Biotechnology Education

Evaluating the Effectiveness of a Practical Inquiry-Based Learning Bioinformatics Module on Undergraduate Student Engagement and Applied Skills[S]

James A. L. Brown*

From the Department of Biochemistry, School of Natural Sciences, National University of Ireland Galway, Ireland and Discipline of Surgery, School of Medicine, Lambe Institute for Translational Research, National University of Ireland Galway, Ireland

Abstract

A pedagogic intervention, in the form of an inquiry-based peer-assisted learning project (as a practical student-led bioinformatics module), was assessed for its ability to increase students' engagement, practical bioinformatic skills and process-specific knowledge. Elements assessed were process-specific knowledge following module completion, qualitative student-based module evaluation and the novelty, scientific validity and quality of written student reports. Bioinformatics is often the starting point for laboratory-based research projects, therefore high importance was placed on allowing students to individually develop and apply processes and methods of scientific research. Students led a bioinformatic inquiry-based project (within a framework of inquiry), discovering, justifying and exploring individually discovered research targets. Detailed assessable reports were produced, displaying data generated and the resources used. Mimicking research settings, undergraduates were divided into small collaborative groups, with distinctive central themes. The module was evaluated by assessing the quality and originality of the students' targets through reports, reflecting students' use and understanding of concepts and tools required to generate their data. Furthermore, evaluation of the bioinformatic module was assessed semi-quantitatively using pre- and post-module quizzes (a non-assessable activity, not contributing to their grade), which incorporated process- and content-specific questions (indicative of their use of the online tools). Qualitative assessment of the teaching intervention was performed using post-module surveys, exploring student satisfaction and other module specific elements. Overall, a positive experience was found, as was a post module increase in correct process-specific answers. In conclusion, an inquiry-based peer-assisted learning module increased students' engagement, practical bioinformatic skills and process-specific knowledge. © 2016 by The International Union of Biochemistry and Molecular Biology, 44:304–313 2016.

Keywords: *bioinformatics; practical; inquiry-led; process; content; undergraduate; project; teaching*

Introduction

The introduction, application, and evaluation of bioinformatic modules using web-based software packages for pedagogic purposes in the biological sciences has been evolving for the last 20 years, requiring constant revision and updating to keep pace with the ever changing scientific technologies and pedagogic techniques [1–6]. This is due to constant advances and availability of tools, databases, and data sets that has allowed these online databases and tools to rapidly grow, evolve, and establish themselves as essential community resources and tools in the last 15 years [7, 8]. This constant evolution has included: massively increased databases (of both size and content), the proliferation of highly specific search and modeling tools, algorithm advances, defining of

Volume 44, Number 3, May/June 2016, Pages 304–313

[S] *Additional Supporting Information may be found in the online version of this article.*

**Address for correspondence to: Discipline of Surgery, School of Medicine, Lambe Institute for Translational Research, National University of Ireland Galway, Ireland.*

E-mail: james.brown@nuigalway.ie

Received 25 May 2015; Revised 20 November 2015; Accepted 8 December 2015

DOI 10.1002/bmb.20954

Published online in Wiley Online Library (wileyonlinelibrary.com)

new biological domains/motifs, and improved interactive graphics [9–11]. In addition, the proliferation of accessible privately maintained databases (i.e., sites offered and maintained by commercial companies) has significantly complemented and enhanced current public research facilities. Supporting this, more advanced, comprehensive, state-of-the-art online databases and tool suites (many on the cutting edge of what is currently possible) often appear rapidly following the introduction of new techniques, technologies and recently large scale "omics" studies [12, 13].

These constantly evolving and growing online databases have become instrumental in defining how scientists organise their thinking and experiments [14, 15]. Exploring the current state of knowledge on a given topic using databases and repositories is the first step in any project [16]. It is also important that multiple databases are queried. The subsequent research steps are then shaped by the knowledge gleaned from these databases. Searches missing a key piece of data can be disastrous for a project, in terms of wasted time and resources.

Furthermore, scientists' thinking can be influenced by the structure of the databases themselves for example, how one subject is linked to another and what it is linked to within the database. This will influence how a researcher interprets this information, particularly if their knowledge of the subject is limited [9, 17].

One of the earliest and probably the most famous and widely used bioinformatic suite of tools is the National Centre for Biotechnology Information (NCBI) [10, 18], established in 1988. The NCBI is part of the United States National Library of Medicine (NLM) administered by the National Institute of Health (NIH). Following the global success of NCBI a number of national institutes followed suit: the Expert Protein Analysis System (ExPASy) [10, 14] suite created by the Swiss Institute of Bioinformatics (SIB) in 1998 and then the European bioinformatics institute (EBI) created by the European Molecular Biology Laboratory (EMBL) in 1992, which has only been accessible online since 2004 [17, [19–22]].

Currently there are now hundreds of databases, tools and software packages available online, with many curated lists available (OBRC: Online Bioinformatics Resources Collection; http://www.hsls.pitt.edu/obrc/and https://en.wikipedia.org/wiki/List_of_open-source_bioinformatics_software).

