This item is the archived peer-reviewed author-version of:

Preparing students for the data-driven life science era through a real-world viral infection case

Reference:

Laukens Kris, Eyckmans Marleen, De Neuter Nicolas, Naulaerts Stefan, Meysman Pieter, van Ostade Xaveer.- Preparing students for the data-driven life science era through a real-world viral infection case
Journal of biological education - ISSN 0021-9266 - Abingdon, Routledge journals, Taylor & Francis Ltd, 2019, 10 p.
Full text (Publisher's DOI): <https://doi.org/10.1080/00219266.2019.1667408>
To cite this reference: <https://hdl.handle.net/10067/1625390151162165141>

Preparing students for the data-driven life science era through a real-world viral infection case

Short title: Teaching data-driven life sciences with a viral infection case

Kris Laukens^{1,2}, Marleen Eyckmans³, Nicolas De Neuter^{1,2}, Stefan Naulaerts^{1,2}, Pieter Meysman^{1,2}, Xaveer Van Ostade⁴

¹. Adrem Data Lab, Department of Mathematics & Computer Science, University of Antwerp, Antwerp, Belgium

². Biomedical informatics network Antwerpen (biomina), University of Antwerp, Antwerp, Belgium

³. Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Antwerp, Belgium

⁴. Laboratory of Protein Science, Proteomics and Epigenetic Signaling (PPES) and Centre for Proteomics and Mass spectrometry (CFP-CeProMa), University of Antwerp, Antwerp, Belgium

* corresponding authors: kris.laukens@uantwerpen.be, xaveer.vanostade@uantwerpen.be

Abstract

Developing practical, hands-on computational skills is an important learning goal in most life science education programs. While computational biology and bioinformatics are nowadays an inherent part of most life science degrees, it remains a challenge to activate students to employ such diverse skills in a practical research context. More specifically, the large-scale, complex datasets that are now omnipresent in research are not trivial to integrate in a life science education curriculum. In this education paper we present a formula to allow students to acquire hands-on skills with the data-driven study of complex biomolecular datasets. The course starts with a series of introductory classes on experimental and computational methods, including a practical session focused on Cytoscape and associated computational network biology techniques. Since we believe that working on your own idea will considerably increase the student's motivation and driving force, we introduce a recent hallmark virus-host protein-protein interaction article from which the database serves as a starting point for the students to develop their individual research goal. This goal is presented and discussed in a first informal meeting with students and teachers. Over the next two months about 4 informal meetings are planned wherein students present their problems, progress and challenges and exchange ideas in order to better reach their goal. The role of the teachers is to assist students to independently collect data and explore new software tools for their research idea. The challenge is based on recent, real world, publicly available (interaction) data and software. We present the concept, and methods and share our experiences from an evaluation of the student learning outcomes.

Introduction

During the last two decades, bio-analytical instruments became increasingly miniaturized and automated, allowing for the simultaneous measurement of hundreds to thousands of biomolecules. A multitude of bioinformatics tools promise to retrieve understandable insights from the overwhelming and exponentially growing amount of resulting 'omics' data. As life

1 sciences become increasingly (big) data-driven, developing skills to handle and interrogate such
2 data has become an indispensable component of the life science curriculum [1,2]. In recent
3 years, life scientist trainers have therefore explored new ways to fulfill training needs imposed
4 by this changing life sciences landscape, accompanied by efforts such as the establishment of
5 bioinformatics trainer networks [3] and the definition of computational biology learning goals for
6 the life science curriculum [4–6].

7
8 A trend that recently received broad attention in the live science community is that
9 computational biology can increasingly rely on large-scale, previously published data to perform
10 fundamentally new biomedical research, inciting a disgruntled editorial in the prestigious New
11 England Journal of Medicine that labeled computational biologists re-analysing public data as
12 "research parasites" [7], which in turn sparked a fierce backlash in the bioinformatics community
13 [8]. Re-analyses and meta analyses of public datasets has already earned a central spot in the
14 computational biology field. It has its merits not only for fundamental new research and
15 assurance of reproducibility, but it also offers an unprecedented opportunity to train life
16 scientists' data-driven computational biology skills. In this paper we present a learning module in
17 which we have embarked in guiding students through this important paradigm shift towards
18 data-driven scientific thinking.

