Valorizing omics visualization for discovery

Christine Chichester

Department of BioSystems Informatics, Nestle Institute of Heath Sciences, Lausanne, Switzerland
E-mail: chchichester@gmail.com; ORCID: https://orcid.org/0000-0001-6818-334X

Abstract. Scientists from diverse backgrounds are joining the field of data science. This leads to advances in data science being actualized in the context of many different domains. Conclusions from datasets using innovative algorithms are obvious aspects but advances in data science can take on many different forms such as new methods for data interpretation, new data integration and processing technologies, or as will be the topic of this editorial, data visualization techniques. The parity and complementary relationship between techniques from all domains provide ways to improve discovery although quantifying the contributions to discovery process from each technique can be elusive. The experiences described here come from visualizing life science multi-omics data, but most of the remarks can be associated with visualization methods in general. From the perspective that visualization serves as an important method for shaping data science interpretations, this paper sets out: 1) some of the necessary requirements for visualization tools due to the nature of multi-omics datasets and, 2) some of the difficulties encountered in creating and valorizing new visualization implementations for scientific discovery.

Keywords: Multidimensional data, valorization, bias

1. Benefits of visualization

All fields and domains require the use and analysis of data; however, not all domain experts are statisticians or algorithm experts. The omics technologies (genomics, transcriptomics, proteomics, metabolomics, lipidomics, etc.) have generated many multifactorial experiments that necessitate effective visual exploration by life science experts to successfully extract knowledge [10,29,37]. The challenge in practical terms is how to present the data at the right level of detail, in a cohesive, insightful manner. In general, transforming spreadsheet data into visual representations can facilitate new knowledge discovery [43]. The discovery often comes from seeing novel and unexpected patterns in datasets by visually interpreting data in a different way. As there is only limited utility in seeing the expected, one often seeks out outliers, oddities, unusual events and patterns, places where the data do not match expectations [4].

Human working memory has limited capacity and transient storage properties for simultaneous interpretation of multiple hypotheses and huge amounts of evidence linked together by numerous relationships [49]. Data for biological systems are organized as complex networks of molecular and functional

interactions making the intuitive interpretation of multi-omics datasets difficult without help. Visual displays provide a method to extend the working memory capacity by establishing a placeholder for information patterns [50]. More evidence can be viewed in concert. Research can advance more quickly if the barrier to the effective exploration by any scientists is minimized. Therefore, insights from the emerging field of visual analytics [21], which specifically studies the role of visualization in the larger process of understanding and interpreting data, can bear significant rewards. Visual analytics methods have begun to be applied to studying the connection between visualization and analytical reasoning in systems biology [13,21].

2. Characteristics of the data

In the field of multi-omics data assemblage and evaluation, common data characteristics surface; the complexity of the data is related to multidimensionality and multivariate nature, where variance in the measurements can be attributed to other numerous explanatory variables and possible confounders. The data complexity and the multitude of questions to be addressed means static visualization is often insufficient. The user needs to explore the data interactively in order to assess a wide range of questions. In addition to the high dimensionality of the data, information overload, data interconnectivity, and pattern extraction pose major hurdles to developing effective visualizations [28]. Here, one of the main difficulties lies in the design of graphical layouts that contain the complete coordinates [39], although there are implementations, for example in variant genomics, that understand and elegantly address these issues [11].

For intuitiveness and usefulness, it is likely that there is no single generic layout that will cover the requirements needed to answer the range of biological questions. Often the better-known representations, bar and pie charts, histograms, line and scatter plots are used to carry out simple statistical visualization and to report trends and summaries [7]. Node-link visualizations can display graphs/networks and trees, such as ontologies, protein interaction networks, or phylogenies [12]. Other visualization methods that have been tested successfully, but typically incorporate just one omics type, include; heat maps and matrices [26]; parallel coordinates [18]; timeline and topology plots [33]; map and landscape views that build on the metaphor of cartography; space-filling visualizations such as tree maps, hive plots [23], icicle, bubble and sunburst plots [15]; iconography, including star and glyph plots [3,17,46,48]. Specific use cases for high-dimensional data may require visualization such as parallel coordinates, while pie charts and scatter plots can be used for associated clinical variables to examine only a small number of dimensions simultaneously. Most novel visualization applications often employ or build on some of the simpler, well-known techniques that are organized together in innovative combinations. Overall, the choice of visualization for multi-omics data needs to reflect the complex organization of biological phenomena and, importantly, the user must have their own internal representation of the biological phenomena in order to reason about it while exploring the data. In general, experts will have built up from extensive experience a set of patterns for exploring the important elements found in their data and these must be taken into account when providing a visualization [34].