Recently, in response to this (almost exponential) proliferation of tools the journal Nucleic Acids Research begun issuing an annual, open access, special issue dedicated to detailing some of the more significant cutting-edge databases or software created [10, [14, 23]]. Issues like this are crucial for compiling and indexing software and web sites that provide professional bioinformatic users, teachers and the public with access to advances in software and data analysis. The terms for inclusion in this issue also assist in establishing a community specific benchmark for the quality and ease of use for such sites [10, 24].

Furthermore, pedagogical publications are appearing that incorporate lists of community specific online services which are either available or were used in their research activities [19–[22, 25]]. Many of these publications employ tools from the NCBI bioinformatic suite- as one of the original sites it has become an essential first point of contact. Importantly, due to its age and continual evolution it is also one of the most user-friendly. Additionally, it has the distinct advantage of being linked to the world's largest database of peer-reviewed publications and curated genetic information.

Significantly, many of the larger online repositories/databases host so many tools that they now have their own published manuals [14, [20, 23]], in addition to their own online frequently asked questions (FAQ) and help sections. These FAQ's and database help sections allow novices and professionals alike to engage in self-directed learning with the software they are exploring, often with helpful examples.

Many of the large scientific community-specific (biology) web-based databases provide tools such as; sequence alignment, Basic Local Alignment Search Tool (BLAST), genomic maps, literature searches, structural/domain analysis and gene expression information. However, in recent years there has been a significant rise in the appearance of highly specialised databases and software tools. Some examples of commonly used specialised databases or research tools available are: tissue or cell line specific gene expression databases (Oncomine, TiGER), cell cycle specific gene expression databases (Cyclebase, GeneCards), interaction networks (String, BioGrid), specific protein posttranslational modification databases (phosphorylation: PhosphoSitePlus, PHOSIDA, NetPhos; Acetylation: ASEB, Scan-x; Ubiquitylation: UbiProt, hUbiquitome), MicroRNA databases (PicTar, HMDD), genetically modified animal strain databases (JMSR, RGD), cancer cell line databases (CCLE, CGAP) and process-specific databases (CMC, CilDB, MetaCore) [14, [26–34]].

This proliferation of specialist/niche databases or software tools emphasises the need for up-to-date training in general software and database use and evaluation-some databases look good but the results are based on very a limited set of data (i.e., Cyclebase), which can significantly affect the interpretation of any results obtained. Many software packages return concise but technical results that often take prior experience or specialised knowledge to interpret. To use most databases they require input in a specific format or search terms to be taken from a specific database and many competing databases label the same item differently.

Recently, many published undergraduate modules have made significant efforts to mimic real world conditions by incorporating real samples, technologies and techniques encountered in industry or research [35–38]. This is indicative of a current pedagogic philosophy and educational strategy, whereby undergraduate students who are familiar and comfortable with real world conditions and technologies are anticipated to be more experienced and ultimately more

employable graduates [39–41]. This includes utilizing generalist skills, such as oral and written communication, teamwork, problem solving and structured report writing [41]."

This module introduces students to the application of key elements of the scientific processes, through the practical real-world application, involved in understanding how to find information, generate new data and evaluate and interpret the usefulness of the results and software packages, producing a valid scientific assessment that can be used further to initiate a wet-lab-based project [22, 24, [42–44]]. The module provided students' with a common frame of reference for community-wide knowledge and a conceptual starting point from which to begin additional scientific investigations. Furthermore, this module required students to employ techniques and skills, in a practical setting, which are valued by employers.

The aim of this teaching intervention was to evaluate the success of a new module in practical bioinformatics through incorporation and practical use (enquiry-based student-led projects) of the most up-to-date, relevant and widely used suites of software tools and databases to generate novel data. Evaluation of the effectiveness of the pedagogic interventions on improving students' knowledge and its application was performed using quantitative (pre- and post-module quiz) and qualitative (questionnaire and evaluation of the project reports) methods.

Materials and Methods

Organisation

This module was deployed to 3rd year undergraduate biotechnology students (class size 20–35 students). The development and deployment of this compulsory 3rd year module (a major, building on pre-requisite 1st and 2nd year modules) was based on the teaching intervention of using small group teaching to facilitate enquiry-based, peer-peer learning [45–48]. The lecturing approach was deliberately minimalist, a short 20–30 min lecture highlighting critical tools (websites), concepts and assessment deliverables, followed by a short question and answer session. The remaining time was allocated for group work, where the lecturer continually circulated providing feedback to each group on progress, answering group specific questions, evaluating strategies or providing short demonstrations or help using software packages.

It has been demonstrated that modifying the learning environment influences students approach to learning, with peer-assisted learning an important example [49]. Furthermore, it has been shown that action-based peer-group projects enhance student engagement [45, 50]. Currently, three main peer-assisted learning methods have been distinguished and shown to improve learning in science: problem-based learning (PBL), process-oriented guided inquiry learning (POGIL) and Peer-Led Team Learning (PLTL) [51]. The module deployed here uses a POGIL based approach, where students work in self-managed teams, guided by lecture content and assignment questions.

