BRIEFINGS IN BIOINFORMATICS. VOL 14. NO 5. 618–632
Advance Access published on 2 May 2013
doi:10.1093/bib/bbt024

Teaching the bioinformatics of signaling networks: an integrated approach to facilitate multi-disciplinary learning

Tamas Korcsmaros, Zsuzsanna A. Dunai, Tibor Vellai and Peter Csermely

Submitted: 31st January 2013; Received (in revised form): 21st March 2013

Abstract

The number of bioinformatics tools and resources that support molecular and cell biology approaches is continuously expanding. Moreover, systems and network biology analyses are accompanied more and more by integrated bioinformatics methods. Traditional information-centered university teaching methods often fail, as (1) it is impossible to cover all existing approaches in the frame of a single course, and (2) a large segment of the current bioinformation can become obsolete in a few years. Signaling network offers an excellent example for teaching bioinformatics resources and tools, as it is both focused and complex at the same time. Here, we present an outline of a university bioinformatics course with four sample practices to demonstrate how signaling network studies can integrate biochemistry, genetics, cell biology and network sciences. We show that several bioinformatics resources and tools, as well as important concepts and current trends, can also be integrated to signaling network studies. The research-type hands-on experiences we show enable the students to improve key competences such as teamworking, creative and critical thinking and problem solving. Our classroom course curriculum can be re-formulated as an e-learning material or applied as a part of a specific training course. The multi-disciplinary approach and the mosaic setup of the course have the additional benefit to support the advanced teaching of talented students.

Keywords: *signaling; network; epistasis; university teaching; protein–protein interaction; domain*

INTRODUCTION

In the past 30 years, we have experienced a rapid development in molecular biology, which resulted in an exponential growth of available biological data. Advances in high-throughput molecular technologies, such as genome-wide microarray platforms [1,2], genome-wide association studies [3,4], signaling pathway analysis [5,6], genome-scale RNA interference (RNAi) profiling [7], next-generation sequence analysis [8] and chromatin interaction analysis [9,10], boosted the amount of available data. While application of the novel wet laboratory techniques has spread all over the world, and their principles are generally taught and known for researchers, knowledge on their bioinformatics background and analysis options has not kept pace. This situation is even more challenging, as not only the amount of available data is growing every week but so does the number of published bioinformatics databases and stand-alone and web applications. For example, as of January 2013, BioCatalogue (http://www.biocatalogue.org/services; [9]), a life-science web server

Corresponding author. Tamas Korcsmaros, Department of Genetics, Eotvos Lorand University, H-1117 Budapest, Pázmány s. 1/C, Hungary. Tel.: +36302686590; E-mail: korcsmaros@gmail.com

Tamas Korcsmaros is a bioinformatician working with signaling networks. He is teaching bioinformatics and systems biology in several university courses and international trainings. He is also involved in talent support initiatives at high school and university level.

Zsuzsanna A. Dunai is a tumor cell biologist. Her research focuses on programmed cell death mechanisms in tumor cells. She is involved in teaching undergraduate students on medical preparatory courses.

Tibor Vellai is an experimental geneticist and head of the Department of Genetics at Eötvös University, Budapest, Hungary. His research work focuses on signaling pathways and cross-talks important in development and age-related processes. He teaches molecular genetics and coordinate practices on genetics.

Peter Csermely is a network scientist with a background of biochemistry, who taught network science and research methods to more than 3000 PhD students. He established a talent support network involving more than 200 000 people, served as a member of the EU Rocard—committee on science education and was the educational advisor of the President of Hungary.

registry, contains 2347 web services; Pathguide (http://pathguide.org; [11]) lists more than 325 pathway-related resources; and the related annual database issues of Nucleic Acids Research have already published more than 1500 articles describing database resources (http://www.oxfordjournals.org/nar/database/a; for our collection of more than 300 biological network–related resources and tools, see http://www.linkgroup.hu/links.php). Although the number of resources and tools is enormous, and we should not expect any researcher to know all of them, in most of the cases, the logic and function of the resources are very similar. Therefore, it is important to teach some key, well-selected and useful resources and tools in undergraduate and postgraduate courses. These resources and tools can be selected based on the teacher's own research or hands-on experience. Resources that can be easily used, linked to each other and generally maintained could be the best to present in a bioinformatics course. In these courses, students and researchers learn the functions and principles of selected bioinformatics methods, which can be applied later, when searching, using or developing a new application. Thus, well-selected and well-described examples are the key assets of a related bioinformatics course.

Teaching bioinformatics methods to scientists having no IT background is a challenging task. Besides processing and analyzing data, a further difficulty arises in the visualization and interpretation of results. In consequence of the conservative nature of high school and university education systems, there is a difference in the number and content of bioinformatics courses compared with the need of bioinformatics methods in current life science research [12,13]. Accordingly, a larger number of teaching courses from the level of high school students up to PhD students are needed with specific content [14,15]. We collected a few examples of university teaching courses, online materials and curricula (Table 1). We found four different types of course materials: (1) self-teaching and e-learning materials, (2) general introductory-type presentations showing the theoretical background of various bioinformatics fields (such as genome analysis, protein structure studies), (3) specific bioinformatics course materials that teach programming to biologists or present a practical guide to use advanced tools, (4) materials specifically designed for bioinformatics teachers (e.g. the Bioinformatics Activity Bank that promotes bioinformatics teaching at high schools (http://teachingbioinformatics.fandm.edu). Postgraduate training courses have been developed faster than university courses, and by providing general or specific learning possibilities for (wet laboratory) researchers, these training courses give experience in the use of variable bioinformatics tools [16–18]. For a collection of such courses, trainers and organizations hosting training programs, see the web page of the Bioinformatics Training Network (http://www.biotnet.org) [19]. Despite these advances, the training of bioinformatics teachers is lagging much behind the needs, and the number of training courses for teachers is much less than that designed for researchers [20]. The lack of institutional commitment to bring bioinformatics teachers and trainers together in one department also makes university course organization more complicated [18].

In the current article, we focus only on teaching bioinformatics at the university level. Our approach covers a methodology of teaching bioinformatics for biology students at Master (MSc) and PhD levels. In addition, the presented teaching approach is suitable for students with other backgrounds (e.g. physics, informatics or chemistry) having sufficient knowledge in molecular biology. Here, we present methods and strategies for teaching bioinformatics, using the topic of signaling network as an example. We aim at highlighting the advantages of teaching the bioinformatics of signaling networks, as a way to develop important competences (e.g. working in teams, presentation skills or problem-solving capabilities) that can be used in further tasks and courses and, later, applied in research work. Based on a detailed methodology and didactical and technical description of a university course one of the authors (T.K.) established in 2010, we aim at sharing our experiences with the wider audience and help teachers to set up similar teaching programs.

Goals of a university course

The major difference between a university-level bioinformatics course and general bioinformatics training is that the former is ahead of the demands. At many bioinformatics trainings, participants have already faced some bioinformatics challenges or worked with *in silico* tools. In contrast, most university students have never faced a scientific problem that requires bioinformatics tools. Thus, we definitely think that making students to be engaged for bioinformatics has a key importance in the success of a university course.

