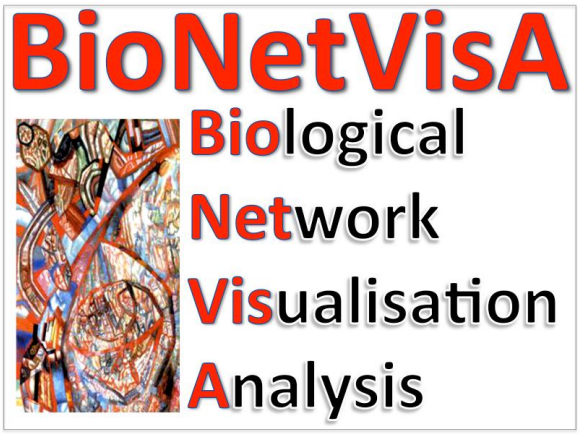

BioNetVisA workshop

From biological network reconstruction to data visualization and analysis in molecular biology and medicine

VIRTUAL

4 September 2020

13.30-16.30 (CEST)

BioNetVisA workshop brings together different actors of network biology from database providers, networks creators, computational biologists, biotech companies involved in data analysis and modeling to experimental biologists, clinicians that use systems biology approaches. The participants are exposed to the different paradigms of network biology and the latest achievements in the field.

Motivation

The goal of BioNetVisA workshop is to build a discussion around various approaches for biological knowledge formalisation, data integration and analysis; compatibility between different methods and biological networks resources available the field; applicability for concrete research and clinical projects depending on scientific question and type of high-throughput data.

The BioNetVisA workshop aims at identifying bottlenecks and proposing short- and long-term objectives for the community as discussing questions about accessibility of available tools for wide range of user in every-day standalone application in biological and clinical labs. In addition, the possibilities for collective efforts by academic researchers, clinicians, biotech companies and future development directions in the field will be discussed during the round table panel.

Audience

The workshop targets computational systems biologists, molecular and cell biologists, clinicians and a wide audience interested in update and discussion around current status of network biology, pathway databases, and related analysis tools, including visualization, statistical analysis and dynamic modelling. No computational background is required to attend the workshop., biotech companies and future development directions in the field will be discussed.

Organizers

Emmanuel Barillot (Institut Curie, France)
Hiroaki Kitano (RIKEN Center for Integrative Medical Sciences, Japan)
Alfonso Valencia (Spanish National Bioinformatics Institute, Madrid, Stain)
Samik Ghosh (Systems Biology Institute, Tokyo, Japan)
Inna Kuperstein (Institut Curie, France)
Luis Cristobal Monraz Gomez (Institut Curie, France)
Robin Haw (Ontario Institute for Cancer Research, Canada)
Andrei Zinovyev (Institut Curie, France)
Minoru Kanehisa (Institute for Chemical Research, Kyoto University, Japan)

Web sites
https://eccb2020.info/ntbw03-bionetvisa-biological-network-reconstruction-data-visualization-and-analysis-in-biology-and-medicine/

https://bionetvisa.github.io/

Webinar link
TBA
Contact

bionetvisa@curie.fr

BioNetVisA 2020 program

Chairs: Inna Kuperstein and Emmanuel Barillot (Institut Curie, Paris, France)

13.30-13.50 Talk 1

COVID-19 Taxila: mining patterns in text in the fight against the pandemic

Sucheendra K. Palaniappan and Samik Ghosh
SBI, Tokyo, Japan
SBX Corporation, Tokyo, Japan

13.50-14.05 Talk 2

Using inter-cellular communication maps to facilitate network medicine

Tamas Korcsmaros
Earlham Institute, Norwich, UK

14.05-14.25 Talk 3

COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms

Marek Ostaszewski
LCSB, Luxembourg

14.25-14.40 Talk 4

AILANI COVID-19 - literature mining and artificial intelligence based question & answering - a scientific assistant for COVID-19 research