TABLE I *Project themes*

Project theme	*Theme focus*
Centrosome clustering	Genome instability
Exosomes	miRNA
Stress granules	mRNA storage
Ciliation	Primary cilia
Nonsense-mediated mRNA decay	mRNA processing

Students were randomly assigned to a group (of 4), to facilitate small group peer-assisted learning and mimic real world situations, where teams are often made up of unacquainted individuals of differing abilities and skills. Furthermore, the random grouping accurately recapitulates the environment found in real laboratories. That this mimics real-world working conditions was highlighted to students, to emphasise that this project facilitates building group skills, learning to interact with a variety of personalities to achieve a given task. Different groups were actively encouraged to assist each other and share discovered resources.

Four tutorials (2-hr long) were given over 3 weeks, with the final assessment due a week after the concluding tutorial. Each student was provided with access to a desktop computer with Internet access, however almost all students chose to use personal laptops with a variety of browser options, primarily Firefox and Chrome.

Project Themes

Each group was assigned a project theme and additionally for each theme a focus area was suggested (Table I). Based around this central theme each group member was then individually required to find and annotate a "novel" or "new" member of this focus area (specific cell-signalling pathway) for further analysis. Themes were chosen based on the lecturers' areas of expertise, to provide a suitable level of base knowledge for advising on the choice of candidate proteins. In addition, research themes were chosen that have recently had renewed levels of interest (i.e., significant yearly increases in indexed publications). The increased research interest in the theme areas provides many recent publications readily identifying/defining new unexplored member proteins for students to choose from.

Peer-assisted Learning

The module was designed to encourage students to actively engage in peer-assisted learning through participation in small group learning. In addition, students were actively

TABLE II *Learning objectives*

Knowledge	*Process*
Understand the terminology	Analyse databases and construct a coherent "research" plan
Defend tools chosen	Use scientific databases to generate data
Identify bioinformatic software packages	Operate the bioinformatic software to produce usable results
Discriminate between "real" results and computational artefacts	Verify the results using secondary software (where possible)
	Arrange results into coherent scientific report summarising your findings
Identify gene or protein features	

encouraged to share knowledge between groups. This was facilitated in two explicit ways: 1) Through the use of an online virtual learning environment (VLE) discussion forum and 2) Interactive class discussions during tutorials.

Assessment

Assessment of the project reports, containing the data generated by students and the resources used, was evaluated against given, key learning objectives. A marking rubric was used, which was provided to the students prior to the submission of the final report.

Ethics Approval

This study was conducted according to the National University of Ireland Galway ethical guidelines and did not require ethics approval.

Results

Lectures

An overall aim of the module was to introduce students to the practical application of the scientific process, by asking students to produce a report that identified and justified a potential new target protein not previously extensively (if at all) studied in the given cellular pathways, upon which a wet lab-based research project could be founded.

From the initial lecture students were given clear explicit learning objectives broken down into either knowledge or process-related (Table II). This was achieved by informing students of the overall guiding scientific concepts and strategic plan. Each group was given an overall theme (in this case specific cellular process), with students required to perform searches of the literature and online databases to define the cellular pathway (e.g., Fig. 1). Well-known established bioinformatic software packages/tools were demonstrated to students in class (Table III). Packages were chosen to facilitate the extraction of informative data. The project deliverables (covering the learning objectives) were listed and the individual objectives covered in each lecture highlighted (Table IV).

Practical Techniques for Project Completion

The report students were required to produce detailed individual results related to their protein of interest/choice. The overall framework/categories for these results were: List/details of the each member of the groups chosen protein and how they relate to the described pathway; Provide images (downloaded or screen captures) from each online database/tool used; Include a short description (justification) of each step/process completed and the database used; Detail the online resources used (tools and web sites) and provide primary scientific references related to the protein chosen (related to the theme cellular process). A more detailed list of the project marking criteria is supplied below in the project reports section (under: student project and module assessment).

A particularly important concept was not that the candidate protein was novel, but that its involvement in the given theme area should be novel. Therefore the ideal candidate proteins for investigation would be referenced in as few publications as possible in relation to the specific pathway/process theme. This was achieved through students examining the number of publications, in the NCBI PubMed database, containing the protein of interest. Publications were investigated in the context of the protein of interest within the given project theme. Well-studied proteins with many publications could be chosen for analysis, provided their involvement in the theme area was relatively unknown ($\leq$2 publications was considered acceptable, as more references likely indicates a considerable amount of study on the target). Students using proteins with a well-defined and known role in the theme area could proceed with the analysis, but would lose significant marks for originality. Following the initial tutorial students were required to post their choice on the discussion forum for evaluation by the lecturer.

