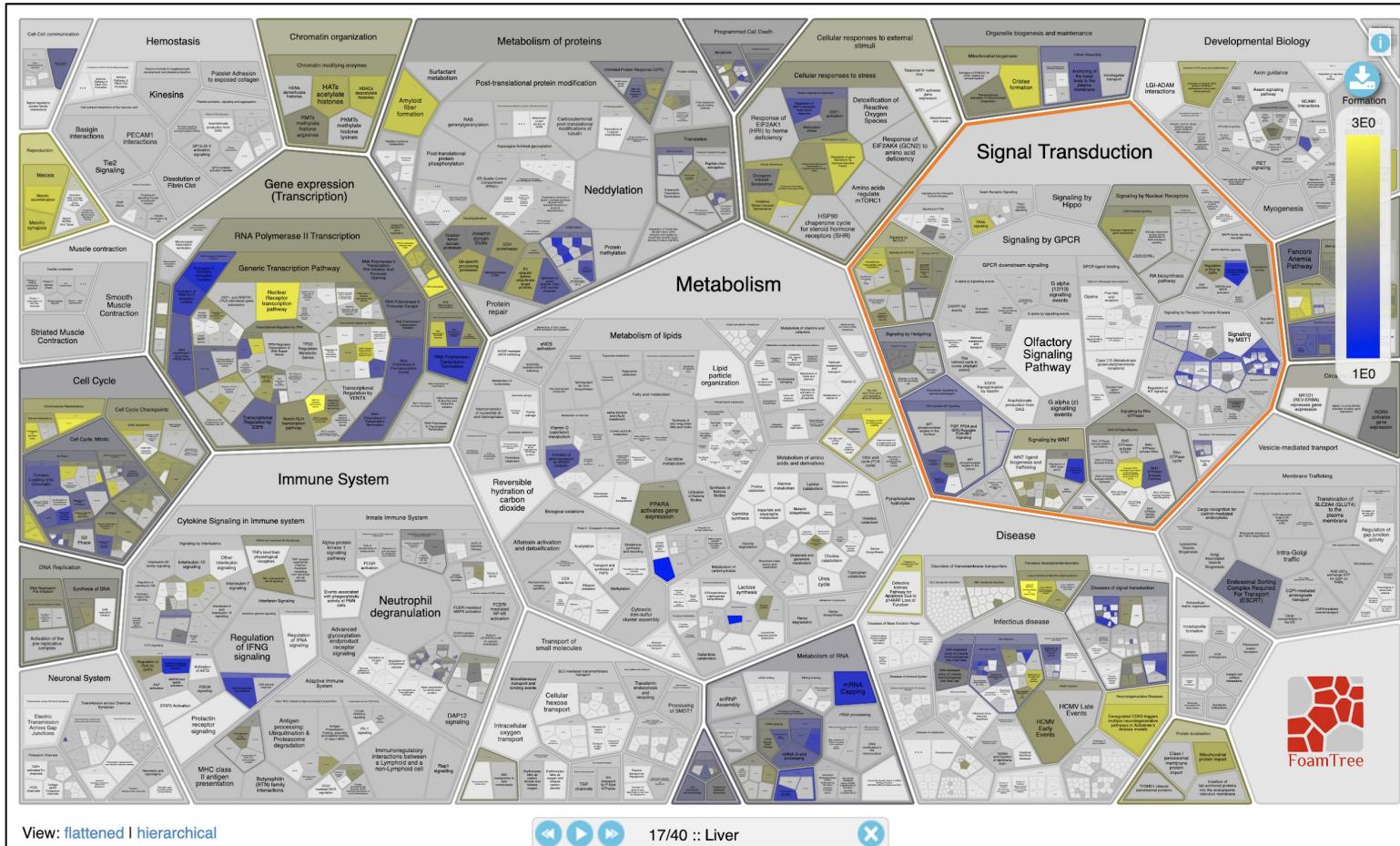
Overlay of TCGA GBM gene list.

Species Comparison



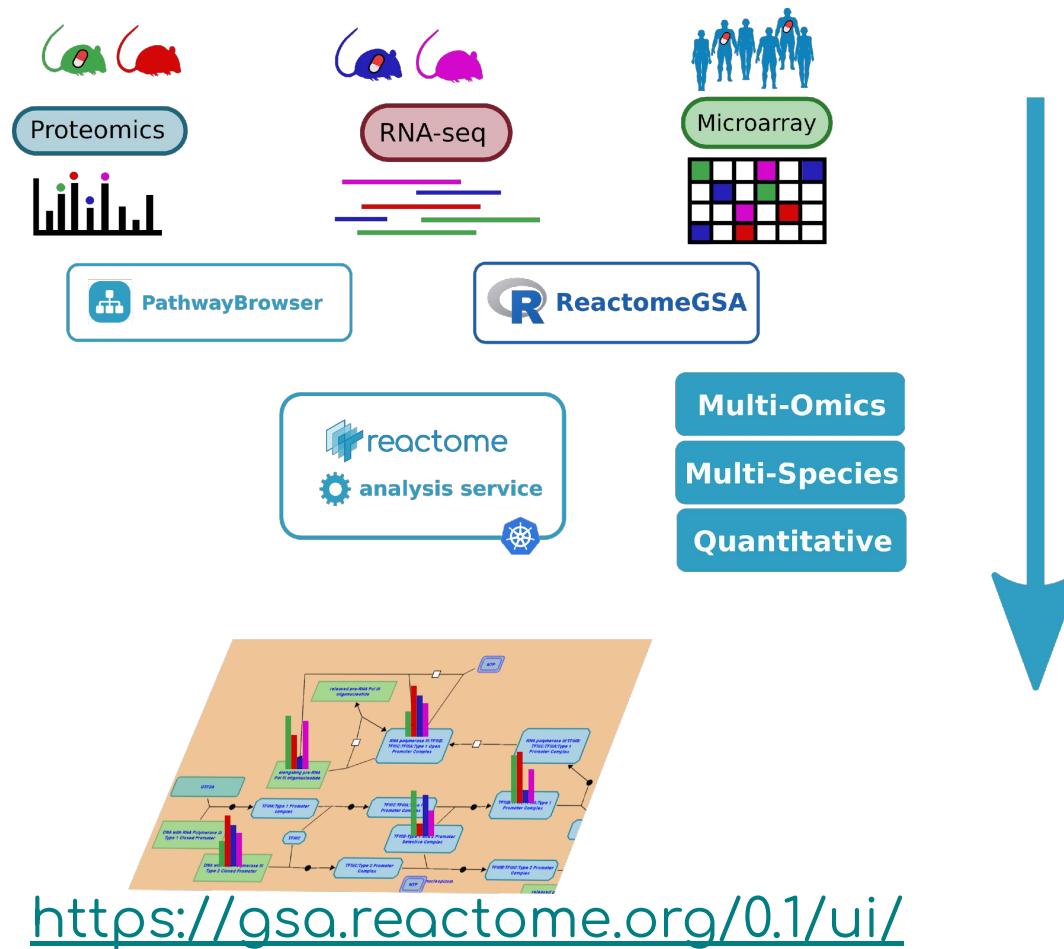
Comparison of Human vs Mouse Signal transduction events

Tissue Distribution

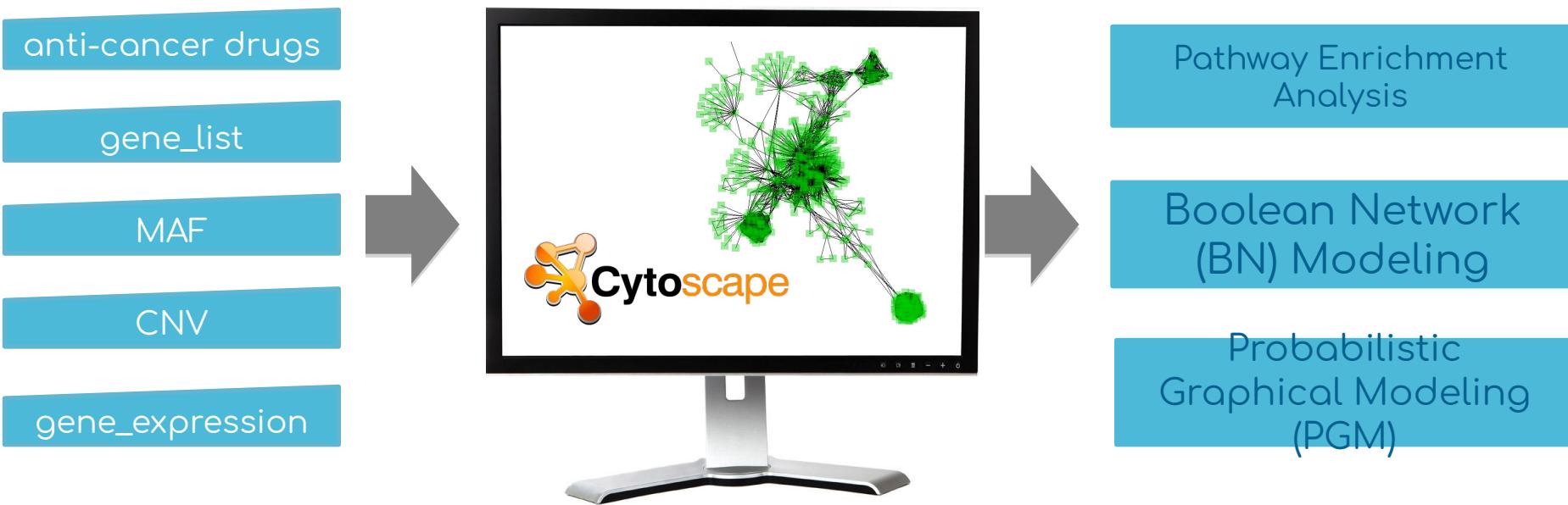


Overlay of Liver-specific proteins expression data

ReactomeGSA – Comparative pathway analysis



ReactomeFIViz: Network Analysis & Pathway Activities



Inputs

*ReactomeFIViz
app*

Analysis Types

<https://reactome.org/tools/reactome-fiviz>

Reactome Python Package: reactome2py



Analysis Service



Content Service



Pathway Downloads



ReactomeFIViz API and
Resources

Reactome2py 1.0.0

- Modules
 - Analysis
 - Content
 - Utils
- Sphinx Documentation
- Jupyter Notebook Demo



Available Install

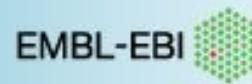


Research Community

<https://reactome.github.io/reactome2py/>

Summary

- Reactome is a highly reliable, curated database of biological pathways.
- Web site provides tools and datasets for visualizing pathway data and interpreting your experimental data.
- All data and software are open to public; no licensing required.



ContentService

- Retrieves data from Reactome.
- Access via RESTful web service interface.
 - Can be used from the command-line with tools such as “curl”.
 - *No special client software is needed!*
- Backed by Neo4j graph database.
- Text searches are powered by a Solr index.
- A great way to access and query Reactome data programmatically!
- URL: <https://reactome.org/ContentService/>

ContentService

https://reactome.org/

The screenshot shows the Reactome website at https://reactome.org/. The header includes the Reactome logo, navigation links for About, Content, Docs, Tools, Community, and Download. A dropdown menu from the 'Tools' link is open, listing Pathway Browser, Analyze Data, Species Comparison, Analysis Service, Content Service (which is highlighted with a red border), ReactomeFIViz, Advanced Data Search, and Site Search.

Find Reactions, Proteins and Pathways

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

Pathway Browser

Visualize and interact with Reactome biological pathways

Analyze Data

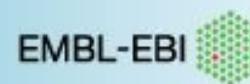
Merges pathway identifier mapping, over-representation, and expression analysis

ReactomeFIViz

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

Documentation

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



ContentService

Content Service 1.1

RESTful service for Reactome content

[Terms of service](#)

[Reactome - Website](#)

[Send email to Reactome](#)

[Creative Commons Attribution 3.0 Unsupported License](#)

database	Reactome Data: Database info queries	>
discover	Reactome Data: Search engines discovery schema	>
diseases	Reactome Data: Disease related queries	>
entities	Reactome Data: PhysicalEntity queries	>
events	Reactome Data: Queries related to events	>
exporter	Reactome Data: Format Exporter	>
interactors	Molecule Interactors	>
mapping	Reactome Data: Mapping related queries	>
orthology	Reactome Data: Orthology related queries	>
participants	Reactome Data: Queries related to participants	>
pathways	Reactome Data: Pathway related queries	>
person	Reactome Data: Person queries	>
query	Reactome Data: Common data retrieval	>
references	Reactome xRefs: ReferenceEntity queries	>
schema	Reactome Data: Schema class queries	>
search	Reactome Search	>
species	Reactome Data: Species related queries	>



ContentService

[orthology](#) Reactome Data: Orthology related queries >

[participants](#) Reactome Data: Queries related to participants >

[pathways](#) Reactome Data: Pathway related queries >

[person](#) Reactome Data: Person queries ▾

GET /data/people/name/{name} A list of people with first or last name partly matching a given string

Retrieves a list of people in Reactome with either their first or last name partly matching the given string.

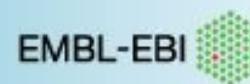
Parameters Cancel

Name	Description
name * required	Person's first or last name
string (path)	Steve Jupe

Execute

Responses Response content type application/json ▾

Code	Description
200	<p>OK</p> <p>Example Value Model</p> <pre>[{ "placeName": "string" }]</pre>



ContentService

GET /interactors/psicquic/molecule/{resource}/{acc}/summary Retrieve a summary of a given accession by resource

Parameters [Try it out](#)

Name	Description
resource * required	PSICQUIC Resource string (path)
acc * required	Single Accession string (path)

Responses [Response content type](#) application/json

Code	Description
200	<p>OK</p> <p>Example Value Model</p> <pre>Interactors <v> { entities <v> [InteractorEntity <v> { acc string This is the interactor accession (or identifier). count integer(\$int32) This is the number of interactions for the given accession. interactors <v> [Interactor <v> { acc string This is the interactor accession (or identifier). accURL string This represents the URL for given accession. } } }</pre>



ContentService

The simplest query: Get the database version.

```
curl -X GET  
"https://reactome.org/ContentService/data/database/  
version" -H "accept: text/plain"
```

Result: “73”

(Version 73 was released in June, 2020)

ContentService

Querying for an entity with a known identifier:

```
curl -X GET  
"https://reactome.org/ContentService/data/query/R-HSA-1640170" -H  
"accept: application/json"
```

Response is a JSON-representation of the entity:

```
{  
    "dbId": 1640170,  
    "displayName": "Cell Cycle",  
    "stId": "R-HSA-1640170",  
    "stIdVersion": "R-HSA-1640170.3",  
    "isInDisease": false,  
    "isInferred": false,  
    ...  
}
```

ContentService

Searching with Solr - `/search/query` endpoint

Takes a number of parameters:

- Query - term to search for
- Species - optional name of species to limit search
- Types - optional set of Reactome Types to limit search
- Compartments - optional set of compartments to limit search
- Cluster - result items will be grouped into “related” cluster

ContentService

Searching for data - search for the entities that exist in the nucleoplasm, containing the term “APEX1” and filter for Human results.

```
curl -X GET  
"https://reactome.org/ContentService/search/query?query=APEX1&species=Homo%20sapiens&compartments=nucleoplasm&cluster=true" -H "accept: application/json"
```

ContentService

Search results

```
{  
  "results": [  
    {  
      "entries": [  
        {  
          "dbId": "50119",  
          "stId": "R-HSA-50119",  
          "id": "R-HSA-50119",  
          "name": "<span class=\"highlighting\">APEX1</span>",  
          "exactType": "ReferenceGeneProduct",  
          "species": [  
            "Homo sapiens"  
          ],  
          "summation": "",  
          "referenceName": "<span class=\"highlighting\">APEX1</span>",  
          "referenceIdentifier": "P27695",  
          "compartmentNames": [  
            "nucleoplasm"  
          ],  
          ...  
        }  
      ]  
    }  
  ]  
}
```



ContentService

Tip: You can use JSON-processing tools like *jq* to handle the web service results.

```
curl -X GET  
"https://reactome.org/ContentService/search/query?qu  
ery=APEX1&species=Homo%20sapiens&compartments=nucleo  
plasm&cluster=true" -H "accept: application/json" |  
jq '.results[] .entries[] | [.stId, .name]'
```

The *jq* expression `'.results[] .entries[] | [.stId, .name]'` filters the results to only show stable identifiers and names.

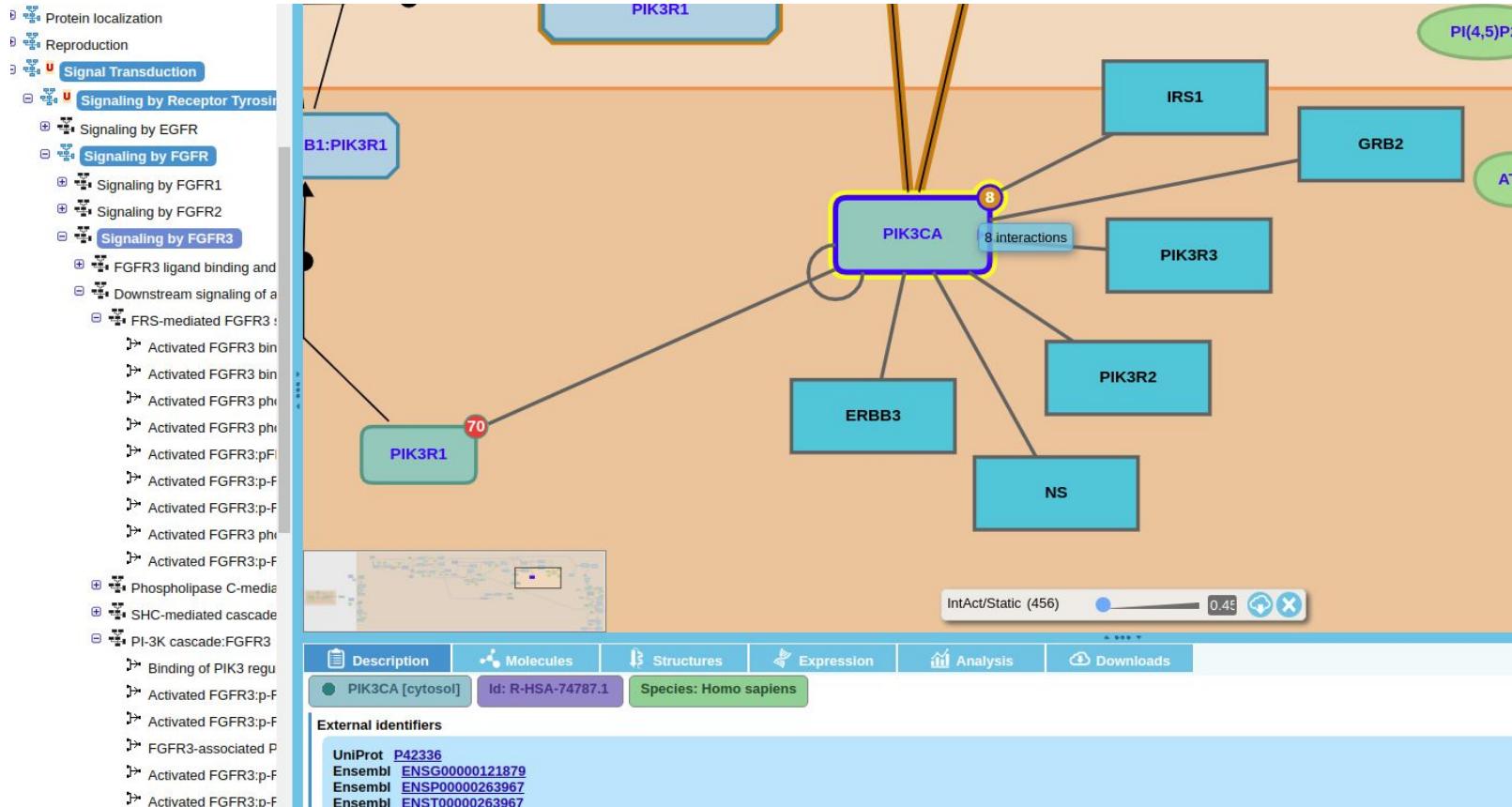
ContentService

Filtered results:

```
[  
  "R-HSA-50119",  
  "<span class=\"highlighting\">APEX1</span>"  
]  
[  
  "R-HSA-110357",  
  "Displacement of DNA glycosylase by <span  
class=\"highlighting\">APEX1</span>"  
]  
[  
  "R-HSA-5649702",  
  "<span class=\"highlighting\">APEX1</span>-Independent  
Resolution of AP Sites via the Single Nucleotide Replacement  
Pathway"  
] ...
```