Angela Bauch and Dieter Maier
Biomax Informatics AG, Planegg, Germany

14.40-14.55 Talk 5

HENA, Heterogeneous Network-Based Data Set for Alzheimer's Disease

Elena Sugis and Ioannis Xenarios
UNIL, Lausanne, Switzerland

14.55-15.10 Talk 6

Contextualization of molecular networks for human diseases

Luana Licata
University of Rome Tor Vergata, Rome, Italy

15.10-15.20 Talk 7

BioKC: a platform for quality controlled curation and annotation of systems biology models

Carlos Vega
LCSB, Luxembourg

15.20-15.30 Talk 8

RA-map: building a state-of-the-art interactive knowledge base for rheumatoid arthritis

Vidisha Singh
University Evry - Paris Saclay, France

15.30-15.40 Talk 9

Comprehensive map of the Regulated Cell Death Signaling Network: a powerful analytical tool

for studying diseases
Cristobal Monraz Gomez
Institut Curie, Paris, France

15.40-15.50 Talk 10
The dynamics of multilayer network community structure
Davide Cirillo
BSC, Barcelona, Spain

15.50-16.05 Talk 11
WikiPathways: Pathway Models for Network Analysis
Martina Kutmon
Maastricht University, Nederlands

16.05-16.25 Talk 12
Reactome Pathway Knowledgebase: Variants, Dark Proteins and Functional Interactions
Robin Haw
Ontario Institute of Cancer Research, Toronto, Canada

16.25-16.30
Closing remarks

BioNetVisA workshop abstracts

Talk 1

COVID-19 Taxila: mining patterns in text in the fight against the pandemic

Sucheendra K. Palaniappan[1,2], Samik Ghosh[1,2]

The Systems Biology Institute, Tokyo, Japan[1], SBX Corporation, Tokyo, Japan[2]

In the context of COVID-19 pandemic, there is an urgent need to empower scientists, doctors and healthcare researchers at the frontline of the fight with access to the dynamic landscape of scientific literature, clinical trials and drug development efforts happening around the world.
We present the COVID-19-Taxila platform, which collects, curates and organizes information on a daily basis from varied sources, including PUBMED, Arxiv, ClinicalTrial.gov, COVID-19 Open Research Datasets and others. It provides an easy to search interface for contextual navigation and a suite of analytics modules to enable the scientific community in obtaining actionable insights. The platform is powered by Taxila, an end-to-end analysis and intelligence platform which combines state-of-the-art natural language processing and natural language understanding (NLP/NLU) algorithms to analyze text in context.
We present our experiences in the development and deployment of the platform for the scientific community at large, and highlight some insights mined from COVID-19 Taxila, which may provide potential directions in the global fight against this pandemic.

References and useful links
https://covid19.taxila.io/
https://medium.com/@sbijapan

Talk 2

Using inter-cellular communication maps to facilitate network medicine

Lejla Potari-Gul[1], Dezso Modos[1,2], Denes Turei[3], Alberto Valdeolivas[3], Matthew Madgwick[1,2], Julio Saez-Rodriguez[3], Tamas Korcsmaros[1,2]

Earlham Institute, Norwich, UK[1]; Quadram Institute, Norwich, UK[2]; Heidelberg University, Heidelberg, Germany[3]

Ulcerative colitis (UC) is an inflammatory bowel disease (IBD), which affects the colon and the rectum. During the pathogenesis of the disease, both the intracellular and the intercellular interactions are rewired. A recent study published single-cell RNA-seq data of healthy, inflamed and non-inflamed UC colon. We combined this dataset with OmniPath, an integrated resource of curated intracellular and the intercellular interactions we developed earlier. We analyzed the intercellular interactions between 5 cell types (dendritic cell, macrophage, regulatory T cell, myofibroblast, Goblet cell), compared healthy colon and non-inflamed UC, and explored the downstream intracellular signaling processes affected by the intercellular communications between cells.

