Reactome: Tools for analysis of biological pathways

Steve Jupe [1]

- Systems
- Intermediate
- 3 hours

This course introduces the tools available via the Reactome pathway database web interface. For a short introduction to Reactome, please visit the Reactome quick tour [2]. For an introduction to the Reactome web interface including the Pathway Browser, please visit the Introduction to Reactome.

Learning objectives:

- Be able to map your own list of gene, protein, or compound identifiers to Reactome pathways and identifying the pathways that are over-represented
- Visualise pathway expression levels using your own microarray or protein expression dataset
- Extend Reactome pathways with protein-protein or protein-compound interactions from relevant databases or your own interactions data
- Compare curated human pathways with computationally predicted pathways in other species

What is Reactome?

Reactome [3] is a freely available, manually-curated database of human biological reactions and pathways. It provides highly detailed pathway diagrams of molecular processes and descriptions, available via the Reactome website and to download in standard formats such as BioPax [4] and SBML [5].

Reactions
Reactome defines a **reaction** as any molecular event; including binding, phosphorylation, biochemical catalysis, molecular transport and spontaneous molecular events. These reactions can involve any biological molecule, but are typically between proteins and small molecules (chemical compounds).

**Pathways**

A **pathway** is defined as a series of connected reactions, with each individual reaction considered as a single step in the pathway. These are represented as pathways in detailed molecular diagrams. Examples of pathways include immune system signalling, the cell cycle, apoptosis, and the host-parasite interactions of HIV (for example see Figure 1).

**Figure 1** Snapshots of various pathway views and analyses provided by Reactome.

**Curation**

The key aspects of Reactome content are **quality** and **attribution**. Pathways are written by expert biologists and independently peer-reviewed before being included in the database. Reactions always cite **scientific literature** that **experimentally verifies** the described event. These **manually curated** human pathways are used as templates to **computationally infer** [6] equivalent pathways in 17 other species, including most commonly-used model organisms.

To learn more about exploring pathways in Reactome have a look at our companion course **Reactome: Exploring biological pathways** [7].

**Reactome tools**

High-throughput technologies allow researchers to identify gene regulation markers and determine gene expression, protein and metabolite levels for many thousands of molecules in a single experiment, generating huge amounts of data. Interpreting these data requires an understanding of the connections between these molecules (Figure 2). An effective way to do this is to integrate experimental information with well-annotated established pathway data. Reactome fulfils this need by providing free access to detailed pathways combined with analytical tools.
Figure 2 Integrating functional data with curated pathways for effective data analysis.

Reactome includes tools for analysing the pathway context of your data, mapping expression or other quantitative data onto pathways or extending pathways with interactions from external databases.

In this tutorial we will explore the following Reactome tools in more detail:

- Pathway over-representation and pathway topology analysis
- Expression data overlay
- Species comparison
- Molecular interaction overlay

Analysis tools

The Analysis tools submission form

There are two ways to access the Analysis submission form. Either clicking the Analyze Data button on the homepage (Figure 3A) or using the Analysis button on the Pathway Browser header bar, highlighted in red (Figure 3B)
Figure 3 The Analyze Data buttons located on (A) the Reactome [8] homepage and (B) on the Pathway Browser header bar.

**The analysis submission form**

From the Analyze Data submission form you can select the analysis you want to perform, paste in or browse for a file containing your data, or use an example data set (Figure 4).

Figure 4 The analysis submission form.
There are two sections to the submission form - Analyse your data and Species Comparison [9]. The Analyse your data section allows you to submit your own data and performs several different types of analysis, depending on the data format, as described later in the course.

**Analysis submission options**

Selecting the Continue button on the Analysis submission form brings you to an options page (Figure 5):

![Figure 5 Analysis submission options form](image)

Figure 5 Analysis submission options form

Several popular proteomics [10] tools generate results that contain a mixture of human and other species identifiers. When the default option of Project to human is selected, all non-human identifiers in your query are converted to their human orthologous [11] equivalents. As Reactome [8] pathways represent human events, this conversion will maximise the chances of a successful match to a Reactome pathway. However, you may want to uncheck this box if your query represents data from an infected tissue, i.e. it consists of a mixture of human and microbial identifiers. To prevent conversion of non-human identifiers to human, uncheck the box.