19
20 We present our experiences with teaching a course on "functional proteome analysis" using a
21 data-driven, interactive and project-centric approach. For this course, each student works on a
22 project, related to an actual (e.g. Ebola in 2015) or highly impactful (e.g. herpes virus host-
23 pathogen interactions in 2016) viral infection theme that strongly captures their attention. The
24 teaching concept builds upon three stages in which the students gradually acquire the desired
25 learning outcomes. First, they get a set of introductory classes on experimental and
26 computational methods for functional proteome analysis. Second, they explore, with

computational network biology approaches, a compilation of recently published data and (meanwhile) study the associated literature. They then use the acquired insights and skills to define and address their own research goal. Third, based on public software tools and public data, they undertake their own data-driven computational study, which they finally describe in a scientific paper and in an oral presentation. We share our method and experiences with this learning module and analyze the change in student thinking through surveys before, during and after the course.

Student background

The course is followed by students of the first year of the master program in biochemistry and biotechnology. General prerequisites of this elective course are passive knowledge of English (the running language of the course is Dutch but international scientific sources such as papers, databases and software are used throughout the course), general knowledge of the use of a PC and the Internet, and a general notion of the basic concepts of Medline (Pubmed). Students following this course also need to fulfill a set of specific prerequisites that are covered by the bachelor program in biochemistry, i.e. they have successfully passed protein biochemistry and basic bioinformatics course exams. The bachelor course also includes the course “Scientific Communication” wherein basic concepts on scientific writing and presenting research outcomes are given.

Learning goals

The learning goals of this course are defined as follows:

- The student has insights in experimental and computational techniques for the analysis of protein interactions
- The student knows and understands the most important public databases for protein interactions, and is able to use them.

- The student has insights in the progress, opportunities, perspectives and limitations of data-driven systems biology and find novel, creative research solutions by using new or updated programs after exploration of the internet.
- The student is able to formulate systems biology insights, and to disseminate these insights in a scientific report and in discussions with the teachers and/or fellow students.
- The student is able to actively and critically work with large scale molecular datasets.
- The student is able to start from public molecular data to generate a research hypothesis that is studied using public software tools in a data-driven approach.

Course outline

A sequence of events of the course is given in Figure 1.

1. origins and preparation of the data

Protein-protein interaction data between hosts and viruses were collected from public data sources and processed according to the following general procedure, before they were supplied to the students. All protein interactions were obtained either directly from the supplementary data of the respective research paper (see below), or from VirHostNet, version 1.0. [9]. From the retrieved dataset, only interactions between viral proteins and human proteins were retained, resulting in the host-pathogen interactome. We then completed the human interactome by adding non-redundant protein-protein information from mIntAct [10] and BIOGRID [11]. Each of the protein identifiers was further enriched with additional information from the Uniprot resource [12].

The choice of the viral data is different every year. In this paper, we briefly describe the human-ebola interaction network (studied in 2015) and a human-herpesvirus interaction network

(studied in 2016). To prepare the Ebola network, we used the full reference proteomes for Zaire Ebola virus (strains Mayinga-76: reference proteome UP000007209/taxid:128952 and Kikwit-95: reference proteome UP000007208/taxid: 128951) from the VirHostNet 1.0 database. For the herpes virus network, a human- Kaposi's sarcoma-associated herpesvirus (KSHV) interaction network was created using Cytoscape's union-based merge tool. The networks used to create this union were BIOGRID's human interactome (v3.4.129) [11] and the high quality human-KSHV interactome as presented by Davis et al. 2015 [13]. Both the Ebola and herpes network were processed in such a way that the networks were ready to be loaded into the Cytoscape network analysis software [14] with the proper accession ids.

