

Extending WikiPathways to Improve Educational Capabilities

Jacob Windsor

ORCID: 0000-0002-0249-7757

March 8, 2017

Supervisor

Egon Willighagen,
Department of Bioinformatics - BiGCaT,
Maastricht University,
Maastricht,
The Netherlands,
egon.willighagen@maastrichtuniversity.nl,
ORCID: 0000-0001-7542-0286

Co-supervisor

Anders Riutta,
The Gladstone Institutes,
San Francisco,
U.S.A,
anders.riutta@gladstone.ucsf.edu,
ORCID: 0000-0002-4693-0591

Internal Supervisor

FeiYian Yoong,
Maastricht Science Programme,
Maastricht University,
Maastricht,
The Netherlands,
feiyian.yoong@maastrichtuniversity.nl

Background

Pathway Diagrams

A pathway is a set of biological interactions between genes, proteins, or metabolites that drive a particular cellular function (Wu, Feng, & Stein, 2010). Pathway diagrams are a useful abstraction to describe these biological processes. They enable simple visualisation of experimentally determined pathways (Bohler et al., 2016). Such diagrams have been used in biological

research for many years and have helped biologists uncover the underlying mechanism for numerous phenomena such as respiration (Warburg), the tricarboxylic acid (TCA) cycle (Krebs), and the importance of ATP in energy transfer (Lipmann) (DeBerardinis & Thompson, 2012).

Pathway diagrams play an important role in understanding the progression of complex diseases, such as cancer, which occur as a result of an intricate interplay between metabolic, genetic and environmental factors. By using pathway diagrams in combination with “omics” data, the differences between biological interactions in normal cells and cancer cells can be elucidated. From this, possible targets for drug therapies can be determined. For example, metabolomic data has shown that cancer cells use macropinocytosis (the uptake of extracellular nutrients) to exploit extracellular proteins as an energy source for tumour growth (Commisso et al., 2013). Macropinocytosis has a role in normal cells but is over-stimulated in the presence of oncogenic RAS (Altman, Stine, & Dang, 2016). Cancer cells use this increased activity to internalize extracellular proteins and degrade them into amino acids such as glutamine, a vital energy source for tumours (Commisso et al., 2013; DeBerardinis & Cheng, 2010). Hence, this metabolic pathway presents many targets for pharmacological inhibition (*figure 1*).

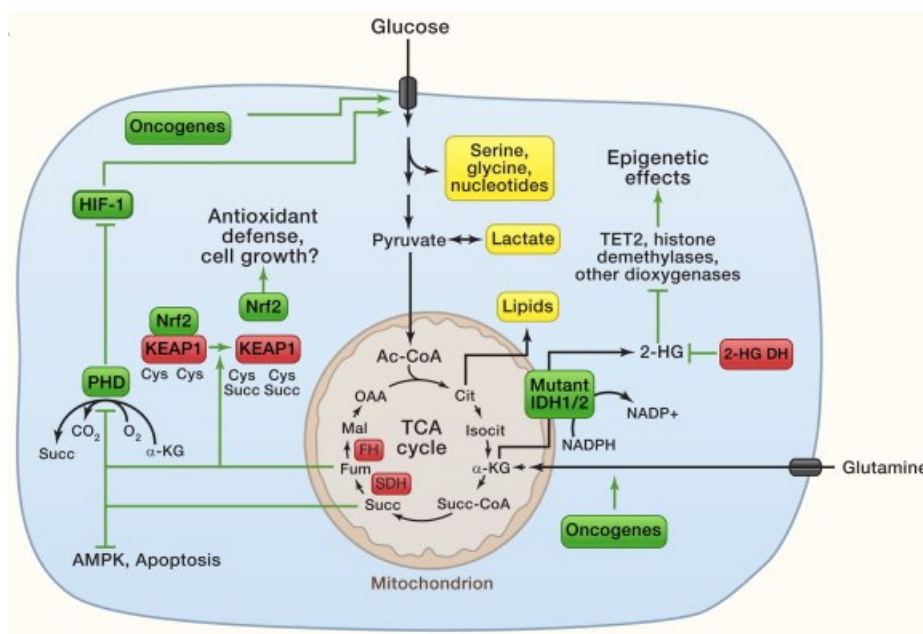


Figure 1. : Cancer cell metabolism, showing the dependence on glutamine for metabolism. The uptake and catabolism of glucose and glutamine is regulated by tumour suppressors (red) and oncogenes (green). Metabolite pools fed by glutamine and glucose are highlighted in yellow. Oncogenes, tumour suppressors and metabolite pools are potential drug targets. (DeBerardinis & Thompson, 2012).

