Pathway annotation and analysis with Reactome: The solute carrier class of membrane transporters

*Bijay Jassal**

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
**Correspondence to:* Tel: +44 (0)1223 494471; Fax: +44 (0)1223 494468; E-mail: bj1@ebi.ac.uk

Date received (in revised form): 18th November 2011

Abstract

Reactome is an expert-authored, peer-reviewed knowledge base of human reactions and pathways that functions as a data-mining resource and electronic textbook. Its current release covers approximately 23 per cent of the complete human proteome from UniProt. The pathway browser, search and data-mining tools facilitate searching and visualising pathway data and the analysis of user-supplied high-throughput datasets. A catalogue of all the solute-carrier (SLC) class of transporters which have known ligands has been annotated in Reactome. Reactome provides a detailed and interactive view of this set of transport reactions. Using the example of the SLC class of transporters, we show how they can be overlaid with protein–protein interaction, protein–drug interaction and gene expression data and compared with equivalent pathways in other species, to facilitate over-representation, expression and other pathway analyses.

Keywords: *solute transport, protein–ligand interaction, protein–protein interaction, pathway analysis, gene expression analysis*

Reactome

Reactome (http://www.reactome.org) is a freely available and open-source database of biological pathways.[1,2] Expert scientists curate information into Reactome, which is then peer-reviewed to obtain a consensus representation of the process or pathway. The data are extensively cross-linked to major protein and nucleotide sequence databases, as well as to the Gene Ontology and PubMed databases. A new website for Reactome was recently released. This includes new functionality for interacting with curated pathways and analysing them with linked or user-supplied datasets. Tools in this new version of Reactome allow users to overlay interaction or expression data, further enriching the pathway information. The results of these analyses can then be downloaded in a range of formats.

Curated SLC-mediated transmembrane transport pathway in Reactome

We have recently completed a catalogue of the solute carrier (SLC) class of transmembrane transporters. These proteins are well conserved in all eukaryotes, as well as most prokaryotes, and play a gate-keeping role for cells and organelles, controlling the uptake and efflux of many types of substrates, such as sugars, inorganic cations and anions, organic anions and carboxylates, amino acids and oligopeptides, fatty acids and lipids, and neurotransmitters and vitamins. The SLC superfamily comprises 55 gene families, with 362 putatively functional protein-coding genes reported.[3,4] Of these, 231 with the criterion that the transporter has a substrate which it transports across the membrane have been catalogued in Reactome. The remainder are orphan proteins, with no characterised substrates at this time. This pathway will be used as the basis for describing the use of Reactome's analysis tools.

Method

Reactome's website can be viewed on PCs, Macs or Linux computers with later versions of the Internet Explorer (IE), Firefox or Safari browsers. Reactome's whole content can be downloaded as a mysql datadump (http://www.reactome.org/download/index.html). Other Reactome datasets and code that can be downloaded can be found here too.

Briefly, a query was constructed in the Universal Protein Resource (UniProt) (http://www.uniprot.org/). UniProt is a comprehensive resource for protein sequence and annotation data.[5] Reactome uses UniProt identifiers as the primary reference for proteins used in their database. The query searched for manually annotated and reviewed human SLC transporters.

gene: SLC* AND organism:human AND reviewed:yes

This query returned 367 results, the 362 proteins reviewed by He *et al.*[4] and five newly characterised ones. These combined results were used as the basis of the information entered into Reactome. Transporters for which substrates were experimentally identified were catalogued in Reactome using an in-house graphical user interface (GUI) called the Curator Tool, an interface which allows curators to structure data around Reactome's data model and commit to a central repository.[1] The base unit of Reactome is the reaction. The basic set of attributes of a reaction that are captured are details of the input and output molecule(s), the modulating protein(s), compartments for these entities, supporting literature reference(s), a textual summary describing this reaction and the species (eg human).

Visualisation

A new feature of Reactome is the pathway diagram (Figure 1). Here, reactions in a pathway are represented as connected, interactive objects. Each reaction displays the entities (proteins or other molecules) that make up the inputs and outputs of the reaction, together with the modulator protein, where appropriate, joined via a central 'reaction node'. Reactions are also compartmentalised on the diagram, to indicate cellular location. This is an important feature, especially for this pathway, where entities are transported from one side of a membrane to another. The entity and reaction nodes are interactive; clicking on them provides context-sensitive information, such as details of the overall reaction or specific details of the entities involved in the reaction.