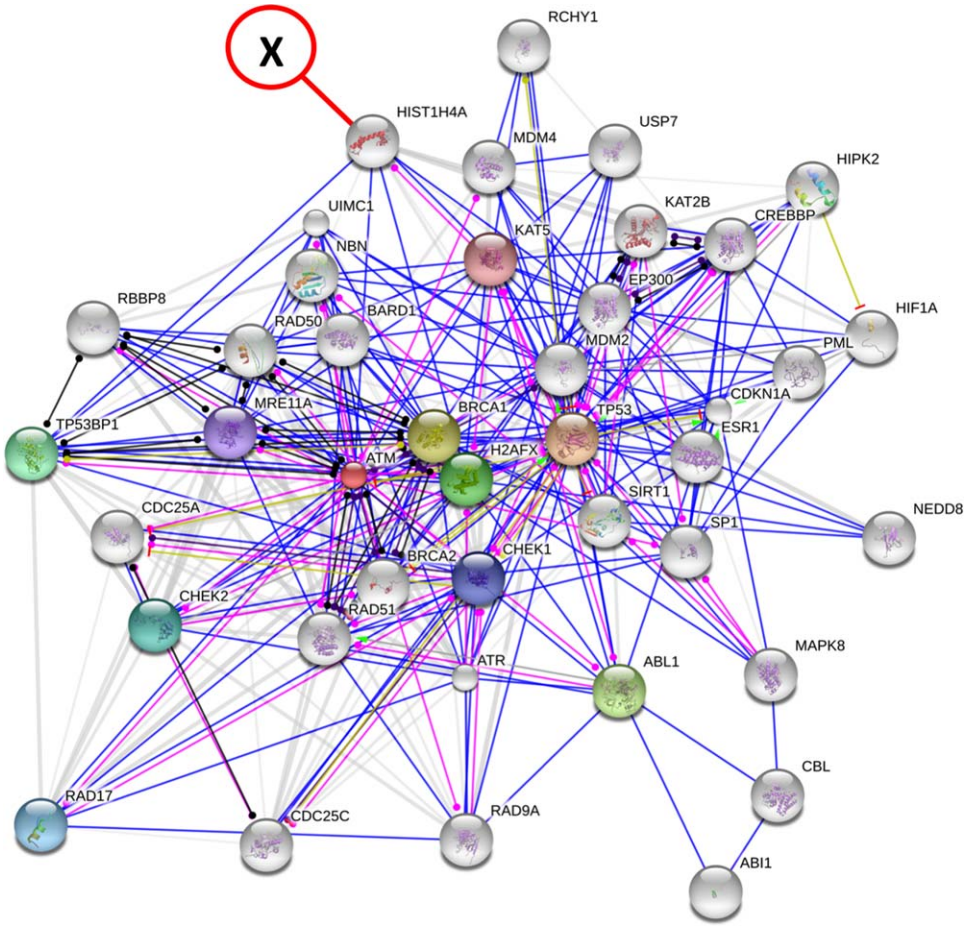

FIG 1 *Example of network mapping. Generated using STRING. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]*

This allowed students to be advised on the inappropriateness of their choice allowing any changes to be made early.

To facilitate diagrammatic representation of the chosen protein and its place in signalling pathways, use of the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [26] was encouraged. STRING results are particularly useful as the results are modifiable on a number of levels, supporting further interaction (Fig. 1). The diagrammatic results returned from STRING can be manipulated, by moving individual nodes (proteins) and by changing the nature of links (interactions) between nodes, to provide additional information to the reader. In addition, STRING diagrams can be exported as high quality images for "publication" (in their project report), an important concept for students to understand (understanding how peer-reviewed publications are prepared).

Once a suitable candidature protein was identified, to fulfil the project deliverables (Table IV) the students would be required to generate their own novel unique data (using online tools) related to this target, by: building a molecular signalling pathway and placing their protein within this pathway, by assessing and analysing the target proteins interactions (known or predicted) with other members of the pathway (constructed from multiple online tools); Define the domain structure of their protein [defined/discovered by students using the online tools]; Compile a list of known or predicted posttranslational modifications and how these relate to the domain structure or biochemical activities of their candidate (in the context of the theme area/pathway); Define any mutations and determine if there are any associated diseases or phenotypes and how these relate to the defined domains, PTM's or activity. All of these are addressed by students directly using the online tools to interrogate/search the protein sequence to discover new or unknown features, often using tools designed to search for specific features (i.e., kinase phosphorylation sites, acetylation sites, nuclear localisation signals, protein binding sites).

Furthermore students were required to create specific sets of primers, for cloning the target gene [with vector (pEGFP-N1) specific restriction enzyme sites, as determined by students analysis of the restrictions sites in their target gene] and for cloning a truncated version of the gene (to facilitate analysis of the functional effects of loss of one of the defined domains or PTMs). In addition, students created

