BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 39, No. 6, pp. 440–447, 2011

Laboratory Exercise

Self-directed Student Research Through Analysis of Microarray Datasets[S]

A COMPUTER-BASED FUNCTIONAL GENOMICS PRACTICAL CLASS FOR MASTERS-LEVEL STUDENTS*

Received for publication, April 17, 2011; accepted 29 June, 2011

Laura J. Grenville-Briggs and Ian Stansfield†

From the School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK

This report describes a linked series of Masters-level computer practical workshops. They comprise an advanced functional genomics investigation, based upon analysis of a microarray dataset probing yeast DNA damage responses. The workshops require the students to analyse highly complex transcriptomics datasets, and were designed to stimulate active learning through experience of current research methods in bioinformatics and functional genomics. They seek to closely mimic a realistic research environment, and require the students first to propose research hypotheses, then test those hypotheses using specific sections of the microarray dataset. The complexity of the microarray data provides students with the freedom to propose their own unique hypotheses, tested using appropriate sections of the microarray data. This research latitude was highly regarded by students and is a strength of this practical. In addition, the focus on DNA damage by radiation and mutagenic chemicals allows them to place their results in a human medical context, and successfully sparks broad interest in the subject material. In evaluation, 79% of students scored the practical workshops on a five-point scale as 4 or 5 (totally effective) for student learning. More broadly, the general use of microarray data as a "student research playground" is also discussed.

Keywords: *Saccharomyces cerevisiae*, DNA damage response, transcript profiling, microarray, bioinformatics.

It is generally accepted that students learn best through inquiry-based approaches, (active learning) and this is particularly true in science courses. Actively constructing knowledge from a combination of experiences, interactions and interpretations optimizes the acquisition of an expert understanding of scientific concepts [1]. Therefore, it is not surprising that social constructivism is the dominant learning theory in science education [2]; the investigative nature of science lends itself well to such effective teaching and learning strategies. Science research now uses a vast array of technology and it is both desirable, and increasingly possible, to incorporate the use of such technology into a University classroom setting to provide methodological training. This allows students to interact with information in new ways, to create their own visualizations and understanding of scientific data. We have been guided by these core principles in the design of a new series of computer workshops centred on functional genomics and systems biology. The workshops, which analyse microarray data, crucially have a strong research theme running through them, taking advantage of the vast information resource represented by transcriptomic datasets to allow students to propose, then test, their own hypotheses. Students thus actively construct knowledge from a self-directed investigation, applying a range of research tools to their own experiments.

Systems biology is a relatively new discipline, in which mathematical modeling of biological systems is used as a tool to understand the structure and regulation of biological systems, as well as the complex network architecture inherent in cellular biochemistry. The application of systems biology to network analysis has in particular grown as transcriptomics, proteomics and metabolomics are used to construct mathematical models of biological pathways in health and disease [3, 4].

As part of their training therefore, it is essential that students of systems biology and other molecular biological subjects understand the theory and practice of high-throughput postgenomic technology such as proteomics and transcriptomics. The computer workshops we have thus developed give students hands on experience of manipulating and interpreting transcriptomic data as the basis for functional genomic investigation. The course is

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

†To whom correspondence should be addressed. Ian Stansfield, School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK. Tel.: +44(0)1224 437318; Fax: +44(01224) 555844.
E-mail: i.stansfield@abdn.ac.uk.

also designed to teach students a range of basic bioinformatic tools, and to closely mimic how such investigations are conducted in a typical research laboratory. Finally, a written report on completion of the practical classes, in the style of a research paper, forms a method of assessing understanding, as well as providing students with training in scientific writing.

In preparation for the practical transcriptomics workshops, students attend lectures on genome structure and evolution, transcriptomics, proteomics and functional genomics. During the functional genomics lecture series students were introduced to the yeast *Saccharomyces cerevisiae* as a model organism for functional genomics. The follow-on practical classes comprise 4 two-hour workshops and derive their starting material from a published *S. cerevisiae* microarray study investigating transcriptomic responses to DNA damage agents [5]. The workshops, described below, are carried out in computer teaching labs over 4 consecutive weeks, with a final report submitted two weeks later.