Tools in Reactome

Interaction data

Interaction data can be overlaid on pathway diagrams to enrich the data already in Reactome. For example, overlaying protein–protein or protein–drug/chemical interaction data on the solute carrier pathway provides a powerful extension to the existing data for identifying heterodimeric partners. This molecular interaction overlay (MIO) feature may be useful — for example, to the pharmaceutical industry, for identifying novel targets that otherwise may have remained elusive — however, finding potential candidates to modulate a pathway experimentally is more likely. An example of protein–protein interactions sourced from the IntAct database (http://www.ebi.ac.uk/intact/)[6] overlaid across the SLC transporter pathway via the PSIQUIC web service (http://code.google.com/p/psicquic/) (Aranda *et al.*, in preparation) is shown in Figure 2. The user clicks on the 'Analyze, Annotate & Upload' button, located at the top left of the page. By default, IntAct is the database showing in the 'Interaction Database' pull-down list. To see all interactors for the pathway, the user clicks on 'Display table of all interactors for pathway' button, and a table appears, displaying the proteins in the pathway and the interactors found.

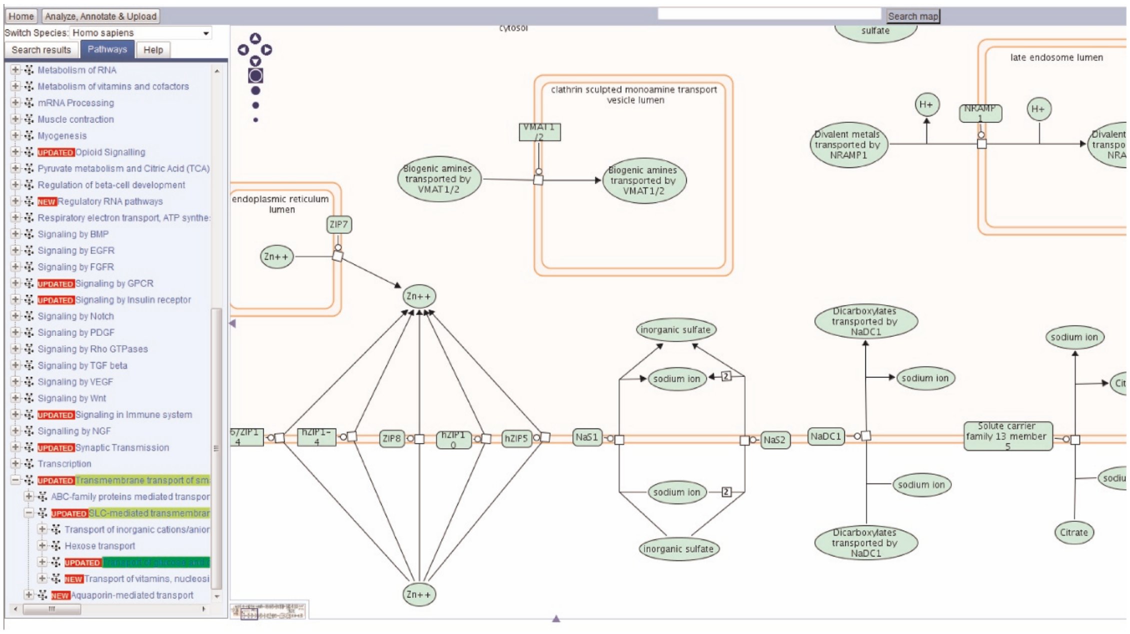
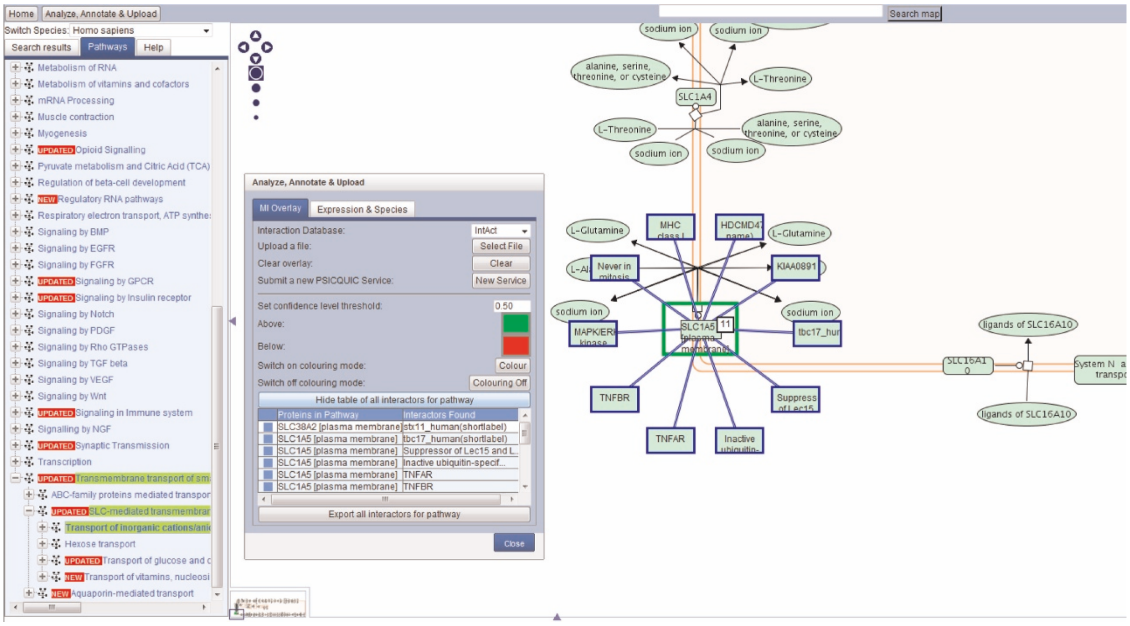