TABLE III *Bioinformatic software packages used*

Database name	*URL.*
NCBI	
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/
CDD	http://www.ncbi.nlm.nih.gov/cdd
BioSystems	http://www.ncbi.nlm.nih.gov/biosystems/
Highwire	http://highwire.stanford.edu/cgi/search
ExPASY	http://www.expasy.org/
PrimerX	http://www.bioinformatics.org/primerx/
STRING	http://string-db.org/
KEGG pathway	http://www.genome.jp/kegg/pathway.html
PhosphoSite	http://www.phosphosite.org/
GeneCards	http://www.genecards.org/
IntAct	https://www.ebi.ac.uk/intact/
Reactomoe	http://www.reactome.org/
NEB cutter V2.0	http://tools.neb.com/NEBcutter2/
NCBI Primer Blast	http://www.ncbi.nlm.nih.gov/tools/primer-blast/
Life Technologies custom primers	http://tools.lifetechnologies.com/content.cfm?pageid=9716
CRISPR design	http://crispr.mit.edu/
Abcam	http://www.abcam.com/
Antibodypedia	http://www.antibodypedia.com/
Cyclebase	http://www.cyclebase.org/
TiGER	http://bioinfo.wilmer.jhu.edu/tiger/
MOPED	https://www.proteinspire.org/MOPED/mopedviews/proteinExpressionDatabase.jsf
JMSR	http://www.shigen.nig.ac.jp/mouse/jmsr/

specific primers to allow site directed mutagenesis (SDM) of a PTM or motif they identified (either activating or inactivating mutations, as chosen and justified by the students). Furthermore students defined any potential Crispr targeting sequence, to allow a knockout cell line to be produced (and justification if this was anticipated to possible, i.e., was this an essential gene that cannot be inactivated without inducing cell death?).

Students were required to define any known or predicted gene regulation (i.e., cell cycle or tissue specific) and relate this to a model organism they would chose to study this protein in (were there any known model systems and what were they i.e., mouse or fly knockouts?). To achieve this, students were required to determine homologous genes in at least 1 other model species (with the human sequence required as a minimum). Students were also required to find any currently available antibodies that could be used for their research (and justify if they were suitable for use in analysis of SDM or truncating mutations).

To complete the project deliverables (by generating their own unique data sets) required students to understand and apply the theory, concepts, techniques and processes underpinning the deliverables (i.e., apply the scientific process). In addition students were required to combine, concisely summarise and reference the tools used and any published data discovered that supported or justified the data they generated.

TABLE IV *Project deliverables*

Knowledge
Images-screen captures of each database used and results returned
Map (of): Pathway Domain PTM
Primers for cloning (with restriction enzyme sites): Whole gene Truncating mutations Site directed mutagenesis of selected PTM
Crispr knockout sequence
Antibodies Protein specifically (or for specific mutated PTM)
Gene expression
Model Organism Mouse, Rat, yeast, chicken?? Are there any known knockouts Show homologous gene is at least 1 other model species

Tutorials

Classes were conducted in a computer lab (with individual computers for students) and consisted of a 2-hr block. Lectures lasted 30–45 min, with the remaining time set aside for tutorials and small peer-group work. Lecture and tutorial content was informed and supplemented by addressing comments and questions from the online discussion forum. The tutorial section of the class began by addressing/demonstrating specific process related questions arising from the previous tutorial, where needed. Otherwise, each group was allocated time with the lecturer to ask questions and have their processes/methods evaluated on an on-going basis.

Troubleshooting

To facilitate troubleshooting of processes, tool manipulation or data generation an online discussion forum was provided using the VLE (Blackboard). Students were able to post anonymously and create threads. In addition, demonstration of techniques and methods to allow students to troubleshooting their own work, were accommodated during the tutorials (without actually doing the work for students). Questions relating to any section of the project (visualization, software manipulation, reporting, database discovery/use) were welcomed and peer-peer answering encouraged (and observed). Peer–peer troubleshooting (within and between groups) was encouraged and observed during the protected tutorial class time.

Student Project and Module Assessment Tools

Student Reports

Although the project was based around a group theme, project reports were submitted individually. Each group member was to identify and report on an individual target protein, fitting within the group theme (examples of target proteins identified, Table V). The students could choose any species to investigate, however human genes/proteins were almost universally (but not exclusively) chosen. Reports were evaluated using a given rubric (Supporting Information Fig. S1), which defined the novelty of chosen protein and compatibility with the project theme. Other assessment criteria covered included: Quality of images produced summarising the data discovered; Number of databases/software packages used; Quality and description of screenshots captured from database enquiries; Additional resources independently discovered and correctly incorporated. Reports could be self-evaluated using the supplied marking rubric (Supporting Information Fig. S1), which was likewise used for the assessment of the final submitted report.

For the project report students were required to provide individual results (unique data they produced) related to their protein of interest/choice and generated using their chosen software package. The specific details for each individual project that each student was marked for were: images (screen captures of each database used); a map (diagram) of the cellular pathway (with the protein from

TABLE V *Student identified target proteins*

Cellular process	*Student target proteins*
Centrosome clustering	MAD2L1 CASC5 ILK
Exosomes	Vps24 SDCBP Rab27a
Stress granules	STYXL1 OGFOD1 TDRD3
Ciliation	RSPH4A RSPH9 DNAH11
Nonsense mediated mRNA decay	XRN2 SMG8 NOM1