For a graphical explanation of this option, click on the plus button underneath the header Project to human

Include Interactors is unchecked by default. With this option selected, Reactome pathways are expanded by including all protein-protein interactors from the IntAct [12] database for every protein represented by the pathway. This can greatly increase the size of the pathway and extend it to include proteins that would not otherwise be represented, which maximises the chances of matching your submitted identifiers to the resulting expanded pathways. However, the risk of this pathway expansion is that the expanded protein set has not been verified by manual curation [13] and review. You should understand that IntAct interactions may be 'false positives', i.e. artefacts resulting from the technology used with no biological relevance.
For a graphical explanation of this option, click on the plus button underneath the header
Include interactors

Analysis data formats

If your data is a single column of identifiers such as UniProt IDs, gene symbols or ChEBI IDs, they are matched
to pathways and over-representation and pathway-topology analyses are run. Over-representation analysis is a
statistical test that determines whether certain Reactome pathways are over-represented (enriched) in the
submitted data. It answers the question ‘Does my list contain more proteins for pathway X than would be expected
by chance?’

Pathway topology analysis considers the connectivity between molecules, which is represented by the pathway
steps (known as reactions in Reactome) in the pathway. It groups all the molecules represented in each reaction as
a pathway ‘unit’. If any of these molecules is represented in your query set, this is considered a match to that
reaction. This may give a better indication of the proportion of the pathway that matches your data, rather than the
number of molecules that are common between your data and the pathway.

If your data has one or more additional columns of numbers, it will be recognised as expression data and an
additional expression overlay process will run. The numbers are used to produce a coloured overlay for Reactome
pathway diagrams. An example of the format required is shown in Figure 6. Note the hash symbol at the beginning
of the header row. Note: Overexpression and pathway topology analyses will run when expression data is
submitted, but the results for these analyses will represent only the pathways that were represented by the
technology used to generate the expression data, not pathway enrichment.
Figure 6 The submission form with a single column of UniProt accessions. Click the 'Analyse' button to begin the analysis.

Identifier mapping

The submission process recognises many types of identifiers. As part of the pre-analysis, they are mapped to Reactome molecules. The ideal identifiers to use are UniProt IDs for proteins, ChEBI IDs for small molecules, and either HGNC gene symbols or Ensembl IDs for DNA/RNA molecules, as these are our main external reference sources for proteins and small molecules.

Many other identifiers are recognised and mapped to appropriate Reactome molecules - a list of these can be found in our user guide.

By default, all non-human identifiers are converted to their human equivalents. If you want to use non-human identifiers to search a computationally-inferred non-human Reactome pathway, uncheck the box. You may also prefer to uncheck this box if your query consists of a mixture of human and microbial identifiers, representing an infection.

Over-representation and pathway topology analyses

Format and results table

If you submit a single-column list of protein or small molecule identifiers, or a mixed list, over-representation and pathway-topology analyses are performed. The results will resemble the example shown in Figure 7.

The names of the columns are explained when you hover over the column name on the Reactome webpage and also in this section of the Reactome User Guide.

Analysis results are shown in the Analysis tab, within the Details panel. All Reactome pathways are shown, in blocks of 20 pathways, ranked by the false discovery rate (FDR) value obtained from over-representation analysis. In addition, the number of molecules matched/total number of molecules and FDR values are added to the right side of pathway names in the Pathway hierarchy panel. Reactions that contain any of the matching molecules are
boxed in orange. When you click on the name of a pathway in the Analysis tab, the Pathway hierarchy will expand to show it and its name will be highlighted in dark blue. If it has sub-pathways, you can use the plus buttons to the left of the pathway name to reveal subpathways and reactions.

By default, all molecules (protein, small molecules, genes, transcripts) are used for over-representation analysis. Subtypes of molecule can be selected from a drop-down list located top-left of the results table. Selecting these subsets will re-display results that consider only the selected molecular subtype.

![Figure 7 Over-representation and pathway topology analysis results table. Hovering over the column names on the Reactome website will give a definition of each column.](image)

Click on a pathway name to open the appropriate Pathway diagram.

**Identifiers not found**

To the right side of the Analysis results details is a button that indicates the number of submitted identifiers that were not successfully matched to molecules in Reactome (red box in Figure 7). Click on the button to produce a list.