ContentService

Interactors - in the PathwayBrowser



ContentService

Interactors - searching with ContentService

```
curl -X GET  
"https://reactome.org/ContentService/interactors/ps  
icquic/molecule/MINT/Q13501/details" -H "accept:  
application/json"
```

ContentService

Interactors search - result:

```
{  
  "resource": "MINT",  
  "entities": [  
    {  
      "acc": "Q13501",  
      "count": 18,  
      "interactors": [  
        {  
          "acc": "Q9GZQ8",  
          "alias": "MLP3B",  
          "id": 1,  
          "evidences": 4,  
          "score": 0.97,  
          "accURL":  
            "http://www.uniprot.org/uniprot/Q9GZQ8",  
          "evidencesURL":  
            "http://www.ebi.ac.uk/intact/pages/interactions/  
              /interactions.xhtml?query=EBI-10107390%20OR%20E  
              BI-7190708%20OR%20EBI-7679811%20OR%20EBI-768039  
            3"  
        }, ...  
      ]  
    }  
  ]  
}
```

```
...  
  {  
    "acc": "O70405",  
    "alias": "ULK1 MOUSE",  
    "id": 18,  
    "evidences": 2,  
    "score": 0.5,  
    "accURL":  
      "http://www.uniprot.org/uniprot/O704  
      05",  
    "evidencesURL":  
      "http://www.ebi.ac.uk/intact/pages/i  
        nteractions/interactions.xhtml?query  
      =EBI-9638162%20OR%20EBI-9638383"  
  }  
]  
]  
}  
]
```

ContentService

Reactome infers orthologies from Human to other species. You can query these with the ContentService:

```
curl -X GET  
"https://reactome.org/ContentService/data/orthology  
/R-HSA-6799198/species/49633" -H "accept:  
application/json"
```

ContentService

Results:

```
{  
  "dbId": 10200760,  
  "displayName": "Complex I  
biogenesis",  
  "stId": "R-SSC-6799198",  
  "stIdVersion":  
  "R-SSC-6799198.1",  
  "isInDisease": false,  
  "isInferred": true,  
  "name": [  
    "Complex I biogenesis"  
  ],...
```

```
  ...  
  "releaseDate": "2020-06-17",  
  "speciesName": "Sus scrofa",  
  "hasDiagram": false,  
  "hasEHLD": false,  
  "schemaClass": "Pathway",  
  "className": "Pathway"  
}
```

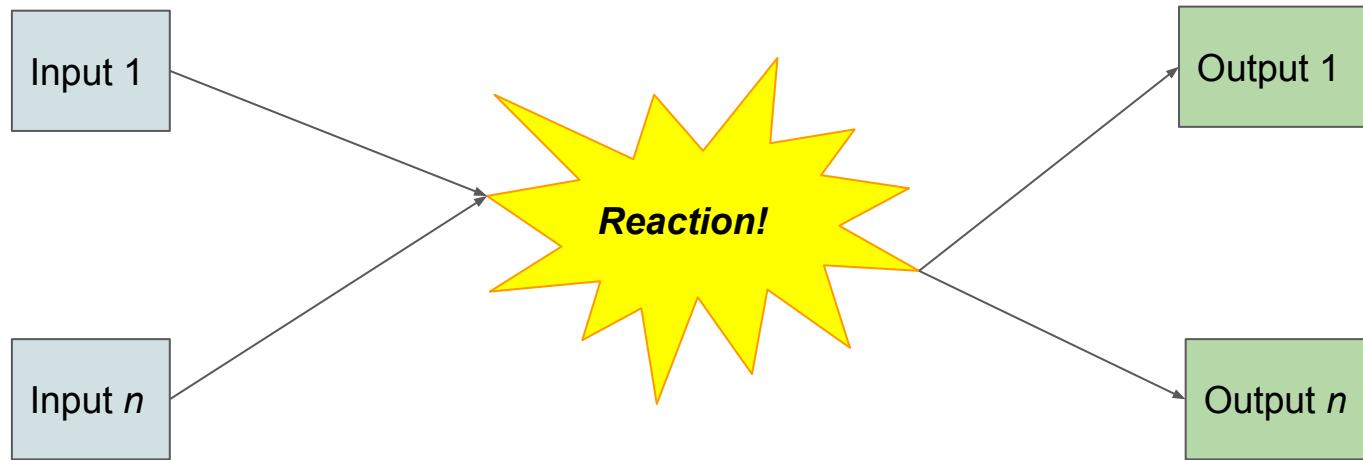
ContentService & Data Schema

Reactome has a rich data schema to describe its contents.

The full data schema is available at
<https://reactome.org/content/schema/>

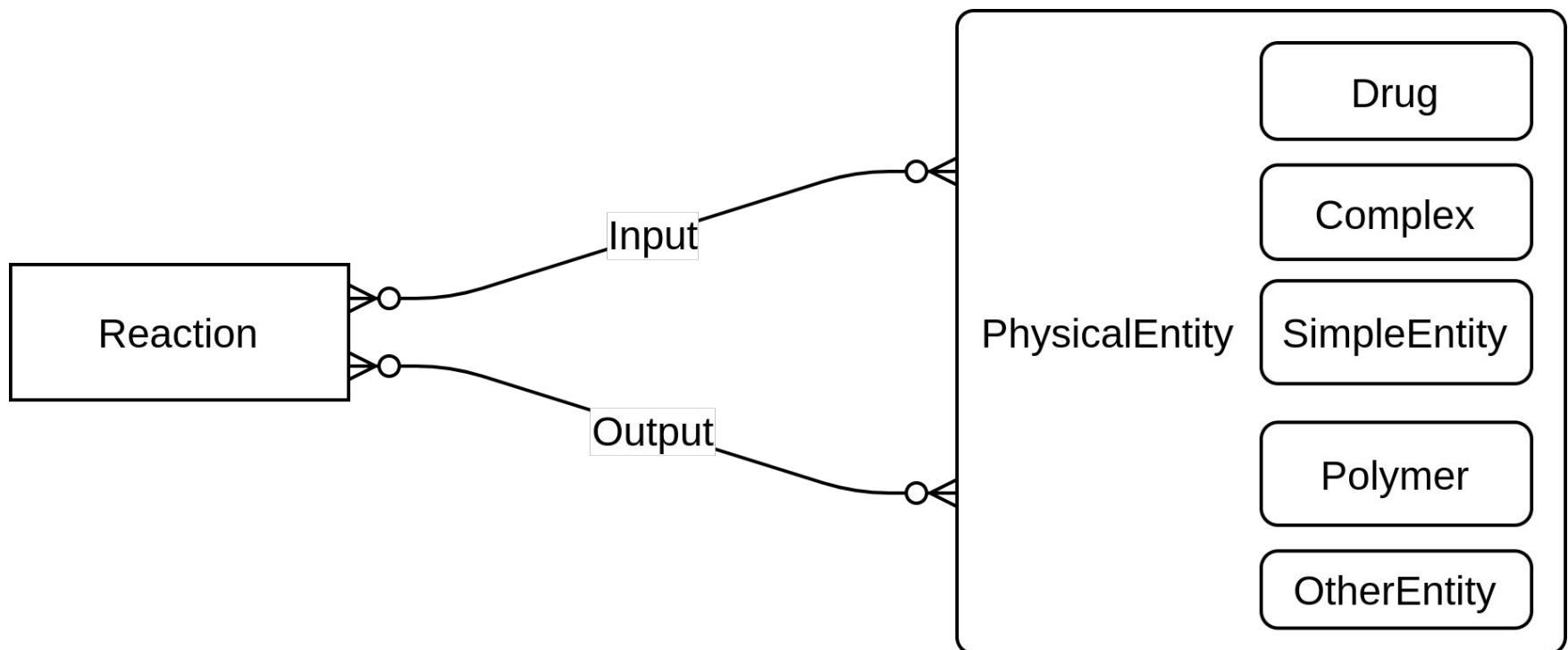
ContentService & Data Schema

The full data schema is quite complex. We will take a simplified view, starting with Reactions.



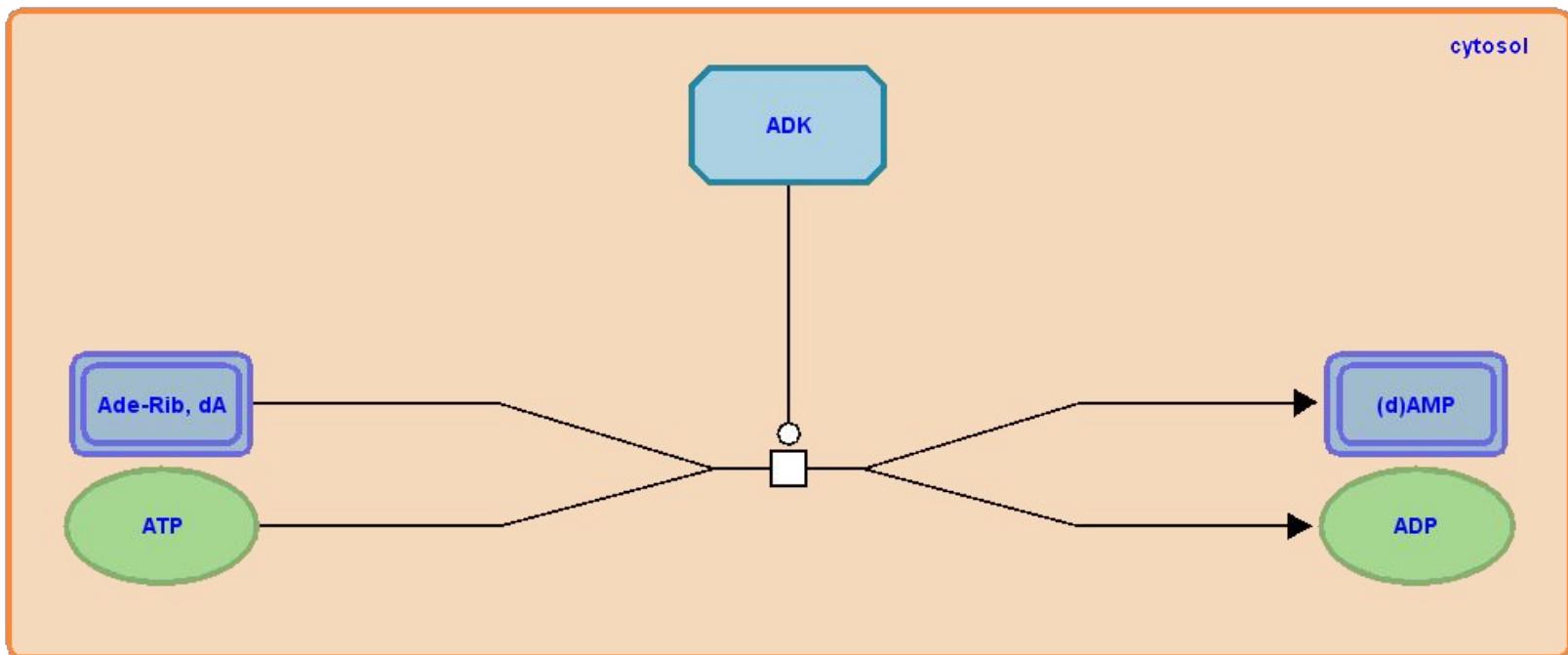
ContentService & Data Schema

In the database, we have this relationship:



ContentService & Data Schema

Reaction in Reactome:



$(2'\text{-deoxy})\text{adenosine} + \text{ATP} \Rightarrow (\text{d})\text{AMP} + \text{ADP}$ (ADK)



<https://reactome.org/content/detail/R-HSA-109624>

ContentService & Data Schema

Querying the same Reaction, and using jq to filter for inputs and outputs:

```
curl -X GET -H "Accept: application/json"  
"https://reactome.org/ContentService/data/query/R-HS  
A-109624" | jq '.input'
```

```
curl -X GET -H "Accept: application/json"  
"https://reactome.org/ContentService/data/query/R-HS  
A-109624" | jq '.output'
```

ContentService & Data Schema

Querying the same Reaction, and using jq to filter for inputs and outputs. Inputs:

```
[  
  {  
    "dbId": 113592,  
    "displayName": "ATP [cytosol]",  
    "stId": "R-ALL-113592",  
    "stIdVersion": "R-ALL-113592.4",  
    "name": [  
      "ATP",  
      "Adenosine 5'-triphosphate",  
      "ATP(4-)"  
    ],  
    "consumedByEvent": [  
      109624  
    ],  
    "referenceType": "ReferenceMolecule",  
    "className": "Chemical Compound",  
    "inDisease": false,  
    "schemaClass": "SimpleEntity"  
  }, ...
```



ContentService & Data Schema

Querying the same Reaction, and using jq to filter for inputs and outputs. Outputs:

```
[  
 {  
   "dbId": 500088,  
   "displayName": "(d)AMP [cytosol]",  
   "stId": "R-ALL-500088",  
   "stIdVersion": "R-ALL-500088.1",  
   "name": [  
     "(d)AMP"  
   ],  
   "producedByEvent": [  
     109624  
   ],  
   "className": "Set",  
   "inDisease": false,  
   "schemaClass": "DefinedSet"  
 },  
 {  
   "dbId": 29370,  
   "displayName": "ADP [cytosol]",  
   "stId": "R-ALL-29370",...
```



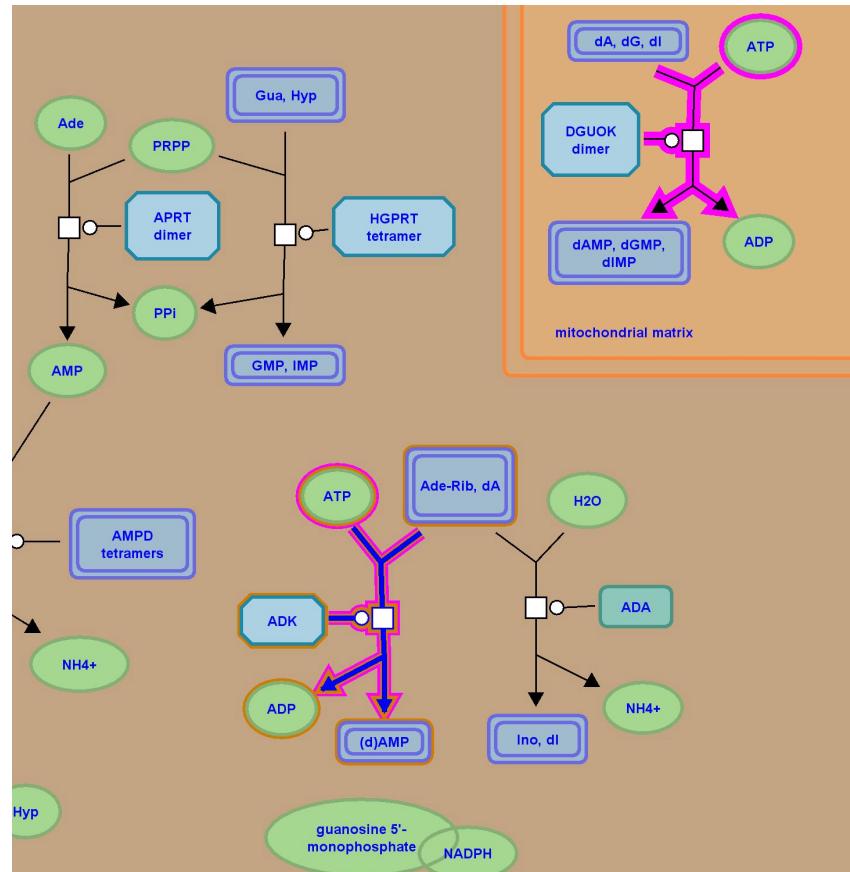
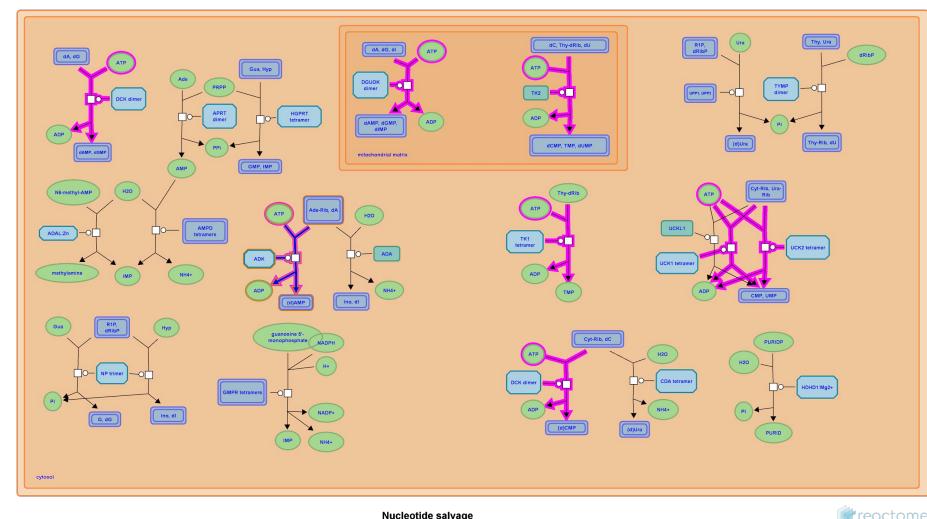
ContentService & Data Schema

Exporting the same Reaction to an image:

```
curl -o R-HSA-109624.svg -X GET  
"https://reactome.org/ContentService/exporter/diagra  
m/R-HSA-109624.svg?quality=10&flgInteractors=true&  
sel=R-HSA-109624&title=true&margin=15&ehld=true&diag  
ramProfile=Modern&resource=TOTAL&analysisProfile=Sta  
ndard&flg=ATP" -H "accept: image/png"
```