Results revealed that cells were connected to each other in both conditions, however, the intercellular signaling diverged among cell types and between the same pair in different conditions as well. In healthy condition, cells were tightly communicating to dendritic cells which are the main antigen-presenting cells. In contrast, during UC, cells were more connected strongly to T reg cells, and activate their receptors, hence causing various downstream signalization.

We examined the interaction between myofibroblast and regulatory T cells and found ligands and receptors expressed only in one condition (healthy or non-inflamed UC - called condition-specific receptors). Analysing the downstream signaling from the condition-specific receptors on T cell surface, MAPK and TLR2/6, TLR7/8 pathways were discovered in healthy while TLR4 and TLR3 pathways in diseased Treg cells. We have built up a pipeline to predict not only cell type but also condition-specific effects on intercellular communication, therefore, facilitate the identification of potential drug targets on cell type level.

References and useful links

http://omnipathdb.org
Turei *et al, Nature Methods*, 2016
Smilie *et al, Cell*, 2019

Talk 3

COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms

Marek Ostaszewski[1]

Luxembourg Centre for Systems Biomedicine, Luxembourg[1]

Due to the ongoing COVID19 pandemic there is an urgent need to understand the nature of SARS-CoV-2 virus infection. The development of more efficient diagnosis and treatment depends heavily on a clear understanding of the multistep and multicellular processes implicated in the disease. However, to grasp the entire picture, the patched pieces of information need to be systematically collected, harmonized and combined together in an integrative picture.

The disease maps community initiated the COVID-19 Disease Map project that aims to develop a comprehensive standardized knowledge repository of mechanisms driving the coronavirus SARS-CoV-2 interactions with the human cell. It will enable domain experts, such as clinicians, virologists, and immunologists, to collaborate with data scientists and computational biologists.

Under this initiative, we are developing novel bioinformatic workflow for precise formulation of COVID-19 computational models, and accurate data interpretation that has the potential to suggest drug repositioning. This workflow integrates expert knowledge of molecular mechanisms of SARS-CoV-2 infection and host cell response, databases and data, and computational modeling. This will serve a basis for a computational model for tests and simulation of the response for drugs and predictions of response depending on patient's risk factor and predispositions.

References and useful links

doi:10.17881/covid19-disease-map
https://disease-maps.org
https://fairdomhub.org/projects/190

Talk 4

AILANI COVID-19 - literature mining and artificial intelligence based question answering – a scientific assistant for COVID-19 research

Angela Bauch[1], Martin Wolff[1], Karsten Wenger[1], Mariana Mondragón-Palomino[1], Wenzel Kalus[1], Dieter Maier[1], Sascha Losko[1]

Biomax Informatics AG, Planegg, Germanty[1]

AILANI is a novel and unique scientific assistant that answers natural language questions such as “How to prevent cytokine storms with COVID-19 infection? “with artificial intelligence (AI) assessed relevant responses as well as a graphical summarization of all responses that enables interactive ontology based exploration and refinement.
The system is driven by an enterprise technology that combines semantic modeling, ontologies, linguistics and AI algorithms. The semantic model allows for mapping and integration of knowledge and data, and enables specific processing pipelines for literature mining.
The COVID-19 specific AILANI integrates a variety of structured resources and provides real-time semantic analysis of literature and breaking news with COVID-19 specific information. As part of the COVID-19 Disease Map Collaboration we automatically mine the COVID-19 literature for information on biomedical concepts such as diseases, symptoms, genes, drugs or cell types which are mapped into an overall knowledge network including structured data and ontologies. The user interface is publically available without registration and, in addition to the question answering system, provides keyword, boolean and chemical structure searches. With the optional registration additional features such as application programming interfaces for batch processing become available as well as personlisation options like a search history, favourite searches, documents, chemical structures and subscription to news feeds.