Table 1: Some bioinformatics teaching courses and resources

Title	Brief description	URL
Teach Yourself Bioinformatics on the Web	Collection of links to online tutorials, online courses, essays, book chapters, course syllabi, glossaries, bibliographies of key papers, etc.	http://www.med.nyu.edu/rcr/rcr/btr/
Distant education and e-learning courses on Computational Systems Biology	Courses from the German/Russian Virtual Network of Bioinformatics at Novosibirsk State University	http://www.bionet.nsc.ru/virtual_network/pages/education_cib_nsu.html
Bioinformatics course materials	Materials from the bioinformatics courses of the Biotechnology Center of the Technische Universität Dresden	http://www.biotec.tu-dresden.de/research/schroeder/teaching.html
Bioinformatics course resources	Basic teaching techniques for teaching assistants in bioinformatics from the University of Santa Cruz	http://courses.soe.ucsc.edu/courses/bme200
Bioinformatics and Functional Genomics teaching curriculum	A bioinformatics teaching curriculum: PowerPoint slides, PDFs, audio files, etc., for an entire course taught at the Johns Hopkins School of Medicine	http://www.bioinfbook.org/
An open-source framework for teaching bioinformatics	Presentation about teaching bioinformatics	http://www.slideshare.net/bosc/an-open-source-framework-for-teaching-bioinformatics#btnNext
Bioinformatics activity bank	Bioinformatics teaching resource for high school students and teachers	http://teachingbioinformatics.fandm.edu/
Open source teaching materials for Bioinformatics	Bioinformatics teaching materials (including textbook, lecture materials, concept tests, homework problems, exam problems, answers, etc.)	http://thinking.bioinformatics.ucla.edu/teaching/
Teaching Bioinformatics and Neuroinformatics by using free web-based tools	Teaching materials such as a detailed student/instructor's manual, PowerPoint slides, sample exams, and links to further resources	https://mdcune.psych.ucla.edu/modules/bioinformatics

Therefore, the teacher of a university course defines the goals of the course collectively with the students at the beginning of the first practice. In addition, during the practical part of a course, the teacher could act as a partner of the students to help answering to questions like 'what' 'why' and 'how' [21]. The teacher assesses students' expectations about the course, and highlights the realistic goals and limits. For example, the purpose of the course we present here is to introduce and familiarize students with computer-aided solutions of molecular biology problems. During the course, students acquire skills necessary for the solution of such problems (e.g. to learn how to use a Linux operating system, to search relevant literature and data sources and to work with novel *in silico* applications). Owing to the frequently changing number and quality of the available bioinformatics tools and databases, the course does not focus on detailed demonstration of tools and resources, or on the mathematical background of applications. We put emphasis on developing competences to handle data sources and tools of different kinds. In addition, much effort is taken to improve the capability of critical assessment of bioinformatics resources. As a result, students learn to formulate adequate questions and to find answers individually using the available resources.

Scheme of a course

In this section, we summarize a university-level bioinformatics course (lecture and hands-on practice) that is compulsory in the majority of biology MSc, and open for PhD students or other participants with adequate background in molecular biology. In the course, we introduce different bioinformatics approaches (e.g. structural bioinformatics, network science, database management), but use signaling network and signaling proteins as a specific example to focus. Each week, a lecture is held followed by a 120 min hands-on practice. During the lecture, the necessary theoretic background is discussed. On the first week, we ask each student to select a signaling protein and use it along the semester in all practices. Each semester, we list 8–10 different proteins for the selection. We have previously tested these proteins and list only those that provide sufficient information for the proceeding lessons (e.g. present in all the resources we will use in the semester). A protein can be allocated to several persons in a larger practice group, but students sitting next to each other should work with different proteins. In this way, they can only

help each other in the methodology, and not in the exact answers.

Optionally, each practice lesson starts with a short test from the topics of the related lecture, followed by a recall of necessary fundamentals. By the teacher, the practice is continued with a short presentation of resources and methods with an example protein. Detailed description, usage and limitations are discussed during the presentation. Next, students have to repeat the same procedure with their own signaling protein received at the first week. Exercises are handed over in an electronic format, and solutions are obtained on the teacher workstations of the computer laboratory. During the hands-on practice, students can get support from the teacher, and there is a chance for open discussions with other students. In this way, students learn how to solve a problem alone, and also how to ask specific questions, and work with the assistance of professional and non-professional colleagues. Each practice is concluded with a summary.

At the end of the course, students are asked to form a team of three to four people. This teamwork includes writing of a short essay on a comprehensive task that involves the subject of multiple lessons and all the signaling proteins distributed to team members at the first lesson. Thus, students are encouraged to use methods/resources they have learnt in the practices, and have to work together in teams to summarize the results. This task should be completed after the course, at home with the students' own computers. In the evaluation of the final works, knowledge of the topics and their application is checked as well as the quality of the short essay. To complete the course, students also have to pass midterm and semester-closing written tests, which check the level of knowledge gained from theoretic lectures.

The topic: signaling networks

Including signaling networks in any bioinformatics course is advantageous for several reasons. Teaching about signaling networks allows presenting resources and methods of different biological disciplines. For example, students get introduced to bioinformatics repositories and approaches regarding (a) protein structures and domains, (b) genetic interaction and model organisms to test them, (c) protein-binding and protein–protein interaction (PPI) networks and (d) signaling pathways. While teaching bioinformatics approaches, we can build on already gained knowledge, as well as point out important connections between classical disciplines (e.g. biochemistry, genetics or cell biology). Meanwhile, students can learn multi-scale thinking when they use data from a variety of levels (i.e. molecular, cellular and physiological). Finally, research-oriented and hands-on teaching of bioinformatics offers problem-solving case studies to answer real biological questions about signaling networks.

Four sample lessons

In this section, we describe four suggested lessons on signaling networks as parts of the university course we mentioned in the previous sections (Table 2): (1) examination of protein sequence, structure and domain composition; (2) genetic interaction and epistasis analysis; (3) construction of PPI networks; and (4) signaling pathways and their integration to networks. The four topics can also be taught in a specific training course. We use signaling networks as an example to show the aims and approaches of these lessons; teachers can freely use these examples as guidelines to design similar lessons with other systems (e.g. metabolic networks, cell cycle or cell death processes). Web resources and tools of the four lessons are summarized in Table 3. Here, we present the practical lessons in detail and leave out the precise topics covered by the preceding lectures where we introduced the necessary theoretical background for the students. During all practical lessons of the course, students work with signaling proteins distributed at the beginning of the course. We suggest the following human proteins to work with, because of the amount and details of available data: Epidermal growth factor receptor (EGFR), Insulin growth factor receptor (IGFR), Kirsten rat sarcoma GTP-ase (RASK), JAK1 (Janus kinase 1), Extracellular signal-regulated kinase 1 (ERK1), Glycogen synthase kinase-3 beta (GSK3-β), mammalian target of rapamycin (mTOR), NOTCH1 and SMAD2. Note the differences among these proteins: most of the proteins are functioning in different signaling pathways, and their malfunctions have known medical relevance. In addition, there are key differences in the function of the proteins: some of them are cell surface receptors (EGFR, IGFR and NOTCH1), while others have kinase activity (EGFR, IGFR, JAK1, ERK1, GSK3-β and mTOR) or have no enzymatic functions (NOTCH1 and SMAD2) [42–45].