Presently, Cytoscape [40], a Java desktop application, has been widely and effectively used for visualizing and analyzing biological networks and omics data. The Cytoscape App Store (http://apps.cytoscape.org/) provides downloads of several plugins, such as CyLineUpi [5], PINA4MS [6], PTMOracle [42] that perform visualizations of omics data on network or pathway maps although most plugins do not permit the overlay of several different data types in one visualization. Furthermore, there are several web-based tools developed for the visualization of omics data on pathway maps. Customized

maps generated by iPath2.0 [52] allow users to ingest their own data in the context of genomics or metagenomics projects. NetGestalt [41] is an advanced tool for integration of multidimensional omics data, exploiting simple and easily readable one-dimensional layouts of gene networks. NaviCom [8] is a novel development that attempts to visualize multi-omics profiles to gain insights into the patterns of regulation of molecular functions. In general, each omics visualization tool has advantages and disadvantages. The examples cited here are only for a very small collection of the available tools, since the volume and variety of open omics data sets is growing quickly, it is recommended to try out several methods and to regularly look for new tools that can be contrasted against user requirements.

3. Building visualizations

Modeling how a scientist thinks about biology plays a big role on how people interpret and interact with an interface. The application of human–computer interaction (HCI) methods enables a process approach to solve the difficult problems of omics visualization. Scientists want to answer questions with their datasets. While detecting trends is important, ultimately researchers want to see the causal relationships of how A has an effect on B. To address these knowledge discovery needs appropriately, it is useful to understand the current discussions pertaining to design study methodology. Sedlmair et al. [38] propose a clear definition of design studies as well as practical guidance for conducting them effectively. They stress the need to understand the contributions design studies can make to visualizations, when design studies are the appropriate method to use, and how design studies are unique from other approaches. Following on from this, design studies should strive to understand the life scientists' usage of multi-omics as applied to a specific real-world problem, validate the visualization design to confirm that it addresses the problem, and then reflect about process in order to refine visualization design guidelines. Based on the design study, it is possible to identify critical areas that are the most important with respect to user issues and plan a research agenda to pursue the most effective solutions. Frequently to be effective, visualizations benefit from a combination of problem-solving research and technique driven research. Although, when the validation criterion depends on calculating the new knowledge derived due to the application of a visualization tool, measuring the impact can be elusive.

4. Quantifying visualization in the scientific discovery process

The power and value of visualization is often described by its ability to foster insight into and improve understanding of data, which then should lead to enabling intuitive, effective knowledge discovery and analytical activity. This can partly be achieved by removing the cognitive load encountered in managing the large amounts of complex, heterogeneous data, which are commonly delivered by multiple omics experiments [1]. More challenging is that knowledge discovery is seldom an instantaneous event, but requires studying and manipulating the data repetitively from multiple perspectives and possibly using multiple tools. Streamlining repetitive tasks may be a benefit that is linked to discovery but the contribution of this may not be easily traceable back to the visualization. The introduction of data visualization tools may trigger changes in work practices, exacerbating the problem of identifying their contribution to discovery. One measure of success for a visualization could be that users can formulate and answer questions they didn't anticipate before looking at the visualization [31]. If users need to look at the same data from different perspectives and over a long time, they must be motivated and actively intellectually engaged in experimenting with the visualization tool [19]. Conducting longitudinal studies that record

each and every finding by the users over a longer period of time to see how visualization tools influence knowledge acquisition can be very valuable [30,35]. These studies should be conducted with scientists analyzing their own experimental results for the first time. Several studies [14,22,32,36] have conducted such longitudinal studies with evaluations that included frequent user interviews, diary studies, and 'Eureka' reports. Overall, measuring the impact of visualizations on discovery is a difficult task but a range of evaluation methods are being tested to measure success [31].