WORKSHOP CONTENT AND STRUCTURE

We chose to base these workshops on the model organism *S. cerevisiae*, as a wide range of functional data is available in this organism. A published microarray dataset, investigating the transcriptomic response of yeast to DNA damage [5] was chosen as the basis of the workshops, since the conservation of DNA repair machinery across phyla allows the data analysis to be easily extended to look at orthologous sequences in other organisms, and to link to human diseases, ensuring a broad appeal across the class. The raw data is freely available via web download, and was provided to the students to allow them to carry out their own analyses on real research data.

The Learning Objectives for the Workshop Series were Broadly

- To understand how microarray data is analysed using hierarchical clustering.
- To understand the concept of gene ontology (GO) classification.
- To utilize a wide range of bioinformatic tools in a research context to understand functional genomics analysis.
- To gain experience and understanding of writing research results in the form of a scientific paper.

Biological Background

The practical workshops were preceded by a discussion with the students of cellular responses to DNA damage, supported by a Powerpoint presentation. DNA damage in a cell compromises genomic stability, and can result from environmental stresses or from cellular processes that occur during normal growth. Thus, cells have evolved complex surveillance mechanisms that monitor genomic integrity, and they orchestrate a multifaceted response to DNA damage to ensure accurate transmission of genetic information [6].

These multiple outputs of the DNA damage response include cell-cycle arrest, alterations in gene expression, DNA damage repair, and cell death. The responses are mediated by a kinase cascade that has been conserved through eukaryotic evolution. At the top of this cascade is a family of phospho-inositol kinase-related proteins, which includes the ATR and ATM kinases in mammals and their homologs in yeast, Mec1p and Tel1p. Downstream of the phospho-inositol kinase-related kinases are two classes of checkpoint kinases, including CHK1 and CHK2 in mammals and Chk1p and Rad53p in yeast. An additional kinase in yeast, named Dun1p, acts downstream of Rad53p and is involved in both cell-cycle arrest and transcriptional regulation in the DNA damage response [5]. Mutations in components of the ATR/Mec1 pathways result in hypersensitivity to DNA-damaging agents and, in higher organisms, predisposition to cancer [7, 8].

The Mec1 pathway also affects gene expression. However, little is known of the transcriptional control exerted by the Mec1 kinase. In the study underpinning the computer workshops [5], DNA microarrays were used to characterize the genomic expression programs in wild-type and *mec1* mutant cells responding to two different DNA-damaging agents: the methylating agent methylmethane sulfonate (MMS) and ionizing (γ) radiation. MMS and ionizing radiation inflict different types of DNA damage by distinct mechanisms, so the study investigated whether each type of DNA damage agent triggered a specific transcriptional response, and, using *mec1* knockout mutants, whether those responses were Mec1-dependent [5].

Without being provided with the Gasch *et al*. publication, [5] the students were then given the broad framework of three central research questions: (i) Does gamma radiation cause a similar transcriptional response to another DNA damaging agent, MMS? (ii) Which genes are induced, which repressed, and are DNA damage response genes induced? (iii) Is the Mec1 kinase essential for the gamma and MMS transcriptional responses, or can a yeast *mec1* gene knockout also exhibit (partial) response to DNA damage?

Materials Required

All data and materials were made available to students through the University of Aberdeen virtual learning environment (WebCT). Students were also given a comprehensive practical guide (Supporting Information S1) containing detailed instructions as to how to carry out all the required analyses, including suggested hypotheses to follow.

Source data from [5] microarray experiments was downloaded from the source journal web site (http://www.molbiolcell.org/) and simplified for the purposes of the computer workshop; students received a limited subsection of the dataset comprising (i) time-course data for wild-type cells treated with either MMS or gamma radiation, and (ii) similar time-course data for *mec1* mutant cells treated with either MMS or gamma radiation.

