The Reactome Pathway Knowledgebase: Variants, Drugs, Dark Proteins and Functional Interactions

Robin Haw<sup>1</sup>, Reactome Team<sup>1,2,3,4,5</sup>, Henning Hermjakob<sup>3</sup>, Peter D'Eustachio<sup>4</sup>, Guanming Wu<sup>5</sup> and Lincoln Stein<sup>1,2</sup>

<sup>1</sup>Ontario Institute for Cancer Research, Informatics and Bio-computing, Toronto, Ontario, M5G 0A3, Canada, <sup>2</sup>University of Toronto, Department of Molecular Genetics, Toronto, Ontario, M5S 1A1, Canada, <sup>3</sup>European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, <sup>4</sup>New York University, School of Medicine, Department of Biochemistry, New York, New York, 10016, USA, <sup>5</sup>Oregon Health & Science University, Portland, OR, USA

Reactome is an open access, open source pathway knowledgebase. Its holdings now comprise 12,986 human reactions organized into 2,423 pathways involving 10,923 proteins, 1,869 small molecules, and 369 drugs. 32,150 literature references support these annotations. The roles of variant forms of some proteins, both germline and somatically arising, have been annotated into disease-variant types of reactions and additional reactions that capture the effects of small molecule drugs on these disease processes. To support different visualization and analysis approaches, we implemented several new features through our website, tools, and ReactomeFIViz-Cytoscape app, such as gene set analysis (GSA), an R interface, a Python client, and an intuitive genome-wide results overview based on Voronoi maps. Furthermore, to increase Reactome adoption within the research community, we developed portals and web services for specific user communities. As part of the Illuminating the Druggable (IDG) program, we have undertaken the role to project understudied proteins into the Reactome pathway context, providing useful contextual information for these understudied proteins for experimental biologists to design experiments to understand these proteins' functions. Reactome thus provides dominant pathway- and network-based tools for analyzing multiple data sets and types.