Dealing with an Abundance of Data

Pathway diagrams are increasingly required to explain the complex biological networks that have arisen from the recent growth in biological data. High throughput “omics” technologies produce vast amounts of biological data that must be stored and understood (Bohler et al., 2016). Hence, many pathway databases have been created to store and facilitate the dissemination of this biological data. Currently, there are over 500 pathway databases in Pathguide, the pathway resource list, up from under 200 in 2005 (Bader, Cary, & Sander, 2006; Cary, Bader, & Sander, 2005). The volume of pathways in these pathway databases makes them

far more useful than pathway diagrams alone in the discovery of drug therapies for systemic diseases such as cancer and heart disease. These diseases are propagated by extensively interconnected pathways with high levels of redundancy (Ryall & Tan, 2015). Thus, it is essential to understand the whole system of pathways leading to oncogenesis in order to find effective drug targets. Moreover, tumour cells are highly adaptive and can quickly develop resistance to any one drug by harnessing compensatory pathways (Dry, Yang, & Saez-Rodriguez, 2016). Due to this, it is essential to develop drug combinations that target pathways that contribute to all aspects of tumour biology such as tumour cell metabolism, the tumour micro-environment, and immune response (Ryall & Tan, 2015). In order to build an understanding of these complex networks of pathways, it is first essential to understand the mechanism of the component pathways individually. Pathway diagrams and databases are a significant aid in achieving this.

Despite the advantage of numerous pathway databases, researchers are finding it increasingly difficult to find and reuse the available data (Dumontier et al., 2016). Crucially, many databases do not provide pathway diagrams, with only 61 pathway diagram databases listed on Pathguide to date (Bader et al., 2006). Furthermore, many databases demand expert knowledge to understand the pathways, meaning they are inaccessible to many scientists. This is in large part due to the fact that many pathways, such as in KEGG PATHWAY, are accompanied with insufficient or no descriptions to convey their role or mechanism of action (Kanehisa, Goto, Kawashima, Okuno, & Hattori, 2004). Pathway curation is also a difficult task in many databases, requiring installation of complex software and specialized training. Thus, the barrier for both viewing and curating pathways is simply too high for many biologists. This leads to pathway curation and use being left to a small amount of specialists who have spent a significant amount of time developing a working knowledge of the available tools.

WikiPathways - Pathways for the People

WikiPathways aims to facilitate the contribution to and use of pathway information by the wider biological community. It is a community-driven, open-source, platform for curating and disseminating biological pathway knowledge, based upon the same wiki model as Wikipedia (Pico et al., 2008). It takes advantage of the wealth of knowledge in the scientific community by allowing anyone to create, edit, and view biological pathways rather than leveraging an internal curation team, such as in Reactome (Joshi-Tope et al., 2005). Furthermore, WikiPathways goes beyond conventional static pathway diagrams by utilising PathVisio, an interactive pathway diagram editor, to edit pathways, and a Javascript version (PVJS) to view pathways (*figure 2*) (Kutmon et al., 2015). Significantly, this interactive editor enables easy curation of pathways by any biologist without extensive training. In addition, the integration of BridgeDb (a database identifier mapping framework) into the diagrams allows for easy access to gene, protein, and metabolite annotations (van Iersel et al., 2010; Kelder et al., 2012). This interactivity drastically improves the educational capacity of pathway diagrams since readers can quickly find supplementary information. Additionally, pathways in WikiPathways are represented in the Graphical Pathway Markup Language (GPML), enhancing machine-readability and allowing for use in several downstream pathway analysis tools, potentially leading to novel insights (Kelder et al., 2012).