References and useful links

https://ailani.ai
Bauch A, Pellet J, Schleicher T, Yu X, Gelemanovic A, Cristella C, et al. Informing epidemic (research) responses in a timely fashion by knowledge management - a Zika virus use case. bioRxiv. 2020; 2020.04.17.044743. doi:10.1101/2020.04.17.044743

Talk 5

HENA, Heterogeneous Network-Based Data Set for Alzheimer's Disease

Elena Sügis[1], Ioannis Xenarios[2]

Institute of Computer Science, University of Tartu, Tartu Estonia[1],
UNIL Departement Formation et Recherche, University of Lausanne, Switzerland[2]

Alzheimer's disease and other types of dementia are the top cause for disabilities in later life and various types of experiments have been performed to understand the underlying mechanisms of the disease with the aim of coming up with potential drug targets. These experiments have been carried out by scientists working in different domains such as proteomics, molecular biology, clinical diagnostics and genomics. The results of such experiments are stored in the databases designed for collecting data of similar types. However, in order to get a systematic view of the disease from these independent but complementary data sets, it is necessary to combine them. In this study we describe a heterogeneous network-based data set for Alzheimer's disease (HENA) [1]. It is accessible via the Network Data Exchange (NDEx) repository [2] and via the figshare repository [3]. HENA integrates Alzheimer's disease-related data from well-known public data collections, as well as novel experimental and computational data sets generated by the members of the AgedBrainSYSBIO consortium. It combines 64 distinct computational and experimental data sets of six data types originating from nine data sources.

Additionally, we demonstrate the application of state-of-the-art graph convolutional networks [4, 5], i.e. deep learning methods for the analysis of such large heterogeneous graph-structured biological data sets. We expect HENA to allow scientists to explore and analyze their own results in the broader context of Alzheimer's disease research.

References and useful links

[1] Sügis, E., Dauvillier, J., Leontjeva, A., Adler, P., Hindie, V., Moncion, T., Collura, V., Daudin, R., Loe-Mie, Y., Herault, Y., Lambert, J.C., Hermjakob, H., Pupko, T., Rain, J.C., Xenarios, I., Vilo, J., Simonneau, M., Peterson, H., HENA, Heterogeneous network-based data set for Alzheimer's disease. Nature Scientic Data vol 6, no. 1 (2019), https://doi.org/10.1038/s41597-019-0152-0

[2] Sügis, E., HENA ver.2: Heterogeneous network-based data set for Alzheimer's disease (with reduced number of coexpression edges). *The Network Data Exchange* (*NDEx*), https://doi.org/10.18119/N97300 (2019)

[3] Sügis, E. HENA: Heterogeneous network-based data set for Alzheimer's disease. *Figshare*, https://doi.org/10.6084/m9.figshare.c.4469240 (2019).

[4] Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Adv*. *Neur*. *In*. **31**, 1024–1034 (2017).

[5] CSIRO data 61 investigative analytics, Stellar-ml v0.2.0: Machine Learning on graphs, https://github.com/stellargraph (2018).

Talk 6

MINT, SIGNOR 2.0 and their ancillary databases, contextualization of molecular networks for human diseases.

Luana Licata[1], Marta Iannuccelli[1], Prisca Lo Surdo[1], Livia Perfetto[2], Francesca Sacco[1], Luisa Castagnoli[1] and Gianni Cesareni[1]

[1]*Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy.*
[1]*European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.*

MINT, the Molecular INTeractions Database (https://mint.bio.uniroma2.it/) is a public database that stores information about experimentally verified protein-protein interactions mined from the scientific literature.
SIGNOR 2.0, the SIGnaling Network Open Resource (https://signor.uniroma2.it/) is a manually curated database that captures, organizes and displays signaling information into binary causal relationships (protein A up-regulates/down-regulates protein B etc.) between biological entities (proteins, chemicals, protein families, complexes, small molecules, phenotypes and stimuli).

SIGNOR 2.0 contains over 24,000 interactions between more than 6,300 biological entities and features tools such as the extraction of networks from a custom lists of entities, as well as access to a collection of pre-assembled signaling pathways. These relationships can be represented as a generic network by capturing observations made in different experimental contexts as such as disease onset. SIGNOR 2.0 also collects information on causal interactions between Human Coronavirus proteins (HCoV) and the human host involved in cellular pathways that are modulated by infection during HCoV infection. This information can be used to retrieve and organize logical models that allow users to infer the perturbations caused by viral infection on cellular networks and to predict the effect of molecules that are candidates for therapeutic treatments.