If a significant portion of the identifiers could not be matched to pathways then:

- Either the identifiers are not good to use with Reactome and you should replace them with something more suitable;
- Or the identifiers represent molecules that are not found in our pathways, but might be connected to a known or unknown definable process.

**Enhanced high level diagram overlay**
Enhanced high-level diagrams represent analysis results within the label for subpathways. The label background changes from blue to white, a yellow band is used to indicate the proportion of the pathway that is represented in the query dataset. A grey bar above the label indicates the number of pathway entities that is represented in the query dataset, the total number of entities in the pathway, and the FDR corrected probability score (Figure 8).

Pathway diagram overlay

Entities in the diagram are coloured yellow if they were represented in the submitted data set. Complexes, sets and sub-pathway icons are coloured to represent the proportion of the molecules they contain that are represented in the submitted identifier [21] list. In Figure 9 below, PIK3R1 is yellow indicating that it was in the submitted list. The complex GRB2:GAB1:PIK3R1 is part yellow, indicating that some molecules in the complex (PIK3R1) were represented in the submitted dataset while others (GRB2 and GAB1) were not.
Figure 9 Colouring of the pathway diagram following analysis of single-column data.

To examine the details of a complex or set, hover the mouse pointer over it. A small blue triangle will appear on the right side. Select this to open the Contextual Information Panel. The uppermost tab on the left size will be selected, the molecules represented by the complex or set are displayed as a table, with a yellow background if they were in the query dataset (Figure 10):

Figure 10 A participating molecules details table following analysis of single-column data.

If the Pathway diagram is an Overview diagram, consisting of sub-pathway icons, it is coloured to represent the proportion of the subpathway that was represented in the submitted identifier list.
See the section Understanding the Pathway browser [22] in our companion course Reactome: Exploring biological pathways [7] to learn how to interpret and navigate these diagrams.

Results, Mapping and Not found

In the bottom left corner of the results panel there are the Results and Mapping buttons, which can be used to download the entire analysis results set or a simple mapping of the submitted identifiers to Reactome [8] pathways, respectively (Figure 11).

![Result and Mapping buttons](image)

**Figure 11** The Results and Mapping buttons.

In the top right corner of the analysis results is an Identifiers not found button, with a number alongside that indicates the number of submitted identifiers that were not successfully matches to molecules in Reactome. Click the button to produce a list. This gives you the opportunity to correct or replace identifiers that were not recognized. If the identifiers are valid, the list represents entities that are not represented in Reactome. This is an important factor to consider when assessing the results.

Expression data overlay

**Expression data - Submission**

If you submit data in a format that includes columns of numeric values, following a first column of protein, compound or other suitable identifiers such as probe [23] IDs, the analysis tool will interpret your data as expression data. The numeric values are used to colour objects in pathway diagrams. This view was created for microarray [24] data, but any dataset that consists of a list of identifiers with associated numeric values can be used, e.g. quantitative proteomics [10], GWAS [25] scores.

Launch this tool by clicking on the the ‘Analyze Data’ button in the Pathway browser header bar. Either paste your data into the submission form or browse to a saved file (or select an example file).

Figure 12 below shows the correct data format. Note that each row must have an identifier [21] in the first column and you must use a header row.
Figure 12 Example of the format to use with expression data.

After column 1, all other columns must contain numbers, representing expression or other values. You can use Microsoft Excel and tab separated value (TSV) files. When submitted, columns of numbers are considered as separate samples or experimental conditions. The values are used to overlay colour onto pathway diagrams. An Experiment Browser tool allows you to select and view overlays for each submitted data column, which is particularly useful for visualising time-points or a disease progression.

- Click on the ‘Analyse’ button to start the analysis. The results may take some minutes to appear.

Expression data - analysis results

The results page is very similar to that seen following submission of a simple one-column list of identifiers, with extra columns in the Analysis details following column 9. These extra columns represent the submitted expression values.

Clicking on a pathway name launches the Pathway browser and displays the relevant Pathway diagram (Figure 13).
Figure 13 Expression data overlaid onto a pathway diagram.