Limitations of WikiPathways

Knowledge dissemination is improved in a multitude of ways in the wiki model over conventional scientific publishing (Hoffmann, 2008), but knowledge accessibility remains a barrier to WikiPathways since it remains difficult for researchers to understand a pathway that is not within their respective field. Thus, WikiPathway's educational capabilities must be improved

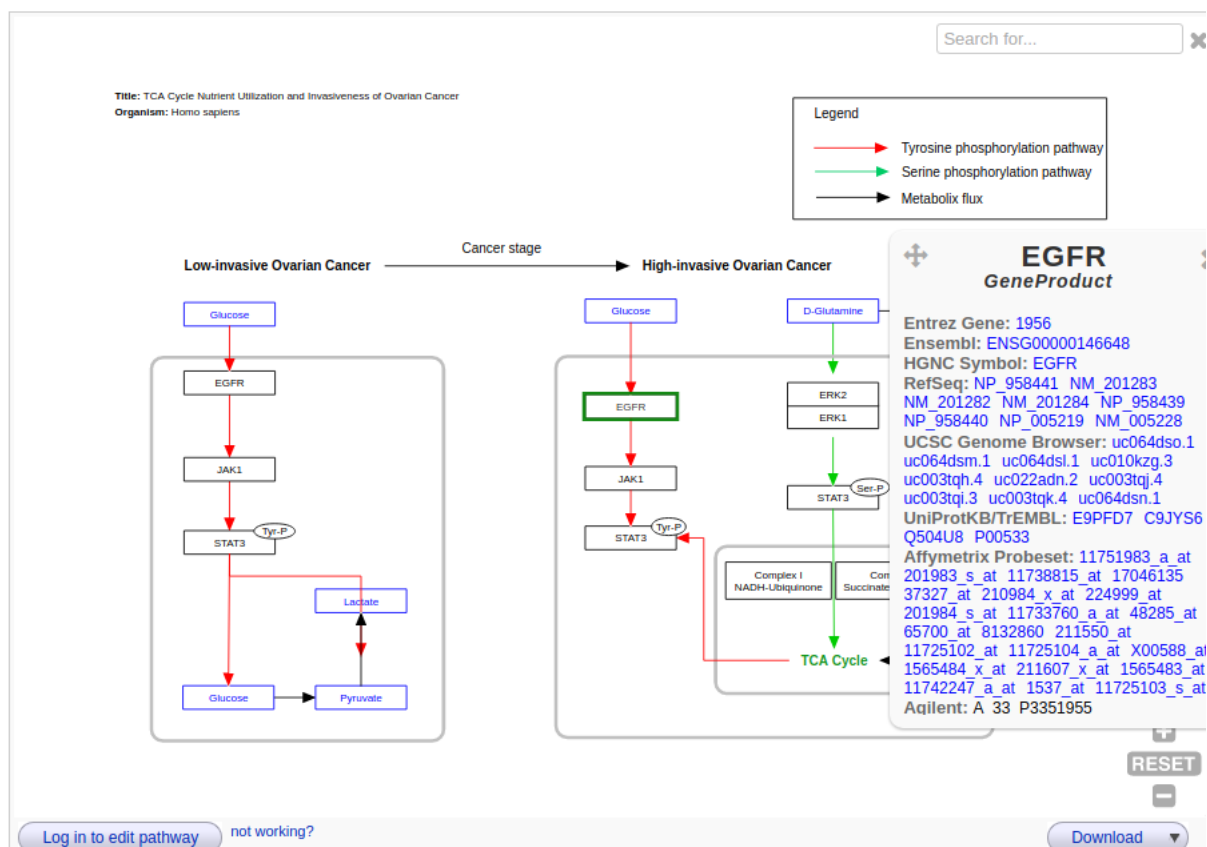


Figure 2. : TCA cycle nutrient utilisation and invasiveness of ovarian cancer in *Homo sapiens* on WikiPathways. Dialogue box for the EGFR enzyme with various annotations shown.

to facilitate an intuitive and simple understanding of all pathways for the broad scientific community. One method for this is to make the available descriptions interactive by including hyperlinks that change the appearance of the diagram produced by PVJS. This method is successfully adopted by Proteopedia, a wiki for structural biology (Hodis et al., 2008). Proteopedia uses “green hyperlinks” within the text that change the three-dimensional visualisation of a biological molecule when clicked. In WikiPathways, the diagram could be altered to zoom in on the described interaction, highlight a particular region, or show annotations for a certain pathway element. Clicking a hyperlink within the description would smoothly transition the accompanying pathway diagram to a different visualisation. These interactive descriptions will improve WikiPathways to convey pathway information in a more accessible manner. Therefore, allowing researchers to build an understanding of biological processes from a field that they are not well versed in.

