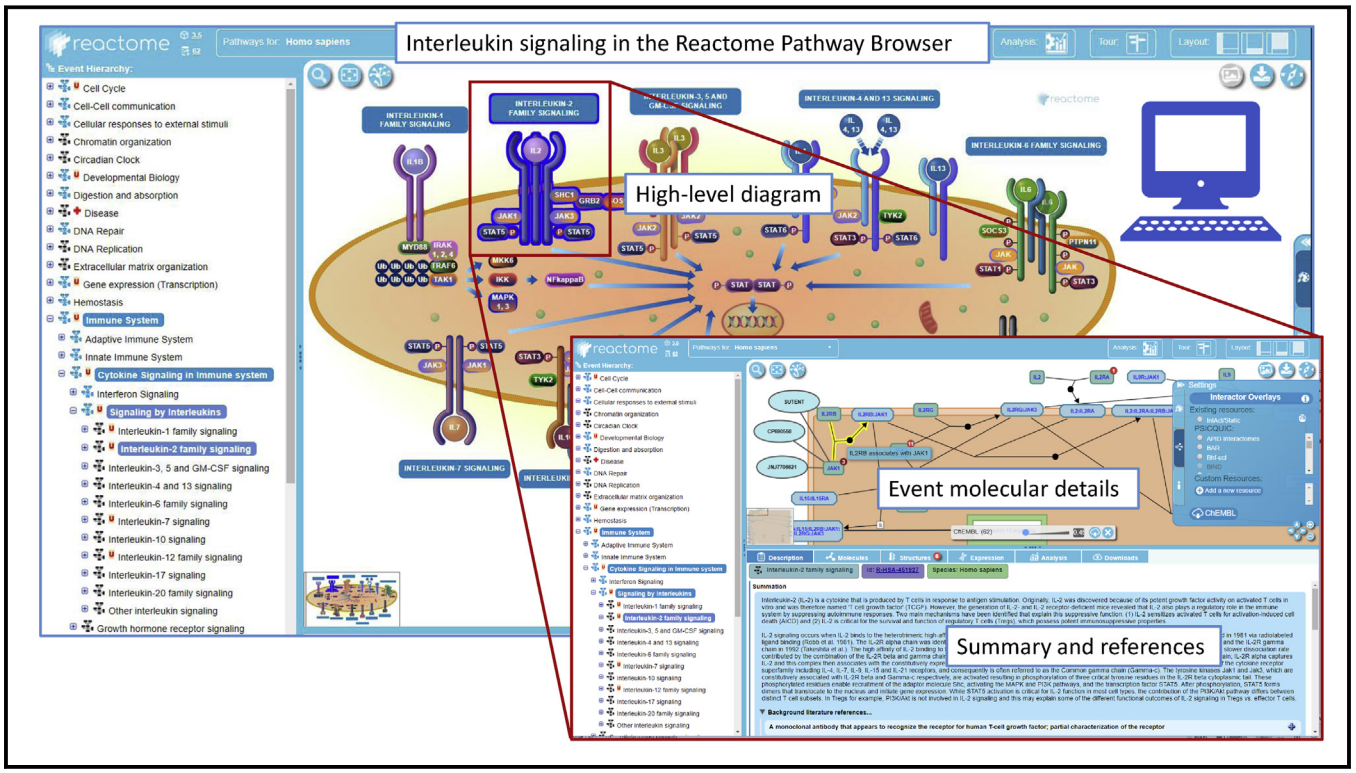
Interleukins and their signaling pathways in the Reactome biological pathway database

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

GRAPHICAL ABSTRACT



Background: There is a wealth of biological pathway information available in the scientific literature, but it is spread across many thousands of publications. Alongside publications that contain definitive experimental discoveries are many others that have been dismissed as spurious, found to be irreproducible, or are contradicted by later results and consequently now considered controversial. Many descriptions and images of pathways are incomplete stylized representations that assume the reader is an expert and familiar with the established details of the process, which are consequently not fully explained. Pathway representations in publications frequently do not represent a complete, detailed, and unambiguous description of the molecules involved; their precise posttranslational state; or a full account of the

From [a]the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton; [b]VHsquared, Cambridge; [c]NYU Langone Medical Center, New York; [d]the Ontario Institute for Cancer Research, Toronto, and the Department of Molecular Genetics, University of Toronto; and [e]the State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences-Beijing (PHOENIX Center), Beijing Institute of Radiation Medicine.

Disclosure of potential conflict of interest: S. Jupe and C. Duenas Roca are employed by EMBL-EBI. V. Shamovsky received National Institutes of Health (NIH) grant BD2K U54GM114833 for this work. L. Stein's institution received a grant from the NIH for this work and grants from the NIH, Canadian Institutes of Health Research, and the Canada Foundation for Innovation for other works and is employed by the Ontario Institute for Cancer Research. P. D'Eustachio's institution received a grant from the NIH for this work. H. Hermjakob's institution received a grant from the NIH for this work and grants from the NIH and European Union for other works and is employed by EMBL-EBI. The rest of the authors declare that they have no relevant conflicts of interest.

Supported by National Institutes of Health (NIH) BD2K grant (U54 GM114833), the National Human Genome Research Institute at the NIH (U41 HG003751), the European Bioinformatics Institute (EMBL-EBI), Open Targets (The target validation platform), and Medicine by Design (University of Toronto). Funding for open access charge was from the NIH (U54 GM114833).

Received for publication September 6, 2017; revised November 27, 2017; accepted for publication December 4, 2017.