2. introductory *ex cathedra* colleges

At the beginning of the course two lectures are given on the physical, chemical, genetic and molecular biological strategies for analysis of protein-protein interactions, with a focus on methods for high-throughput screening such as co-purification-LC/MS and yeast-two-hybrid methods. Several landmark articles and/or reviews are given for background information [15–20]. Next, two lectures are given regarding the main principles of graph theory (definitions, graph types, graph / node / edge properties, essential algorithms) and its applications in network biology (inference, analysis and interpretation of biomolecular networks) [21,22]. This is followed by a practical hands-on introduction to the use of Cytoscape [14] for the analysis of biomolecular networks. However, since we also consider as a learning goal in this course the exploration of the internet for (continuously renewed) tools in computational analysis of biomolecular networks, we limit our introduction to established methods such as Cytoscape and associated plugins/apps and ... as to provide the student with a basis which could serve for further exploration towards more specialized software.

3. assignment instructions, work colleges & intermediary reports

At the end of the *ex cathedra* sessions, students are asked to read the paper that describes the virus-host protein interaction dataset that will be studied, and they are invited to formulate one or more hypotheses they want to explore. At this stage, students are encouraged to think “out-of-the-box” and to have an open mind towards new ideas. These hypotheses are discussed and adjusted or reformulated when needed in the first of a series of interactive work colleges.

Work colleges start several weeks after the *ex cathedra* colleges and serve to inform, guide and troubleshoot the students during the learning process. During these colleges students are asked to present their progress to fellow students and teachers with a short slideshow, in which they can describe and discuss - in group - the problems they encountered. These work colleges are organized as often as required, depending on the progress (typically there are 3 to 4). Students that get ‘stuck’ meanwhile can also directly contact the lecturers. Most frequent questions were related to the correct conversion of datasets, instructions on the use of Cytoscape, feedback on the feasibility of a chosen strategy or the relevance of a literature search etc.

4. use of software tools

The students are encouraged to look around and use any available software tool that was useful for their study. Most students mainly relied on Cytoscape [14] and a range of Cytoscape Apps / plugins. An overview of the tools that were successfully employed is presented in Table 1.

Table 1: Overview of student projects and used tools for the academic years 2015 and 2016. Projects were categorized according to focus of analysis (T: therapeutic analysis, I: infection analysis) and type of study (1: comparison of networks, 2: single network focused, 3: literature study). Tools explored by students, in addition to the basic Cytoscape tools provided, include both

- 1 Cytoscape plug-in applications as well as other bioinformatics resources, and are listed for each
2 project.

Category	Project title	Tools used
T1	Ebola versus HIV	BiNGO app BioGrid Uniprot
I1	Analysis of a putative Ebola virus – host PPIN suggests reduced interference with transcription and nucleosomal integrity mechanisms	ClueGo app
I3	Analysis of the interaction of the ebola proteins VP24 and VP35 with importins and dynein	BiNGO app Swiss PDB Viewer
I2	Ebola virus: an interaction study	ClueGo app
I1	Viruses and the human interferon network	ClueGO app CluePedia
I2	Validation of <i>in silico</i> predictions based on the KSHV PPI-network	CyKEGGParser app Uniprot ID converter BiNGO app WordCloud
I3	The pathogenesis of multicentric Castleman's disease by Kaposi's sarcoma-associated herpesvirus	BiNGO app Reactome pathways app
T3	Human proteins involved with KS pathogenesis as caused by KSHV: possible therapeutic targets	BiNGO app Uniprot Muscle (MSA) Interpro
I3	Comparison of the oncogenic mechanism of KSHV with other oncogenic viruses	/
I2	Discrimination of lytic and latent oncogenicity by way of an extended KSHV interactome	DisGeNET database + app cBioPortal FunDO GeneAlaCart GeneCards BioMart OMIM LabWorm DAVID WebGestalt Uniprot ID converter

11	KSHV and EBV infection of epithelial cells share similar targets in oncogenic pathways.	StringDB DyNet KEGG pathways
----	---	------------------------------------

5. Final presentation of the research paper

Students are asked to write a paper on their investigation, which must be no longer than 10 pages including figures and excluding references. Supplemental data can be added. The paper must include 1. an introductory part, 2. information on used methodology, 3. results and 4. discussion. In addition, at the end of the course they are required to present their progress during a 20 minute presentation, that is discussed within the group.

6. Quotation of student performances

Students are assessed on the basis of several criteria (Table 2). We do not base the evaluation on the final research outcomes. Instead the emphasis lies on the way the student handles, reports and presents the data, and the diversity and originality of the approach.