Things to look for:

- Objects in the Pathway diagram are coloured according to the numeric values submitted. There are several colour schemes, selected using the Settings popout panel on the right side. The colours are based on a scale, which is represented in a bar on the right hand side. Using the default colour scheme, the highest values are represented in bright yellow, through green to pale blue, to dark blue to represent the lowest values. The scale (see figure above) automatically adjusts to fit the range of values in the dataset.
- When zoomed out, objects are coloured to reflect the average of the values submitted for the molecules that the object represents. When zoomed in, complexes and sets will resolve to coloured bands, representing the values of individual molecules.
- To view details of the components of a complex or set, hover the mouse pointer over it and a small blue triangle will appear on the right side. Select this to open the Contextual Information Panel. This has three tabs, Molecules, Pathways and Interactions. Molecules shows the participating molecules as a table. The background colour represents their expression values. Multiple CIPs can be opened and pinned so they remain visible when other entities are selected. The browser remembers CIPs that were pinned on the last visit (for up to 5 diagrams).
- The Experiment Browser toolbar (grey panel at the bottom of the pathway diagram panel) is used to step through the columns of your data, e.g. time-points or disease progression. Move between them by clicking the arrow buttons. The Pathway diagram will re-colour to reflect the new values.

Species comparison

Species Comparison I

Reactome uses the manually curated human pathways to computationally infer their equivalents in 17 other species. This inference process is based on gene orthology, using the Ensembl Compara [27] database.

The Species comparison tool allows you to compare human pathways with these predicted pathways, to see what
is common to both or perhaps missing in human or in the model organism.

- Species comparison is launched using the ‘Analyse Data’ button in the Pathway browser header bar or the Analyze button at the top of the Pathway Browser.
- In the Species comparison section, select one of the species in the dropdown list. Click the ‘Go’ button (Figure 14 below).

![Species selection dropdown](image)

**Figure 14** Species selection dropdown.

### Species Comparison II

The results are ready to view when analysis results appear (or are updated if already present) in the Pathway Overview.

Clicking on a pathway node displays the relevant Pathway diagram (Figure 15).

The colour of reaction objects indicates the result of the comparison:

- Yellow indicates that the protein has an inferred equivalent in the comparison species;
- Blue indicates that no equivalent was identified. This protein may not exist in the comparison species;
- No colour indicates that inference was not possible. This is always the case for small molecules [16], DNA and other objects that have no UniProt [14] entry (or did not at the time the pathway was constructed);
- Objects with bands of colour represent complexes or sets containing more than one molecule. The bands of colour reflect the inference success for the molecules within the complex/set;
- The grey species comparison bar at the bottom of the Pathway diagram (Figure 15) can be used to turn off Species comparison colouring.
Figure 15 A pathway diagram colour-coded using the species comparison tool.

To see the Species comparison results for the components of a complex (boxes with bands of colour), hover your mouse over the object and a small blue triangle will appear on its right side. Select this to open the Contextual information Panel. This reveals a table representing all the proteins involved in the complex/set (Figure 16). Each row in the table represents a single component of the complex, coloured as described above.

Figure 16 Grid representation of a complex for species comparison.

To learn more about the pathway diagrams have a look at Understanding Pathway diagrams [28] in our companion course Reactome: Exploring biological pathways [7].

Molecular Interactions
Molecular interaction (MI) overlay - Part I

The Molecular interaction (MI) overlay tool is built into the Pathway browser. It allows protein-protein or protein-compound interactions to be overlaid onto the Pathway diagram.

The default source of interaction data is IntAct [12] (Static). This provides fast access to a locally-hosted copy of IntAct. Other sources of interaction data (protein-protein and protein-small molecule) can be selected using the Interactor Overlays tab, the middle button of the Settings panel in the middle-right of the Pathway diagram (Figure 17). The databases listed are automatically populated by querying the PSICQUIC [29] Registry, any databases that are offline will be greyed out. It is also possible to use your own interactors data, by selecting the button Add a New Resource.

When a new database is selected it becomes the source for interactions. If interactors are currently being displayed on the pathway diagram, the interactors will automatically update.