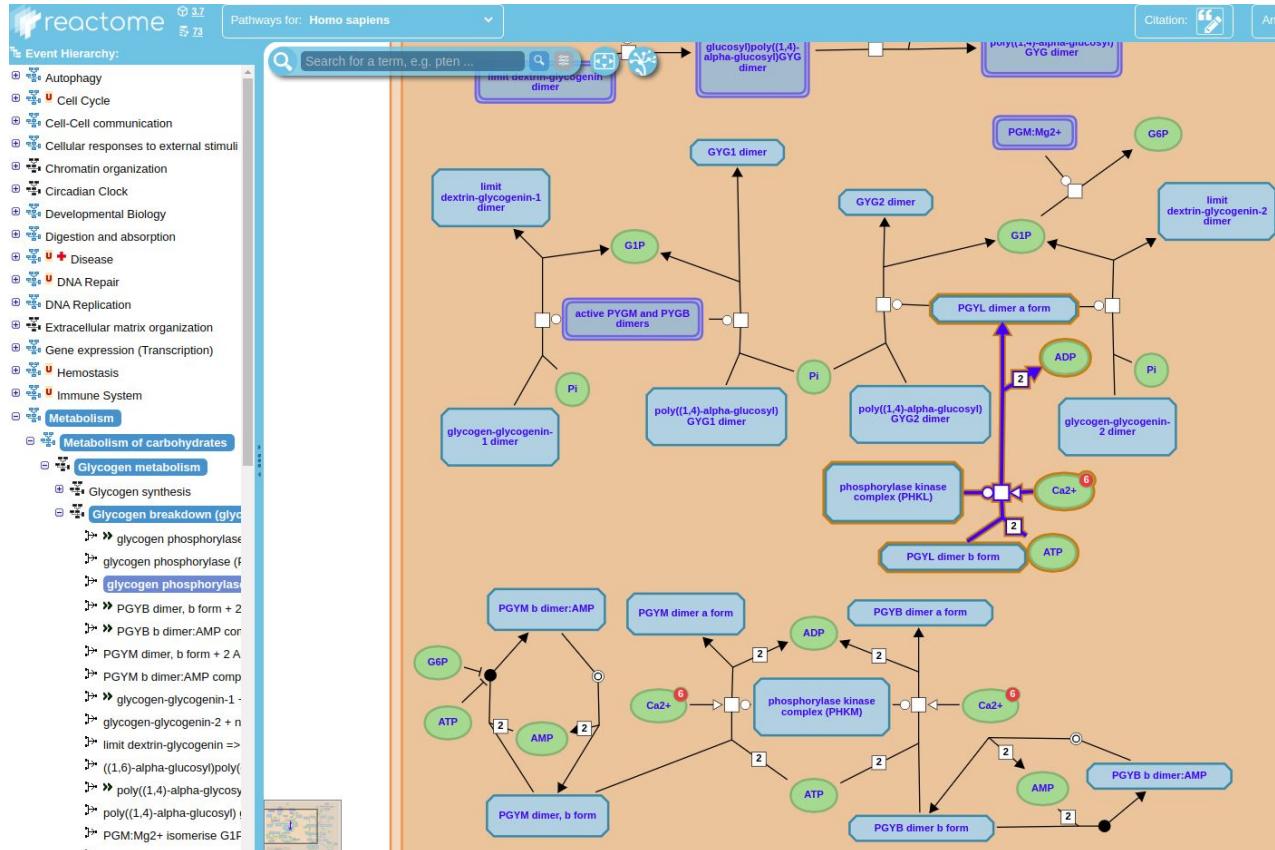
ContentService & Data Schema

Exporting the same Reaction to an image:



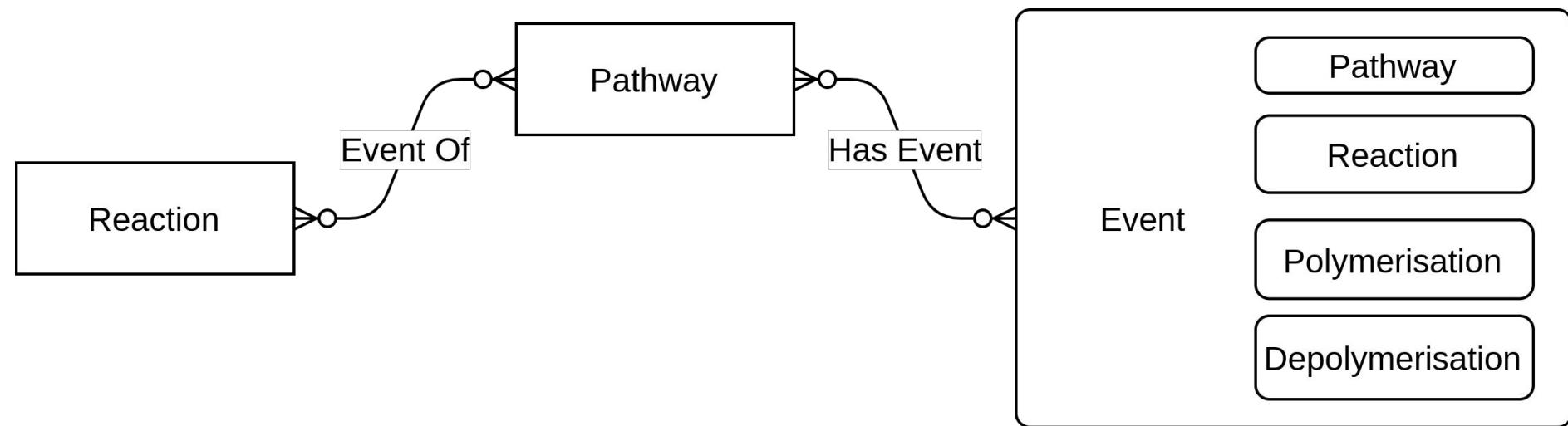
ContentService & Data Schema

Pathways, in the PathwayBrowser



ContentService & Data Schema

Data model for Pathways



ContentService & Data Schema

To see all of the Events (Reactions, Pathways and other Event types) contained within a Pathway:

```
curl -X GET  
"https://reactome.org/ContentService/data/pathway/R  
-HSA-5673001/containedEvents" -H "accept:  
application/json"
```

ContentService - Summary

Reactome's ContentService has a variety of endpoints that can help you easily get Reactome's data into your application.



Analyzing Data in Reactome

Justin Cook



Reactome Analysis

Overview

Reactome offers a variety of bioinformatics analyses that can be performed on user-submitted data:

- Single-column lists of identifiers
 - Including, but not limited to, gene symbols (eg: TP53), protein ids (eg: UniProt) and small molecules ids (eg: ChEBI)
 - Lists can be combinations of types
- Tab-separated expression data

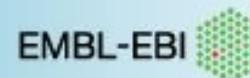
Reactome also offers comparative analyses that make use of our knowledge base without requiring user-submitted data:

- Species comparison
- Tissue distribution

These analyses can be run through the Reactome website or programmatically using a RESTful web service

- Analyses are usable for at least a week using a 'token' system

*Covered today



Reactome Analysis

Tutorial Overview

Given a single-column list of **identifiers** (gene, protein, mixed, etc.), the analysis system will:

- Perform Over-representation analysis and output the results into a table
- Find any **PhysicalEntities** (gene, protein, etc.) in the knowledge base that contain any of the **identifiers** and map these hits to their associated **Pathways** (Pathway-topology analysis)
- (Browser) Highlight any **Pathways** and **PhysicalEntities** from pathway-topology in the Pathway Browser
 - Pathway highlighting can be viewed in either 'Fireworks' or 'Reacfoam' views

TP53
PIK3CA
PTEN
APC
VHL
KRAS
MLL3
MLL2
ARID1A
PBRM1
NAV3
EGFR
NF1
PIK3R1
CDKN2A
GATA3
RB1
NOTCH1
FBXW7
CTNNB1
DNMT3A
MAP3K1
FLT3
MALAT1
TSHZ3
KEAP1
CDH1
ARHGAP35
CTCF
NFE2L2
SETBP1
BAP1
NPM1
RUNX1
NRAS
IDH1
TBX3
MAP2K4
RPL22
STK11
CRIPAK
CEBPA
KDM6A
EPHA3
AKT1
STAG2
BRAF

Snippet of gene list that will be used today

Reactome Analysis

Outline

Topics

1. Overview of Analysis results
2. Tutorial - Analyse gene list via Analysis Tools
(browser)
3. Tutorial - Analyse gene list via Analysis Service
(command line)

Analysis Tools

Results views - Pathways

reactome.org/PathwayBrowser/#/DTAB=AN&ANALYSIS=MjAyMDA3MTQxNzM0MTdfNTQ0MjY%253D

Paused

Event Hierarchy:

- FC epsilon receptor (FCER1) signaling
- C-type lectin receptors (CLRs)
- Antimicrobial peptides
- Neutrophil degranulation (4/480)
- ROS and RNS production in phagocytosis
- Alpha-protein kinase 1 signaling
- Cytokine Signaling in Immune system
- Metabolism (213,650) FDR: 1E0
- Metabolism of proteins (36/2,356) FDR: 1E-0
- Metabolism of RNA (10/782) FDR: 6.14E-0
- Muscle contraction
- Neuronal System (6/498) FDR: 6.66E-0
- Organelle biogenesis and maintenance
- Programmed Cell Death (8/197) FDR: 1E-0
- Protein localization (3/170) FDR: 4.03E-0
- Reproduction (9/123) FDR: 3.18E-0
- Signal Transduction (81/3,396) FDR: 1E-0
- Signaling by Receptor Tyrosine Kinases
- Signaling by EGFR (7/61) FDR: 1E-0
- Signaling by FGFR (14/107) FDR: 1E-0
- Signaling by NTRKs (10/166) FDR: 1E-0
- Expression and Processing of mRNAs
- Signaling by NTRK1 (TRKA)
- Signaling by NTRK2 (TRKB)
- Signaling by NTRK3 (TRKC)

Pathways for: Homo sapiens

Citation: Analysis: Tour: Layout:

Search for a term, e.g. pten ...

OVERREPRESENTATION Showing: pValue

Description | Molecules | Structures | Expression | Analysis 942 | Downloads

Overrepresentation analysis results for TOTAL Data submitted with no name

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
Signaling by NTRKs	10	166	0.011	1.01E-4	6.08E-4	31	163	0.013	Homo sapiens
Negative regulation of FGFR3 signaling	5	34	0.002	1.11E-4	6.67E-4	12	14	0.001	Homo sapiens
Signaling by VEGF	9	137	0.009	1.19E-4	7.14E-4	12	84	0.006	Homo sapiens
Ub-specific processing proteases	11	206	0.014	1.31E-4	7.15E-4	10	40	0.003	Homo sapiens
NOTCH1 Intracellular Domain Regulates Transcription	6	57	0.004	1.41E-4	7.15E-4	16	18	0.001	Homo sapiens
SHC1 events in EGFR signaling	4	19	0.001	1.43E-4	7.15E-4	4	4	0	Homo sapiens
Aberrant regulation of mitotic cell cycle due to RB1 defects	5	36	0.002	1.45E-4	7.23E-4	3	3	0	Homo sapiens
Diseases of mitotic cell cycle	5	36	0.002	1.45E-4	7.23E-4	3	3	0	Homo sapiens
	161-180 of 942								

Analysis Tools

Results views - Pathways

Reactome Pathway hierarchy in classical 'fireworks' view with analysis results overlaid



Statistical results table

Analysis Tools

Results Views - Reactions

← → C reactome.org/PathwayBrowser/#/R-HSA-1221632&PATH=R-HSA-1640170,R-HSA-1500620,R-HSA-1221632&DTAB=AN&ANALYSIS=MjAyMDA3MTQxNzM0MT... Paused :

Pathways for: Homo sapiens

Citation: Analysis: Tour: Layout:

Event Hierarchy:

- Mitotic G1 phase
- S Phase (10/180) FDR
- Mitotic G2-G2/M phase
- M Phase (8/416) FDR
- Regulation of mitc
- Chromosome Mainte
- Meiosis (992) FDR: 5E-**
- Meiotic recombination
 - PRDM9 binds
 - PRDM9 trimel
 - SPO11 hydroly
 - MRN:CtIP end
 - MRN:CtIP exo
 - Formation of r
 - Formation of r
 - Formation of r
 - Resolution of
- Meiotic synapsis (7)
 - Formation of a
 - Telomeres clu

Search for a term, e.g. pten ...

OVERREPRESENTATION

Description Molecules Structures Expression Analysis 942 Downloads

Analysis Tools

Results Views - Reactions

The screenshot shows the Reactome Pathway Browser interface. On the left, there's a sidebar titled "Event Hierarchy" listing various biological processes like Mitotic G1 phase, S Phase, and Meiosis. A red bracket on the left side of the main panel groups the sidebar and the search bar. The main panel displays a complex pathway diagram with nodes representing molecules and interactions. A red bracket on the right side groups three annotations:

- A red arrow points to a cluster of nodes in the top right labeled "Physical Entities not containing submitted identifiers".
- A red arrow points to a cluster of nodes in the bottom right labeled "Physical Entities containing submitted identifiers".
- A red arrow points to a cluster of nodes in the center labeled "Pathway diagram containing multiple Reactions".

At the bottom of the main panel, there are tabs for Description, Molecules, Structures, Expression, Analysis (with a count of 942), and Downloads.



Reactome Analysis

Outline

Topics

1. Overview of Analysis results
2. Tutorial - Analyse gene list via Analysis Tools (browser)
3. Tutorial - Analyse gene list via Analysis Service (command line)

Analysis Tools

Accessing Analysis Tools

The screenshot shows the Reactome website interface. At the top, there is a navigation bar with links for About, Content, Docs, Tools, Community, and Download. Below the navigation bar is a search bar with placeholder text "e.g. O95631, NTN1, signaling by EGFR, glucose" and a "Go!" button. The main content area features several icons:

- Pathway Browser**: Represented by a cluster icon with a dashed red border.
- Analyze Data**: Represented by a bar chart icon with a solid red border.
- ReactomeFIViz**: Represented by a network icon.
- Documentation**: Represented by a document icon.

A large red arrow points from the "Analyze Data" icon towards the detailed view of the Pathway Browser below.

Pathway Browser (Detailed View):

- Visualize and interact with Reactome biological pathways.
- Merges pathway identifier mapping, over-representation, and expression analysis.

Analyze Data (Detailed View):

- Search for a term, e.g. pten ...
- Event Hierarchy:

 - Autophagy
 - 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 expression (Transcription)
 - Hemostasis
 - Immune System
 - Metabolism
 - Metabolism of proteins
 - Metabolism of RNA
 - Muscle contraction
 - Neuronal System

- Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events.

The detailed view of the Pathway Browser shows a complex network of biological pathways. A specific node in the network is highlighted with a red box, and a callout box indicates it is being analyzed. The network includes nodes for Cell Cycle, Metabolism, Signal Transduction, and various cellular processes.



EMBL-EBI

NYU Langone
MEDICAL CENTER

GHEU

OICR
Ontario Institute
for Cancer Research

Analysis Tools

Tutorial

Initial Analysis Tools page

The screenshot shows the 'Analysis tools' section of the Reactome website. On the left, there's a sidebar with a tree view of biological pathways categorized by type (e.g., Event, Process, Entity, Location). The main area has tabs: 'Your data' (selected), 'Options', and 'Analysis'. A sidebar on the right lists examples for analysis, such as UniProt accession list, Gene name list, and Small molecules (ChEBI).

Your data Options Analysis

Step 1: Select a file from your computer or paste your own data and click on the corresponding "Continue" button.

Select data file for analysis: No file chosen

Paste your data to analyse or try example data sets:
Paste your data here or select an example from the right >>

Some examples:

- UniProt accession list
- Gene name list
- Gene NCBI / Entrez list
- Small molecules (ChEBI)
- Small molecules (KEGG)
- Microarray data
- Metabolomics data
- Cancer Gene Census (COSMIC)
- Tissue Specific Expression (HPA)

Reactome v73
Click to learn more about our analysis tools

Analysis Tools

Tutorial - Submitting data

Upload gene list file via 'Choose File' or paste list into the box then press 'Continue'.

The screenshot shows the 'Analysis tools' interface on the Reactome website. On the left, there's a sidebar with various pathway categories like 'Event', 'Metabolism', 'Protein', etc. The main area has tabs: 'Your data' (highlighted), 'Options', and 'Analysis'. A red box highlights the 'Select data file for analysis:' input field, which currently says 'No file chosen'. Below it is a text input for pasting data, with a list of genes: TPS3, PIK3CA, PTEN, APC, VHL, KRAS, MLL3, MLL2, ARID1A, PBRM1, NAV3, EGFR, NF1, PIK3R1, CIN82A, GATA3, RB1, NOTCH1. To the right, there's a 'Some examples:' section with links to UniProt accession list, Gene name list, Gene NCBI / Entrez list, Small molecules (ChEBI), Small molecules (KEGG), Microarray data, Metabolomics data, Cancer Gene Census (COSMIC), and Tissue Specific Expression (HPA). At the bottom, there are 'Clear' and 'Continue' buttons.

Analysis Tools

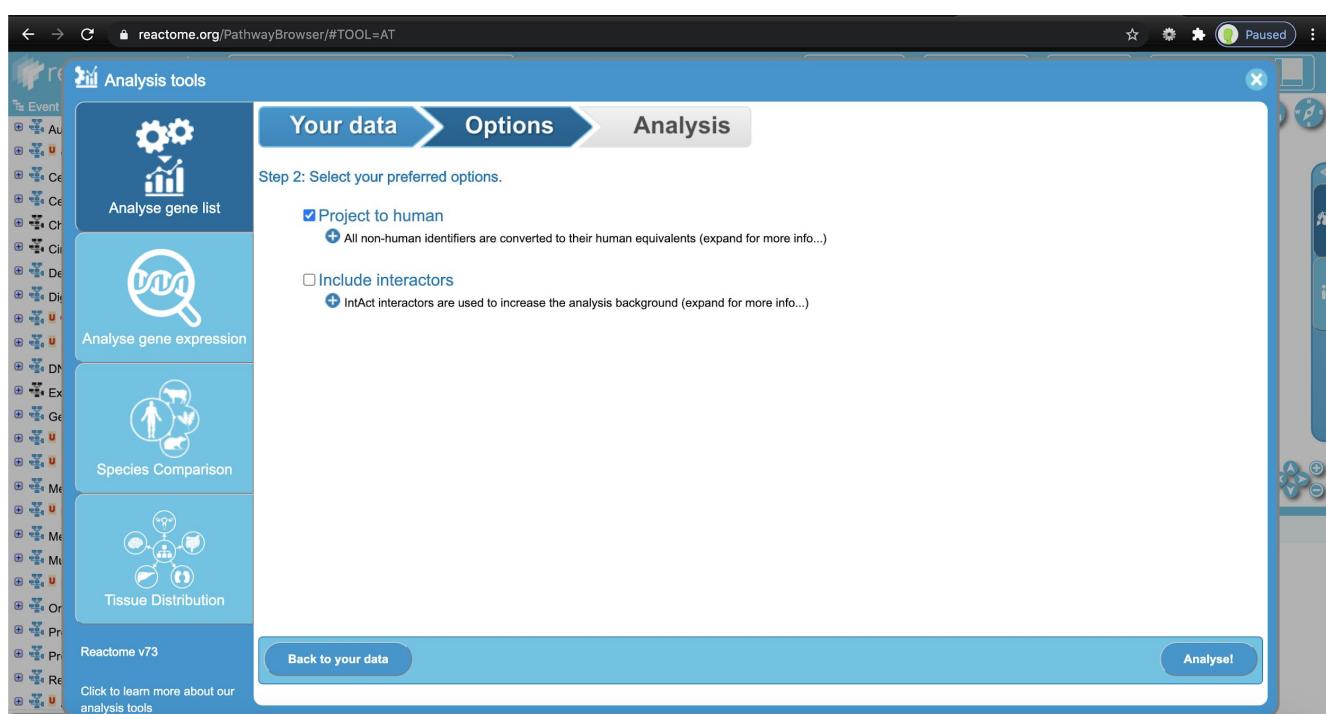
Tutorial - Submitting data

Leave the 'Project to human' option checked.