Through the analysis of the signaling events whose disruption causes pathological phenotypes, MINT, SIGNOR 2.0 and their ancillary databases (DISNOR, http://signor.uniroma2.it/disease_browser.php and CancerGeneNet, https://signor.uniroma2.it/CancerGeneNet/) can be utilised as a platform for precision medicine and their data can be exploited to build an a priori “disease logic model” useful for diseases analysis.

References and useful links

Calderone et al. Using the MINT Database to Search Protein Interactions. PMID: 31945268
Licata et al. SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. PMID: 31665520
Iannuccelli et al. CancerGeneNet: linking driver genes to cancer hallmarks. PMID: 31598703
Lo Surdo et al. DISNOR: a disease network open resource. PMID: 29036667
MINT: https://mint.bio.uniroma2.it/
SIGNOR 2.0: https://signor.uniroma2.it/
DISNOR: http://signor.uniroma2.it/disease_browser.php
CancerGeneNet: https://signor.uniroma2.it/CancerGeneNet/

Talk 7

BioKC: a platform for quality controlled curation and annotation of systems biology models

Carlos Vega

Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, Luxembourg

Standardisation of biomedical knowledge into systems biology models is essential for the study of the biological function. However, biomedical knowledge curation is a laborious manual process aggravated by the ever increasing growth of biomedical literature. High quality curation currently relies on pathway databases where outsider participation is minimal.

The increasing demand of systems biology knowledge presents new challenges regarding curation, calling for new collaborative functionalities to improve quality control of the review process. These features are missing in the current systems biology environment, whose tools are not well suited for an open community-based model curation workflow. On one hand, diagram editors such as CellDesigner or Newt provide limited annotation features. On the other hand, most popular text annotations tools are not aimed to biomedical text annotation or model curation. Detaching the model curation and annotation tasks from diagram editing improves model iteration and centralizes the annotation of such models with supporting evidence.

In this vain, we present BioKC, a web-based platform for systematic quality-controlled collaborative curation and annotation of biomedical knowledge following the standard data model from Systems Biology Markup Language.

References and useful links
http://biokb.lcsb.uni.lu/

Talk 8

RA-map: building a state-of-the-art interactive knowledge base for rheumatoid arthritis

Vidisha Singh[1], George D. Kalliolias[2,3], Marek Ostaszewski[4], Maëva Veyssiere[1], Eleftherios Pilalis[5], Piotr Gawron[4], Alexander Mazein[4], Eric Bonnet[6], Elisabeth Petit-Teixeira[1] and Anna Niarakis[1]

[1]Laboratoire Européen de Recherche pour la Polyarthrite Rhumatoïde - Genhotel, Univ Evry, Université Paris-Saclay, 91057, Evry, France
[2]Arthritis and Tissue Degeneration Program, Hospital for Special Surgery, New York, NY 10021, USA
[3]Department of Medicine, Weill Cornell Medical College, New York, NY 10021, USA
[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, L-4367 Belvaux, Luxembourg
[5]e-NIOS PC, Kallithea-Athens, Greece
[6]Centre National de Recherche en Génomique Humaine (CNRGH), CEA, 91015, Evry

Rheumatoid arthritis (RA) is a progressive, inflammatory autoimmune disease of unknown aetiology. The complex mechanism of aetiopathogenesis, progress and chronicity of the disease involves genetic, epigenetic and environmental factors. To understand the molecular mechanisms underlying disease phenotypes, one has to place implicated factors in their functional context. However, integration and organization of such data in a systematic manner remains a challenging task. Molecular maps are widely used in biology to provide a useful and intuitive way of depicting a variety of biological processes and disease mechanisms. Recent large-scale collaborative efforts such as the Disease Maps Project demonstrate the utility of such maps as versatile tools to organize and formalize disease-specific knowledge in a comprehensive way, both human and machine-readable.

