

Managing biological data in pathways and networks

Citation for published version (APA):

Kutmon, M. (2015). *Managing biological data in pathways and networks*. [Doctoral Thesis, Maastricht University]. Uitgeverij BOXPress. <https://doi.org/10.26481/dis.20150122mk>

Document status and date:

Published: 01/01/2015

DOI:

[10.26481/dis.20150122mk](https://doi.org/10.26481/dis.20150122mk)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

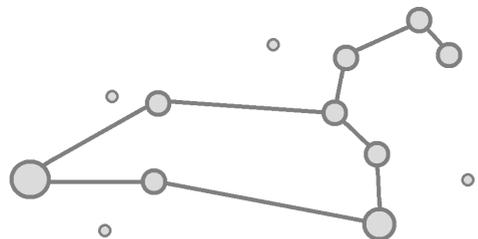
Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Valorization



Introduction

Biological systems are highly complex and bioinformaticians have to deal with vastly increasing amount of data. In this PhD thesis on “Managing biological data in pathways and networks”, I present new network-based approaches to organize, analyze and visualize biological data. The developed methods and tools help other researchers to make sense of their large and complex datasets and put them into a biological context. Because of the huge amount of diverse biological data produced everyday the integration, analysis and interpretation of data is one of the most difficult aspects in biomedical research projects nowadays. The developed approaches will help researchers to better understand and visualize their own data, to integrate it with existing knowledge and therefore to develop better strategies to improve health as well as understand and cure diseases.

In this thesis, I demonstrate the power of pathway and network analysis in different examples, studying the diabetic liver (chapter 7), combining knowledge about regulation of cholesterol biosynthesis and uptake by SREBP (chapter 3) or investigating the effect of starvation in a mouse animal study (chapter 4). One of the developed tools enables the study of drug effects on biological processes or networks (chapter 6). In the diabetic liver study (chapter 7), it is shown how this could be used to find new drug-targets, advantageous drug combinations or possible interferences with the action or efficiency of a drug caused by other drugs.

Open Source Software

Three chapters (2,5,6) in this thesis present new software tools and their applications. To ensure that others can benefit from the tools and methods developed, everything is published under an open source license. Because of the collaborative nature of open source development, effective, scalable and adaptive software can be produced more quickly. We deliberately choose an open source license with very few restrictions to maximize reusability. Although I am leading the development of all three tools, there are many other developers around the world who contribute source code or give input. The methods described in this thesis are already used by many scientists in their research and the numbers of downloads as well as publications confirm the wide adoption of the tools in the research community.

The methods used are highly generic and therefore they can be used in many different research fields. While PathVisio has a more biological focus, Cytoscape can be used for any network visualization or analysis project including social networks, transportation maps or business process modeling. Therefore the apps developed for Cytoscape can reach a much wider audience also outside the scientific community.

As part of the National Resource for Network Biology (NRNB), we also participate in large Open Source events which reach a very broad audience. For example, in the Google Summer of Code in which Google supports Open Source organizations

by providing money for students around the world to work within such an organization, we yearly get between two and five students who are paid to work on our tools during the summer. Although this money does not directly reach the department or university, it definitely results in an improvement and therefore an increased value of the tools we develop and consequently an increased visibility and reputation of the university.

Sharing Knowledge and Data

Besides the tools and methods developed, there is another key result of the research in this PhD thesis, the organization of data. Here we are not only talking about experimental data but also knowledge that has to be structure and visualized. Pathway diagrams, like the SREBP pathway described in chapter 2, are intuitive visual representations of the processes happening in our bodies and in nature around us. In some complex diseases, like Parkinson's Disease, those diagrams might be used by clinicians to explain the disease to their patients and their families. A visual illustration is much more powerful than saying that gene X is not functional. With pathway tools like PathVisio, we can go one step further towards personalized medicine and visualize the patients data on the pathway. This provides the clinician with a tool to give a unique, personalized view of the disease to the specific patient. Although high-quality pathway diagrams for many complex diseases are not yet available and further research is necessary, for some diseases this could possibly already be used in the near future.

In this thesis, we are using publicly available datasets from different online repositories. Therefore, I would like to shortly mention the value of open data and how all parties involved can benefit when data is findable, accessible, interoperable and re-usable (guide to FAIRness of data, www.datafairport.org). Repositories like ArrayExpress or GEO enable researchers to use and combine existing datasets that might lead to new insights and generate new hypotheses. The findings can also be organized and structured for example in pathways which further increases the value of data.

Spreading / Workshops

One important, probably the most important, aspect of valorization is the spreading of the new tools and methods developed. The shown statistics and number of citations already indicate that the tools are downloaded and used by many researchers around the world but we envision to further increase those numbers and make sure that researchers are using the tools in the correct, intended ways. In a first step it is important to present the tools and methods at different conferences, not only bioinformatics focused but also more biological focused ones, to inform scientists about how these tools can be used in their research.

During this 4-year research project, the tools and methods developed have been presented at numerous Bioinformatics and Systems Biology conferences in the Netherlands and abroad. Many of the approaches to analyse biological data with

the tools we developed can become very complex. This leads to a market for tutorials, courses and such. We already developed several workshop and tutorial sessions on ‘How to draw a biological pathway?’, ‘How to use pathway and network analysis to interpret biological data’ or more focused on the different tools like ‘How to use PathVisio/Cytoscape to understand biological processes?’. These workshops can be further extended and improved. Then we can offer them, if possible even as online web seminars, to biomedical research groups, pharmaceutical companies and/or as educational tools to students. This could definitely result in the acquisition of money to further develop the tools and carry out new research projects.

Conclusion

In conclusion, I believe that in scientific research, collaborative approaches allow us to build on each others work and expertise and move forward faster. This is only possible with an open attitude towards sharing data and knowledge. This thesis really follows this principle by creating open source software, using open data to analyse and publishing the results in open access journals. As a final statement, I would also like to mention that having the tools and the expertise to apply them in different research fields can also lead to more traditional valorization in the future for example by participation in research directed towards health improvement or drug development.