Users adopt applications that have intuitive interfaces and deliver appropriate context and personalization via a rich end-user interaction. This usually means that the application has been perfectly simplified. The tasks being performed via the interface are streamlined. Irrelevant features or uncertainty does not distract user focus over where to click for the information for answering the next question. Real-time interactive features bring engaging, time-sensitive, or contextual biological information to the forefront [16]. The mental model that users build up whilst interacting feels natural to the way they think without realizing it. Creating this type of visualization takes time, much trial and error, and an attention to psychological as well as the scientific detail. Measuring these attributes has been a current focus in evaluation practices [24].

Finally, Dörk et al. [9], have outlined an approach for HCI that promotes: disclosure of bias and decisions made about the visualization (disclosure), the enabling of multiple interpretations (plurality), a range of possible ways to interact with the visualization (contingency), and allowing users to derive their own hypotheses (empowerment). The principles of disclosure and plurality largely address insight by promoting comprehensible representations, while contingency and empowerment are guiding principles driving impact through flexible interactions and empowering user experiences [9].

5. Bias as a confounding issue

As with any domain of data science, visualizations are to some extent subjective and interpretive. No visualization captures all aspects of a particular dataset from all possible perspectives. Each visualization encompasses some assumptions of the developer and it is important to avoid potentially biasing users with a particular line of thought [45]. With high dimensional data there may be many reasonable approaches to analyzing it. The scientist's perception is biased towards interpretation of information into existing (internal) models of biology and existing expectations. However, human reasoning is subject to a variety of well-documented heuristics and biases [47] that cause people to deviate from how they should rationally make decisions. Therefore, a major challenge to any scientist is to be open to new and important insights while simultaneously avoiding being misled by the tendency to see structure in randomness and to find meaningful patterns in meaningless noise, such that confirmation bias leads to false conclusions [25]. There appears to be little guidance and material that teaches people how to do actual exploratory analysis work [51], let alone with an understanding of their biases. People are fixated with complex statistical models and blindly applying machine learning to data problems when in fact what we need to improve and perfect is our ability to reason with data and make rational decisions under conditions of uncertainty. Complementarily, visualizations are challenged to incorporate a notion of confidence or certainty because the factors that influence the certainty or uncertainty of data vary with the type of information and the type of decisions being made [44]. Statisticians see the world in the light of confirmatory analysis and regard exploration as an inferior approach to analysis. Visualization researchers, too busy building innovative implementations to cope with the new data overload, have done little to teach users how to run actual data exploration methods. Part of the solution to this conundrum may depend on the visualization researchers adopting the philosophy that their implementations

must teach as well as **systematically guide exploratory data analysis** in ways that make the process as effective, reliable, and rational as possible.

6. Visualization as a valuable asset to be rewarded

As discussed above, many aspects must be taken into consideration when developing an interface. A good multidimensional omics visualization tool must maximize simplicity, familiarity, intuitiveness, effectiveness, data correctness [2] as well as minimize bias from both the developer and end user. Even when doing all this, visualization tools can be overlooked and not interpreted as a valuable publishable scientific effort in the context of data science. Clearly, visualizations are necessary for the adoption, use, and efficacy of uptake of computational methods in data science. Major efforts have been made in recent years to create visualization tools that can extract useful knowledge from the vast amount of data generated by high-throughput technologies [10,29,37]. However, more progress is required to create new tools to meet the changing needs of the field. Incremental improvement of visualization software is highly important, but requires great effort from developers for low scientific reward when compared to the development of new methods. There must be acknowledgement that the investment to the study and effort dedicated to the development and maintenance of new tools, as well as user training and support, will be adequately compensated to encourage advancement of the field. Long-term investment and funding are needed to guarantee the maintenance, improvement, and evolution of visualization tools beyond their first publication [37].