Data was provided to students in the form of an Excel workbook.

The open-source microarray analysis program, MeV (Multiexperiment Viewer part of the TM4 Microarray Software Suite [9] was downloaded from http://www.tm4.org/mev/ onto classroom PCs to allow students to perform their own hierarchical cluster analyses of transcriptome data.

Hyperlinks to all other web-based bioinformatic tools for the workshops were also provided through WebCT. These included: FuncAssociate, used for GO function analysis, http://llama.mshri.on.ca/funcassociate/ [10]; The GO website itself, http://www.geneontology.org/ [11]; The *Saccharomyces cerevisiae* genome database, http://www.yeastgenome.org/ [12]; BLAST, for similarity searches, http://blast.ncbi.nlm.nih.gov/Blast.cgi [13]; InterProScan, for domain and pattern searches within proteins, http://www.ebi.ac.uk/Tools/InterProScan/, [14]; On-line Mendelian Inheritance in Man (OMIM) for examining human disease linkage http://www.ncbi.nlm.nih.gov/omim; and ClustalW, for multiple sequence alignments, http://www.ebi.ac.uk/Tools/msa/clustalw2/, [15].

Rationale and Structure of the Workshops

Each computer workshop lasted 2 hours, with each student having access to his or her own PC. The four workshops comprised the practical element of a single course; 8 courses similar to this, spanning a range of subjects, typically comprise an MSc program. Students worked individually or in small groups, receiving technical guidance and help with experimental interpretation from the instructors. Each workshop began with an introduction including aims, hypotheses and where appropriate discussion of the previous week's material. Students were told they were dealing with a real biological dataset, subject to experimental variation. The reproducibility and limitations of microarray and genome-wide datasets were discussed in the class. In future iterations of this computer practical, to fully exemplify this point, it would be desirable to then ask students to "hands-on" explore for themselves the magnitude of variation that can typically be exhibited between replicate microarray datasets. Although the published, web-download DNA damage response dataset that forms the focus of this practical [5] does not contain duplicate experimental datasets, there are other published examples of yeast microarrays that do contain replicate experimental datasets [16, 17], and these provide an excellent source of data using which to explore the issue of variability, using the tools and techniques described below in the descriptions of workshops 1 and 2. They were further reminded that transcriptome profiles identified variations in mRNA level, but that gene expression could also be controlled at the level of mRNA translation, thus additional proteomic analysis would ideally be required to develop a full picture of gene expression responses to DNA damaging agents. Further Supporting Information was provided in the form of a glossary of bioinformatic terms (Supporting Information S2) and a detailed Appendix containing guides to the major bioinformatics programs used (Supporting Information S3). This stand-alone guide contained detailed information on the algorithms used in each program, how to input data and how to interpret results.

Practical science courses are often designed in a simplistic manner, to cope with the demands of teaching large classes, and follow discrete ordered steps to guide students' inquiry through a precise path to complete an investigation [18]. However, this does not necessarily reflect how scientists carry out their own investigations; neither does it engender critical thinking in students. To develop critical thinking skills, it is essential to provide students with "open-ended" tasks, where there is no one right answer, and to provide an environment for interactive learning amongst students [19]. Brownstein [20] reflects that learners should be constantly challenged with tasks that refer to skills and knowledge just beyond their current mastery, to capture motivation, build on previous successful experiences and to enhance the confidence of students. Accordingly, and because these workshops were designed as a taught postgraduate (Masters) module, the concepts, data analysis and interpretation was designed to facilitate critical thinking and challenge students in this area. This computer practical in particular encourages such critical analysis, and presents students with surprising conclusions; the principal response of baker's yeast to exposure to gamma radiation or MMS treatment is not, as might be expected, composed primarily of genes that respond to DNA damage, but instead of a broad and apparently nonspecific heat-shock stress response [5]. Thus the students are compelled to readjust their expectations of the experimental outcome.