Table 2: Summary of assessment criteria.

Quotation criteria			
	Cooperation and commitment during sessions	Written report	Oral presentation
percentage of total:	30%	40%	30%
mainly assessed on:	Attendance and participation	Scientific accuracy	Clarity of the structure and the slides
	Provide ideas / creativity	Thoroughness	Clarity of oral presentation
	Knowledge of various techniques	Clarity of the report	Answers to Questions
	Critical attitude	Innovativeness and originality of the approach	

	Data collection		
--	--------------------	--	--

Outcomes of the course

The topics selected by the students broadly fell into two categories: the elucidation of new therapeutics or the discovery of trends within the viral infection mechanism pathways. A list of topics selected during the academic years 2015 and 2016 is presented in table 1. A selection of figures from the papers can be seen in Figure 2 showing the variety of approaches used by the students. In all cases, students independently searched for complementary data from either the literature or other public databases to support their study.

Most students remained within the comfort of the Cytoscape environment and its plugin ecosystem. As such, most projects had at least one gene ontology enrichment analysis using either BinGO [23] or ClueGO [24]. Some students looked at other tools but often had difficulty using or interpreting the tool and its results. This prompted most students to take one of three potential paths i.e. i) a comparison between two or more networks; ii) analysis of the single virus-host interaction network and iii) a network-inspired literature search. Interestingly, once one of the students found an interesting new analysis or tool that was easy to use, other students would quickly learn from these experiences and adapted such methods for their own study.

In the discussion of their work, the students often endeavored to interpret their findings and potential implications within the bigger picture. While some students tended to overestimate the significance of their result, most remained critical of their findings and potential implications, indicating that they have gained some insight into the limitations of the studied approaches and methods.

Course evaluation

While courses at the University of Antwerp are routinely evaluated by a standardized quality assurance cycle, the students who followed this course in 2016 were asked to participate in a more elaborate evaluation procedure before, during and after the program.

1. At the beginning of the course, students shared some individual characteristics regarding learning and motivation traits (based on the LEMO test [25,26]), e.g. the reasons why they chose this specific course, how they assess their own relevant background and what they expect of the course.
2. During the course, students wrote down the time they had spent on the assignment on one hand and on attending lectures on the other hand.
3. At the end of the course, two surveys were sent out. The first one repeated some of the individual characteristics of the students from the survey at the beginning of the course combined with an enquiry about the acquired learning outcomes and skills. The second survey entailed the standard evaluation form of a course evaluation, validated by the university [27] and provided insights in the quality of teaching, study material, feedback possibilities, and so on.

Since the number of students participating in the course was limited (6 students in 2016), the results of the evaluations only give an indication of trends within participating students. Therefore, results must be interpreted with this aspect in mind. Some interesting outcomes are presented in Figure 3. In the first weeks of the course, most of the students struggled with the openness of the assignment, with finding a suitable topic and with getting started with Cytoscape, probably since they were not familiar to this degree of interactivity and freedom. The self-doubt and self-criticism students experienced to a lesser or more extent, might be reflected in the amount of time they spent on the assignment (ranging from 36 to 134 hours) as well. At

1 the end of the course, students tended to be less discouraged when faced with large datasets
2 and were more inclined to search for and use different software tools. They also showed a
3 tendency to have less difficulties in dealing with ill-defined problems. Prior to the course, none of
4 the students indicated that they would search literature for ideas when confronted with a large
5 dataset. However, after finishing the course, three students indicated that they would. As a
6 general conclusion, all learning outcomes were, according to the students, met at the end of the
7 course. As an overall feeling, students have found the course and its specific didactic form
8 challenging but worthwhile.

10 **Conclusions**

11 To anticipate upon the continuously emerging global health threats, it is important to instruct
12 new generations of life scientists not only to analyze their own experimental data, but also to
13 exploit the vast amount of publicly available data. Through the learning format described in this
14 paper, we have successfully shown that publicly sharing data not only benefits research
15 purposes but also presents exciting training opportunities. Indeed, public data offered the
16 students a chance to work with data that is otherwise very expensive, complex and time-
17 consuming to generate. We propose that learning to perform life sciences from a data-driven
18 perspective becomes part of life science curricula and express our hope that it can also
19 contribute to an evolution in which, in a not too-distant future, thoughtful but critical data-reuse
20 will be broadly considered a virtue instead of an act of parasitism.