Table 2: Suggested syllabus of four sample lessons about the bioinformatics of signaling networks, showing titles and topics and listing previous knowledge that the given lesson is built on

Lesson title	Topics covered	Background needed
Examination of a protein sequence, structure and function	Basic bioinformatics data on proteins UniProt resource Gene Ontology Protein domain resources	Biochemistry and protein structures Cell biology
Genetic interactions, epistasis analysis and pathways	Wormbase and Flybase resources Basics of epistasis analysis Functional genomics and orthology Orthology-based predictions	Molecular biology Genetics
Constructing a protein–protein interaction network	Small- and large-scale approaches to study protein–protein interactions (PPIs) Prediction of PPIs based on structural data High-throughput experiments to detect PPIs PPI databases (MINT, BioGRID, STRING) Optional: Cytoscape and network analysis	Molecular biology Cell biology Optional: biophysics
Signaling pathways and networks	Literature curation and search tools Pathway resources Definition of pathways and cross-talks Multi-layered signaling networks (SignaLink) Optional: regulatory networks	Cell biology Molecular biology Genetics

Table 3: Resources and tools suggested for teaching the bioinformatics of signaling networks

Lesson title	Name of the resource or tool	URL	Reference
Examination of a protein sequence, structure and function	UniProt	http://uniprot.org	[22]
	InterPro	http://www.ebi.ac.uk/interpro	[23]
	PFAM	http://pfam.sanger.ac.uk	[24]
	Gene Ontology	http://www.geneontology.org	[25]
Genetic interactions, epistasis analysis and pathways	Wormbase	http://wormbase.org	[26]
	Flybase	http://flybase.org	[27]
	InParanoid	http://inparanoid.sbc.su.se	[28]
Constructing a protein–protein interaction network	DOMINE	http://domine.utdallas.edu	[29]
	ELM	http://elm.eu.org	[30]
	Phosphosite	http://www.phosphosite.org	[31]
	MINT	http://mint.bio.uniroma2.it	[32]
	STRING	http://string-db.org	[33]
	BioGRID	http://thebiogrid.org	[34]
	Cytoscape	http://cytoscape.org	[35]
Signaling pathways and networks	iHOP	http://www.ihop-net.org	[36]
	Chilibot	http://www.chilibot.net	[37]
	KEGG	http://www.kegg.jp	[38]
	Reactome	http://www.reactome.org	[39]
	SignaLink	http://signalink.org	[40,41]

Lesson 1. Examination of protein sequence, structure and function

At the beginning, it is important to allow the students to explore the major properties of the signaling protein they got. Using the UniProt Knowledgebase (http://uniprot.org; [22]), students can easily find the main features of the protein (e.g. synonyms, brief description about its function, subcellular localization and sequence data) and direct hyperlinks to other specific databases. Further information about the assigned protein can be found by using links to genome, 3D structure, proteomic, phylogenomic and gene expression databases. Moreover, information on protein domain

composition can be reached by the application of hyperlinks, for example, to InterPro (http://www.ebi.ac.uk/interpro; [23]) or PFAM resources (http://pfam.sanger.ac.uk; [24]). Information on the domain composition will be used later, in the third lesson. UniProt also lists and links Gene Ontology data (GO; http://www.geneontology.org; [25]) containing cellular localization, molecular function and biological process annotations about the given protein.

During this lesson, students are asked to gather basic information about the proteins and use a few specific external databases. Specific questions from the teacher could guide students on this exploration. Finally, students should write a few sentences with their own words about the protein examined.

This lesson could serve as a basic introduction to biological databases, as well as to specific web resources such as GO and PFAM. While using the different resources, students learn that each database uses different IDs (or aliases) and contains both complementary and sometimes equivalent information. A further comparative practice could be to ask students to assess and evaluate different databases developed for the same reason (like PFAM, InterPro and SMART). By examining two to three similar databases, students will learn differences in the content and utility of databases, and may encounter contradictory information.

Lesson 2. Genetic interactions, epistasis analysis and genetic pathways

Protein-coding genes are often functionally linked. This functional connection (i.e. genetic interaction) can more easily be examined in model organisms, such as the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster.* Wormbase (http://wormbase.org; [26]) and Flybase (http://flybase.org; [27]) are two web resources that contain information on worm and fly genes, respectively. Using phylogenomic services such as InParanoid (http://inparanoid.sbc.su.se; [28]) at the previously introduced UniProt site, students can find human orthologs of any worm and fly gene, if it exists (orthologs are homologous sequences with a shared evolutionary path, separated by a speciation event). Next, students can use Wormbase or Flybase to search for the orthologous gene, while the teacher could present the main features of these species-specific resources. Students can examine the genomic location, expression pattern and genetic interactions of the gene of interest.

Before students examine the list of known genetic interactions, the teacher could explain the experimental detection of genetic interactions and how genetic interactions can help the reconstruction of pathways. To identify the upstream and downstream components in a signaling pathway (i.e. to define which gene is regulated by another), epistatic analysis should be performed (Figure 1). Epistatic analysis determines the sign (effect) of a genetic interaction: that is, either activation or inhibition. For example, in case of an inhibitory interaction between the genes *a* and *b*, supposing that the phenotypes of two single mutants are different (X and Y), while the phenotype of double mutants shows an epistatic relation, the phenotype of the double mutant is identical with that of the 'downstream' component. If the double mutant phenotype is X, then the gene *a* inhibits gene *b* (if the phenotype of the double mutant is different from both single mutants, then the two genes function in two separate pathways). Note that in case of a metabolic (also called biochemical) pathway, where genes encode functionally connected enzymes, the phenotype of double mutants is the same as the phenotype of the 'upstream' component. After discussing a few examples of different genetic interactions, students can explore the genetic interactions of their genes and draw a genetic/signaling pathway in their notepads.

This lesson allows the teacher to introduce phylogenomic resources, the concept of orthology, functional genomics and the genetic background of signaling pathway reconstruction. In addition, the predictive power of orthology can also be presented. For example, interactions can be predicted based on orthology. These interactions are called 'interologs', and they represent two proteins that possibly interact, if their orthologs interact in another organism [46]. Similarly, 'regulogs' are orthology-based predictions of regulatory interactions between a protein (i.e. a transcription factor) and a corresponding DNA sequence (i.e. a transcription factor–binding site). Recently, 'phenologs' were used as predictors of disease-associated genes [47] and 'signalogs' as orthology-based predictions of signaling pathway memberships in model organisms and in humans [48].