7. Conclusion

As the size and complexity of omics datasets continues to increase, the development of user interfaces and interaction techniques that expedite the process of exploring that data must receive new attention. Novel approaches also need to take into consideration the technological challenges and opportunities given by new interaction contexts, ranging from mobile, touch [19,20], and gesture interaction to visualizations on large displays, and encompassing highly responsive web applications. Regardless of the speed of rendering and context, it is important to coherently organize the visual process of exploration to give insight about the data to a user and address psychological aspects of the user experience. Measures to access impact of visualizations remain a challenge and so it follows valorization may not be proportional to the effort put in for development [27]. Overall, to quote Nils Gehlenborg [13]: "The challenge is to create clear, meaningful and integrated visualizations that give biological insight, without being overwhelmed by the intrinsic complexity of the data".

Acknowledgements

References

[1] E.W. Anderson, Evaluating scientific visualization using cognitive measures, in: *BELIV Workshop: Beyond Time and Errors – Novel Evaluation Methods for Visualization BELIV*, 2012. doi:10.1145/2442576.2442581.

[2] A. Bertini, D. Tatu and A. Keim, Quality metrics in high-dimensional data visualization: An overview and systematization, *IEEE Trans. Vis. Computer Graphics* **17**(12) (2011), 2203–2212. Available at: https://bib.dbvis.de/uploadedFiles/350.pdf. doi:10.1109/TVCG.2011.229.

[3] S.K. Card, J.D. Mackinlay and B. Shneiderman, *Reading in Information Visualization*, Morgan Kaufmann Publishers, Inc., 1999. ISBN-13:978-1558605336.

[4] K. Cook, R. Earnshaw and J. Stasko, Guest editors' introduction: Discovering the unexpected, *Computer Graphics and Applications, IEEE* **27**(5) (2007), 15–19. PMID:17913020.

[5] M.C.D. Costa, T. Slijkhuis, W. Ligterink, H.W.M. Hilhorst and D. de Ridder, CyLineUp: A cytoscape app for visualizing data in network small multiples, *F1000Research* **5** (2016), 635. doi:10.12688/f1000research.8402.1.

[6] M.J. Cowley, M. Pinese, K.S. Kassahn et al., PINA v2.0: Mining interactome modules, *Nucleic Acids Res.* **40** (2012), D862–D865. doi:10.1093/nar/gkr967.

[7] A.S. Dadzie and M. Rowe, Approaches to visualizing linked data: A survey, *Semantic Web* **2**(2) (2011), 89–124. doi:10.3233/SW-2011-0037.

[8] M. Dorel, E. Viara, E. Barillot, A. Zinovyev and I. Kuperstein, NaviCom: A web application to create interactive molecular network portraits using multi-level omics data, *Database* **2017** (2017), bax026. doi:10.1093/database/bax026.

[9] M. Dörk, P. Feng, C. Collins and S. Carpendale, Critical InfoVis: Exploring the politics of the visualization, in: *CHI '13 Extended Abstracts of Human Factors on Computing Systems (CHI EA '13)*, 2013, pp. 2189–2198. doi:10.1145/2468356.2468739.

[10] W. Dunn, A. Burgun, M.O. Krebs and B. Rance, Exploring and visualizing multidimensional data in translational research platforms, *Brief Bioinform.* (2016). doi:10.1093/bib/bbw080.

[11] J.A. Ferstay, C.B. Nielsen and T. Munzner, Variant view: Visualizing sequence variants in their gene context, *IEEE Transactions on Visualization and Computer Graphics* **19**(12) (2013), 2546–2555. doi:10.1109/TVCG.2013.214.

[12] T.C. Freeman, L. Goldovsky, M. Brosch, S. van Dongen et al., Construction, visualisation, and clustering of transcription networks from microarray expression data, *PLoS Comput. Biol.* **3**(10) (2007), 2032–2042. doi:10.1371/journal.pcbi.0030206.

[13] N. Gehlenborg, S.I. O'Donoghue, N.S. Baliga and A. Goesmann, Visualization of omics data for systems biology, *Nature Methods* **7**(3 Suppl.) (2010), S56–S68. doi:10.1038/nmeth.1436.

[14] J. Gerken, P. Bak and H. Reiterer, Longitudinal evaluation methods in human–computer studies and visual analytics, in: *InfoVis*, 2007. Available at: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-47547.