Research Statement

The educational capabilities of WikiPathways will be improved by the addition of interactive descriptions that manipulate the visualisation of the accompanying diagram.

Proposed methods

The interactive descriptions proposed here will first be tested in an educational cancer metabolism application before implementing them in WikiPathways. Two application flows to implement the interactive descriptions can be followed (figure 4). These flows require initial changes to the application programming interfaces (APIs) of WikiPathways’ component

libraries, PVJS and Kaavio. The latter is a Javascript library used to render diagrams (Riutta, Pico, Bumbu, Hanspers, et al., 2013), and the former is an abstraction on top of Kaavio to render WikiPathways specific pathways (Riutta, Pico, Bumbu, & Summer-Kutmon, 2013). The choice of application flow is dependent on time constraints, with one being a fallback if the other fails to provide results in sufficient time. The methods for both flows are described herein.

A collaboration with Anders Riutta and others at the Gladstone Institutes will take place throughout the duration of this thesis.

Kaavio Manipulation API

The preferred application flow utilises a “manipulation API” that provides Kaavio with methods to alter the diagram visualisation (*figure 3a*). To this end, Kaavio will be extended with this API before usage in any application. The various methods, their parameters, and their functionalities are described in *table 1*.

Although Kaavio is written in JavaScript, extensions will be written in Typescript, a typed superscript of JavaScript that compiles into regular JavaScript (Microsoft, 2016). This is to aid in debugging during development and to ensure robust software.

Once complete, this API will allow developers to easily adapt the visualisation of pathways rendered in Kaavio by calling the proposed methods.

Table 1:: Proposed manipulation API methods and their parameters.

Method		Parameters	
Name	Description	Name	Description
toggleHighlight	Toggle highlighting of a target node (gene, metabolite, etc.)	node_id	The identifier of the target node.
		colour	The desired colour to highlight with.
resetHighlight	Reset all highlights.	-	-
zoom	Zoom in/out to a particular region.)	zoom_perc	The percentage to zoom (0-1).
		origin	Object containing X and Y coordinates.
zoomOn	Zoom into a particular node or group of nodes.	node_id(s)	The identifier(s) of the target node(s).
resetZoom	Reset zoom.	-	-
pan	Pan to a coordinate.	coordinates	Object containing X and Y coordinates.
panTo	Pan to a particular node or group of node.	node_id(s)	The identifier(s) of the target node(s).
resetPan	Reset pan to center.	-	-
toggleAnnotations	Toggle the annotations for a particular node.	node_id	The identifier of the target node.

Direct Diagram Manipulation

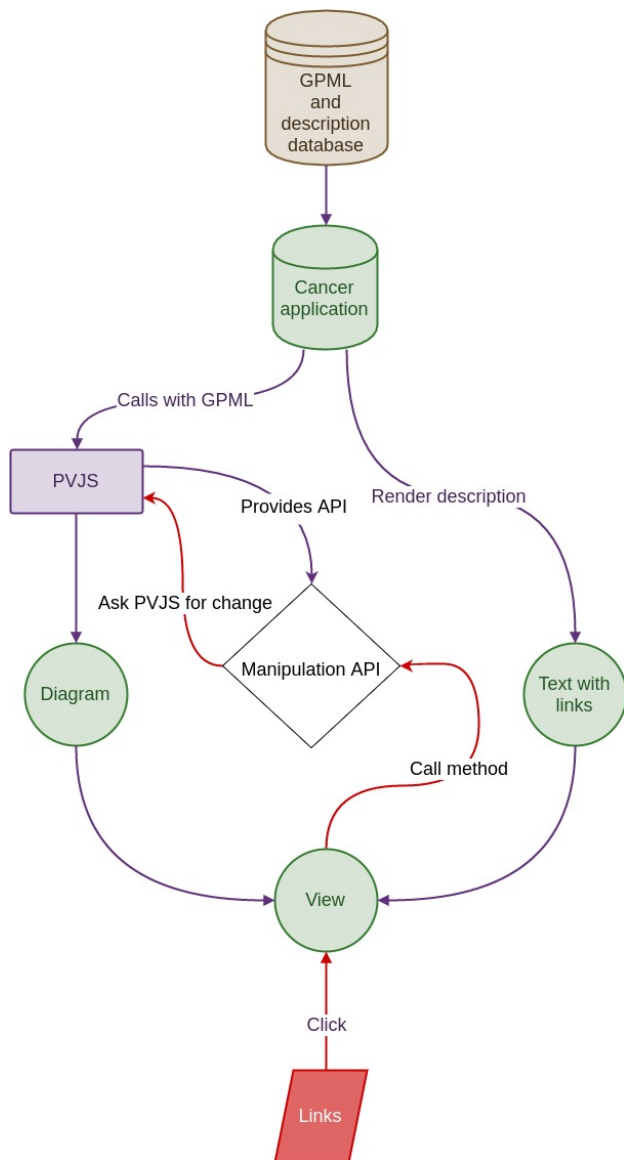
The development of the manipulation API within Kaavio may take a significant amount of time to complete. Thus, an alternative approach is proposed. Rather than altering Kaavio, the output pathway diagram from PVJS can be directly manipulated using cascading style sheets (CSS) (*figure 3b*). This method only allows for the highlighting of data nodes and not other operations such as zooming and panning. If this approach is followed, maintainability is significantly reduced since the single-responsibility principle is not followed (Martin, 2003). Moreover, the difficulty of adding these new features to WikiPathways will be greatly increased.