Home
Analyze, Annotate & Upload
Search map
Switch Species: Homo sapiens
Search results
Pathways
Help
Metabolism of RNA
Metabolism of vitamins and cofactors
mRNA Processing
Muscle contraction
Myogenesis
UPDATED Opioid Signalling
Pyruvate metabolism and Citric Acid (TCA)
Regulation of beta-cell development
NEW Regulatory RNA pathways
Respiratory electron transport, ATP synthe
Signaling by BMP
Signaling by EGFR
Signaling by FGFR
UPDATED Signaling by GPCR
UPDATED Signaling by Insulin receptor
Signaling by Notch
Signaling by PDGF
Signaling by Rho GTPases
Signaling by TGF beta
Signaling by VEGF
Signaling by Wnt
UPDATED Signaling in Immune system
Signaling by NGF
UPDATED Synaptic Transmission
Transcription
UPDATED Transmembrane transport of sm
ABC-family proteins mediated transpor
UPDATED SLC-mediated transmembrar
Transport of inorganic cations/anior
Hexose transport
UPDATED
NEW Transport of vitamins, nucleosi
NEW Aquaporin-mediated transport
cytosol
sulfate
late endosome lumen
clathrin sculpted monoamine transport vesicle lumen
VMAT1/2
Biogenic amines transported by VMAT1/2
endoplasmic reticulum lumen
ZIP7
Zn++
H+
NRAMP1
Divalent metals transported by NRAMP1
Dicarboxylates transported by NaDC1
inorganic sulfate
sodium ion
2
NaS1
NaS2
NaDC1
ZIP8
hZIP1-4
hZIP10
hZIP5
Solute carrier family 13 member 5
Citrate

Figure 1. Pathway diagram of a section of the solute carrier (SLC) family of the transmembrane transporter pathway. A hierarchy is visible on the left and the diagram in the main window to the right. 'Mousing' over names shows the name and cellular location. The navigation section to the top left of the diagram allows the user to scroll around the whole pathway, as well as zooming options.