22 **Acknowledgements**

23 The authors wish to thank the students who have followed this course in 2015 and 2016 for their
24 active collaboration and willingness to participate in the evaluation surveys. The colleagues of
25 the Antwerp School of Education are gratefully acknowledged for their valuable input within the

- 1 framework of the "docentenopleiding". We dedicate this paper to our colleague, Prof. Sandra
- 2 Apers, who passed away much too early on February 5th, 2017.

References

1. Greene AC, Giffin KA, Greene CS, Moore JH. Adapting bioinformatics curricula for big data. *Brief Bioinform.* 2016;17: 43–50. doi:10.1093/bib/bbv018
2. Carvalho BS, Rustici G. The challenges of delivering bioinformatics training in the analysis of high-throughput data. *Brief Bioinform.* 2013;14: 538–547. doi:10.1093/bib/bbt018
3. Atwood TK, Bongcam-Rudloff E, Brazas ME, Corpas M, Gaudet P, Lewitter F, et al. GOBLET: The Global Organisation for Bioinformatics Learning, Education and Training. *PLOS Comput Biol.* 2015;11: e1004143. doi:10.1371/journal.pcbi.1004143
4. Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, Clase K. A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE-Life Sci Educ.* 2014;13: 607–623. doi:10.1187/cbe.13-10-0193
5. Via A, Blicher T, Bongcam-Rudloff E, Brazas MD, Brooksbank C, Budd A, et al. Best practices in bioinformatics training for life scientists. *Brief Bioinform.* 2013;14: 528–537. doi:10.1093/bib/bbt043
6. Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, et al. Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinform.* 2010;11: 544–551. doi:10.1093/bib/bbq021
7. Longo DL, Drazen JM. Data Sharing. *N Engl J Med.* 2016;374: 276–277. doi:10.1056/NEJMe1516564
8. McNutt M. #IAmAResearchParasite. *Science.* 2016;351: 1005–1005. doi:10.1126/science.aaf4701
9. Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, André P, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus–host interaction networks. *Nucleic Acids Res.* 2009;37: D661–D668. doi:10.1093/nar/gkn794
10. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 2014;42: D358–D363. doi:10.1093/nar/gkt1115
11. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucl Acids Res.* 2006;34: D535–539.
12. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015;43: D204–D212. doi:10.1093/nar/gku989
13. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, et al. Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Mol Cell.* 2015;57: 349–360. doi:10.1016/j.molcel.2014.11.026
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003;13: 2498–2504. doi:10.1101/gr.1239303