Educational Cancer Metabolism Application

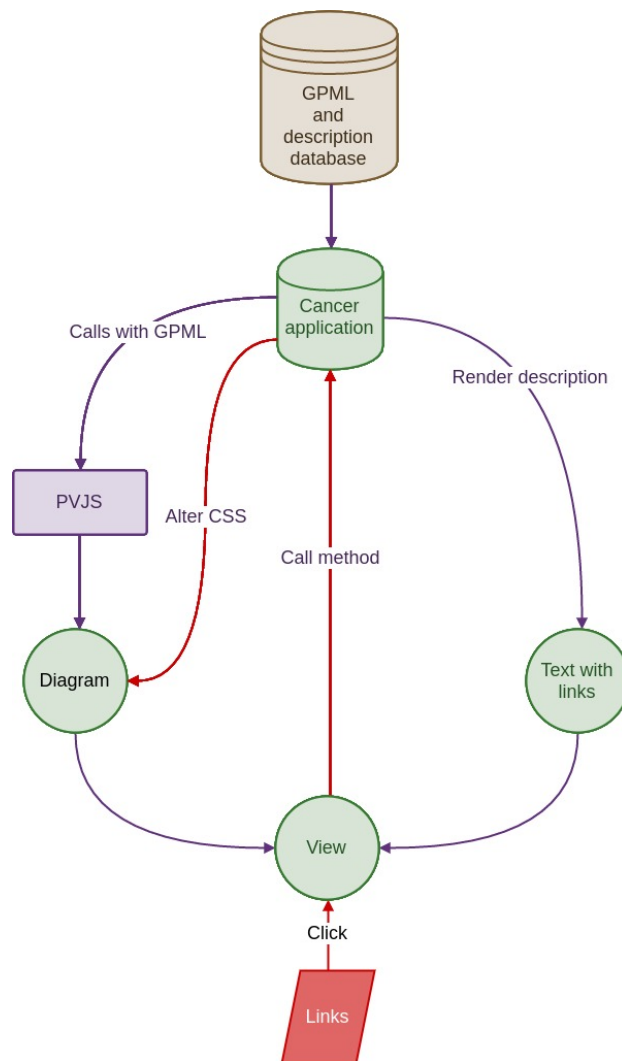
In order to ensure the quality of the proposed additions, an educational cancer metabolism application will be developed. This application will resemble WikiPathways and will detail different metabolic pathways and how they differ in oncogenic cells compared to normal cells. For example, the pathway for nutrient utilisation in the TCA cycle, as shown in *figure 2*, would be displayed alongside text describing the pathway and its differences in normal and cancer cells. PVJS will be used to display the pathways. Additionally, accompanying pathway

descriptions will contain interactive links that alter the displayed diagram.