This option takes any submitted **non-human** identifiers and checks if any orthologous human identifiers exist in the database. This maximizes the chance of finding a hit for an identifier.

Leave the 'Include interactors' option unchecked.

Click 'Analyse!'.



Analysis Tools

Tutorial - Pathway Browser

The default results view!

The screenshot shows the Reactome Pathway Browser interface. At the top, there's a navigation bar with a back arrow, forward arrow, a search field containing 'reactome.org/PathwayBrowser/#/DTAB=AN&ANALYSIS=MjAyMDA3MTQxNzM0MTdfNTQ0MjY%253D', and a 'Paused' button. Below the navigation is a header with the Reactome logo, a dropdown for 'Pathways for: Homo sapiens', and buttons for 'Citation', 'Analysis', 'Tour', and 'Layout'. A search bar with placeholder text 'Search for a term, e.g. pten ...' is also present.

The main area features a hierarchical pathway map. Nodes are represented by icons and text, such as 'Autophagy', 'Cell Cycle', 'Cell-Cell communication', etc. Edges represent interactions between pathways. A color scale on the right indicates pathway enrichment, ranging from 0 (light blue) to 0.05 (dark yellow). A legend at the bottom of the map identifies various pathway types.

Below the map is a table titled 'Overrepresentation analysis results for TOTAL Data submitted with no name'. The table has columns for Pathway name, Entities found, Entities Total, Entities ratio, Entities pValue, Entities FDR, Reactions found, Reactions total, Reactions ratio, and Species. The first few rows include:

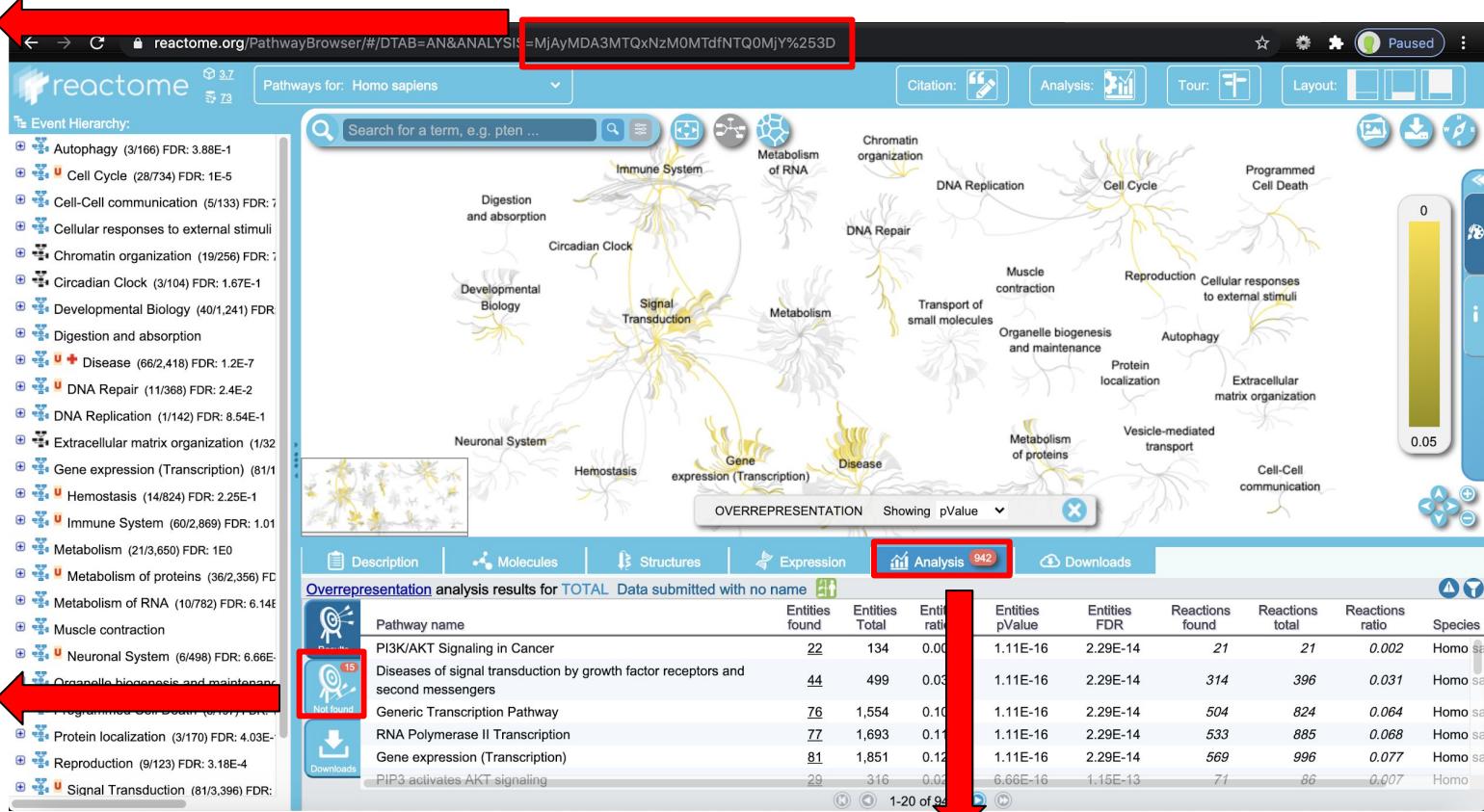
Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sa
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sa
Generic Transcription Pathway	76	1,554	0.016	1.11E-16	2.29E-14	504	824	0.064	Homo sa
RNA Polymerase II Transcription	77	1,693	0.015	1.11E-16	2.29E-14	533	885	0.068	Homo sa
Gene expression (Transcription)	81	1,851	0.126	1.11E-16	2.29E-14	569	996	0.077	Homo sa
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sa

At the bottom, there are tabs for 'Description', 'Molecules', 'Structures', 'Expression', 'Analysis' (with a count of 942), and 'Downloads'. On the left, there are sections for 'Results' (15 items), 'Not found', and 'Downloads'.

Analysis Tools

Tutorial - Pathway Browser

Results can be accessed for at least 7 days using the token string in the URL.



The screenshot shows the Reactome Pathway Browser interface. At the top, a red arrow points to the URL bar containing `reactome.org/PathwayBrowser/#/DTAB=AN&ANALYSIS=MjAyMDA3MTQxNzM0MTdfNTQ0MjY%253D`. Below the URL bar, a red box highlights the search term `Homo sapiens` in the search input field. The main area displays a hierarchical network of biological pathways. A red arrow points to the 'Analysis' tab, which is highlighted with a red border and shows a count of 942 hits. Another red arrow points to the 'Downloads' section, where a blue button labeled 'Not found' is highlighted with a red border. The bottom of the screen shows a table of overrepresentation analysis results.

Pathway name	Entities found	Entities Total	Entity ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
PI3K/AKT Signaling in Cancer	22	134	0.00	1.11E-16	2.29E-14	21	21	0.002	Homo sa
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.03	1.11E-16	2.29E-14	314	396	0.031	Homo sa
Generic Transcription Pathway	76	1,554	0.10	1.11E-16	2.29E-14	504	824	0.064	Homo sa
RNA Polymerase II Transcription	77	1,693	0.11	1.11E-16	2.29E-14	533	885	0.068	Homo sa
Gene expression (Transcription)	81	1,851	0.12	1.11E-16	2.29E-14	569	996	0.077	Homo sa
PIP3 activates AKT signaling	29	316	0.00	6.66E-16	1.15E-13	71	86	0.007	Homo sa

Identifiers in submitted list without hits

Pathways with at least one hit

Analysis Tools

Tutorial - Pathway Browser

(2 dashed boxes):
Legends for
Pathway Browser
highlighting

(Solid box)
The default
colours can be
changed by
clicking the
'painters palette'
icon on the right
side of the
screen.

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a tree view of biological pathways under 'Event Hierarchy'. The main area displays a complex network of pathways, with nodes representing various biological processes like 'Signal Transduction', 'Metabolism', and 'Cell Cycle'. A color scale legend on the right indicates pathway enrichment levels, with a solid blue box highlighting the first few categories. A dashed red box highlights the 'OVERREPRESENTATION' section of the interface, which contains a table of results. The table lists pathways such as 'PI3K/AKT Signaling in Cancer' and 'Diseases of signal transduction by growth factor receptors and second messengers', along with their respective statistics like pValue and FDR.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.106	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens
RNA Polymerase II Transcription	77	1,693	0.115	1.11E-16	2.29E-14	533	885	0.068	Homo sapiens
Gene expression (Transcription)	81	1,851	0.126	1.11E-16	2.29E-14	569	996	0.077	Homo sapiens
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sapiens

Analysis Tools

Tutorial - Pathway Browser

The windows can be shaped to your liking by clicking and dragging on the borders (red boxes).

As well, windows can be hidden/revealed by selecting the appropriate 'Layout' button at the top right (red arrow).

The screenshot shows the Reactome Pathway Browser interface. At the top, there's a search bar with the placeholder "Search for a term, e.g. pten ...". Below the search bar is a navigation bar with tabs: Citation, Analysis, Tour, and Layout. The Layout tab is highlighted with a red arrow pointing to it from the top right. The main content area displays a network diagram of biological pathways like Immune System, Metabolism of RNA, and Cell Cycle. Below the diagram is a table titled "OVERREPRESENTATION Showing pValue". The table has columns for Pathway name, Entities found, Entities Total, Entities ratio, Entities pValue, Entities FDR, Reactions found, Reactions total, Reactions ratio, and Species name. The first few rows of the table are:

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.06	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens
RNA Polymerase II Transcription	77	1,693	0.015	1.11E-16	2.29E-14	533	885	0.068	Homo sapiens
Gene expression (Transcription)	81	1,851	0.126	1.11E-16	2.29E-14	569	996	0.077	Homo sapiens
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sapiens
Negative regulation of the PI3K/AKT network	20	134	0.009	4.55E-15	6.69E-13	8	10	0.001	Homo sapiens
Intracellular signaling by second messengers	29	363	0.025	2.28E-14	2.94E-12	71	114	0.009	Homo sapiens
Transcriptional Regulation by TP53	32	486	0.033	1.66E-13	1.89E-11	196	259	0.02	Homo sapiens
PIP5, PP2A and IER3 Regulate PI3K/AKT Signaling	18	126	0.009	2.29E-13	2.36E-11	5	7	0.001	Homo sapiens
Transcriptional regulation by RUNX3	17	118	0.008	9.52E-13	8.85E-11	35	47	0.004	Homo sapiens
Constitutive Signaling by Aberrant PI3K in Cancer	16	103	0.007	1.49E-12	1.28E-10	2	2	0	Homo sapiens
MAPK1/MAPK3 signaling	26	352	0.024	3.18E-12	2.51E-10	54	82	0.006	Homo sapiens
FLT3 Signaling	26	362	0.025	5.87E-12	4.29E-10	69	97	0.007	Homo sapiens
IRS-mediated signalling	13	65	0.004	9.12E-12	5.8E-10	3	9	0.001	Homo sapiens
ESR-mediated signaling	22	256	0.017	9.67E-12	5.8E-10	60	110	0.008	Homo sapiens

Analysis Tools

Tutorial - Pathway Browser

You can see which identifiers were found in specific Pathways by clicking the corresponding number under the 'Entities found' column.

This will also focus on the corresponding Pathway node in the Pathway Browser.

The screenshot shows the Reactome Pathway Browser interface. At the top, there's a search bar and a navigation bar with various icons. Below the header, a network diagram of biological pathways is displayed, with nodes representing different biological processes like 'Immune System', 'Metabolism of RNA', and 'Cell Cycle'. In the center, a table titled 'OVERREPRESENTATION Showing pValue' lists pathway results. The first few rows include:

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.106	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens
RNA Polymerase II Transcription	77	1,693	0.115	1.11E-16	2.29E-14	533	885	0.068	Homo sapiens
Gene expression (Transcription)	81	1,851	0.126	1.11E-16	2.29E-14	569	996	0.077	Homo sapiens
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sapiens
Negative regulation of the PI3K/AKT network	20	134	0.009	4.55E-15	6.69E-13	8	10	0.001	Homo sapiens
Intracellular signaling by second messengers	29	363	0.025	2.28E-14	2.94E-12	71	114	0.009	Homo sapiens
Transcriptional Regulation by TP53	32	486	0.033	1.66E-13	1.89E-11	196	259	0.02	Homo sapiens
PIP3, PP2A and IER3 Regulate PI3K/AKT Signaling	18	126	0.009	2.29E-13	2.36E-11	5	7	0.001	Homo sapiens
Transcriptional regulation by RUNX3	17	118	0.008	9.52E-13	8.85E-11	35	47	0.004	Homo sapiens
Constitutive Signaling by Aberrant PI3K in Cancer	16	103	0.007	1.49E-12	1.28E-10	2	2	0	Homo sapiens
MAPK1/MAPK3 signaling	26	352	0.024	3.18E-12	2.51E-10	54	82	0.006	Homo sapiens
FLT3 Signaling	26	362	0.025	5.87E-12	4.29E-10	69	97	0.007	Homo sapiens
IRS-mediated signalling	13	65	0.004	9.12E-12	5.8E-10	3	9	0.001	Homo sapiens
ESR-mediated signaling	22	256	0.017	9.67E-12	5.8E-10	60	110	0.008	Homo sapiens

Analysis Tools

Tutorial - Pathway Browser

This image shows matching identifiers for the 'PI3K/AKT Signaling in Cancer' Pathway.

Clicking 'Back to results overview' will return you to the default statistics table.

The screenshot shows the Reactome Pathway Browser interface. At the top, there's a search bar with the placeholder "Search for a term, e.g. pten ...". Below the search bar, a large network diagram illustrates various biological pathways, with a specific cluster highlighted in yellow. A tooltip for "Gene expression (Transcription)" is visible near one of the nodes. On the right side of the screen, there's a circular diagram showing "OVERREPRESENTATION" with a p-value of 942. The main content area displays a table of "Overrepresentation analysis results for TOTAL Data submitted with no name". The table has three columns: "Identifiers found", "Resource UNIPROT", and "Resource ENSEMBL". The "Identifiers found" column lists 15 entries, each with a small icon next to it. The "Resource UNIPROT" column lists corresponding protein identifiers, and the "Resource ENSEMBL" column lists their respective ENSEMBL IDs. The table includes navigation links like "Back to results overview" and "Matching identifiers for: PI3K/AKT Signaling in Cancer". At the bottom of the table, there are links for "Results", "Not found", and "Downloads". The footer of the browser window shows the URL "reactome.org/PathwayBrowser/#/R-HSA-221952&DTAB=AN&ANALYSIS=MjAyMDA3MTQxNzM0MTdfNTQ0MjY%253D".

Identifiers found	Resource UNIPROT	Resource ENSEMBL
PTPN11	Q06124	
PIK3R1	P27986	
ERBB4	Q15303-2, Q15303-1	
PIK3CA	P42336	
KIT	P10721	
PDGFRA	P16234	
AKT1	Q9Y243, P31749	
FGFR2	P21802-18, P21802-1, P21802-5, P21802-3	
FGFR3	P22607-2, P22607-1	
PTEN	P60484	
HGF	P14210	
CDKN1A	P38936	
CDKN1B	P46527	

Analysis Tools

Tutorial - Pathway Browser

This image shows matching identifiers for the 'PI3K/AKT Signaling in Cancer' Pathway.

Clicking 'Back to results overview' will return you to the default statistics table.

The screenshot shows the Reactome Pathway Browser interface. At the top, there's a search bar with the placeholder "Search for a term, e.g. pten ...". Below the search bar, a large network diagram is visible, with a specific node labeled "Gene expression (Transcription)" highlighted. To the right of the diagram, a tooltip displays "OVERREPRESENTATION Showing pValue". The main content area shows a table titled "Overrepresentation analysis results for TOTAL Data submitted with no name". The table has three columns: "Identifiers found", "Resource UNIPROT", and "Resource ENSEMBL". A red arrow points to the "Back to results overview" link in the table header. The table lists 15 entries, each with a protein name and its corresponding UniProt ID. The bottom of the table shows a page number "1-15 of 15".

Overrepresentation analysis results for TOTAL Data submitted with no name		
Identifiers found	Resource UNIPROT	Resource ENSEMBL
PTPN11	Q06124	
PIK3R1	P27986	
ERBB4	Q15303-2, Q15303-1	
PIK3CA	P42336	
KIT	P10721	
PDGFRA	P16234	
AKT1	Q9Y243, P31749	
FGFR2	P21802-18, P21802-1, P21802-5, P21802-3	
FGFR3	P22607-2, P22607-1	
PTEN	P60484	
HGF	P14210	
CDKN1A	P38936	
CDKN1B	P46527	

Analysis Tools

Tutorial - Statistics

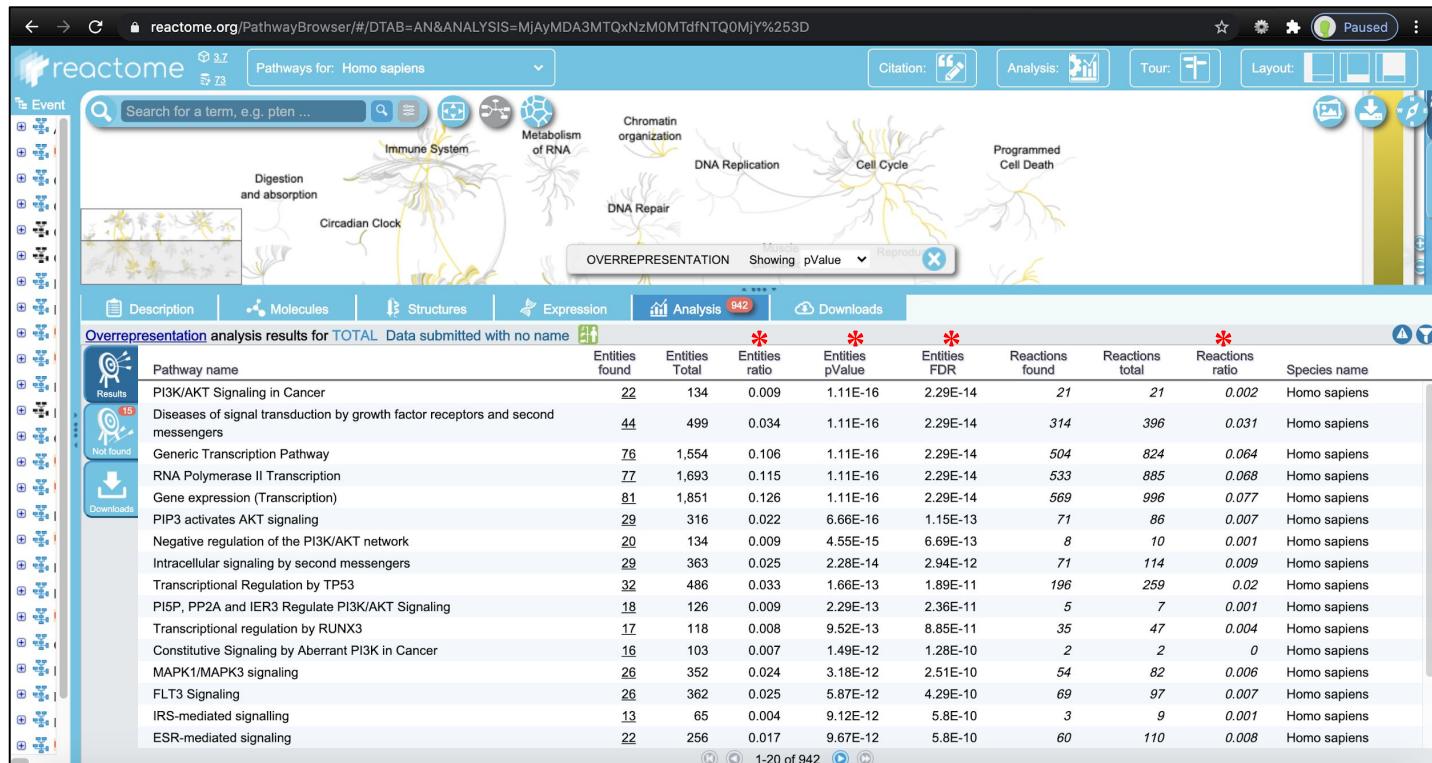
Stats Table Overview

Entities ratio:
Total Entities in Pathway/Total Entities in all Pathways.