We present a systematic effort to construct a fully annotated, expert validated, state-of-the-art knowledge base for RA in the form of a molecular map using the software CellDesigner. The RA map illustrates critical molecular and signaling pathways implicated in the disease. Signal transduction is depicted from receptors to the nucleus using the Systems Biology Graphical Notation (SBGN) standard representation. The knowledgebase interfaces with various other databases for content annotation and enrichment and can also serve as a template for omic data visualization. Furthermore, topological analysis of the underlying biological network can help reveal structural hubs while functional enrichment analysis can highlight possible disease comorbidities.

The RA map is available online at ramap.elixir-luxembourg.org

References and useful liks

Funahashi, Akira, et al. "CellDesigner: a modeling tool for biochemical networks." *Proceedings of the 2006 Winter Simulation Conference*. IEEE, 2006.

Gawron, Piotr, et al. "MINERVA—a platform for visualization and curation of molecular interaction networks." *NPJ systems biology and applications* 2.1 (2016): 1-6.

Talk 9

Comprehensive signaling network of regulated cell death: comparison of cell death modes in Alzheimer's neurodegenerative disease and cancer

L. Cristobal Monraz Gomez[1], Jean-Marie Ravel[2,3], Emmanuel Barillot[1], Andrei Zinovyev[1] and Inna Kuperstein[1]

Institut Curie, 26 rue d'Ulm, F-75005 Paris, France, PSL Research University, F-75005 Paris, France, Inserm, U900, F-75005, Paris France, Mines Paris Tech, F-77305 cedex Fontainebleau, France[1], Laboratoire de génétique, Centre Régional Hospitalier Universitaire de Nancy, Vandœuvre-lès-Nancy[2], INSERM UMR 954, Université de Lorraine, Vandœuvre-lès-Nancy[3]

Based on experimental data retrieved from literature, an integrated signalling network of Regulated Cell Death (RCD map) has been constructed. The RCD map is composed of three layers; the "Initiation" layer covers biochemical triggers, input signals and mechanisms that initiate RCD. The "Signalling" layer, recipient of inputs, is the level where the decision about cell death mode is made choosing among Apoptosis, Necroptosis, Ferroptosis and Parthanatos and Pyroptosis. The "Execution" layer depicts the mechanisms activated by one of the five signalling RCD modes and represents the decomposition and degradation mechanisms of the cell. The RCD map is divided into 26 functional modules that can be visualized in the context of the whole map or as individual diagrams. The map contains about 1200 proteins and genes, 2020 biochemical reactions and is based on 600 scientific papers. The map is an open source platform facilitated by the NaviCell web-tool (https://navicell.curie.fr/pages/maps_rcd.html). The RCD network map was applied for interpreting the functional differences in cell death regulation between Alzheimer's disease and non-small cell lung cancer based on gene expression data that allowed emphasizing the molecular mechanisms underlying the inverse comorbidity between the two pathologies. Furthermore, the map was employed for the analysis of genomic and transcriptomic data from ovarian cancer patients that provided RCD map-based signatures of four distinct tumor subtypes and highlighted the difference in regulations of cell death molecular mechanisms.

Talk 10

The dynamics of multilayer network community structure

Davide Cirillo[1,*], Iker Nuñez Carpintero[1], Alfonso Valencia[1,2]

Barcelona Supercomputing Center (BSC), C/ Jordi Girona 29, 08034, Barcelona, Spain [1]
ICREA, Pg. Lluís Companys 23, 08010, Barcelona, Spain [2]
* presenting author

By accounting for the complex and diverse nature and scales of clinical and molecular data, biomedical multilayer networks offer a range of research challenges that still require substantial investigation, while providing the means to achieve a comprehensive view of human diseases. Biomedical multilayer networks have proven striking analytical advantages for heterogeneous data integration [1], especially the effective detection of communities of genes to infer functional associations and drug targets based on multiple evidence. Nevertheless, community structure determination in networks is an open problem to such an extent that the preferred formulations of network communities are often domain-specific [2]. We implement several approaches to study the community structure of biological multilayer network, with special emphasis on the detection of stable partitions at different scales of modularity resolution. Modules of genes that are consistently found in such partitions can be deemed functionally related as supported by the multiple association evidence of the multilayer network. We examine this approach to the study of multilayer community structure of multi-omics data in rare diseases, namely Congenital Myasthenic Syndromes [3] and medulloblastoma [4].