The application will make use of multiple well-known application development tools to aid rapid development. Angular, a web application framework written in Typescript, will be used to structure and build the application logic (Google, 2017). Bootstrap, a popular user-interface framework, will be used to provide rapid prototyping for the application interface (Twitter, n.d.).



Application flow with Manipulation API. Upon clicking a link in the description, a manipulation API method is called which causes PVJS to alter the diagram visualisation.



Application flow with direct diagram manipulation. Clicking a link in the description calls a method within the main application, which then alters the diagram's styling. PVJS has no knowledge of visualisation changes.

Figure 4. : Flow diagrams for educational cancer metabolism application structures. Note that these flows can also be used in any application, including WikiPathways.

Testing

The finished application will be tested amongst researchers and students with a background in the biomedical field but with varying degrees of experience in cancer metabolism. A set of

questions, as outlined in *table 2*, will be asked to determine the efficacy of the improvements made. The testing will be an iterative process with feedback being taken into account, further improvements made if needed, and then further testing until satisfactory feedback is given. Once this process is complete, it can be concluded that the improvements are successful and ready to be used in WikiPathways itself.

Table 2:: Proposed questions for the cancer metabolism application testing. Two sets of questions will be given, before and after usage of the application.

Before use	After use
Do you have prior experience with WikiPathways? If so, for how long and what have you done (pathway creation, editing, viewing etc.)?	How easy was the application to use? In particular, did you find the interactive links within the description easily?
Do you have any experience with other biological pathway databases? If so, which one(s)?	In general, do you think you have learnt about cancer metabolism via this application?
Ranking from 0-10, how much knowledge do you have of cancer metabolism?	If you've used WikiPathways before, how much of an improvement are these interactive descriptions?
	Do you have any further ideas for improvement?

Refactoring

Aside from enabling interactive descriptions in WikiPathways, time will be spent refactoring the existing PVJS codebase. Currently, PVJS is poorly structured with a weak distinction between Kaavio and PVJS. Thus, an attempt to rectify this will be carried out by re-writing both PVJS and Kaavio in Typescript, using the React JavaScript library to build the user-interface (Facebook, 2017; Microsoft, 2016). This will enhance the maintainability of PVJS and Kaavio and allow for faster development in the future since understandability will be increased for new developers. Note that refactoring will be done only as time allows.

Timeline

The projected timeline for these methods is shown in *table 3*.

Table 3:: Proposed timeline for the project. Dates approximate.

Date	Goal
20th March	Assess if manipulation API can be completed and switch to direct diagram manipulation if not.
17th April	Complete manipulation API or direct diagram manipulation.
8th May	Complete educational cancer metabolism applicaton & begin user testing.
29th May	Finish user testing and perform any final improvements.
Post-thesis	Add new improvements to WikiPathways

Implications

Biological pathway information is only valuable to the broad scientific community when it is part of a larger effort to communicate its role, mechanism, and implications. WikiPathways goes a great way to communicating pathway information to the general scientific community by integrating text-based descriptions and an interactive pathway diagram (Pico et al., 2008). The addition of interactive descriptions proposed here will greatly increase the educational capabilities of WikiPathways, allowing researchers to better understand and visualize their data. The potential consequences of this research project reach beyond the WikiPathways and bioinformatics communities to the general biomedical research community. Interactive descriptions will aid information accessibility so that researchers can more easily understand a pathway from outside of their respective field. This is exemplified by Proteopedia, where the combination of descriptions and interactive 3D renderings of biological molecules aid in communicating complex information (Hodis et al., 2008). Ultimately, these improvements this will aid in various biological fields, particularly in the understanding of complex, systemic biological phenomena such as cancer progression. By enabling researchers to easily understand the parts of a system, new insights into how best to manipulate the system as a whole can be made.

References

- Altman, B. J., Stine, Z. E., & Dang, C. V. (2016). From krebs to clinic: glutamine metabolism to cancer therapy. *Nature Reviews Cancer*, 16(10), 619–634.
- Bader, G. D., Cary, M. P., & Sander, C. (2006). Pathguide: a pathway resource list. *Nucleic acids research*, 34(suppl 1), D504–D506.
- Bohler, A., Wu, G., Kutmon, M., Pradhana, L. A., Coort, S. L., Hanspers, K., . . . Evelo, C. T. (2016). Reactome from a wikipathways perspective. *PLoS Comput Biol*, 12(5), e1004941.
- Cary, M. P., Bader, G. D., & Sander, C. (2005). Pathway information for systems biology. *FEBS letters*, 579(8), 1815–1820.