Lesson 3. Constructing a PPI network

After introducing and discussing general protein structural features and functional interactions, students can learn how to build a PPI network with

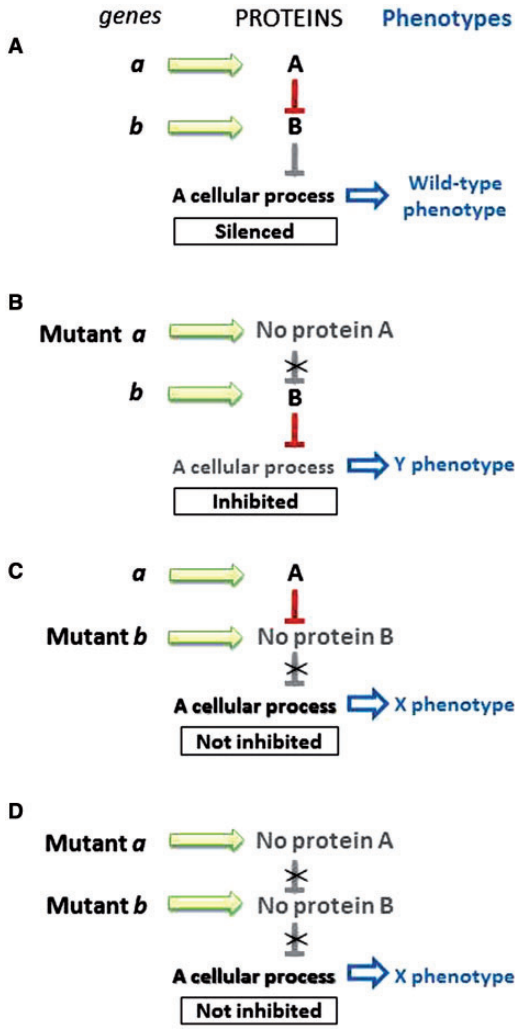

Figure 1: Epistasis analysis. Proteins A and B are two arbitrary signaling components acting in one pathway, and are encoded by genes *a* and *b*, respectively. (**A**) In a wild-type system, protein A inhibits protein B but not completely. Thus, protein B can only silence a regulated cellular process. (**B**) If gene *a* is mutated, there is no protein A, and the lack of repression hyperactivates protein B (resulting in the inhibition of the regulated cellular process). Hyperactive protein B shows a phenotype Y. (**C**) If gene *b* is mutated, there is no protein B. Thus, the regulated cellular process is not inhibited, causing a phenotype X. (**D**) If both genes are mutated, both proteins are missing, and lack of B, as previously, results in phenotype X. Observing such connection between genes shows that protein A has an upstream, while protein B has a downstream, position in the pathway, and protein B has an epistatic effect on protein A. T-arrows with red (darker) color represent inhibition, while T-arrows in light gray show reduced or missing activity. Crosses on arrows show dysfunctional processes.

signaling proteins. Teachers could introduce two approaches to construct a PPI network: small-scale building and high-throughput detection [49]. Both approaches require different bioinformatics resources, and the structural context of an interaction should also be discussed.

While studying small-scale approaches, students learn how to examine the structural properties of a potentially interacting protein pair. For example, students can investigate an interolog prediction from the previous lesson. An interolog contains two (human) proteins, the orthologs of which were found interacting in a model organism. We suggest that one of the potential interacting protein pairs should be the student's protein distributed at the first lesson. To confirm whether the two proteins do indeed interact in humans, students can apply the domain composition of the proteins from PFAM (introduced on the first lesson) and DOMINE resource. DOMINE is a database of possible domain–domain interactions (http://domine.utdallas.edu; [29]). A possible domain–domain interaction between the domains of the proteins can point to a possible PPI (Figure 2A). A PPI could indicate an adaptor or anchoring function between the two proteins that could be important in the localization of the protein of interest. Note that besides interaction validation, investigating domain–domain interactions could point out structural details also for known PPIs. Signaling proteins often participate in enzymatic reactions as enzymes or substrates. An approach to discover such directed PPIs involves linear motifs of proteins. Linear motifs are short conservative components of proteins, and provide a low-affinity interaction interface [30]. ELM web server contains linear motifs (http://elm.eu.org; [30]) and allows students to identify enzymes that can modify or regulate the queried protein of interest (Figure 2B). Finally, if the signaling protein of interest has enzymatic activity, students could analyze the potential substrates with the Phosphosite web portal (http://www.phosphosite.org; [31]). Phosphosite contains enzyme data for kinases, phosphatases, peptidases and several other enzyme types.

PPIs can also be detected in a large-scale high-throughput manner, which was applied to many signaling proteins. Before analyzing large-scale studies and their databases, first, the teacher could briefly describe major experimental detection methods, such as yeast two-hybrid assays, fluorescence microscopy–based techniques and co-immunoprecipitation

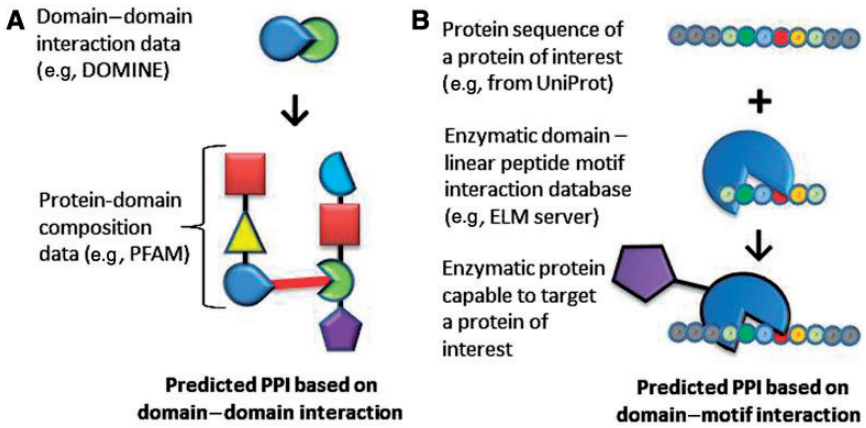

Figure 2: Protein–protein interaction prediction. PPIs can be predicted based on (**A**) domain–domain or (**B**) domain–motif interactions. See the main text for details.

followed by mass spectrometry [50]. Next, the teacher could present PPI databases that contain large collections of PPIs detected by high-throughput techniques. For example, by introducing the functions of MINT (http://mint.bio.uniroma2.it; [32]), STRING (http://string-db.org; [33]) and BioGRID (http://thebiogrid.org; [34]) resources, the teacher could point out the advantages and disadvantages of the resources. Here, critical discussions about the detection methods, the compilation of the databases and the applied confidence scores of the PPIs can improve and contribute to the knowledge of the students and facilitate their future individual work.

This lesson allows the teacher to explain and compare experimental detection methods and *in silico* predictions and introduce small-scale and large-scale approaches to build PPI networks. As an extra module, this lesson could be enriched by or complemented with a separate part on network visualization and basic network analysis. For this purpose, we suggest using Cytoscape, a widely used network analysis tool (http://cytoscape.org; [35]). A recent protocol [51], which could be integrated in this course, shows a 2-h-long practice to interactively visualize biological networks and protein structures for gaining insight into biological processes. Network analysis methods and resources were recently summarized in our comprehensive review [52].