References and useful liks

1. Halu A, De Domenico M, Arenas A, Sharma A. The multiplex network of human diseases. NPJ Syst Biol Appl. 2019;5: 15.
2. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. J Complex Netw. 2014;2: 203–271.
3. Lochmüller, Hanns, Dorota M. Badowska, Rachel Thompson, Nine V. Knoers, Annemieke Aartsma-Rus, Ivo Gut, Libby Wood, et al. 2018. “RD-Connect, NeurOmics and EURenOmics: Collaborative European Initiative for Rare Diseases.” European Journal of Human Genetics: EJHG 26 (6): 778–85.
4. Forget A, Martignetti L, Puget S, Calzone L, Brabetz S, Picard D, et al. Aberrant ERBB4-SRC Signaling as a Hallmark of Group 4 Medulloblastoma Revealed by Integrative Phosphoproteomic Profiling. Cancer Cell. 2018;34: 379–395.e7.

Talk 11

WikiPathways: Pathway Models for Network Analysis

Martina Kutmon[1,2], Anders Riutta[3], Denise Slenter[1], Egon Willighagen[1], Kristina Hanspers[3], Chris T Evelo[1,2], Alexander R Pico[3]

Department of Bioinformatics – BiGCaT, NUTRIM, Maastricht University, Maastricht, the Netherlands[1], Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, the Netherlands[2], Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA[3]

WikiPathways (www.wikipathways.org) is a community curated pathway database that enables researchers to capture rich, intuitive models of biological pathways. Importantly, pathway models from WikiPathways are also a valuable source for network analysis and the content is provided in different formats including RDF [1], via dedicated apps like for Cytoscape [2], and on the network data exchange platform, NDEx [3]. This enables simple integration of pathway and interaction data in network analysis as highlighted in recent publications [4-6].

In addition to ongoing curation efforts to grow and maintain the database, we have identified publication figures as a valuable resource. We estimate ~1000 pathway figures are published and indexed by PubMed Central each month [7]. These figures contain novel pathway content not present in the text nor captured in pathway databases. We identified 64,643 pathway figures published over the past 25 years and performed optical character recognition (OCR) to extract over a million gene symbols mapping to 13,464 unique human NCBI Genes [8]. Pathway figure-based gene sets can be used to index and annotate the literature, to perform enrichment analysis, and to prioritize curation of new pathway models for downstream network analysis.

References

https://doi.org/10.1371/journal.pcbi.1004989
https://doi.org/10.12688/f1000research.4254.2
https://dx.doi.org/10.1016%2Fj.cels.2015.10.001
https://doi.org/10.3389/fgene.2019.00059
https://doi.org/10.1167/iovs.61.4.24
https://doi.org/10.1016/j.jsbmb.2019.01.003
https://doi.org/10.1101/379446
https://doi.org/10.1101/2020.05.29.124503

Talk 12

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

Robin Haw[1], Reactome Team[1,2,3,4,5], Henning Hermjakob[3] , Peter D'Eustachio[4,] , Guanming Wu[5] and Lincoln Stein[1,2]

[1]Ontario Institute for Cancer Research, Informatics and Bio-computing, Toronto, Ontario, M5G 0A3, Canada, [2]University of Toronto, Department of Molecular Genetics, Toronto, Ontario, M5S 1A1, Canada, [3]European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, [4]New York University, School of Medicine, Department of Biochemistry, New York, New York, 10016, USA, [5]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.