Entities pValue:
Over-representation analysis result.

Entities FDR:
False discovery rate.
Corrected over-representation analysis result.

Reactions ratio:
Total Reactions in Pathway/Total Reactions in all Pathways.



The screenshot shows the Reactome Pathway Browser interface. At the top, there's a search bar with the placeholder "Search for a term, e.g. pten ...". Below the search bar, a navigation menu includes "Citation", "Analysis", "Tour", and "Layout". The main area displays a network diagram of biological pathways, with major clusters labeled: Immune System, Metabolism of RNA, Chromatin organization, DNA Replication, Cell Cycle, and Programmed Cell Death. A sidebar on the left shows a tree view of the pathway hierarchy. In the center, a table titled "OVERREPRESENTATION Showing pValue" lists 19 pathway results. The columns include: Pathway name, Entities found, Entities Total, Entities ratio, Entities pValue, Entities FDR, Reactions found, Reactions total, Reactions ratio, and Species name. The first few rows show results for PI3K/AKT Signaling in Cancer, Diseases of signal transduction by growth factor receptors and second messengers, Generic Transcription Pathway, and RNA Polymerase II Transcription.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.046	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens
RNA Polymerase II Transcription	77	1,693	0.044	1.11E-16	2.29E-14	533	885	0.068	Homo sapiens
Gene expression (Transcription)	81	1,851	0.043	1.11E-16	2.29E-14	569	996	0.077	Homo sapiens
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sapiens
Negative regulation of the PI3K/AKT network	20	134	0.009	4.55E-15	6.69E-13	8	10	0.001	Homo sapiens
Intracellular signaling by second messengers	29	363	0.025	2.28E-14	2.94E-12	71	114	0.009	Homo sapiens
Transcriptional Regulation by TP53	32	486	0.033	1.66E-13	1.89E-11	196	259	0.02	Homo sapiens
PIP5, PP2A and IER3 Regulate PI3K/AKT Signaling	18	126	0.009	2.29E-13	2.36E-11	5	7	0.001	Homo sapiens
Transcriptional regulation by RUNX3	17	118	0.008	9.52E-13	8.85E-11	35	47	0.004	Homo sapiens
Constitutive Signaling by Aberrant PI3K in Cancer	16	103	0.007	1.49E-12	1.28E-10	2	2	0	Homo sapiens
MAPK1/MAPK3 signaling	26	352	0.024	3.18E-12	2.51E-10	54	82	0.006	Homo sapiens
FLT3 Signaling	26	362	0.025	5.87E-12	4.29E-10	69	97	0.007	Homo sapiens
IRS-mediated signalling	13	65	0.004	9.12E-12	5.8E-10	3	9	0.001	Homo sapiens
ESR-mediated signaling	22	256	0.017	9.67E-12	5.8E-10	60	110	0.008	Homo sapiens

Analysis Tools

Tutorial - Statistics

Entities pValue: The result of the binomial test using the sample size and the previously explained entities ratio.

- Reports statistical likelihood that Pathway is over-represented in the submitted identifiers list

Entities FDR: The FDR (Benjamini Hochberg*) method is the default p-value adjustment method in the analysis tool. In this method, the p-values are first sorted and ranked. The smallest value gets rank 1, the second rank 2, and the largest gets rank N. Then, each p-value is multiplied by N and divided by its assigned rank to give the adjusted p-values.

- Corrects for false-positives in over-representation analysis. More reliably reports likelihood that Pathway is over-represented in submitted identifiers list.

Fabregat, Antonio, et al. "Reactome pathway analysis: a high-performance in-memory approach." *BMC bioinformatics* 18.1 (2017): 142.

* Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995): 289-300.

Analysis Tools

Tutorial - Reacfoam

You can also view the Pathway hierarchy with an alternative 'Reacfoam' (Voronoi tessellation) view. This view will also be overlaid with analysis results.

Click the 'Open voronoi view' button (red box).

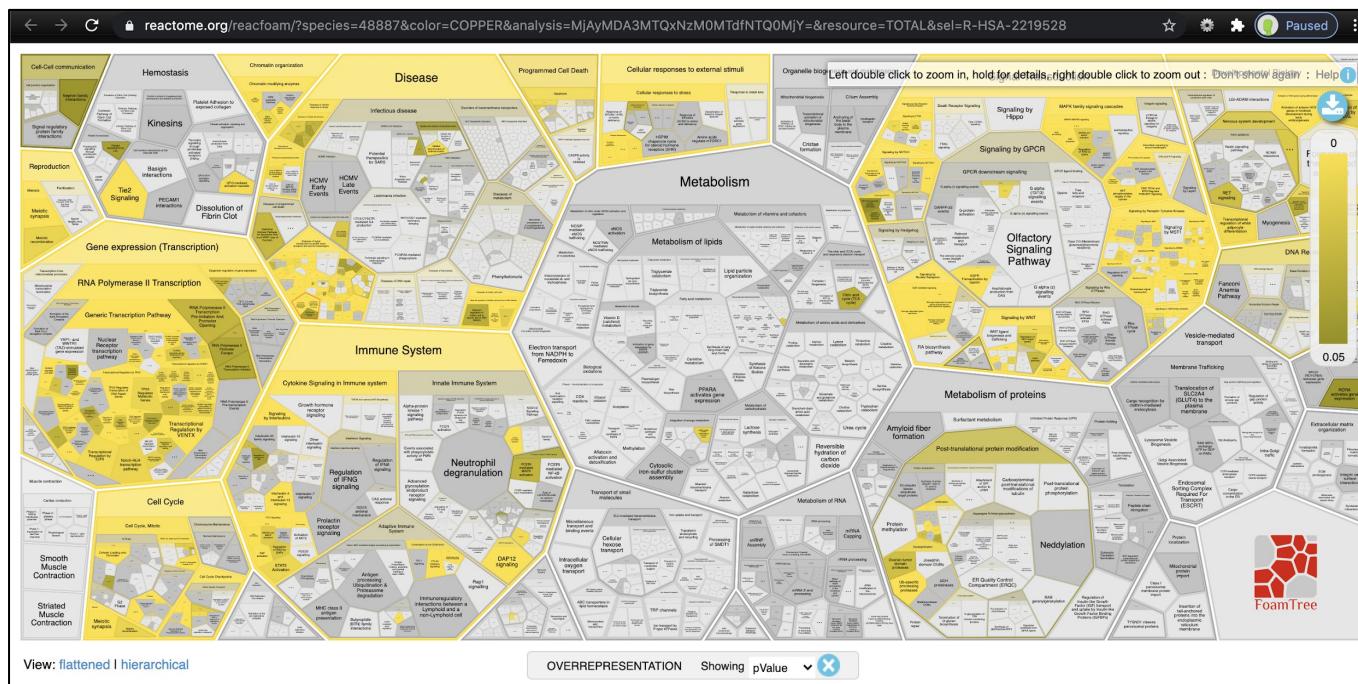
This opens another window/tab to the Reacfoam view.

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a sidebar titled 'Event Hierarchy' listing various biological pathways. In the center, a large network diagram (Reacfoam view) shows the relationships between different biological processes. At the bottom, there is a table titled 'OVERREPRESENTATION' showing analysis results for various pathways. The 'Analysis' tab is selected in the navigation bar. A red box highlights the 'Open voronoi view' button in the top right corner of the main interface.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
Signaling by NTRKs	10	166	0.011	1.01E-4	6.08E-4	31	163	0.013	Homo Sapiens
Negative regulation of FGFR3 signaling	5	34	0.002	1.11E-4	6.67E-4	12	14	0.001	Homo Sapiens
Signaling by VEGF	9	137	0.009	1.19E-4	7.14E-4	12	84	0.006	Homo Sapiens
Ub-specific processing proteases	11	206	0.014	1.31E-4	7.15E-4	10	40	0.003	Homo Sapiens
NOTCH1 Intracellular Domain Regulates Transcription	6	57	0.004	1.41E-4	7.15E-4	16	18	0.001	Homo Sapiens
SHC1 events in EGFR signaling	4	19	0.001	1.43E-4	7.15E-4	4	4	0	Homo Sapiens
Aberrant regulation of mitotic cell cycle due to RB1 defects	5	36	0.002	1.45E-4	7.23E-4	3	3	0	Homo Sapiens
Diseases of mitotic cell cycle	5	36	0.002	1.45E-4	7.23E-4	3	3	0	Homo Sapiens

Analysis Tools

Tutorial - Reacfoam



Reacfoam allows you to see results overlaid throughout the Pathway hierarchy instead of just the eagle-eye view offered by the traditional ‘fireworks’ Pathway view.

This view divides each Top Level Pathway into a 'cell' that contains its sub-Pathways.

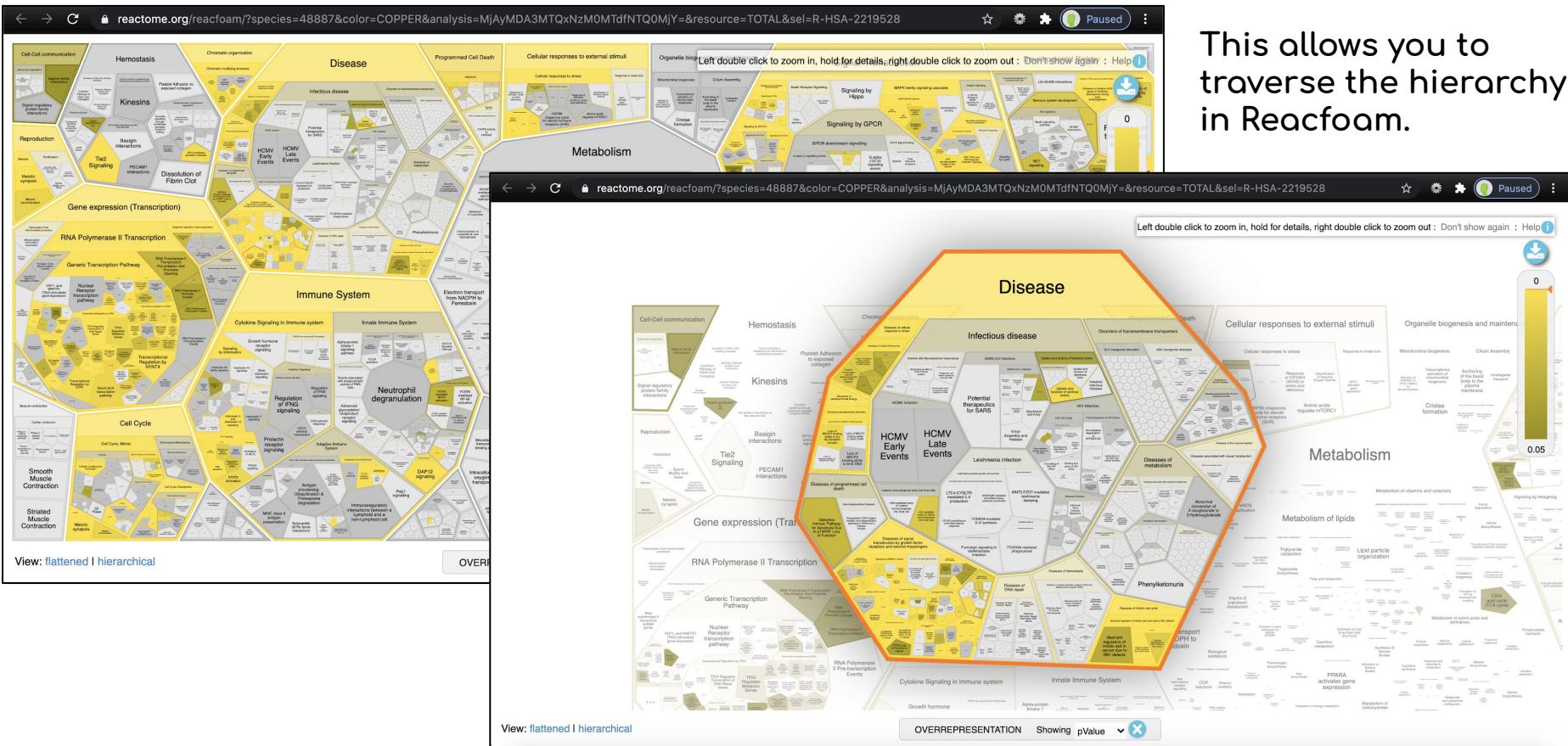


Analysis Tools

Tutorial - Reacfoam

You can click on a cell's 'header' to expand your view of it.

This allows you to traverse the hierarchy in Reacfoam.



Analysis Tools

Tutorial - Report download

You can also download the results in CSV or JSON format, as well as a PDF Analysis report.

Select the 'Downloads' button (red box).

Click on 'Analysis report' to view the PDF report.

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a sidebar titled 'Event Hierarchy' with a tree view of biological processes. The main area displays a network diagram with nodes like 'Gene expression (Transcription)', 'Meiotic synapsis', and 'Disease'. A color scale on the right indicates p-values from 0 to 0.05. At the bottom, there are several download options:

- Pathway analysis results**: Download pathway analysis results in CSV format.
- Complete analysis results**: Download complete analysis results in JSON format.
- Identifier mappings**: Download identifier mappings between data and resources in CSV format.
- Analysis report**: Download a detailed report in PDF format (highlighted with a red box and arrow).
- Not found identifiers**: Download a CSV file of unmapable identifiers.

Analysis Tools

Tutorial - Analysis Report



Pathway Analysis Report

This report contains the pathway analysis results for the submitted sample ". Analysis was performed against Reactome version 73 on 15/07/2020. The web link to these results is:

<https://reactome.org/PathwayBrowser/#/ANALYSIS=MjAyMDA3MTQxNzM0M0TdfNTQ0MjY%3D>

Please keep in mind that analysis results are temporarily stored on our server. The storage period depends on usage of the service but is at least 7 days. As a result, please note that this URL is only valid for a limited time period and it might have expired.

Table of Contents

1. Introduction
2. Properties
3. Genome-wide overview
4. Most significant pathways
5. Pathways details
6. Identifiers found
7. Identifiers not found

4. Most significant pathways

The following table shows the 25 most relevant pathways sorted by p-value.

Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
PI3K/AKT Signaling in Cancer	22 / 134	0.009	1.11e-16	2.29e-14	21 / 21	0.002
Diseases of signal transduction by growth factor receptors and second messengers	44 / 499	0.034	1.11e-16	2.29e-14	314 / 396	0.031
Generic Transcription Pathway	76 / 1,554	0.106	1.11e-16	2.29e-14	504 / 824	0.064
RNA Polymerase II Transcription	77 / 1,693	0.115	1.11e-16	2.29e-14	533 / 885	0.068
Gene expression (Transcription)	81 / 1,851	0.126	1.11e-16	2.29e-14	569 / 996	0.077
PIP3 activates AKT signaling	29 / 316	0.022	6.66e-16	1.15e-13	71 / 86	0.007
Negative regulation of the PI3K/AKT network	20 / 134	0.009	4.55e-15	6.69e-13	8 / 10	7.71e-04
Intracellular signaling by second messengers	29 / 363	0.025	2.28e-14	2.94e-12	71 / 114	0.009
Transcriptional Regulation by TP53	32 / 486	0.033	1.66e-13	1.89e-11	196 / 259	0.02
PIP3, PP2A and IER3 Regulate PI3K/AKT Signaling	18 / 126	0.009	2.29e-13	2.36e-11	5 / 7	5.40e-04
Transcriptional regulation by RUNX3	17 / 118	0.008	9.52e-13	8.85e-11	35 / 47	0.004
Constitutive Signaling by Aberrant PI3K in Cancer	16 / 103	0.007	1.49e-12	1.28e-10	2 / 2	1.54e-04
MAPK1/MAPK3 signaling	26 / 352	0.024	3.18e-12	2.51e-10	54 / 82	0.006
FLT3 Signaling	26 / 362	0.025	5.87e-12	4.29e-10	69 / 97	0.007
IRS-mediated signalling	13 / 65	0.004	9.12e-12	5.80e-10	3 / 9	6.94e-04

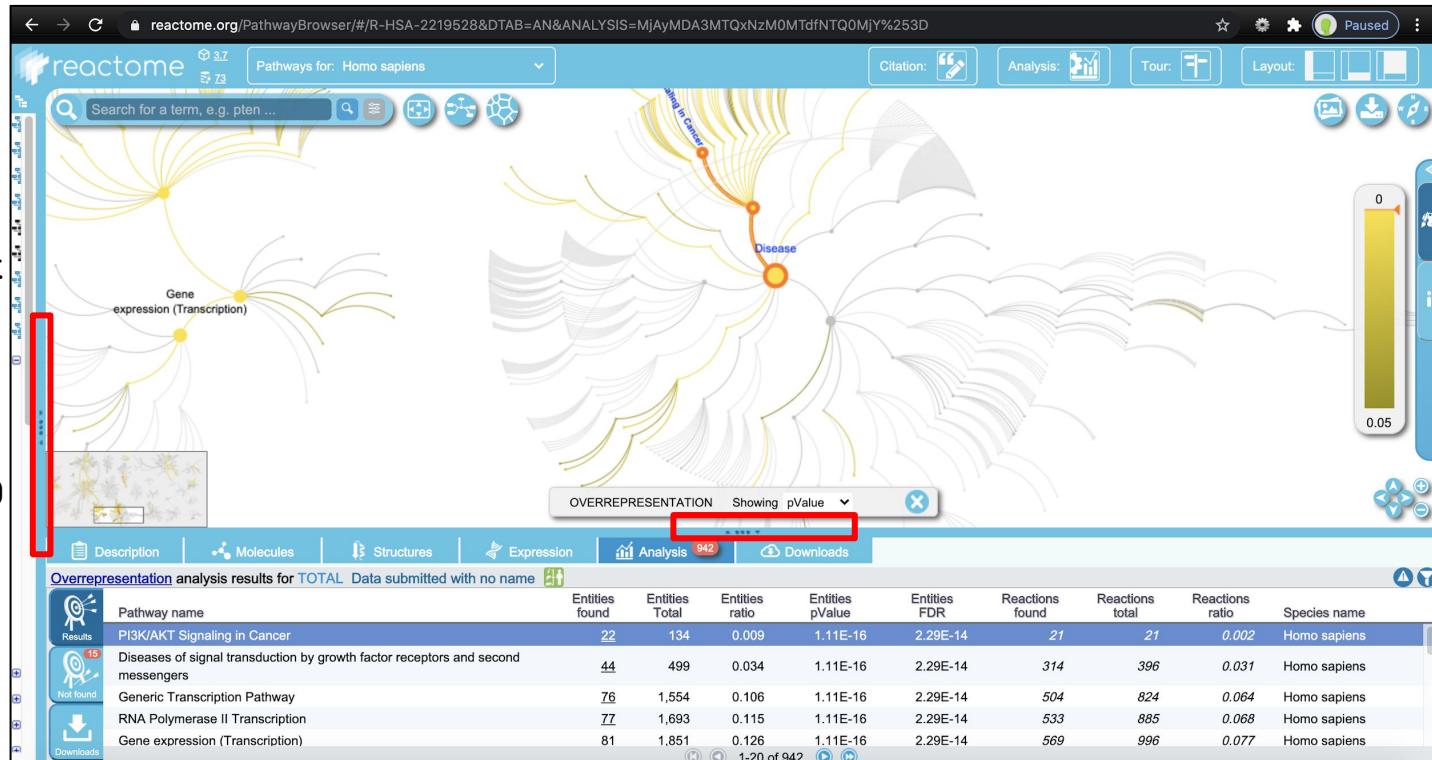


Analysis Tools

Tutorial - Traversing hierarchy

Now we'll traverse a Pathway hierarchy to look at a specific Reaction to see how it looks with a results overlay.