Lesson 4. Signaling pathways and networks

Signaling pathway resources also contain interactions of signaling proteins. Most of the interactions in signaling pathway resources have been detected by small-scale experimental methods, and were curated manually by experts or developers of a resource. Therefore, before introducing pathway resources, teachers could talk about the challenges and drawbacks of manual curation of the literature. Also, tools for finding interactions in scientific literature can be presented. Examples of these advanced text-mining tools are iHOP (http://ihop-net.org; [36]) and Chilibot (http://www.chilibot.net; [37]), which are two user-friendly tools that help researchers and users without sufficient background to interpret an interaction or a sentence about an interaction. KEGG (http://www.kegg.jp; [38]) is one of the most well-known pathway resource, although it contains less detailed information on interactions, and its curation protocol is not standardized. Reactome (http://www.reactome.org; [39]) is a pathway resource with a well-documented and standardized curation protocol, and it contains molecular details on each interaction. After introducing pathway resources, students examine the pathway membership annotations and signaling reactions of their protein of interest. In addition, students are asked to compare the number of interactors, pathway annotations and interaction details found in the two resources.

Signaling pathways are not linear but heavily interlinked by cross-talks [53] (Figure 3). Before examining cross-talks and proteins that form these inter-pathway connections, the teacher could explain the importance of standardized curation protocols as well as discuss the different pathway definitions. Currently, several different pathway definitions exist: (1) canonical (e.g. MAPK), (2) function-based (e.g. inflammation), (3) inferred (e.g. from gene expression data), (4) cellular process regulating (e.g. autophagy induction), (5) organ related (e.g. vulva development), (6) disease related (e.g. list of

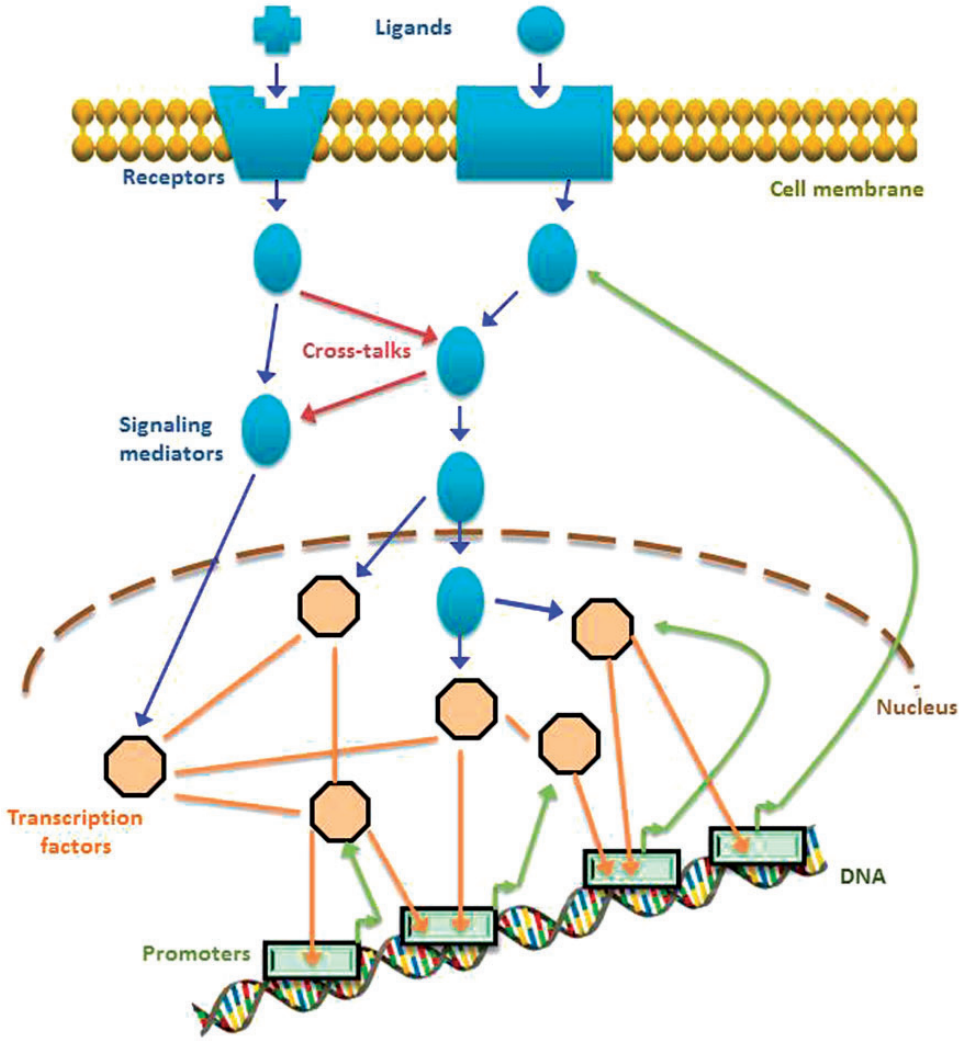

Figure 3: Signaling pathways. Signaling pathways are formed by ligands, receptors, signaling mediators (e.g. kinases) and transcription factors. Interactions between signaling components can be analyzed from genetic (functional) and biochemist (structural) point of view: with epistatic analysis, the order of the components (i.e. upstream and downstream components) can be determined (see Lesson 2 and Figure 1 for details), while integration of protein domain and motif data can uncover underlying enzymatic and binding effect. Signaling pathways often interact with each other and form cross-talks. Terminal transcription factors bind to the promoters of target genes (arrows between octagons and squares) to regulate their expression. Regulatory circuits are shown with arrows pointing up-wards to up-stream components.

connected proteins affected by a mutation in breast cancer, Alzheimer's disease, etc.) and (7) drug related (e.g. pharmacologically affected list of connected proteins). It is essential to teach students to understand that cross-talks can only be identified between uniformly defined pathways. Cross-talks can be examined and identified with the SignaLink resource (http://signalink.org; [40,41,54]), where canonical pathways were uniformly curated. In addition, SignaLink allows the inclusion of transcriptional and post-transcriptional regulatory mechanisms to the analysis. Thus, students can also examine the transcription factors and microRNA that can regulate their signaling protein. With the direct hyperlinks to UniProt embedded in SignaLink, students could discover the properties and functions of the interactors or regulators of their protein of interest. Students are asked to collect information on the pathway functions of their protein of interest and also of those interactors/regulators that are members of another pathway. Thus, with SignaLink, students could list possible cross-talks between the pathway of

their protein and other pathways through PPIs or transcriptional or post-transcription regulations [40,41,54].

This lesson allows the teacher to introduce pathway resources, concepts and tools about literature curation as well as to extend pathways to cross-talking networks. The lesson could be completed with further work on regulatory networks. Regulatory resources and web tools, such as the ENCODE project and JASPAR, can be introduced to broaden the knowledge of the students [54,55].