Clicking on the blue box next to a protein in the list takes the user to that protein in the diagram and displays all the interactors for it. For clarity, only ten interactors are displayed. A small white box with a number appears next to the protein, to indicate how many interactors there are. A more



Analyze, Annotate & Upload

MI Overlay | Expression & Species

Interaction Database: IntAct
Upload a file: Select File
Clear overlay: Clear
Submit a new PSICQUIC Service: New Service

Set confidence level threshold: 0.50
Above: 🟩
Below: 🟥
Switch on colouring mode: Colour
Switch off colouring mode: Colouring Off

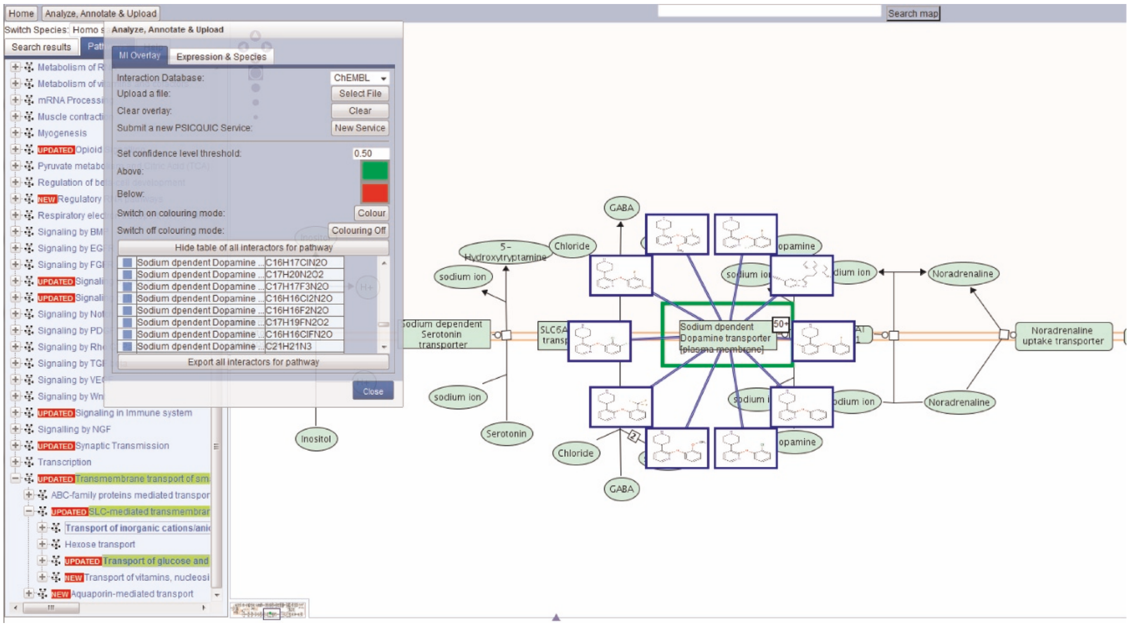
Hide table of all interactors for pathway

	Proteins in Pathway	Interactors Found
■	SLC38A2 [plasma membrane]	stx11_human(shortlabel)
■	SLC1A5 [plasma membrane]	tbc17_human(shortlabel)
■	SLC1A5 [plasma membrane]	Suppressor of Lec15 and L...
■	SLC1A5 [plasma membrane]	Inactive ubiquitin-specif...
■	SLC1A5 [plasma membrane]	TNFAR
■	SLC1A5 [plasma membrane]	TNFBR

Export all interactors for pathway

Close

Figure 2. The protein SLC1A5 in the pathway is showing 11 interactions, overlaid with blue outlines. This interaction information can be exported as a tsv file by clicking the 'export all interactors for pathway' button.