[15] M. Glueck, P. Hamilton, F. Chevalier, S. Breslav, A. Khan, D. Wigdor and M. Brudno, PhenoBlocks: Phenotype comparison visualizations, *IEEE Transactions on Visualization and Computer Graphics* **22**(1) (2016), 101–110. doi:10.1109/TVCG.2015.2467733.

[16] V. Gonzales and A. Kobsa, A workplace study of the adoption of information visualization systems, in: *Proceeding of IKNOW'03: 3rd International Conference of Knowledge Management*, 2003, pp. 92–102. Available at: http://www.manchester.ac.uk/escholar/uk-ac-man-scw:1b7449.

[17] J. Heer, M. Bostock and V. Ogievetsky, A tour through the visualization zoo, *Communications of the ACM* **53**(6) (2010), 59–67. doi:10.1145/1743546.1743567.

[18] A. Inselberg, The plane with parallel coordinates, *The Visual Computer* **1** (1985), 69–91. Available at: http://www.springerlink.com/index/X3P504736MU14661.pdf.

[19] T. Isenberg, Position Paper: Touch interaction in scientific visualization, in: *Proceedings of the Workshop on Interactive Surfaces*, 2011, pp. 24–27. Available at: https://hal.inria.fr/hal-00781512.

[20] D.F. Keefe, Integrating visualization and interaction research to improve scientific workflows, *IEEE Computer Graphics and Applications* **30** (2010), 8–13. doi:10.1109/MCG.2010.30.

[21] D. Keim, G. Andrienko, J.D. Fekete, C. Görg, J. Kohlhammer and G. Melançon, Visual analytics: Definition, process, and challenges, in: *Information Visualization: Human-Centered Issues and Perspectives*, A. Kerren et al., eds, Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-70956-5_7.

[22] A. Kobsa, An empirical comparison of three commercial information visualization systems, in: *Proceedings of InfoVis*, 2001, pp. 123–130. doi:10.1109/INFVIS.2001.963289.

[23] M. Krzywinski, I. Birol, S.J. Jones and M.A. Marra, Hive plots – Rational approach to visualizing networks, *Brief Bioinform.* **13**(5) (2012), 627–644. doi:10.1093/bib/bbr069.

[24] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale, Empirical studies in information visualization: Seven scenarios, *IEEE Transactions on Visualization and Computer Graphics* **9**(18) (2012), 1520–1536. doi:10.1109/TVCG.2011.279.

[25] M.R. Munafo, B.A. Nosek, D.V.M. Bishop et al., A manifesto for resproducible science, *Nature Human Behavior* **1** (2017), 21. doi:10.1038/s41562-016-0021.

[26] C. Nielsen and B. Wong, Points of view: Managing deep data in genome browsers, *Nature Methods* **9** (2012), 512. doi:10.1038/nmeth.2049.

[27] C. North, Toward measuring visualization insight, *Computer Graphics and Applications* **26**(3) (2006), 6–9. doi:10.1109/MCG.2006.70.

[28] M. Oghbaie, M.J. Pennock and W.B. Rouse, Understanding the efficacy of interactive visualization for decision making for complex systems, in: *Systems Conference (SysCon) Annual IEEE*, 2016, pp. 1–6. doi:10.1109/SYSCON.2016.7490526.

[29] G.A. Pavlopoulos, D. Malliarakis, N. Papanikolaou, T. Theodosiou, A.J. Enright and I. Iliopoulos, Visualizing genome and systems biology: Technologies, tools, implementation techniques and trends, past, present and future, *Gigascience* **4**(1) (2015), 1–27. doi:10.1186/s13742-015-0077-2.

[30] A. Perer and B. Shneiderman, Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines, *IEEE Computer Graphics and Applications* **29**(3) (2009), 39–51. doi:10.1109/MCG.2009.44.

[31] C. Plaisant, The challenge of information visualization evaluation, in: *Proceedings of the Working Conference on Advanced Visual Analytics*, 2004, pp. 109–116. doi:10.1145/989863.989880.

[32] J. Rieman, A field study of exploatory learning strategies, *ACM Transactions on the Computer–Human Interaction* **3** (1996), 189–218. doi:10.1145/234526.234527.