Evaluation and details of the original university course

The presented course, we initiated in 2010, replaced a former lecture-based bioinformatics course. In the former course, only sequence-based techniques, phylogenetic tools and study of domain structures were taught. Our aim was to provide a broader picture of the applicability of bioinformatics resources and tools. The novelty of the developed course is that we applied a multi-disciplinary approach with integrated practical lessons. Focus on a specific signaling protein allowed the students to perform research activities with their protein of interest. In addition, as students use the same signaling protein along the semester in all practices, it provides coherence among the practical lessons. The course material (i.e. lecture slides, practical task lists and final test questions) is mostly in Hungarian, owing to the educational requirements in Hungary, but specific subparts are available in English on request. In the past 3 years, we taught this course to 100–120 students in each autumn semester. Twenty-five to thirty-five percent of the students were BSc students in biology, 60–70% of the students were first-year MSc students in biology and 5% were MSc students in biophysics. We note that as a pilot event, the presented four lessons focusing on signaling networks were also separately taught for eight graduate students and four MSc students. Each semester, after the last lesson, we asked the students to fill out a questionnaire with 20 questions to grade each lecture and practical lesson. We also asked feedbacks on positive and negative experiences in the course (e.g. tools that were extremely useful and topics that were difficult to understand, respectively). Ninety percent of the students voted that they would recommend the course for other students. We have received numerous suggestions to improve the course after the first semester. We have implemented most of these suggestions in the proceeding courses that resulted in substantially less suggestions and higher satisfactory index. We also asked the students to evaluate the development of their own skills (e.g. digital skills, communication). To illustrate the development of their skills, we presented a comparison about a task from their first lesson and a task from one of their final lessons. We showed the students how their problem-solving and creative-thinking capabilities had been improved during the course: the students were able to manage a more complex task with less help in a shorter time frame in the final lessons.

DISCUSSION

We presented a university bioinformatics course, where undergraduate students get insight into the everyday challenges of a researcher, while post-graduate students could learn novel approaches and techniques that will be important in their future research work. Thus, both undergraduate and graduate students can participate in this course. However, graduate students have more practical questions and research experience that necessitates more complex tasks to be solved and more theoretical preparations from the teacher. Through the prepared sample exercises, students learn the theoretical backgrounds of protein structures, genetic interactions, PPIs and signaling pathways from a bioinformatics point of view. As probably most of these topics have already been covered in other university courses, teachers of this course can focus on the bioinformatics relevance of the topics. The research-like practices introduce the UniProt portal and several different databases containing data on protein domains, genetic and genomic information of model organisms, and PPI and pathway datasets. Useful web servers and search engines are also presented, such as phylogenomic, linear motif and literature curation search tools.

Nevertheless, there are further topics to be included in bioinformatics studies on signaling networks, beyond the presented four sample lessons. For example, the integration of signaling pathway data with tissue-, development-, stage- or diseased state-specific expression datasets can point out the dynamics and medically important aspects of signaling networks [56]. Another interesting multi-disciplinary outlook option is to show the applicability

of bioinformatics and network tools to identify key components of signaling networks important in evolution or in pharmacological targeting of the given signaling system. Based on the teacher's knowledge and experience in related areas of bioinformatics or molecular genetics, the presented course can be extended with other topics such as introducing the genomic information or genetic map of a gene of interest as well as advanced tools like hierarchical clustering of expression datasets. We also note that the presented exercises with signaling networks could point out useful strategies for teachers to use them as guidelines to design similar lessons with other systems (e.g. metabolic networks, immunology). We anticipate that the presented approach can be applied to other fields of bioinformatics, allowing the teacher to decide the topic to focus on.

In addition to the bioinformatics curricula, in the presented course, students also learn how to solve a problem, ask relevant questions, evaluate the results and select, rank and use the most efficient bioinformatics methodology. Thus, the structure and topic of the course has advantageous indirect effects on the competences of the students. The course facilitates the development and improvement of several skills important in future research work: communication and digital skills, collaborative work and problem-solving capabilities. Here, we briefly describe how teaching bioinformatics in the presented way can improve these skills (Figure 4).

For a successful scientist, it is highly important to have good 'communication skills'. Therefore, most of the university courses should help in the development and improvement of communication skills. As lectures can mostly be characterized by one-way communication, the practice lessons are the right opportunity for the teacher to encourage active participation of the students. Accordingly, in the presented course, we try to build a collaborative connection with the students, and work with them as partners to solve a scientific problem [21]. The course is structured in a way that enforces active communication and cooperation with the students during the practical occasions and during the teamwork to be done after the course. We also welcome questions and comments from the students throughout and after the practices. Students regularly have to perform tasks that can be solved together with their neighbors (i.e. students sitting directly next to them) working with a different protein. These tasks contribute to a high-level improvement of problem-solving, cooperative and social competences.

A bioinformatics course inherently contains *in silico* data and computational applications to be learned. Moreover, it gives an option for the teacher to improve not only specific bioinformatics but also general 'digital skills' of the students. During the course, students gather different data on their own protein assigned at the beginning of the course (such as its genomics, structural and systems-level properties, etc.). Students learn to be aware of the reliability of the available information or how the confidence of the data could be evaluated. Critical assessment of the available resources and tools is a key point in the practice lessons. Students acquire skills to analyze, judge and distinguish experimentally verified and *in silico* predicted datasets.

The course structure allows the formation of two different types of 'collaborative skills': teamwork in randomly assembled and in self-organized groups. In some of the hands-on practices, groups are selected based on the members working on the same protein. Students, therefore, have no direct impact on putting the groups together, as the proteins of the students have previously been distributed randomly. In this way, for the time of the practice, students are expected to adapt to other members of the group and to cooperate during group work. The better their cooperation is, the faster they will complete the exercise. In addition, an efficient cooperation could help the students to find better solutions. On the contrary, at the end of the semester, students are asked to form self-organized teams to perform a comprehensive project together. In this case, the team members are working together for days, with several internal meetings. For a better efficiency, students should not compete but have to settle a joint working plan, define the main solution steps and distribute the tasks of the project.

Practices are structured to promote effective 'problem-solving capabilities'. In the first part of a hands-on practice, students get acquainted with the use of databases, major software and web applications of the given topic. Then, in the second part of the practice, students can apply the introduced resources and tools to solve a problem. During the problem-solving part, students get a continuous support. Promoting the problem-solving capabilities is strengthened by the course structure: the tasks to be done are getting gradually more and more