Analyze, Annotate & Upload

MI Overlay | Expression & Species

Interaction Database: ChEMBL

Upload a file: Select File

Clear overlay: Clear

Submit a new PSICQUIC Service: New Service

Set confidence level threshold: 0.50

Above: 🟩

Below: 🟥

Switch on colouring mode: Colour

Switch off colouring mode: Colouring Off

Hide table of all interactors for pathway

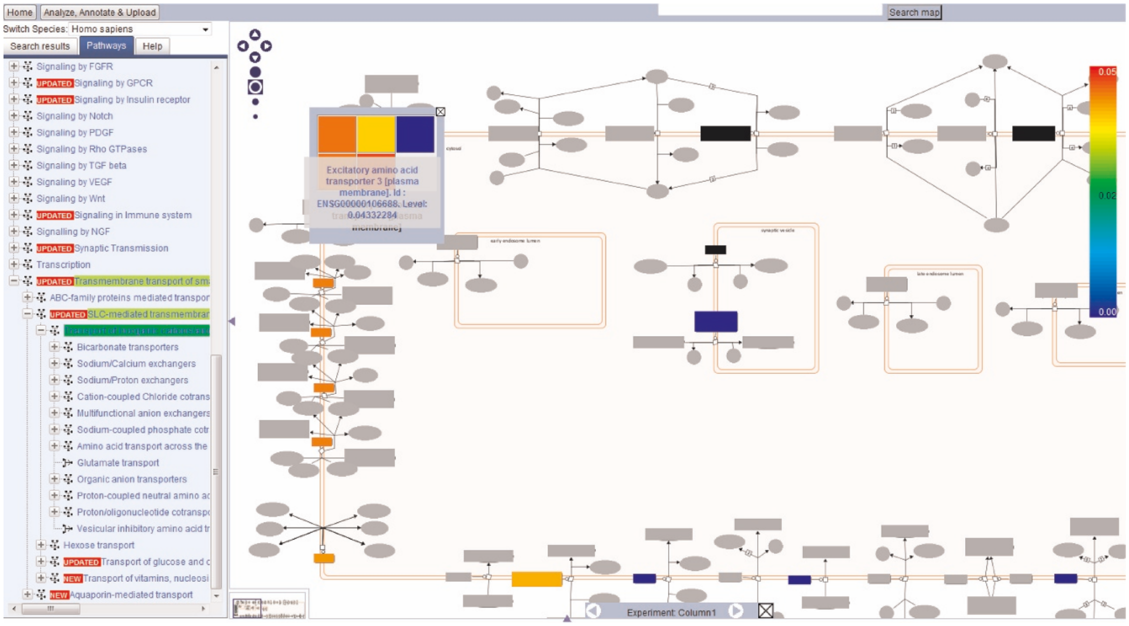
■	Sodium dpendent Dopamine ...	C16H17ClN2O
■	Sodium dpendent Dopamine ...	C17H20N2O2
■	Sodium dpendent Dopamine ...	C17H17F3N2O
■	Sodium dpendent Dopamine ...	C16H16Cl2N2O
■	Sodium dpendent Dopamine ...	C16H16F2N2O
■	Sodium dpendent Dopamine ...	C17H19FN2O2
■	Sodium dpendent Dopamine ...	C16H16ClFN2O
■	Sodium dpendent Dopamine ...	C21H21N3

Export all interactors for pathway

Close

Figure 3. The sodium-dependent dopamine transporter in the pathway interacts with more than 50 chemical compounds, overlaid with blue outlines. The table can be exported as a tsv file by clicking on the 'Export all interactors for pathway' button.

detailed description can be found in Reactome's user guide (http://www.reactome.org/userguide/Usersguide.html#Molecular_Interaction_Overlay). The MIO feature may also help in predicting the function or location of an unknown protein through 'guilt by association', a term used to



Home
Analyze, Annotate & Upload
Search map
Switch Species: Homo sapiens
Search results
Pathways
Help
Signaling by FGFR
UPDATED Signaling by GPCR
UPDATED Signaling by Insulin receptor
Signaling by Notch
Signaling by PDGF
Signaling by Rho GTPases
Signaling by TGF beta
Signaling by VEGF
Signaling by Wnt
UPDATED Signaling in Immune system
Signalling by NGF
UPDATED Synaptic Transmission
Transcription
UPDATED Transmembrane transport of sm
ABC-family proteins mediated transpor
UPDATED SLC-mediated transmembran
Bicarbonate transporters
Sodium/Calcium exchangers
Sodium/Proton exchangers
Cation-coupled Chloride cotrans
Multifunctional anion exchangers
Sodium-coupled phosphate cotr
Amino acid transport across the
Glutamate transport
Organic anion transporters
Proton-coupled neutral amino ac
Proton/oligonucleotide cotransp
Vesicular inhibitory amino acid tr
Hexose transport
UPDATED Transport of glucose and
NEW Transport of vitamins, nucleosi
NEW Aquaporin-mediated transport
0.05
0.02
0.00
Experiment: Column1

Excitatory amino acid transporter 3 [plasma membrane]. Id : ENSG00000106688. Level: 0.04332284

Figure 4. Gene expression dataset overlaid upon an SLC transporter pathway. The data indicate those SLC proteins which are upregulated in the human brain and act as symporters. The colour chart on the right indicates the range of regulation, from blue (downregulation) to red (upregulation). Blacked-out hits indicate complexes. Right-clicking on the complex will show all proteins in that complex, individually coloured. Here, the group of excitatory amino acid transporters show varying expression in the brain, indicated by the colour scheme. Proteins not hit by the dataset are greyed out.

describe annotation transfer based on analysis of interacting partners.