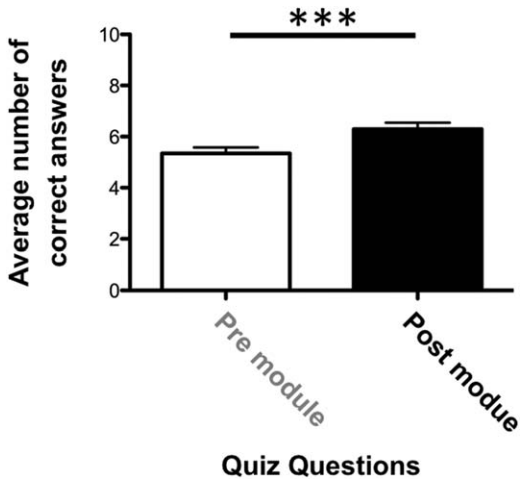

FIG 2

Average number of correct quiz answers. Comparison of (average) numbers of correct answers to the Pre- or Post- module quiz. Graph represents N = *2 [53 paired responses (individual students pre- and post quiz responses)]. Students* T *test, two tailed. A* p *value <0.001(***) was deemed highly significant. Graph displays mean (±SEM).*

each member of their group included); a diagram of the domain structure of their protein as determined using the online tools/databases; discovered post-translational modifications (PTM) of their protein; produce primers for cloning their gene, with choice and justification of specific restriction enzyme sites for cloning their gene into the vector pEGFP-N1 (truncating mutations to remove important domains identified); site directed mutagenesis against identified specific PTM; produce a "Crispr" targeting knockout sequence against their gene; identify an antibody for their protein (or for their identified specific PTM); demonstrate the expression profile of their gene (tissue or cell cycle); describe and justify a model organism system suitable for studying the identified protein (i.e., Mouse, Rat, Yeast, Chicken, Nematode; including any known knockout systems) and demonstrate homologous genes in at least 1 other model species (related to the model organism system chosen).

For all categories marks were awarded if a search (database search, online tool use) was performed and their working demonstrated (by screen capture or images), regardless of the result returned (which was entirely dependent on each individual protein and database/tool chosen).

Module Assessment

Assessment of the effectiveness of the module was gauged through three distinct methods: 1) primarily through the quality of the submitted project reports, 2) a non-assessable (not contributing to their grade) pre- and post-module short answer quiz (SAQ), and 3) a student feedback survey. A custom-made student feedback survey was used for student evaluation of the module (Supporting Information Fig. S2). The SAQ was designed as a short, simple quantitative test of both knowledge and application. The SAQ consisted of 30 MCQs, 15 each related to knowledge (Content) or application (Process) (Supporting Information Fig. S3).

Pre/Post Module Assessment of Learning

Overall there was a small but highly significant ($p = 0.0003$) increase in the number of correct answers answered post-module, compared to premodule (Fig. 2). Investigating this further by segregating questions as either process or content related we observed upward trend (not significant) in the number of correct content-related question answers (Fig. 3, left). Interestingly, we found a very significant increase ($p = 0.0057$) in the number of correct process related questions post-module (Fig. 3, right side). In addition, subtle changes between the 1st and 2nd deployment led to a highly significant increase in the overall marks students gained [an increase in the mean from 58% (SD ± ~11) to ~76% (SD ± ~19)]. However, this could be attributed to a more academically capable 2nd class, better delivery of the module the 2nd time or another factor. Additional research would be required to determine if this effect was due to improvements in the module content.

Post Module Questionnaire and Comments

Generally, the post module evaluation (Supporting Information Fig. S2) was completed by >50% of students. From the qualitative section, all responders found the module very challenging. The top "likes" comments were (in order of occurrence): "Topic is interesting" (always wanted to do bioinformatics); "The discussion forum is good" and "This

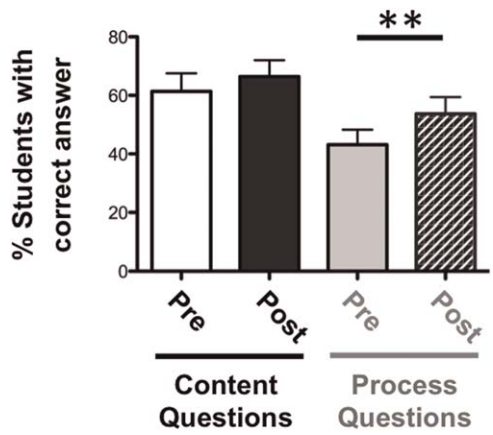

FIG 3

Correct pre- and post-module quiz answers grouped by category. Comparison of (average) numbers of correct answers to the Pre- or Post-module quiz, grouped as either content or process related. Graph represents N = *2 (40 pre-module, 28 post-module responses). Students* T *test, two tailed. Students* T *test, two tailed. A* p *value <0.01(**) was deemed highly significant. Graph displays mean (±SEM).*

subject will be useful for 4th year." The top "dislikes" comments were (in order of occurrence): "Would have preferred a more comprehensive run through of featured software packages" and "would have a preferred longer tutorial."