You can also change the colour of interactors in the ‘Interactors Colour Profile:’ drop-down menu, within the ‘Colour Palette’ feature of the Settings panel.
Molecular interaction (MI) overlay - part II

When a pathway diagram is displayed, availability of interactors is indicated by a small red circle with a white-lettered number in the top right corner of the entity. The number represents the number of known interactors for the selected interaction source. Clicking on the red circle will display the interactors, to a maximum of 10, as a ring of boxes (protein interactors) or ovals (small molecule interactors) connected by black lines to the selected protein. If the same interactor is connected to more than one entity the box is re-used, i.e. connected to all the selected entities in the diagram. Furthermore, if the interactor is a protein or small molecule entity that pre-exists within the pathway diagram, a black line connects the two entities. Pressing the red circle a second time will remove the interactors from the pathway diagram. If an entity interacts with itself, a loop-link is included. Details of all interactor for every protein in the displayed pathway can be downloaded as a table by selecting the 'cloud' button, on the Interactors tab of the Selection pop-out panel on the right side of the Pathway panel.
The ‘Slider’ feature of the Molecular Overlay toolbar, a light grey bar with dark grey slider at the bottom of the Pathway Diagram (see Figure 18), can be used to set the confidence level cutoff for displaying interactors. Moving the Slider to the right will progressively hide lower confidence interactors. Clicking the ‘X’ in the Molecular Overlay toolbar will remove the toolbar and all the interactors from the viewport. Note: This feature is only available for interaction databases that have confidence scores.

**Figure 18** Molecular interactions overlaid for the proteins PLCG2 and TLR3, showing P85A as an interactor of both. Some interactors with a low confidence score have been hidden using the slider on the Molecular Overlay toolbar.

A number of clickable actions provide additional information about the interactor and its relationship to the protein entity. These include:

- Hovering the mouse pointer over a protein interactor produces a tooltip containing the name and [UniProt](https://www.uniprot.org) [14]

- Clicking on an interactor opens the database entry for the selected protein or chemical in a new window.
- Clicking on the line that connects the interactor to the pathway item opens a new window containing details of the interaction at the source database.
- Right-clicking the protein entity or selecting the blue triangle that appears on the right side when hovering the mouse pointer over it will open the Contextual Information Panel. Select the bottom Interactors tab to display a table of all the interactors for the selected entity. The interactive and scrollable table provides access to additional interactor information. Clicking the 'Interactor' name or accession [30] identifier with connect to the UniProt or the interaction database, respectively. Clicking the 'id' button in the table will toggle the table to display either the interactor name or database identifier. Clicking the pin will lock the interactors table to the viewport. Clicking the 'X' in the table will close the table view.

Contributing to Reactome

The goal of Reactome is to represent the molecular details of all human pathways. In Reactome version 61 (June 2017) 10,846 human proteins are represented, over half of all known human protein coding genes and roughly three quarters of proteins that have an identified function.

All our content is sourced from and reviewed by expert biologists who generously volunteer their time and expertise. We welcome contributions of any size from the biological community. Significant additions can be assigned Digital Object Identifiers and therefore qualify as online publications.

If you are interested in contributing to Reactome, please help [at] reactome.org (contact us) for an informal discussion.

Summary

Reactome pathways

- Reactome is a free, curated pathway database that represents human biological processes as interconnected molecular events or 'reactions'.
- Reactions are the 'steps' in pathways and can be any molecular event in biology.
- Pathways are organized hierarchically and often have sub-pathways. At highest hierarchical level pathways are represented in an Overview. At intermediate and lower levels they are often represented as interactive illustrations and detailed Pathway Diagrams.
- Pathways in 17 other species are computationally inferred from the human pathways

You can learn more in Reactome: Exploring biological pathways [7]

Reactome tools

- Reactome includes tools for analysing the pathway context of your data, mapping expression or other quantitative data onto pathways or extending pathways with interactions from external databases.
- Using Reactome tools you can:
  - Perform pathway over-representation and pathway topology analyses on lists of IDs, for instance
the results of a proteomics study.
- Compare your dataset with Reactome curated pathways, or with pathways extended to include interaction data from IntAct.
- Overlay your gene expression data onto Reactome pathways to assess the likelihood that the pathway is relevant to your study.
- Compare Reactome’s curated human pathways with pathways that are computationally inferred to exist in other species.
- Reactome pathways can be extended using interactive tools that graphically represent interactors from IntAct and other interaction databases, or your own interactions data.