The MIO feature using the PSIQUIC server function can be used to overlay protein–drug/chemical interactors on a pathway diagram (Figure 3). ChEMBL (http://www.ebi.ac.uk/chembldb/) is a database of bioactive drug-like small molecules. Again, the user clicks on the 'Analyze, Annotate & Upload' button, located at the top left of the page. By default, IntAct is the database showing in the 'Interaction Database' pull-down list. Clicking on the pull-down menu displays all interactor databases available. The user can choose ChEMBL and click on the 'Display table of all interactors for pathway' button. A list appears, showing the proteins in the pathway and the interactors found for them. Clicking on the blue box next to a protein takes the user to that protein in the diagram and displays all interactors for it. For clarity, only ten interactors are displayed. A small white box with a number appears next to the protein to indicate how many interactors there are. A more detailed description can be found in Reactome's Userguide (http://www.reactome.org/userguide/Usersguide.html#Molecular_Interaction_Overlay). The example shows drug/chemical interactors of the sodium-dependent dopamine transporter. Neuronal reuptake of dopamine is the primary means of regulating synaptic availability of dopamine. Altering the function of this transporter may be implicated in the pathophysiology or treatment of several neuropsychiatric disorders. The various small molecules displayed here may be useful starting points for novel drug design.

Expression data

Gene expression datasets derived from microarray studies can be overlaid on pathway diagrams. Proteins which have expression values are coloured on the pathway diagram, allowing the user to visualise expression over the whole pathway at a glance. Proteins which are not 'hit' by the dataset are highlighted in grey. Figure 4 shows part of a diagram in which a dataset from the Gene Expression Atlas database (http://www.ebi.ac.uk/gxa/).[7] A query was performed in the Gene Expression Atlas which searched for SLC genes which are upregulated in the brain and act as symporters. The results can be downloaded as a tab separated file, which is a format that Reactome can accept. The columns of identifiers and their expression values were

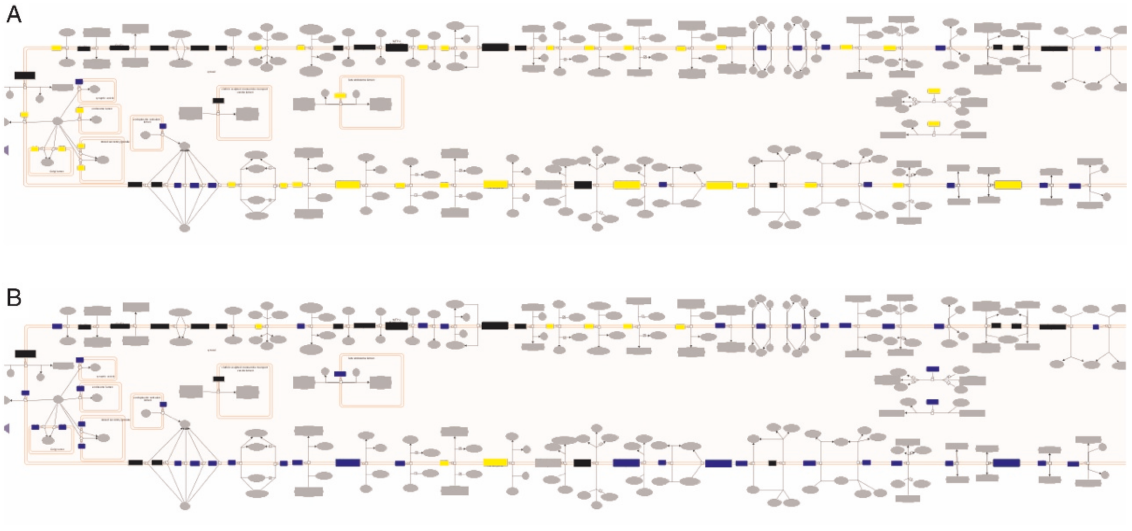


Figure 5. (A) Human pathway compared with that of *Mus musculus*. (B) Human pathway compared with that of *Escherichia coli*. Key: yellow: orthologous protein in other species; blue: protein only known in human; black: complex, right-click to open revealing grid representing components of that complex; grey: either inference not possible or small molecules.