- Commisso, C., Davidson, S. M., Soydaner-Azeloglu, R. G., Parker, S. J., Kamphorst, J. J., Hackett, S., ... others (2013). Macropinocytosis of protein is an amino acid supply route in ras-transformed cells. *Nature*, 497(7451), 633–637.
- DeBerardinis, R. J., & Cheng, T. (2010). Q’s next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*, 29(3), 313–324.
- DeBerardinis, R. J., & Thompson, C. B. (2012). Cellular metabolism and disease: what do metabolic outliers teach us? *Cell*, 148(6), 1132–1144.
- Dry, J. R., Yang, M., & Saez-Rodriguez, J. (2016). Looking beyond the cancer cell for effective drug combinations. *Genome medicine*, 8(1), 125.
- Dumontier, M., Gray, A. J., Marshall, M. S., Alexiev, V., Ansell, P., Bader, G., ... others (2016). The health care and life sciences community profile for dataset descriptions. *PeerJ*, 4, e2331.
- Facebook. (2017). *A javascript library for building user interfaces - react*. Retrieved from <https://facebook.github.io/react/>
- Google. (2017). *Angular - one framework. mobile desktop*. Retrieved 2017-02-14, from <https://angular.io/>
- Hodis, E., Prilusky, J., Martz, E., Silman, I., Moul, J., & Sussman, J. L. (2008). Proteopedia-a scientific’wiki’bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome biology*, 9(8), R121.
- Hoffmann, R. (2008). A wiki for the life sciences where authorship matters. *Nature genetics*, 40(9), 1047–1051.
- Joshi-Tope, G., Gillespie, M., Vastrik, I., D’Eustachio, P., Schmidt, E., de Bono, B., ... others (2005). Reactome: a knowledgebase of biological pathways. *Nucleic acids research*, 33(suppl 1), D428–D432.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The kegg resource for deciphering the genome. *Nucleic acids research*, 32(suppl 1), D277–D280.
- Kelder, T., van Iersel, M. P., Hanspers, K., Kutmon, M., Conklin, B. R., Evelo, C. T., & Pico, A. R. (2012). Wikipathways: building research communities on biological pathways. *Nucleic acids research*, 40(D1), D1301–D1307.
- Kutmon, M., van Iersel, M. P., Bohler, A., Kelder, T., Nunes, N., Pico, A. R., & Evelo, C. T. (2015). Pathvisio 3: an extendable pathway analysis toolbox. *PLoS Comput Biol*, 11(2), e1004085.
- Martin, R. C. (2003). *Agile software development: principles, patterns, and practices*. Prentice Hall PTR.
- Microsoft. (2016). *Typescript - javascript that scales*. Retrieved 2017-02-14, from <http://www.typescriptlang.org>
- Pico, A. R., Kelder, T., Van Iersel, M. P., Hanspers, K., Conklin, B. R., & Evelo, C. (2008). Wikipathways: pathway editing for the people. *PLoS Biol*, 6(7), e184.
- Riutta, A., Pico, A., Bumbu, & Summer-Kutmon, M. (2013). *Pvjs*. <https://github.com/wikipathways/pvjs>.
- Riutta, A., Pico, A., Bumbu, A., Hanspers, K., Windsor, J., Willighagen, E., & Summer-Kutmon, M. (2013). *Kaavio*. <https://github.com/wikipathways/kaavio>.
- Ryall, K. A., & Tan, A. C. (2015). Systems biology approaches for advancing the discovery of effective drug combinations. *Journal of cheminformatics*, 7(1), 7.
- Twitter. (n.d.). *Bootstrap - the world’s most popular mobile-first and responsive front-end framework*. Retrieved 2017-02-14, from <http://getbootstrap.com/>
- van Iersel, M. P., Pico, A. R., Kelder, T., Gao, J., Ho, I., Hanspers, K., ... Evelo, C. T. (2010). The bridgedb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC bioinformatics*, 11(1), 5.
- Wu, G., Feng, X., & Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. *Genome biology*, 11(5), R53.