Discussion

Assessment of Goals

The goal of the pedagogic intervention (in the form of a new module) was to engage students and increase their process-specific knowledge (application) of their content-based knowledge using the scientific process, through a research-led project. Based on the submitted student reports (including the novelty of targets chosen), the qualitative and quantitative data gathered, this goal was achieved. A highly significant increase was observed when answering questions that required process-specific (applied) knowledge. Furthermore, the high quality of the submitted reports and the positive comments received in the post module questionnaire indicated that students were very engaged in the project/scientific process. Each year the students independently "discovered" and described >3 previously unknown or new (to the author as well) valuable bioinformatic web based packages. Furthermore, several target proteins chosen (i.e., RSPH4A) now have several recent primary research publications (last 12–18 months), validating the students' decision to choose (recommend) further investigation of these proteins in the indicated pathways. There is clearly room for improvement, such as longer tutorials and increased tutorial demonstrations of some software packages, but overall students enjoyed, learnt from and valued the skills gained from the module.

Conclusion

Project-based learning applied to real (not simulated), student led research projects enabled both student engagement and learning. This module has also provided students with working knowledge of current community databases and software packages, knowledge and application of which are required in both the workforce and in academia.

Acknowledgements

The author thanks Dr Andrew Flaus (Biochemistry, NUI Galway) for allowing me to implement these new undergraduate modules. The author thanks Dr Simon Warren (Centre for Excellence in Learning and Teaching, NUI Galway) and Dr Emer Bourke (Pathology Department, NUI Galway) for critical discussions of the manuscript.

References

[1] Chumley-Jones, H. S. Dobbie, A., and Alford, C. L. (2002) Web-based learning: Sound educational method or hype? A review of the evaluation literature. Acad. Med. 77, S86–S93.

[2] Cohen, L., Manion, L., and Morrison, K. (2013) Research Methods in Education, Routledge. ISBN: 9780415583350

[3] Gott, R. and Duggan, S. (2002) Problems with the assessment of performance in practical science: Which way now? Cambridge J. Educ. 32, 183–201.

[4] Furge, L. L. Stevens-Truss, R. Moore, D. B., and Langeland, J. A. (2009) Vertical and horizontal integration of bioinformatics education: A modular, interdisciplinary approach. Biochem. Mol. Biol. Educ. 37, 26–36.

[5] Howard, D. R. Miskowski, J. A. Grunwald, S. K., and Abler, M. L. (2007) Assessment of a bioinformatics across life science curricula initiative. Biochem. Mol. Biol. Educ. 35, 16–23.

[6] Inlow, J. K. Miller, P., and Pittman, B. (2007) Introductory bioinformatics exercises utilizing hemoglobin and chymotrypsin to reinforce the protein sequence-structure-function relationship. Biochem. Mol. Biol. Educ. 35, 119–124.

[7] Curcin, V., Ghanem, M., and Guo, Y. (2005) Web services in the life sciences. Drug Discov. Today 10, 865–871.

[8] Sutcliffe, I. C. and Cummings, S. P. (2007) Making bioinformatics projects a meaningful experience in an undergraduate biotechnology or biomedical science programme. Biosci. Educ. 10(1), 1–2.

[9] Gelbart, W. M. (1998) Databases in genomic research. Science 282, 659–661.

[10] Fernández-Suárez, X. M. and Galperin, M. Y. (2013) The 2013 nucleic acids research database issue and the online molecular biology database collection. Nucleic Acids Res. 41, D1–D7.

[11] Fernández-Suárez, X. M. Rigden, D. J., and Galperin, M. Y. (2014) The 2014 nucleic acids research database issue and an updated NAR online molecular biology database collection. Nucleic Acids Res. 42, D1–D6.

[12] Gomez-Cabrero, D, *et al*. (2014) Data integration in the era of omics: current and future challenges. BMC Syst Biol 8(Suppl 2), I1

[13] Galperin, M. Y., Rigden, D. J., and Fernández-Suárez, X. M. (2015) The 2015 nucleic acids research database issue and molecular biology database collection. Nucleic Acids Res. 43, D1–D5.

[14] Gasteiger, E., Gattiker A, Hoogland C, Ivanyi I. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788.

[15] Hine, C. (2006) Databases as scientific instruments and their role in the ordering of scientific work. Social Stud. Sci. 36, 269–298.

[16] Stevens, R., Goble, C., Baker, P., and Brass, A. (2001) A classification of tasks in bioinformatics. Bioinformatics 17, 180–188.

[17] McWilliam, H., *et al*. (2013) Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 41, W597–W600.

[18] Cates, S. NCBI: National center for biotechnology information. 2006.

[19] Grisham, W., Schottler, N. A., Valli-Marill, J., Beck, L., and Beatty, A. J. (2010) Teaching bioinformatics and neuroinformatics by using free web-based tools. CBE Life Sci. Educ. 9, 98–107.

[20] Bednarski, A. E., Elgin, S. C., and Pakrasi, H. B. (2005) An inquiry into protein structure and genetic disease: Introducing undergraduates to bioinformatics in a large introductory course. Cell Biol. Educ. 4, 207–220.

[21] Campbell, A. M. (2003) Public access for teaching genomics, proteomics, and bioinformatics. Cell Biol. Educ. 2, 98–111.