[33] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, T. Turic and F. Drexler, Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation, in: *Symposium of the Austrian HCI and Usability Engineering Group*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-25364-5_22.

[34] D. Sacha, A. Stoffel, F. Stoffel, B.C. Kwon, G. Ellis and D.A. Keim, Knowledge generation model for visual analytics, in: *IEEE Transactions on Visualization and Computer Graphics*, 2014. doi:10.1109/TVCG.2014.2346481.

[35] P. Saraiya, C. North and K. Duca, An evaluation of microarray visualization tools for biological insight, in: *INFOVIS 04: Proceedings of the IEEE Symposium on Information Visualization*, 2004. doi:10.1109/INFVIS.2004.5.

[36] P. Saraiya, C. North and K. Duca, An insight-based methodology for evaluating bioinformatics visualizations, *IEEE Trans. Vis. Comput. Graph.* **11**(4) (2005), 443–456. doi:10.1109/TVCG.2005.53.

[37] M.P. Schroeder, A. Gonzalez-Perez and N. Lopez-Bigas, Visualizing multidimensional cancer genomics data, *Genome Medicine* **5** (2013), 9. doi:10.1186/gm413.

[38] M. Sedlmair, M. Meyer and T. Munzner, Design study methodology: Reflections from the trenches and the stacks, *IEEE Transactions on Visualization and Computer Graphics* **18** (2012), 2431–2440. doi:10.1109/TVCG.2012.213.

[39] H.X. Self, J. Zeitz, L. House, S. Leman and C. North, Designing usable interactive visual analytics tools for dimension reduction, in: *Human Centered Machine Learning at CHI*, 2016. Available at: https://infovis.cs.vt.edu/sites/default/files/Self_design_paper_final.pdf.

[40] P. Shannon, A. Markiel, O. Ozier et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.* **13** (2003), 2498–2504. doi:10.1101/gr.1239303.

[41] Z. Shi, J. Wang and B. Zhang, NetGestalt: Integrating multidimensional omics data over biological networks, *Nat. Methods* **10**(7) (2013), 597–598. doi:10.1038/nmeth.2517.

[42] A.J. Tay, C.M.I. Pang, D.L. Winter and M.R. Wilkins, PTMOracle: Cytoscape app for co-visualising and co-analysing post-translational modifications in protein interaction networks, *J. Proteome Res.* **16** (2017), 1988–2003. doi:10.1021/acs.jproteome.6b01052.

[43] J.J. Thomas and K.A. Cook, A visual analytics agenda, *Computer Graphics and Applications, IEEE* **26**(1) (2006), 10–13. PMID:16463473.

[44] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan and M. Pavel, A typology for visualizing uncertainty, *Proceedings SPIE, Visualization and Data Analytics* **5669** (2005), 146–157. doi:10.1117/12.587254.

[45] M. Tory and T. Möller, Human factors in visualization research, *IEEE Trans. Vis. Comput. Graph.* **10**(1) (2004), 72–84. doi:10.1109/TVCG.2004.1260759.

[46] E. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, 2001. ISBN:0-9613921-0-X.

[47] A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and bias, *Science* **185** (1974), 1124–1131. doi:10.1126/science.185.4157.1124.

[48] J.M. Villaveces, P. Koti and B.H. Habermann, Tools for visualization and analysis of molecular networks, pathways, and -omics data, *Adv. Appl. Bioinform. Chem.* **8** (2015), 11–22. doi:10.2147/AABC.S63534.

[49] E.K. Vogel and M.G. Machizawa, Neural activity predicts individual differences in visual working memory capacity, *Nature* **428** (2004), 748–751. doi:10.1038/nature02447.

[50] C. Ware, *Information Visualization: Perception for Design*, 2012. ISBN:9780123814654.

[51] R.W. White, B. Kules, S.M. Drucker and M.C. Schraefel, Supporting exploratory search, introduction, *Communications of the ACM* **49**(4) (2006), 36–39. doi:10.1145/1121949.1121978.

[52] T. Yamada, I. Letunic, S. Okuda, M. Kanehisa and P. Bork, iPath2.0: Interactive pathway explorer, *Nucleic Acids Res.* **39** (2011), W412–W415. doi:10.1093/nar/gkr313.