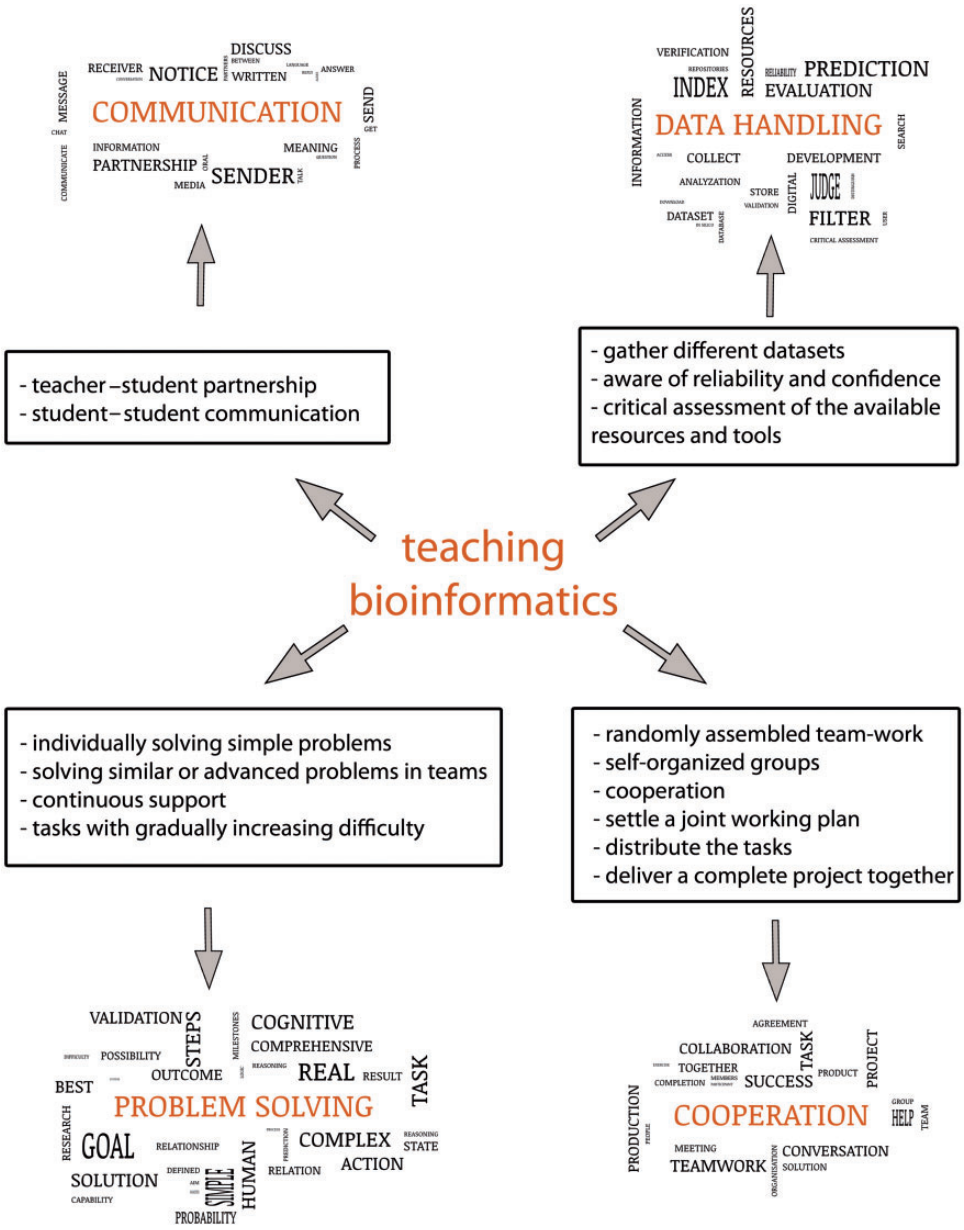

Figure 4: Competences and skills developed during the bioinformatics course by the suggested teaching approach (major points are listed in boxes). Word clouds contain the words associated to the presented four skills. The word collection was done by the authors.

difficult with decreasing previous support. In other words, at the beginning of the course, the whole solution procedure is described for the students in advance, while in the last practices of the semester, only the milestones are identified for the students. By this time, students are capable to solve complex problems with reduced support. The successful completion of a final project to be done at home after the university course indicates that the students acquired substantial problem-solving capabilities.

Outlook and further applications: e-learning and talent support

We presented here a classroom course with lectures and hands-on practices. However, the curriculum can be re-formulated as an e-learning material or applied as a part of a training course. We believe that the presented material or a part of it is suitable for several purposes besides its original university application. It can be used as a training material of a few-day training course held for researchers or at high school level [57], after a proper adaptation of the contents.

Some students are more talented than others. It is the responsibility and role of the teacher to recognize the level of skills of talented students and pay a special attention for their adequate development. In several cases, this involves extra work or additional more complex versions of the generally applied problem [58]. The presented course and teaching approach provide the teachers with the possibility of specific support provision for talented students owing to its mosaic setup. Talented students—mainly during hands-on practices—can be grouped together to make a faster progress with additional more complex problems provided to them. These additional complex questions may require both a deeper understanding and more investigation from the students. The level of complexity can be matched to the needs and abilities of the individual talent. As we discussed before, the teacher supports and facilitates the learning process, while treating the students as equivalent partners in their communications. This becomes an especially important asset when working together with talented students of the class. The e-learning material mentioned in the previous paragraph may also help talented students, as in e-learning, the student may set the speed of learning in a personalized manner.

CONCLUDING REMARKS

In this publication, we presented the aims, methods, advantages and experiences of a bioinformatics university course we developed for undergraduate and PhD students. In four sample lessons, we illustrated how the topic of signaling networks can be used to teach students to use bioinformatics resources and tools as well to facilitate the multi-disciplinary thinking of the students. With the examination of one signaling protein in various levels, students learn how to integrate knowledge of biochemistry, cell biology, genetics and network sciences with bioinformatics approaches and gain more information for more complex questions. During the course, competences of the students are improved by hands-on practices and discussions about the scientific background and limitations of the bioinformatics resources and tools. The research-like tasks that should be solved individually and in teams help the students to develop senses and skills essential to elucidate more complex questions in their future work. We hope that the presented methodological approach and collected resources will help other bioinformatics teachers to train students to become successful young scientists capable to perform complex research work.

Key Points

- Teaching the bioinformatics of signaling networks necessitates the integration of background topics about biochemistry, genetics, molecular and cell biology, while allows introducing the general features of bioinformatics resources and tools.
- The presented research-like real-life problem-solving tasks can improve key competences of the students, such as teamworking, creative and critical thinking and problem solving.
- The methodological approach, collected resources and teaching experiences of the presented university course can help bioinformatics teachers to prepare their own (course/training) material.
- The mosaic setup and multi-disciplinary nature of the presented course allow the specific support of talented students, who need more complex tasks and have grossly differing time allocation to solve them.

Acknowledgement

We are grateful for the discussions and advices of Eszter Ari, David Fazekas, Dezso Modos and Denes Turei. The structure of the presented course was based on a full-semester course of Eszter Ari and T.K. We thank the technical help of Janos Kubisch in the preparation of the manuscript.

FUNDING

Work of the authors was supported by the European Union and the European Social Fund [TAMOP-4.2.2/B-10/1-2010-0013] and the Hungarian Scientific Research Fund [OTKA K83314, NK78012]. T.K. is a grantee of the Janos Bolyai Scholarship.

References

1. Hegde P, Qi R, Gaspard R, *et al*. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* 2001;**61**:7792–7797.
2. Whitfield ML, Sherlock G, Saldanha AJ, *et al*. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002;**13**: 1977–2000.
3. Bush WS, Moore JH. Chapter 11: genome-wide association studies. *PLoS Comput Biol* 2012;**8**:e1002822.
4. Gerber MM, Hampel H, Schulz NP, *et al*. Evaluation of allele-specific somatic changes of genome-wide association study susceptibility alleles in human colorectal cancers. *PLoS One* 2012;**7**:e37672.
5. Wei P, Tang H, Li D. Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. *PLoS One* 2012;**7**:e46887.