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Learn more

Find out more

To learn more about exploring pathways using the Reactome [8] website have a look at Reactome: exploring biological pathways [7].

You can also find more information in the Reactome user guide [31] and other documentation including the Developers Guide [32] at the Reactome homepage [3].

Recommended courses

Reactome’s experts have contributed to several courses at EMBL-EBI, including:

Networks and pathways [33].

See the EMBL-EBI Training pages [34] for details of all our up-coming courses

Following Reactome

You can follow Reactome on Facebook [35], LinkedIn [36] and Twitter [37].

Help with Reactome

For help contact our help [at] reactome.org (support team).

References

Key Reactome references
Reactome: Tools for analysis of biological pathways
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)


Publications that refer to Reactome
A list [41] is maintained by Reactome staff.

Citing Reactome
If you would like to cite Reactome in your own publications, you can find out how to do so on Reactome's publications [41] page.

Cited references

Contributors

Steve Jupe [1]
EMBL-EBI
Biology Project Coordinator: Reactome (EBI)

Steve Jupe is the Biology Coordinator for the Reactome project in the Molecular Systems team [43] at the European Bioinformatics Institute. His role is primarily the identification and addition of new topics to Reactome, particularly in the areas of immune signalling and haemostasis. Steve has a BSc in Biology. For his PhD he researched the growth-regulating activity of the plant hormones gibberellin and ethylene at UCW Aberystwyth, obtaining his PhD in 1989. This was followed by postdoctoral research positions at the Institute for Arable Crops Research, Rothamsted and later at Long Ashton, where he became a molecular biologist as part of the AFRC plant molecular biology initiative. He worked on plant-pathogen interactions at Royal Holloway, University of London before moving away from the lab in 1996 to join the Bioinformatics group at SmithKline Beecham, later merging with Glaxo to form GlaxoSmithKline, where he worked on numerous drug discovery projects and maintained a database of GPCRs. Steve joined the Reactome group in 2009 as a Curator, extending Reactome's coverage predominantly in the areas of Immune signaling and Hemostasis.
Source URL: https://www.ebi.ac.uk/training/online/course/reactome-tools-analysis-biological-pathways

Links
[1] https://www.ebi.ac.uk/training/online/trainers/sjupe
[2] https://www.ebi.ac.uk/training/online/course/reactome-quick-tour
[6] https://www.ebi.ac.uk/training/online/glossary/computationally-infer
[7] https://www.ebi.ac.uk/training/online/course/reactome-exploring-biological-pathways
[8] https://www.ebi.ac.uk/training/online/glossary/reactome
[10] https://www.ebi.ac.uk/training/online/glossary/proteomics
[11] https://www.ebi.ac.uk/training/online/glossary/orthologous
[12] https://www.ebi.ac.uk/training/online/glossary/intact
[13] https://www.ebi.ac.uk/training/online/glossary/curation
[14] https://www.ebi.ac.uk/training/online/glossary/uniprot
[15] https://www.ebi.ac.uk/training/online/glossary/chebi
[16] https://www.ebi.ac.uk/training/online/glossary/small-molecules
[18] https://www.ebi.ac.uk/training/online/glossary/ensembl
[21] https://www.ebi.ac.uk/training/online/glossary/identifier
[22] https://www.ebi.ac.uk/training/online/course/reactome-exploring-biological-pathways/understanding-pathway-browser-0
[23] https://www.ebi.ac.uk/training/online/glossary/probe
[24] https://www.ebi.ac.uk/training/online/glossary/microarray
[25] https://www.ebi.ac.uk/training/online/glossary/gwas
[26] https://www.ebi.ac.uk/training/online/glossary/tsv
[28] https://www.ebi.ac.uk/training/online/course/reactome-exploring-biological-pathways/understanding-pathway-diagrams-0
[29] https://www.ebi.ac.uk/training/online/glossary/psicquic
[30] https://www.ebi.ac.uk/training/online/glossary/accession
[33] http://www.ebi.ac.uk/training/course/networks-and-pathways-bioinformatics-biologists-0
[34] http://www.ebi.ac.uk/training
[37] http://twitter.com/reactome
[38] http://europepmc.org/abstract/MED/26656494
[40] http://europepmc.org/abstract/MED/21067998
[43] http://www.ebi.ac.uk/proteomics/