Expand the Pathway Browser window using the borders so that it takes up most of the screen.



Analysis Tools

Tutorial - Traversing hierarchy

In the Top-Left corner of the Pathway Browser, type 'Cell Cycle' in the search and press Enter or double-click it in the search results window that pops up.

This should bring you to the 'Cell Cycle' Pathway node.

The screenshot shows the Reactome Pathway Browser interface. In the top-left corner, there is a search bar with a magnifying glass icon. A red arrow points from the text 'In the Top-Left corner of the Pathway Browser, type 'Cell Cycle'' to this search bar. Below the search bar, a dropdown menu is open, listing various search terms related to 'Cell Cycle', such as 'cell cycle-associated', 'cell-associated cycle', etc. The main central area displays a hierarchical pathway map for the 'Cell Cycle'. On the right side, there is a large network diagram showing interactions between pathways like 'PI3K/AKT Signaling in Cancer' and 'Disease'. At the bottom, there is a table titled 'Overrepresentation analysis results for TOTAL Data submitted with no name'.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.106	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens

Analysis Tools

Tutorial - Traversing hierarchy

Once you can see the 'Cell Cycle' node in the Pathway Browser, double-click it.

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a tree view of the 'Event Hierarchy' under the 'Pathways for: Homo sapiens' tab. The 'Cell Cycle' node is expanded, showing its sub-categories: Cell Cycle, Cell Cycle Ch, Cell Cycle, I, Mitotic G, S Phase, Mitotic G2, M Phase, Regulatory, Chromosome, Meiosis, and Cell-Cell communication. A red arrow points to the 'Cell Cycle' node in the hierarchy. In the center, there is a search bar with 'Cell Cycle' and a results panel showing 'Results (5797)'. Below the search bar, a detailed view of the 'Cell Cycle' pathway is shown, including its ID (R-HSA-1640170) and a 'GO' button. To the right, a large network graph displays various pathways and their interactions. At the bottom, there is an 'OVERREPRESENTATION' analysis table with tabs for 'Description', 'Molecules', 'Structures', 'Expression', 'Analysis 942', and 'Downloads'. The 'Analysis' tab is selected, showing results for 'TOTAL' data submitted with no name. The table includes columns for Pathway name, Entities found, Entities Total, Entities ratio, Entities pValue, Entities FDR, Reactions found, Reactions total, Reactions ratio, and Species name. The 'Cell Cycle' pathway is highlighted in blue in this table. A color scale on the right indicates a value from 0 to 0.05.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Diseases	1	1	0.001	0.0001	0.0001	0	0	0	Homo sapiens
PI3K events in ERBB2 signaling	6	22	0.001	6.89E-7	8.26E-6	7	7	0.001	Homo sapiens
Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors	8	52	0.004	6.95E-7	8.34E-6	15	44	0.003	Homo sapiens
Cell Cycle	28	734	0.05	7.9E-7	1E-5	163	449	0.035	Homo sapiens

Analysis Tools

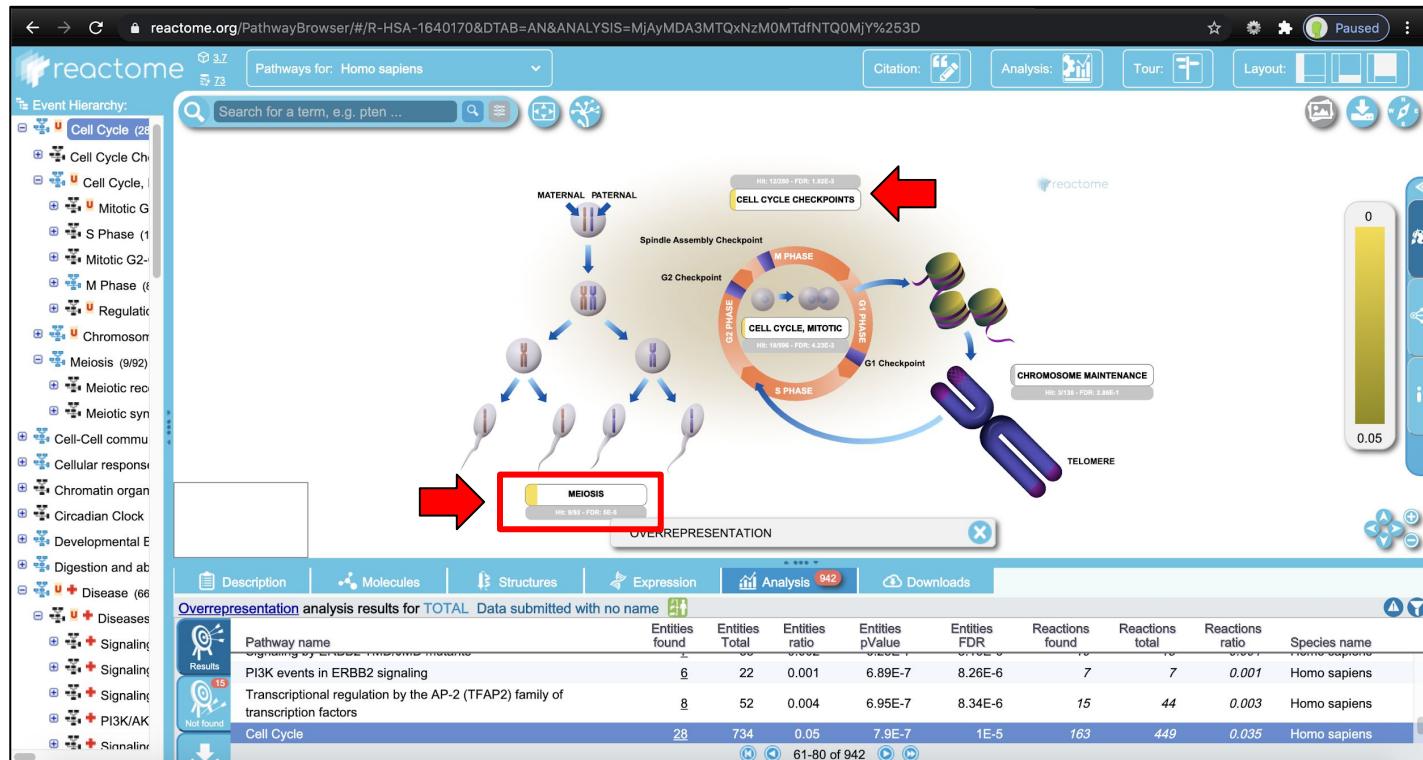
Tutorial - Traversing hierarchy

After double-clicking the 'Cell Cycle' node, you will see the 'Cell Cycle' EHLD* with Analysis results overlaid (red arrows).

The white and yellow boxes show the proportion of submitted identifiers (yellow) found in each Pathway.

Double-click 'Meiosis'.

(9/92 hits, FDR 5E-5)



*EHLD = Enhanced High Level Diagram

Analysis Tools

Tutorial - Traversing hierarchy

You should now be viewing the 'Meiosis' Pathway EHLD.

We need to traverse one more level to see our analysis results at the Reaction level.

Double-click 'Meiotic Recombination'.

(4/58 hits, FDR 2.4E-2)

The screenshot shows the Reactome Pathway Browser interface. On the left, the 'Event Hierarchy' tree is expanded to show the 'Meiosis' branch under 'Meiosis (9/92)'. A red box highlights the 'Meiotic recombination' node. The main panel displays a diagram of meiotic synapsis, showing chromosomes pairing and exchange. A callout box labeled 'MEIOTIC RECOMBINATION' is shown. Below the diagram, an 'OVERREPRESENTATION' section displays a table of analysis results:

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Interleukin-4 and Interleukin-13 signaling	14	211	0.014	1.39E-6	1.67E-5	5	46	0.004	Homo sapiens
Transcriptional activation of p53 responsive genes	4	6	0	1.63E-6	1.79E-5	5	5	0	Homo sapiens
Transcriptional activation of cell cycle inhibitor p21	4	6	0	1.63E-6	1.79E-5	5	5	0	Homo sapiens
Negative regulation of FGFR2 signaling	7	41	0.003	1.78E-6	2E-5	12	14	0.001	Homo sapiens

Analysis Tools

Tutorial - Reactions view

This is the Reaction-level view of the 'Meiotic recombination' Pathway with analysis results overlaid on PhysicalEntities (red arrows; yellowish boxes).

You can zoom in using a mouse scroll or the buttons at the bottom-right (red box).

Zoom in towards the centre of the diagram.

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a tree view of the Event Hierarchy, with the 'Meiosis (9/92)' node expanded. In the center, a reaction network diagram for the 'Meiotic recombination' pathway is displayed, showing various reactions and entities. Red arrows and yellowish boxes highlight specific entities of interest. At the bottom, an 'Overrepresentation' analysis table is shown, with a red box highlighting the zoom controls in the bottom-right corner of the table header.

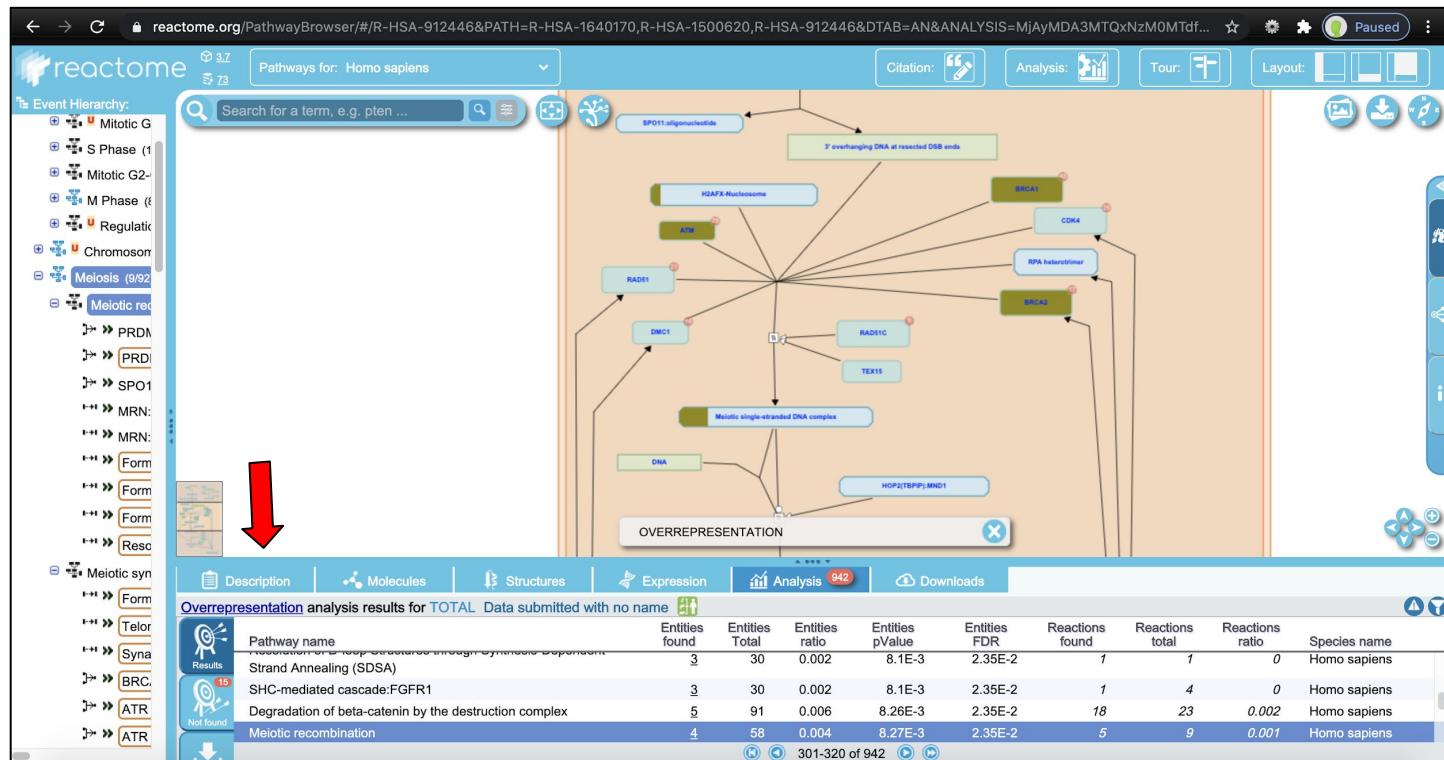
Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Resolution of D-loop structures through Synthesis-Dependent Strand Annealing (SDSA)	3	30	0.002	8.1E-3	2.35E-2	1	1	0	Homo sapiens
SHC-mediated cascade:FGFR1	3	30	0.002	8.1E-3	2.35E-2	1	4	0	Homo sapiens
Degradation of beta-catenin by the destruction complex	5	91	0.006	8.26E-3	2.35E-2	18	23	0.002	Homo sapiens
Meiotic recombination	4	58	0.004	8.27E-3	2.35E-2	5	9	0.001	Homo sapiens

Analysis Tools

Tutorial - Reactions view

Click on the central Reaction ('Formation of meiotic single-stranded DNA invasion complex'). The entire Reaction will become highlighted in blue.

No statistical results are displayed at the Reaction level, but you can click on the 'Description' tab to see the Reaction's curated information (red arrow).



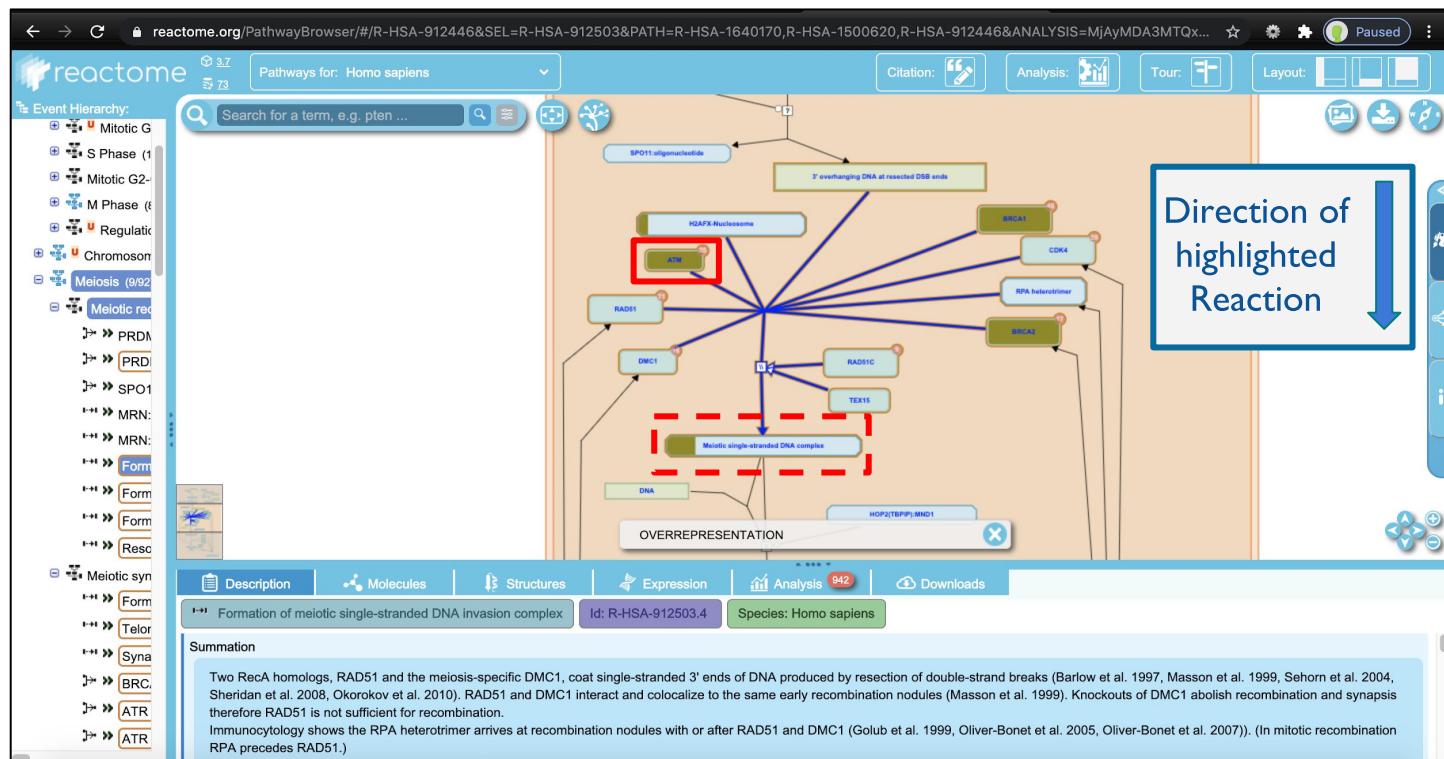
Analysis Tools

Tutorial - Reactions view

We made it! Now we are looking at a specific Reaction that contains some identifiers from our submitted list.

A PhysicalEntity is coloured fully (red box) when all contained entities had at least 1 identifier present in the submitted list.

An entity is coloured proportionally (dashed box) according to the number of identifiers it contains from the submitted list.