[22] Cattley, S. and Arthur, J. W. (2007) BioManager: The use of a bioinformatics web application as a teaching tool in undergraduate bioinformatics training. Briefings Bioinform. 8, 457–465.

[23] McEntyre, J. and Ostell, J. (2002–2005): *The NCBI Handbook.* Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information (NCBI).

[24] Schneider, M. V., Watson J, Attwood T, Rother K, Budd A, McDowall J, *et al*. (2010) Bioinformatics training: A review of challenges, actions and support requirements. Briefings in bioinformatics. 2, 544–551.

[25] LeClair, E. E. (2008) Teaching the toolkit: A laboratory series to demonstrate the evolutionary conservation of metazoan cell signaling pathways. Biochem. Mol. Biol. Educ. 36, 225–233.

[26] Szklarczyk, D., *et al*. (2011) The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39, D561–D568.
[27] Stark, C., *et al*. (2011) The BioGRID interaction database: 2011 update. Nucleic Acids Res. 39, D698–D704.
[28] Barretina, J., *et al*. (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603.
[29] Arnaiz, O., *et al*. (2009) Cildb: A knowledgebase for centrosomes and cilia. Database (Oxford) 2009, bap022.
[30] Rhodes, D. R., *et al*. (2007) Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9, 166–180.
[31] Liu, X., Yu, X., Zack, D. J., Zhu, H., and Qian, J. (2008) TiGER: A database for tissue-specific gene expression and regulation. BMC Bioinform. 9, 271.
[32] Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. (1998) GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics 14, 656–664.
[33] Stark, C, *et al*. (2006) BioGRID: A general repository for interaction datasets. Nucleic Acids Res. 34, D535–D539.
[34] Gnad, F., *et al*. (2007) PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol. 8, R250.
[35] Zhang, B., Wang, Y., Xu, X., Guan, X., and Bai, Y. (2013) Using PCR-RFLP technology to teach single nucleotide polymorphism for undergraduates. Biochem. Mol. Biol. Educ. 41, 262–266.
[36] Brownell, S. E., Kloser, M. J., Fukami, T., and Shavelson, R. (2012) Undergraduate biology lab courses: Comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. J. Coll. Sci. Teach. 41, 36–45.
[37] Dopico, E., Linde, A. R., and Garcia-Vazquez, E. (2014) Learning gains in lab practices: Teach science doing science. J. Biol. Educ. 48, 46–52.
[38] Smit, J., Heath, D. D., and Walter, R. P. (2014) A molecular genetic lab to generate inclusive and exclusive forensic evidence: Two suspects, a victim, and a bloodstained T-shirt. Biochem. Mol. Biol. Educ. 42, 230–236.
[39] Peat, M., Taylor, C. E., and Franklin, S. (2005) Re-engineering of undergraduate science curricula to emphasise development of lifelong learning skills. Innovat. Educ. Teach. Int. 42, 135–146.
[40] Rodrigues, S., *et al*. (2007) The usefulness of a science degree: The "lost voices" of science trained professionals. Int. J. Sci. Educ. 29, 1411–1433.
[41] Perdigones, A., García, J. L., Valiño, V., and Raposo, C. (2009) Assessing heterogeneous student bodies using a methodology that encourages the acquisition of skills valued by employers. Assess. Eval. Higher Educ. 34, 389–400.
[42] Schneider, M. V., *et al*. (2012) Bioinformatics training network (BTN): A community resource for bioinformatics trainers. Brief. Bioinform. 13, 383–389.
[43] Maloney, M., *et al*. (2010) Bioinformatics and the undergraduate curriculum. CBE Life Sci. Educ. 9, 172–174.
[44] Boyle, J. A. (2004) Bioinformatics in undergraduate education. Biochem. Mol. Biol. Educ. 32, 236–238.
[45] Hodgson, Y., Benson, R., and Brack, C. (2013) Using action research to improve student engagement in a peer-assisted learning programme. Educ. Action Res. 21, 359–375.
[46] Springer, L., Stanne, M. E., and Donovan, S. S. (1999) Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Rev. Educ. Res. 69, 21–51.
[47] Quitadamo, I. J., Brahler, C. J., and Crouch, G. J. (2009) Peer-led team learning: A prospective method for increasing critical thinking in undergraduate science courses. Sci. Educ. 18, 29–39.
[48] Shapiro, C., Ayon, C., Moberg-Parker, J., Levis-Fitzgerald, M., and Sanders, E. R. (2013) Strategies for using peer-assisted learning effectively in an undergraduate bioinformatics course. Biochem. Mol. Biol. Educ. 41, 24–33.
[49] Micari, M. and Light, G. (2009) Reliance to independence: Approaches to learning in peer-led undergraduate science, technology, engineering, and mathematics workshops. Int. J. Sci. Educ. 31, 1713–1741.
[50] Tariq, V. N. (2005) Introduction and evaluation of peer-assisted learning in first-year undergraduate bioscience. Biosci. Educ. 6(3). 1–19.
[51] Eberlein, T, *et al*. (2008) Pedagogies of engagement in science. Biochem. Mol. Biol. Educ. 36, 262–273.