Corresponding author: Henning Hermjakob, MSc, EMBL-EBI European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom. E-mail: hhe@ebiac.uk.

0091-6749

https://doi.org/10.1016/j.jaci.2017.12.992

molecular events they undergo while participating in a process. Although this might be sufficient to be interpreted by an expert reader, the lack of detail makes such pathways less useful and difficult to understand for anyone unfamiliar with the area and of limited use as the basis for computational models.
Objective: Reactome was established as a freely accessible knowledge base of human biological pathways. It is manually populated with interconnected molecular events that fully detail the molecular participants linked to published experimental data and background material by using a formal and open data structure that facilitates computational reuse. These data are accessible on a Web site in the form of pathway diagrams that have descriptive summaries and annotations and as downloadable data sets in several formats that can be reused with other computational tools. The entire database and all supporting software can be downloaded and reused under a Creative Commons license.
Methods: Pathways are authored by expert biologists who work with Reactome curators and editorial staff to represent the consensus in the field. Pathways are represented as interactive diagrams that include as much molecular detail as possible and are linked to literature citations that contain supporting experimental details. All newly created events undergo a peer-review process before they are added to the database and made available on the associated Web site. New content is added quarterly.
Results: The 63rd release of Reactome in December 2017 contains 10,996 human proteins participating in 11,426 events in 2,179 pathways. In addition, analytic tools allow data set submission for the identification and visualization of pathway enrichment and representation of expression profiles as an overlay on Reactome pathways. Protein-protein and compound-protein interactions from several sources, including custom user data sets, can be added to extend pathways. Pathway diagrams and analytic result displays can be downloaded as editable images, human-readable reports, and files in several standard formats that are suitable for computational reuse. Reactome content is available programmatically through a REpresentational State Transfer (REST)-based content service and as a Neo4J graph database. Signaling pathways for IL-1 to IL-38 are hierarchically classified within the pathway "signaling by interleukins." The classification used is largely derived from Akdis et al.
Conclusion: The addition to Reactome of a complete set of the known human interleukins, their receptors, and established signaling pathways linked to annotations of relevant aspects of immune function provides a significant computationally accessible resource of information about this important family. This information can be extended easily as new discoveries become accepted as the consensus in the field. A key aim for the future is to increase coverage of gene expression changes induced by interleukin signaling. (J Allergy Clin Immunol 2018;■■■:■■■-■■■.)

Key words: *Interleukins, pathways, signaling, database, Reactome, diagram, illustration*

At the cellular level, life is a network of molecular events that includes processes, such as signal transduction, transport, and metabolism. Reactome represents these processes as pathways, connected events termed reactions that define changes in the state of biological molecules. Reactions include binding; translocation; posttranslational modifications, such as phosphorylation; and biochemical reactions. The molecular events defined by reactions connect to represent ordered transformation networks. Reactome represents human biological events in graphic pathway diagrams supported by descriptions, references, and links to external resources. The result is an extension of the concept underlying classic metabolic maps, diagrams that represent a broad range of physiologic processes underpinned by data that are freely available in a formally defined and consistent structure that is computationally reusable.[1]

By placing human proteins in pathways, Reactome characterizes their molecular functions, providing a resource that is both an archive of biological processes and a tool for discovering functional relationships within user data sets. Reactome pathways include (December 2017, version 63) 10,996 human proteins. Reactome pathway descriptions, images, and underlying data can be downloaded easily in a variety of human-readable and computationally reusable formats. In addition to the Web site, Reactome provides programmatic access tools through content and analysis services.

Reactome is a free database of human biological pathways available through the Reactome Web site (https://reactome.org), where content can be searched, viewed, and downloaded. Pathways are annotated manually by professional curators and peer reviewed by expert biologists. The aim is to represent the consensus view for each area of biology in a reliable and consistent manner such that the underlying data have a structure amenable to use both as a reference knowledge base and as a computational resource for high-throughput querying and reuse. Pathways are represented in as much mechanistic detail as possible, detailing every established molecular event. Events are termed reactions but are not limited to classical biochemical events; binding, dissociation, and translocation between cellular compartments and all forms of posttranslational modification, such as phosphorylation, are also considered reactions.

On the Reactome Web site, pathways are organized hierarchically, following, where possible, the Gene Ontology Biological Process classification. The uppermost levels of the hierarchy represent broad topics in biology, such as the immune system, and serve to group together as subpathways the narrower topics contained within them. At lower hierarchical levels, the pathway topics become more focused and are represented in full molecular detail in interactive pathway diagrams. All pathways are represented as graphics with associated descriptions, literature references, and links to further annotation sources covering areas such as expression, structure, genetics, and disease. Reactome also provides analytic tools that can be used to submit a user data set for overrepresentation analysis, visualize expression data as an overlay on Reactome pathway graphics, or compare Reactome's human pathways with predicted pathways in model organisms.

All of Reactome's content, including graphics and results of analyses, can be downloaded in editable formats for reuse in reports and publications, and the images of molecules and cellular components are available as an icon library for use and extension by the community.

Reactome's coverage of interleukins, their receptors, and signaling pathways has recently been extended to cover IL-1 to IL-38. The classification used is largely derived from Akdis et al.[2] Human interleukins, which were first identified as mediators of signaling between leukocytes (white blood cells),[3] are a family of secreted proteins that function as ligands for receptors that mediate signaling among cells of the immune system.[2] Certain interleukins have chemokine- or interferon-like properties, but their