Reactome Analysis

Outline

Topics

1. Overview of Analysis results
2. Tutorial - Analyse gene list via Analysis Tools
(browser)
3. Tutorial - Analyse gene list via Analysis Service
(command line)

Analysis Service

API Tutorial

Reactome's analysis tools can also be accessed via a RESTful web service. This makes it easy to run analyses from the command-line or your own application

A unique token is generated for every submitted analysis

- Allows users to view their results over multiple visits
- Results can be viewed in the Pathway Browser or retrieved programmatically
- Results associated with a token will persist for at least 7 days

Analysis types are run automatically depending on the format of data submitted

- Single-column lists: Over-representation analysis
- Expression matrix: Expression analysis

Analysis Service

API Tutorial - Submitting requests

Submitting a single identifier as a GET request to the '/identifier/{identifier}' endpoint using curl

```
$ curl -X GET "https://reactome.org/AnalysisService/identifier/TP53" -H "accept: application/json" > TP53-analysis-results.json
```

Submitting multiple identifiers requires the '-d' (--data) flag for the identifiers and a POST request to the '/identifiers/' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/" -H "accept: application/json" -H "content-type: text/plain" -d "TP53 PTEN" > TP53-PTEN-analysis-results.json
```

Submitting an online dataset requires the '-d' flag for the dataset URL and a POST request to the '/identifiers/url' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/url" -H "accept: application/json" -H "content-type: text/plain" -d "https://www.ebi.ac.uk/pride/ws/archive/protein/list/assay/27929.acc" > url-analysis-results.json
```

Submitting a local file requires 'content-type' to be 'multipart/form-data' and the '-F' (--form) flag for the file as a POST request to the '/identifiers/form/' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```

Analysis Service

API Tutorial - Submitting requests

Submitting a single identifier as a GET request to the '/identifier/{identifier}' endpoint using curl

```
$ curl -X GET "https://reactome.org/AnalysisService/identifier/TP53" -H "accept: application/json" > TP53-analysis-results.json
```

Submitting multiple identifiers requires the '-d' (--data) flag for the identifiers and a POST request to the '/identifiers/' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/" -H "accept: application/json" -H "content-type: text/plain" -d "TP53 PTEN" > TP53-PTEN-analysis-results.json
```

Submitting an online dataset requires the '-d' flag for the dataset URL and a POST request to the '/identifiers/url' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/url" -H "accept: application/json" -H "content-type: text/plain" -d "https://www.ebi.ac.uk/pride/ws/archive/protein/list/assay/27929.acc" > url-analysis-results.json
```

Submitting a local file requires 'content-type' to be 'multipart/form-data' and the '-F' (--form) flag for the file as a POST request to the '/identifiers/form/' endpoint

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```



Analysis Service

API Tutorial - URL parameters

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```

Results can be filtered or tweaked using URL parameters:

- interactors=false|true → Whether to include interactors in the analysis (default: false)
- includeDisease=false|true → Whether to include Disease Pathway in analysis (default: false)
- pValue=# → Sets pValue threshold. Only Pathways with pValue at or below value are kept.
- sortBy=PROPERTY → Which property the returned dataset can be sorted by (eg: pValue)
- order=ASC|DSC → Whether the 'sortBy' is ascending/descending
- pageSize=#, page=# → # of Pathways per page and which page # to return, respectively

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/?pValue=1" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```



Analysis Service

API Tutorial - URL building

To replicate the analysis that was run during the Pathway Browser tutorial, we will use:

- The /identifiers/form/projection endpoint
- pValue threshold of 1
- Results sorted ascendingly by pValue using 'ENTITIES_PVALUE':

Original URL submission:

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```

New URL submission with '/projection/?pValue=1&sortBy=ENTITIES_PVALUE&order=ASC` added:

```
(base) jcook@justin-oicr:~$ curl -X POST "https://reactome.org/AnalysisService/identifiers/form/projection?pValue=1&sortBy=ENTITIES_PVALUE&order=ASC" -H "accept: application/json" -H "content-type: multipart/form-data" -F file=@TwelveTumorsNaturePaper127genes.txt > analysis-TumoursNaturePaper127genes.json
```



Analysis Service

API Tutorial - Results

Results from the Analysis Service are returned in JSON format*.

The 'summary' section contains information such as specified parameters, the type of analysis run, and the token string.

```
analysis-TumoursNaturePaper127genes.j...
1  {
2    "summary": {
3      "token": "MjAyMDA3MTUxNzQzMfNTk0NzE%3D",
4      "projection": true,
5      "interactors": false,
6      "type": "OVERREPRESENTATION",
7      "sampleName": "",
8      "text": false,
9      "includeDisease": true,
10     "fileName": "TwelveTumorsNaturePaper127genes.txt"
11   },
12   "expression": {=},
15   "identifiersNotFound": 15,
16   "pathwaysFound": 942,
17   "pathways": [=],
25453   "resourceSummary": [=],
25470   "speciesSummary": [=],
25479   "warnings": []
25480 }
25481
```

*JSON prettified in Atom

Analysis Service

API Tutorial - Results

The 'pathways' section contains information about Pathways that had any hits.

This entails basic Pathway instance information (dbId, name, species, inDisease) as well as Reaction/Entity statistics.

```
analysis-TumoursNaturePaper127genes.j...
15 "identifiersNotFound": 15,
16 "pathwaysFound": 942,
17 "pathways": [
18   {
19     "stId": "R-HSA-2219528",
20     "dbId": 2219528,
21     "name": "PI3K/AKT Signaling in Cancer",
22     "species": {
23       "dbId": 48887,
24       "taxId": "9606",
25       "name": "Homo sapiens"
26     },
27     "llp": false,
28     "entities": {
29       "resource": "TOTAL",
30       "total": 134,
31       "found": 22,
32       "ratio": 0.0091324200913242,
33       "pValue": 1.1102230246251565e-16,
34       "fdr": 2.2870594307278225e-14,
35       "exp": []
36     },
37     "reactions": {
38       "resource": "TOTAL",
39       "total": 21,
40       "found": 21,
41       "ratio": 0.001620120351797562
42     },
43     "inDisease": true
44   }
]
```

Analysis Service

API Tutorial - Results

The ‘pathways’ section in the JSON results contains the same information as the statistical table seen in the Pathway Browser view!

The screenshot shows the Reactome PathwayBrowser interface. At the top, a search bar contains the query "Gene expression (Transcription)". Below the search bar, a network diagram highlights this pathway. The main panel displays an "OVERREPRESENTATION" analysis for "TOTAL" data, showing a pValue of 942. The results table lists various biological processes, each with its entity count, total entities, ratio, and pValue. The table includes columns for Species name, Reactions found, Reactions total, and Reactions ratio. A sidebar on the left provides navigation links for Description, Molecules, Structures, Expression, Analysis (with 942 items), and Downloads.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
PI3K/AKT Signaling in Cancer	22	134	0.009	1.11E-16	2.29E-14	21	21	0.002	Homo sapiens
Diseases of signal transduction by growth factor receptors and second messengers	44	499	0.034	1.11E-16	2.29E-14	314	396	0.031	Homo sapiens
Generic Transcription Pathway	76	1,554	0.06	1.11E-16	2.29E-14	504	824	0.064	Homo sapiens
RNA Polymerase II Transcription	77	1,693	0.015	1.11E-16	2.29E-14	533	885	0.068	Homo sapiens
Gene expression (Transcription)	81	1,851	0.126	1.11E-16	2.29E-14	569	996	0.077	Homo sapiens
PIP3 activates AKT signaling	29	316	0.022	6.66E-16	1.15E-13	71	86	0.007	Homo sapiens
Negative regulation of the PI3K/AKT network	20	134	0.009	4.55E-15	6.69E-13	8	10	0.001	Homo sapiens
Intracellular signaling by second messengers	29	363	0.025	2.28E-14	2.94E-12	71	114	0.009	Homo sapiens
Transcriptional Regulation by TP53	32	486	0.033	1.66E-13	1.89E-11	196	259	0.02	Homo sapiens
PIP3, PP2A and IER3 Regulate PI3K/AKT Signaling	18	126	0.009	2.29E-13	2.36E-11	5	7	0.001	Homo sapiens
Transcriptional regulation by RUNX3	17	118	0.008	9.52E-13	8.85E-11	35	47	0.004	Homo sapiens
Constitutive Signaling by Aberrant PI3K in Cancer	16	103	0.007	1.49E-12	1.28E-10	2	2	0	Homo sapiens
MAPK1/MAPK3 signaling	26	352	0.024	3.18E-12	2.51E-10	54	82	0.006	Homo sapiens
FLT3 Signaling	26	362	0.025	5.87E-12	4.29E-10	69	97	0.007	Homo sapiens
IRS-mediated signalling	13	65	0.004	9.12E-12	5.8E-10	3	9	0.001	Homo sapiens
ESR-mediated signalling	22	256	0.017	9.67E-12	5.8E-10	60	110	0.008	Homo sapiens
RAF/MAP kinase cascade	25	345	0.024	1.29E-11	6.94E-10	52	75	0.006	Homo sapiens

analysis-TumoursNaturePaper127genes.i

```
"identifiersNotFound": 15,
"pathwaysFound": 942,
"pathways": [
  {
    "stId": "R-HSA-2219528",
    "dbId": 2219528,
    "name": "PI3K/AKT Signaling in Cancer",
    "species": {
      "dbId": 48887,
      "taxId": "9606",
      "name": "Homo sapiens"
    },
    "llp": false,
    "entities": {
      "resource": "TOTAL",
      "total": 134,
      "found": 22,
      "ratio": 0.0091324200913242,
      "pValue": 1.102230246251565e-16,
      "fdr": 2.2870594307278225e-14,
      "exp": []
    },
    "reactions": {
      "resource": "TOTAL",
      "total": 21,
      "found": 21,
      "ratio": 0.001620120351797562
    },
    "inDisease": true
```

Analysis Service

API Tutorial - Token

The token system allows you to view your results in the Pathway Browser or retrieve them programmatically for at least 7 days after submitting

Token in Pathway Browser URL:

- ANALYSIS={token string}
- Can add 'DTAB=AN' to URL so that the details panel is set to the 'Analysis' tab

```
https://reactome.org/PathwayBrowser/#/DTAB=AN&ANALYSIS=MjAyMDA3MTUxNzQzMZfNTk0NzE%3D
```

Retrieving analysis results programmatically using token:

- Retrieve original results
- Can also filter results further using many other endpoints

```
(base) jcook@justin-oicr:~$ curl -X GET "https://reactome.org/AnalysisService/token/%20MjAyMDA3MTUxNzQzMZfNTk0NzE%253D"  
-H "accept: application/json" > results-token.json
```



Reactome Analysis

Summary

Today, we went over the submission of a single-column file that contained gene symbols to Reactome's Analysis Tools (browser) and Analysis Service (RESTful API).

Running Analysis Tools through the Reactome website:

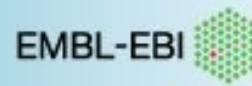
- Pathway-topology analysis
 - Results are overlaid in Reactome's Pathway Browser
 - Fireworks
 - Reacfoam
 - Reaction-level diagrams
- Over-representation analysis
 - Shows which Pathways are over-represented in the submitted identifiers list

The Analysis Service, queried programmatically, has nearly all of the functionality of the Analysis Tools except, of course, for the Pathway Browser overlays.

There is a lot more to Reactome's Analysis system! I recommend you check it all out!

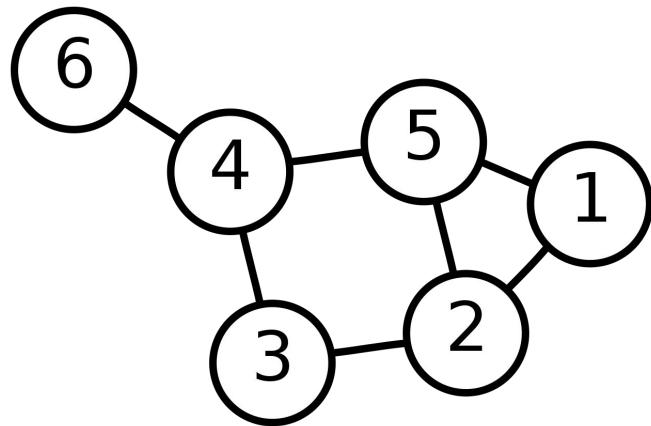


Reactome's Graph Database

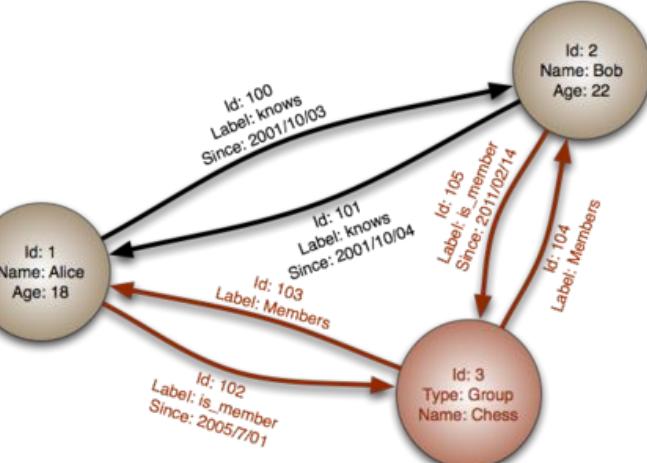
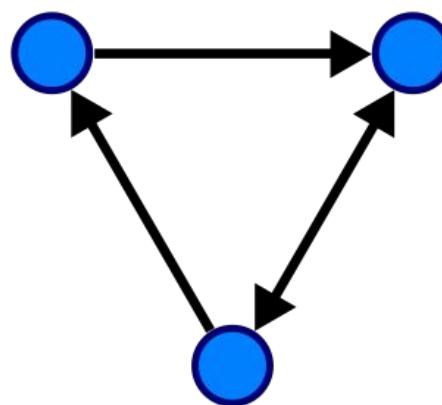


Graph Databases

- Based on relationships as the primary consideration
- Represent data as a graph
 - Nodes
 - Edges (relationships in Neo4j)



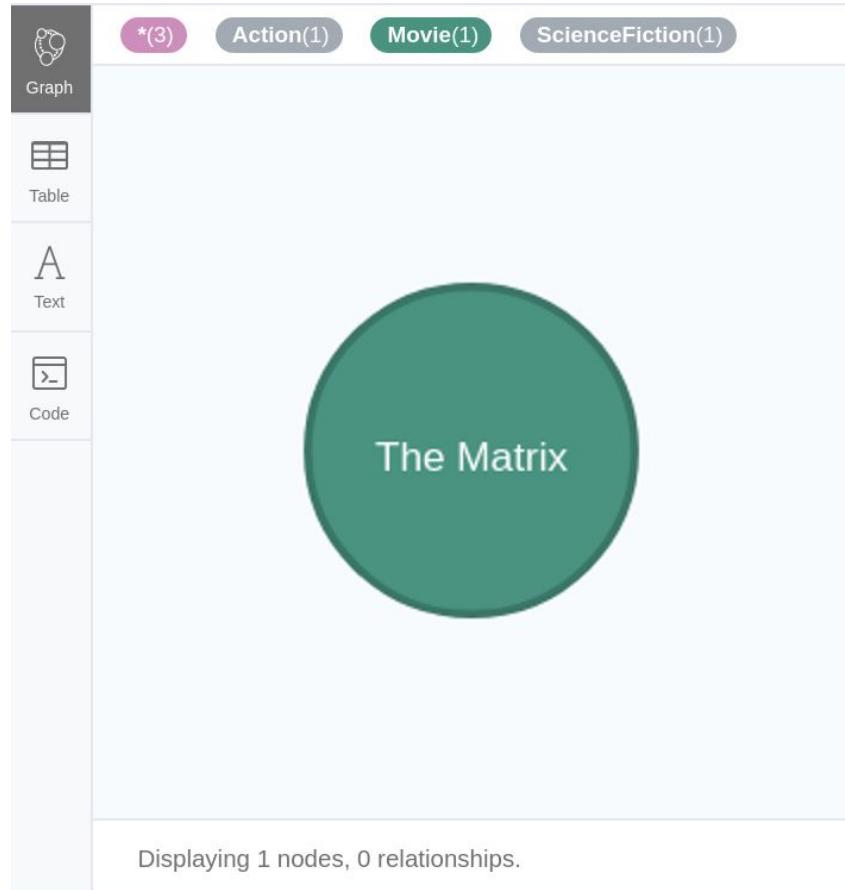
https://en.wikipedia.org/wiki/Graph_theory



https://en.wikipedia.org/wiki/Graph_database

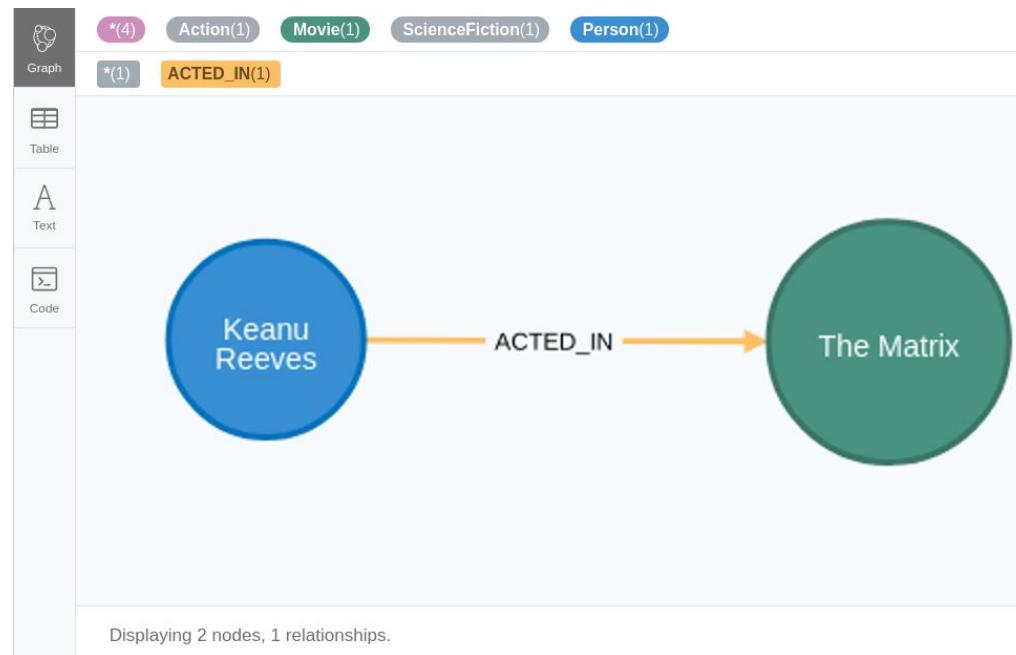
Graph Databases - Nodes

- Entities
 - Usually nouns (person, place, thing, category)
- Zero or more labels to define their type
 - Used to group them
 - e.g. Select all “Movie” nodes