- 1 15. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey
2 reveals modularity of the yeast cell machinery. *Nature*. 2006;440: 631–636.
3 doi:10.1038/nature04532
- 4 16. Gavin A-C, Bösch M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional
5 organization of the yeast proteome by systematic analysis of protein complexes. *Nature*.
6 2002;415: 141–147. doi:10.1038/415141a
- 7 17. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive
8 visualization of biological networks and protein structures. *Nat Protoc*. 2012;7: 670–685.
9 doi:10.1038/nprot.2012.004
- 10 18. Meyniel-Schicklin L, de Chassey B, André P, Lotteau V. Viruses and interactomes in
11 translation. *Mol Cell Proteomics MCP*. 2012;11: M111.014738.
12 doi:10.1074/mcp.M111.014738
- 13 19. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Deinhardt K, Darie CC. Protein-
14 protein interactions: switch from classical methods to proteomics and bioinformatics-based
15 approaches. *Cell Mol Life Sci CMLS*. 2014;71: 205–228. doi:10.1007/s00018-013-1333-1
- 16 20. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Loo JA, Darie CC. Investigation of
17 stable and transient protein-protein interactions: Past, present, and future. *Proteomics*.
18 2013;13: 538–557. doi:10.1002/pmic.201200328
- 19 21. Laukens K, Naulaerts S, Vanden Berghe W. Bioinformatics approaches for the functional
20 interpretation of protein lists: From ontology term enrichment to network analysis.
21 *PROTEOMICS*. 2015;15: 981–996. doi:10.1002/pmic.201400296
- 22 22. Meysman P, Titeca K, Eyckerman S, Tavernier J, Goethals B, Martens L, et al. Protein
23 complex analysis: From raw protein lists to protein interaction networks. *Mass Spectrom*
24 *Rev*. 2015; n/a-n/a. doi:10.1002/mas.21485
- 25 23. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation
26 of Gene Ontology categories in Biological Networks. *Bioinformatics*. 21: 3448–3449.
27 doi:10.1093/bioinformatics/bti551
- 28 24. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a
29 Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation
30 networks. *Bioinformatics*. 2009;25: 1091–1093. doi:10.1093/bioinformatics/btp101
- 31 25. Mosselaer H Van de, Donche V, Jansen N, Petegem P Van. De Lemo-test : wat, waarom,
32 hoe? Goesting in leren en werken / Mosselaer, Van de, H [edit]; e.a. Garant; 2012. pp. 31–
33 60. Available: <http://hdl.handle.net/10067/969670151162165141>
- 34 26. Vanthournout G, Mosselaer H Van de, Donche V, Vansteenkiste M. Discovering and
35 strengthening learning strategies and motivation using the LEMO-instrument. Ten times the
36 first year : reflections on ten years of the European First Year Experience Conference /
37 Bonne, PJ [edit]; et al. Lannoo Campus; 2016. pp. 53–70. Available:
38 <http://hdl.handle.net/10067/1328120151162165141>

- 1 27. Spooren P, Mortelmans D, Denekens J. Student evaluation of teaching quality in higher
2 education: development of an instrument based on 10 Likert-scales. *Assess Eval High*
3 *Educ.* 2007;32: 667–679. doi:10.1080/02602930601117191

4

Figure legends

Figure 1. Sequence of events during the course of Functional Proteomics. Teachers are preparing a suitable virus-host PPI dataset. After a series of lessons on wet lab techniques and computational methods to analyse protein-protein interactions and their networks (blue arrows), a given virus-host network, extensively described in a recent article, is presented to the students. Soon thereafter, students follow a practical course on the use of Cytoscape (pink arrow), during which the virus-host dataset can be used to explain the different features. With this information, students start their search for a topic that is close to their interest, using literature combined with their Cytoscape skills (green arrows). The first results of this search are presented in a first meeting during which the teachers advise on innovativeness and feasibility concerning positive outcome, workload, time, available software, accesibility of data etc.. With this information, the students are asked to (partially) reconsider or built and further explore their research goal which must be assured the next meeting. They then use the different data sets and bioinformatics tools they have chosen which, in combination with literature, must allow them to finalise their project. During this period one or two meetings were organised to further guide the students in this proces. Finally, students are asked to report on this proces in a written report, which is submitted a few days in advance of their final presentation (yellow arrows). This presentation summarises their research goal (introduction and aims), the bioinformatics tools they used (materials and methods), the outcome(s) of their research, including the problems they encountered (results), and a discussion/future perspectives part.

Figure 2. Examples of figures generated by students for their final report. Figures were selected to show the variability in the types of analyses performed by students. A) From “Ebola virus, an interaction study”; analysis of the network properties using NetworkAnalyzer, a built-in Cytoscape tool. B) From “Viruses and the human interferon network”; visual representation of several virus-host protein-protein interaction networks. In this particular case, the networks were limited to human interferon proteins and their interaction partners. C) From “Analysis of the interaction of the ebola proteins VP24 and VP35 with importins and dynein”; figure highlighting both structural and pathway-related characteristics for one of the proteins under investigation. D) From “Human proteins involved with KS pathogenesis as caused by KSHV: possible therapeutic targets”; bioinformatics analysis of two related human proteins of interest, including protein sequence alignment and comparison of protein domains using data from InterPro.

Figure 3: Relevant results of the survey completed by the students before (pre-course) and after (post-course) the course. Number of students are plotted, from “Totally Disagree” to “Totally Agree”.

Figure 1.

Figure 2.

1 Figure 3

2
3