extracted and used in the 'Analyze, Annotate & Upload' toolbar. There, the 'Expression & Species' tab is opened and the file containing the expression values uploaded. Check the box 'Expression painting on' and click the submit button. The pathway is painted according to the expression values from the submitted file. The most upregulated gene from this dataset is *SLC1A1* (also known as excitatory amino acid carrier 1 [*EAAC1*], or excitatory amino acid transporter 3 [*EAAT3*]), which is particularly abundant in the brain.

Species comparison

Reactome uses manually curated human pathways electronically to 'infer' their equivalents in 20 other species. The 'Expression & Species' tab on the control panel allows a user to view these predicted pathways to see what is common to the human pathway or perhaps missing in the model organism. This may be a useful way to determine the extent of conservation of biological processes across species. Figures 5A and 5B show a human pathway compared with that of *Mus musculus* (mouse) and *Escherichia coli*. As one might expect, there is far more conservation between human and *M. musculus* than between human and *E. coli*.

Conclusion

Reactome is a freely available database of pathways. The SLC family of transporters plays a vital role in mediating the movement of essential metals, ions, drugs and many endogenous compounds into and out of the cell and cellular organelles. Information about the SLC family of transporters has been systematically annotated in Reactome and this provides a basis for a number of analyses which can be performed on these data. These analyses include interactions, expression data, over-representation analysis and species comparison. The results of such analyses can be starting points for further investigations using systems biology. The value of the database to users should continue to grow as additional pathways are annotated and new software for data analysis and integration are developed. Work is now under way to improve the visual overview of expression data and provide closer integration with Cytoscape (http://www.cytoscape.org/). Cytoscape is an open-source platform for the visualisation and data integration of biological pathways and networks. Tools are being developed to support additional analysis of interactors, including functional interactors; to pull data from other omics sources, such as expression data or transcription factors; and to support integration with medical data.

Acknowledgments

Development of the Reactome data model and database is a collaborative project and this work benefited greatly from interactions with fellow Reactome team members (http://www.reactome.org/about.html), as well as colleagues at the EBI from GO, ChEBI, ChEMBL, UniProt, ArrayExpress and IntAct. This work was supported by grants from the US National Human Genome Research Institute, NIH (P41 HG003751) and the EU 6th Framework Programme 'ENFIN' (LSHG-CT-2005-518254).

References

1. Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G. *et al.* (2007), 'Reactome: A knowledge base of biologic pathways and processes', *Genome Biol*. Vol. 8, p. R39.
2. Matthews, L., Gopinath, G., Gillespie, M., Caudy, M. *et al.* (2009), 'Reactome knowledgebase of human biological pathways and processes', *Nucleic Acids Res*. Vol. 37, pp. D619–D622.
3. Hediger, M.A., Romero, M.F., Peng, J.B., Rolfs, A. *et al.* (2004), 'The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins Introduction', *Pflugers Arch*. Vol. 447, pp. 465–468.
4. He, L., Vasiliou, K. and Nebert, D.W. (2009), 'Analysis and update of the human solute carrier (SLC) gene superfamily', *Hum. Genomics*. Vol. 3, pp. 195–206.
5. UniProt Consortium (2010), 'The Universal Protein Resource (UniProt) in 2010', *Nucleic Acids Res*. Vol. 38, pp. D142–D148.
6. Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I. *et al.* (2009), 'The IntAct molecular interaction database in 2010', *Nucleic Acids Res*. Vol. 38, pp. D525–D531.
7. Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P. *et al.* (2010), 'Gene expression atlas at the European bioinformatics institute', *Nucleic Acids Res*. Vol. 38, pp. D690–D698.