Graph Databases - Relationships

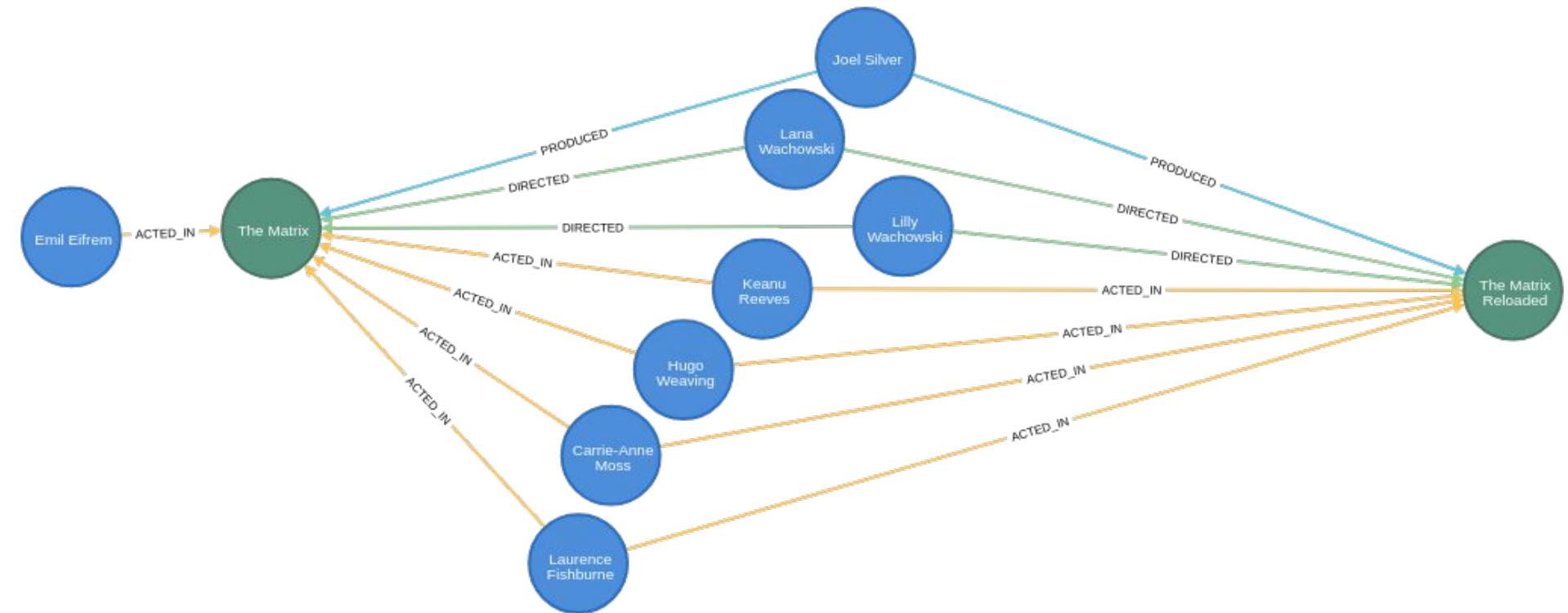
- Connection/interaction between two nodes
 - Usually verbs (e.g. creates, owns, watches, etc.)
- Defined directionally
- Queried directionally or non-directionally



Graph Databases - Properties

- Neo4j is a labeled-property graph
- Data added for description
- Can be added to nodes or relationships
- Examples
 - Movie Node
 - Title
 - Summary
 - “ACTED_IN” Relationship
 - Salary

Graph Databases



Neo4j

- Neo4j: “New or Contemporary” for Java
 - Originally implemented as a Java library
- NoSQL Database
- Native graph storage and processing <https://matrix.fandom.com/wiki/Neo4j>
- ACID Compliant (as are most relational databases)
 - Atomicity
 - transactions executed as all or none
 - Consistency
 - transactions move from one legal database state to another
 - Isolated
 - concurrent transactions give same results as sequential transactions
 - Durability
 - committed transactions guaranteed if there is a system failure



Neo4j - How to use it

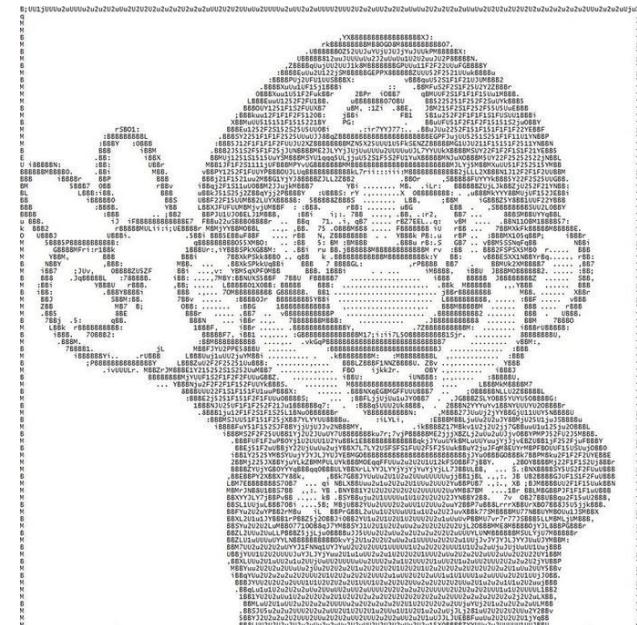
- Cypher Query Language
 - Created by Neo4j
 - Queries nodes and relationships
 - Has been open sourced
 - Based on ASCII art



<https://matrix.fandom.com/wiki/Cypher>



<https://blog.hubspot.com/marketing/shrug-emoji>



<https://knowyourmeme.com/photos/592787-ascii-art>

Neo4j - How to use it

- Cypher Query Language

```
MATCH (p:Person)-[:ACTED_IN]->(m:Movie)  
WHERE m.title = "The Matrix"  
RETURN p, m
```

Neo4j - How to use it

- Software
 - Download and run on Command line

```
jweiser@LL4063-JWEISER:~/neo4j-community-3.5.8$ bin/neo4j start
Active database: graph.db
Directories in use:
home:      /home/jweiser/neo4j-community-3.5.8
config:     /home/jweiser/neo4j-community-3.5.8/conf
logs:       /home/jweiser/neo4j-community-3.5.8/logs
plugins:    /home/jweiser/neo4j-community-3.5.8/plugins
import:     /home/jweiser/neo4j-community-3.5.8/import
data:       /home/jweiser/neo4j-community-3.5.8/data
certificates: /home/jweiser/neo4j-community-3.5.8/certificates
run:        /home/jweiser/neo4j-community-3.5.8/run
Starting Neo4j.
WARNING: Max 1024 open files allowed, minimum of 40000 recommended. See the Neo4j manual.
Started neo4j (pid 9140). It is available at http://localhost:7474/
There may be a short delay until the server is ready.
See /home/jweiser/neo4j-community-3.5.8/logs/neo4j.log for current status.
jweiser@LL4063-JWEISER:~/neo4j-community-3.5.8$
```

<https://neo4j.com/download-center/#community>

Neo4j - How to use it

- Software
 - Docker Container from Command line

```
jweiser@LL4063-JWEISER:~$ docker run --publish=7474:7474 --publish=7687:7687 --volume=$HOME/neo4j/data:/data neo4j
Active database: graph.db
Directories in use:
home:          /var/lib/neo4j
config:        /var/lib/neo4j/conf
logs:          /logs
plugins:       /var/lib/neo4j/plugins
import:        /var/lib/neo4j/import
data:          /var/lib/neo4j/data
certificates: /var/lib/neo4j/certificates
run:          /var/lib/neo4j/run
Starting Neo4j.
2019-07-31 00:03:43.241+0000 WARN Unknown config option: causal_clustering.discovery_listen_address
2019-07-31 00:03:43.244+0000 WARN Unknown config option: causal_clustering.raft_advertised_address
2019-07-31 00:03:43.244+0000 WARN Unknown config option: causal_clustering.raft_listen_address
2019-07-31 00:03:43.244+0000 WARN Unknown config option: ha.host.coordination
2019-07-31 00:03:43.245+0000 WARN Unknown config option: causal_clustering.transaction_advertised_address
2019-07-31 00:03:43.245+0000 WARN Unknown config option: causal_clustering.discovery_advertised_address
2019-07-31 00:03:43.245+0000 WARN Unknown config option: ha.host.data
2019-07-31 00:03:43.245+0000 WARN Unknown config option: causal_clustering.transaction_listen_address
2019-07-31 00:03:43.262+0000 INFO ===== Neo4j 3.5.8 =====
2019-07-31 00:03:43.275+0000 INFO Starting...
2019-07-31 00:03:46.072+0000 INFO Bolt enabled on 0.0.0.0:7687.
2019-07-31 00:03:47.994+0000 INFO Started.
2019-07-31 00:03:49.094+0000 INFO Remote interface available at http://localhost:7474/
```

Neo4j - How to use it

- Software
 - Neo4j Browser or Internet Browser

neo4j@bolt://localhost:7687 - Neo4j Browser

```
1 MATCH (p:Person)-[:ACTED_IN]->(m:Movie)
2 WHERE m.title = "The Matrix"
3 RETURN p, m
```

\$ MATCH (p:Person)-[:ACTED_IN]->(m:Movie) WHERE m.title = "The Matrix" RETURN p, m

Graph

*(6) Person(5) Movie(1)

(5) ACTED_IN(5)

Table

A Text

Code

Displaying 6 nodes, 5 relationships.

Reactome Graph Database

- Reactome Data Model Hierarchy
 - Pathway
 - A *biological process*
 - ReactionlikeEvent
 - A *change/interaction among biological molecules*
 - PhysicalEntity
 - A *representation of a biological molecule (i.e. has a distinct location, modifications, etc.)*
 - ReferenceEntity
 - A *canonical form of a biological molecule in an external database (e.g. UniProt, ChEBI, etc.)*

Reactome Graph Database

- Reference Entity Query

```
MATCH (rgp:ReferenceGeneProduct)  
WHERE rgp.identifier = "P04637"  
RETURN DISTINCT rgp
```



Reactome Graph Database

- Reference Entity Query

```
MATCH (rgp:ReferenceGeneProduct)
WHERE rgp.identifier = "P04637"
RETURN DISTINCT rgp.dbId, rgp.identifier,
rgp.geneName, rgp.sequenceLength
```

rgp.dbId	rgp.identifier	rgp.geneName	rgp.sequenceLength
69487	"P04637"	["TP53", "P53"]	393

Reactome Graph Database

- Reference Entity to Protein Query

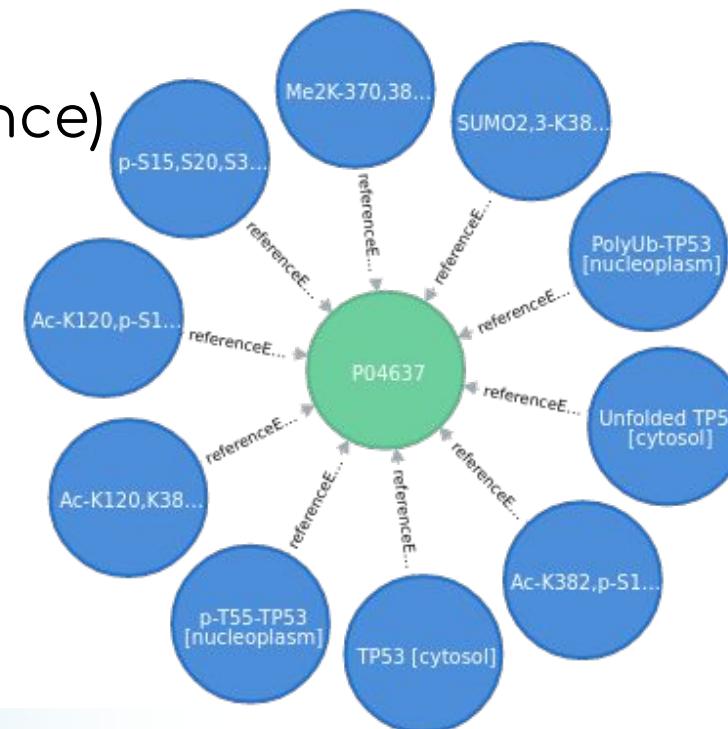
```
MATCH (rgp:ReferenceGeneProduct)
```

```
<-[:referenceEntity]-
```

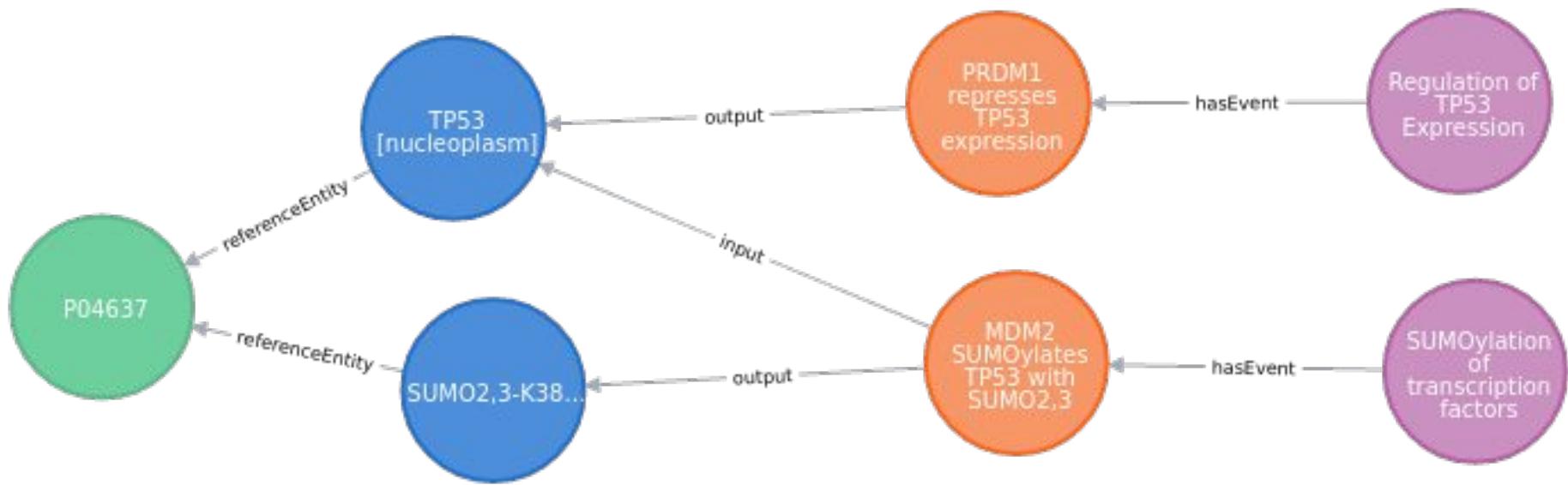
```
(ewas:EntityWithAccessionedSequence)
```

```
WHERE rgp.identifier = "P04637"
```

```
RETURN DISTINCT rgp, ewas
```



Demo!



Learning Neo4j

- Official Neo4j YouTube Channel

The screenshot shows the official Neo4j YouTube channel page. At the top, there's a search bar and a sign-in button. The channel logo features three green circles. It has 21.9K subscribers. A red "SUBSCRIBE" button is visible. Below the header, there are navigation links for HOME, VIDEOS, PLAYLISTS, COMMUNITY, CHANNELS, and ABOUT. A magnifying glass icon is also present.

What is a graph database? (in 10 minutes)
3,000 views • 1 month ago
High-level overview of 'what is Neo4j'.
<https://neo4j.com/developer>

Intro to Graph Databases Series ► PLAY ALL
This Intro to Graph Databases series is a collection of short videos getting folks started on Graph Databases. You'll learn why graph databases are special, how they're an antibiotic for some use cases.

Video Title	Length	Views	Published	Description
What is a graph database? (in 10 minutes)	10:58	3K	1 month ago	High-level overview of 'what is Neo4j'.
Intro to Graph Databases Episode #1 - Evolution of DBs	5:12	113K	4 years ago	Introduction to the evolution of databases and the emergence of graph databases.
Intro to Graph Databases Episode #2 - Properties of...	13:59	115K	4 years ago	Explains the properties of graph databases and how they differ from traditional relational databases.
Intro to Graph Databases Episode #3 - Property Grap...	8:33	67K	4 years ago	Introduces the concept of property graphs and their applications.
Intro to Graph Databases Episode #4 - (RDBMS+SQL)...	17:06	77K	4 years ago	Compares graph databases with traditional RDBMS and SQL.
Intro to Graph Databases Episode #5 - Cypher, the...	19:23	44K	2 years ago	Introduces the Cypher query language used for querying graph databases.

Featured Videos ► PLAY ALL

Video Title	Length	Views	Published	Description
Why Neo4j	1:19	937	1 month ago	Introduction to Neo4j.
What Is Neo4j?	1:44	22K	9 months ago	Introduction to Neo4j.
neo4j Graph Database 4.1	1:02:20	540	1 month ago	Whirlwind tour of Neo4j 4.1.
The World's Leading Companies Use Neo4j	0:38	79A	1 month ago	Companies using Neo4j.
NODES 2019 - Neo4j Online Developer Summit	36:38	7.8K	4 months ago	Summit coverage.
NODES 2019 - Neo4j Online Developer Summit	42:32	747	4 months ago	Summit coverage.

Learning Neo4j

● Neo4j Tutorials

The screenshot shows the Neo4j GraphAcademy Online Training landing page. At the top, there's a navigation bar with links for News, Blog, Support, Company, Contact Us, and a prominent red 'Download Neo4j' button. Below the navigation is a search bar. The main header reads 'GraphAcademy Online Training'. Underneath, a sub-header says 'Kick off with online training' followed by the text 'Get an Intro to Neo4j and then Launch Your Graph Database Application into Production'. There are four course cards: 'Introduction to Neo4j', 'Neo4j Administration', 'Data Science with Neo4j', and 'Applied Graph Algorithms'. Each card features a thumbnail image, the course name, and a brief description.

The screenshot shows a Neo4j Match Clause tutorial page from TutorialsPoint. The top navigation includes links for Library, Videos, Q/A, and eBooks. The main content area has a heading 'Neo4j - Match Clause' and a sidebar with 'Advertisements'. The main content starts with a sub-section 'Get All Nodes Using Match' with the text 'Using the MATCH clause of Neo4j you can retrieve all nodes in the Neo4j database.' It then moves to an 'Example' section with the text 'Before proceeding with the example, create 3 nodes and 2 relationships as shown below.' Below this is a block of Neo4j CQL code. Further down, it says 'Following is the query which returns all the nodes in Neo4j database.' with the code 'MATCH (n) RETURN n'. The sidebar on the right lists other Neo4j tutorial topics.

<https://neo4j.com/graphacademy/online-training/>

<https://www.tutorialspoint.com/neo4j>



Summary

- Reactome Graph Database
 - Implemented with Neo4j
 - Nodes represent the actors in Reactome
 - *E.g. Pathways, Reactions, Physical Entities, Reference Entities*
 - Relationships represent how the actors are connected
 - *E.g. A pathway containing reactions; A reference entity referred to by a physical protein*
 - The database can be downloaded and run in a Neo4j docker container or local installation
 - Data is retrieved by using the declarative, ASCII Art inspired Cypher Query Language

Acknowledgements

- OICR
 - Lincoln Stein
 - Justin Cook
 - Marc Gillespie
 - Robin Haw
 - Bijay Jassal
 - Bruce May
 - Marija Orlic-Milacic
 - Karen Rothfels
 - Solomon Shorser
 - Joel Weiser
 - NYU
 - Peter D'Eustachio
 - Lisa Matthews
 - Veronica Shamovsky
 - EMBL-EBI
 - Henning Hermjakob
 - Chuqiao Gong
 - Yusra Haider
 - Cristoffer Sevilla
 - Thawfeek Varusai
 - Guilherme Viteri
 - OHSU
 - Guanming Wu
 - Liam Beckman
 - Tim Brunson
 - Nasim Sanati
 - Funding
 - US NIH NHGRI P41 HG003751
 - US NIH NCI IDG 1U01CA239069
 - EMBL
 - Websites
 - reactome.org
 - idg.reactome.org
 - spp.reactome.org
 - reactome.ncpsb.org
- Help Desk: help@reactome.org
Twitter: @reactome