6. Zhang M, Liang L, Xu M, *et al*. Pathway analysis for genome-wide association study of basal cell carcinoma of the skin. *PLoS One* 2011;**6**:e22760.
7. Kittler R, Pelletier L, Heninger AK, *et al*. Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nat Cell Biol* 2007;**9**:1401–1412.
8. Ding D, Lou X, Hua D, *et al*. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PloS Genet* 2012;**8**:e1003065.
9. Bhagat J, Tanoh F, Nzuobontane E, *et al*. BioCatalogue: a universal catalogue of web services for the life sciences. *Nucleic Acids Res* 2010;**38**:W689–W694.
10. Sandhu KS, Li G, Poh HM, *et al*. Large-scale functional organization of long-range chromatin interaction networks. *Cell Rep* 2012;**2**:1207–1219.
11. Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. *Nucleic Acids Res* 2006;**34**:D504–D506.
12. Cummings MP, Temple GG. Broader incorporation of bioinformatics in education: Opportunities and challenges. *Brief Bioinformatics* 2010;**11**:537–543.
13. Pevzner P, Shamir R. Computing has changed biology–biology education must catch up. *Science* 2009;**325**: 541–542.
14. Counsell D. A review of bioinformatics education in the UK. *Brief Bioinformatics* 2003;**4**:7–21.
15. Form D, Lewitter F. Ten simple rules for teaching bioinformatics at the high school level. *PLoS Comput Biol* 2011;**7**: e1002243.
16. Fernandes PL. The GTPB training programme in Portugal. *Brief Bioinformatics* 2010;**11**:626–634.
17. Wright VA, Vaughan BW, Laurent T, *et al*. Bioinformatics training: selecting an appropriate learning content management system-an example from the European Bioinformatics Institute. *Brief Bioinformatics* 2010;**11**:552–562.
18. Zatz MM. Bioinformatics training in the USA. *Brief Bioinformatics* 2002;**3**:353–360.
19. Schneider MV, Walter P, Blatter MC, *et al*. Bioinformatics Training Network (BTN): a community resource for bioinformatics trainers. *Brief Bioinformatics* 2012;**13**:383–389.
20. Schneider MV, Watson J, Attwood T, *et al*. Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinformatics* 2010;**11**:544–551.
21. Tabak I, Baumgartner E. The teacher as partner: exploring participant structures, symmetry, and identity work in scaffolding. *Cogn Instr* 2004;**22**:393–429.
22. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011; **2011**:bar009.
23. Hunter S, Jones P, Mitchell A, *et al*. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* 2012;**40**:D306–D312.
24. Finn RD, Mistry J, Tate J, *et al*. The Pfam protein families database. *Nucleic Acids Res* 2010;**38**:D211–D222.
25. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–29.
26. Yook K, Harris TW, Bieri T, *et al*. WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res* 2012; **40**:D735–D741.
27. Marygold SJ, Leyland PC, Seal RL, *et al*. FlyBase: improvements to the bibliography. *Nucleic Acids Res* 2013;**41**: D751–D757.
28. Ostlund G, Schmitt T, Forslund K, *et al*. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 2010;**38**:D196–D203.
29. Yellaboina S, Tasneem A, Zaykin DV, *et al*. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res* 2011;**39**:D730–D735.
30. Dinkel H, Michael S, Weatheritt RJ, *et al*. ELM–the database of eukaryotic linear motifs. *Nucleic Acids Res* 2012;**40**: D242–D251.
31. Hornbeck PV, Kornhauser JM, Tkachev S, *et al*. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 2012;**40**:D261–D270.
32. Ceol A, Chatr Aryamontri A, Licata L, *et al*. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res* 2010;**38**:D532–D539.
33. Szklarczyk D, Franceschini A, Kuhn M, *et al*. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011; **39**:D561–D568.
34. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, *et al*. The BioGRID interaction database: 2013 update. *Nucleic Acids Res* 2013;**41**:D816–D823.
35. Smoot ME, Ono K, Ruscheinski J, *et al*. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;**27**:431–432.
36. Hoffmann R, Valencia A. A gene network for navigating the literature. *Nat Genet* 2004;**36**:664.
37. Chen H, Sharp BM. Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics* 2004;**5**:147.
38. Kanehisa M, Goto S, Sato Y, *et al*. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012;**40**:D109–D114.
39. Croft D, O'Kelly G, Wu G, *et al*. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011;**39**:D691–D697.
40. Korcsmaros T, Farkas IJ, Szalay MS, *et al*. Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* 2010;**26**: 2042–2050.
41. Fazekas D, Koltai M, Turei D, *et al*. SignaLink 2—a signaling pathway resource with multi-layered regulatory networks. *BMC Syst Biol* 2013;**7**:7.
42. Corvaia N, Beck A, Caussanel V, *et al*. Insulin-like growth factor receptor type I as a target for cancer therapy. *Front Biosci* 2013;**1**:439–450.
43. Greenwald I. Notch and the awesome power of genetics. *Genetics* 2012;**191**:655–659.
44. Mayo C, Bertran-Alamillo J, Molina-Vila MA, *et al*. Pharmacogenetics of EGFR in lung cancer: perspectives and clinical applications. *Pharmacogenomics* 2012;**13**:789–802.
45. McCormick F. Cancer therapy based on oncogene addiction. *J Surg Oncol* 2011;**103**:464–467.
46. Yu H, Luscombe NM, Lu HX, *et al*. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res* 2004;**14**:1107–1118.

47. McGary KL, Park TJ, Woods JO, *et al*. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci USA* 2010;**107**:6544–6549.
48. Korcsmaros T, Szalay MS, Rovo P, *et al*. Signalogs: orthology-based identification of novel signaling pathway components in three metazoans. *PLoS One* 2011;**6**:e19240.
49. De Las Rivas J, Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* 2010;**6**:e1000807.
50. Xia Y, Yu H, Jansen R, *et al*. Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem* 2004;**73**:1051–1087.
51. Doncheva NT, Assenov Y, Domingues FS, *et al*. Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protocol* 2012;**7**:670–685.
52. Csermely P, Korcsmaros T, Kiss HJ, *et al*. Structure and dynamics of biological networks: a novel paradigm of drug discovery. A comprehensive review. *Pharmacol Ther* 2013;**pii**: S0163-7258(13)00028-4.
53. Papin JA, Hunter T, Palsson BO, *et al*. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 2005;**6**:99–111.
54. Gerstein M. Genomics: ENCODE leads the way on big data. *Nature* 2012;**489**:208.
55. Portales-Casamar E, Thongjuea S, Kwon AT, *et al*. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2010;**38**: D105–D110.
56. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol* 2012;**8**:565.
57. Csermely P. Recruiting the younger generation to science. *EMBO Rep* 2003;**4**:825–828.
58. Csermely P, Korcsmáros T, Lederman L (eds). *Science Education: Best Practices of Research Training for Students under 21'*, NATO Science Series Vol. 47. Amsterdam, The Netherlands: IOS Press, 2005.