The acquisition of knowledge, and how that is translated into comprehension, is profoundly influenced by individual learning styles, categorized as primarily visual, auditory or kinesthetic [21]. We sought to tailor workshop content so that each type of learner could successfully access knowledge and understanding. Therefore, the workshops were designed to allow students to explore the data with as much freedom as possible. Students were then prompted to propose their own hypotheses (e.g., "MMS induces an identical transcriptional response to that of gamma radiation" or "MMS responses are Mec1-dependent" or "MMS elicits a more rapid transcriptional response than gamma radiation") and test these using self-selected datasets or groups of genes. They were also free to choose how to approach the research problem they themselves had identified, within the framework of three overarching questions (see "Biological Background" section). Thus, we hoped to mimic a research environment, within the constraints of the classroom. The caveat to this was that use of a range of bioinformatic tools should be taught; at least some meaningful experimental conclusions should be drawn within the context of the workshop; and ultimately, that structured assessment of student achievement should be carried out. Guidance and guidelines for marking the final assessment report are given in Supporting Information S4.

Workshop 1—The aims of this workshop were to show how transcript profile (microarray) information is analysed, and how information derived from this analysis can be used as the basis for a functional genomic

investigation. In essence, students were asked to identify the functional classes (GO) of genes that are induced or repressed by DNA damage. Students were introduced to the concept of GO classification (http://www.geneontology.org/GO.doc.shtml; [11] for assignment of proteins to functional groups. They then identified the GO function of each of these genes in yeast, using the FuncAssociate web server and looked for trends in the categories of genes induced. In the second task, students compared the responses of MMS and gamma treatment, by correlating levels of induction of genes in response to one DNA damage agent (e.g., MMS), with the response of those same genes to the other agent, gamma radiation.

Workshop 2—The aims of this workshop were to use hierarchical clustering software MeV to answer the questions (i) are genes with a DNA repair GO category induced at the transcriptional level by MMS or gamma radiation? (ii) more generally, can genes be clustered by distinct and individual responses to MMS and gamma radiation treatment time-courses. Students were provided with a list of all the genes in yeast that have GO category 6218: process = DNA repair. After a demonstration of the use of the Microsoft Excel functions MATCH and INDEX, students were able to fetch back the MMS time-course data for all the 168 DNA repair genes in their list. MeV was then used to view the expression profiles and perform hierarchical clustering of these 168 genes to examine their induction profiles. Then, using the whole microarray dataset, MeV was applied to investigate the induction of genes by MMS and gamma radiation treatments, in both wild-type and *mec1* genetic backgrounds. The typical "colored block" microarray representations generated by MeV allowed students to readily identify through visual inspection clusters of genes showing different classes of response to MMS and gamma radiation (e.g., induced by gamma, unaffected by MMS).

Workshop 3—Workshop 3 aimed to teach the students a range of bioinformatic tools to investigate protein similarity and predict protein function. Students were also given the freedom to choose the direction of their own research from this point on; they chose one of the genes identified in the previous workshops for further characterization. They were asked to identify homologous sequences in a variety of organisms using BLAST. They then performed multiple sequence alignments using ClustalW and predictions of protein function using InterProScan. Combining these analyses, they were able to identify conserved regions in the sequences, as well as those domains that are essential for protein function.

Workshop 4—The final workshop consolidated knowledge and understanding from the previous three. Students continued their investigation of homologues of DNA damage response genes in other organisms, particularly investigating human homologues, and identifying human genetic diseases linked to those genes. This exemplified the power of model genetic systems for the study of human disease. Finally, using the Saccharomyces Genome Database (SGD)[1] students identified other proteins interacting with their DNA damage response protein of interest, and then returned to the use of the microarray viewer MeV to answer the question "are groups of DNA damage response proteins, known to interact and form a stoichiometric complex, also coregulated at the transcriptional level by DNA damage?"

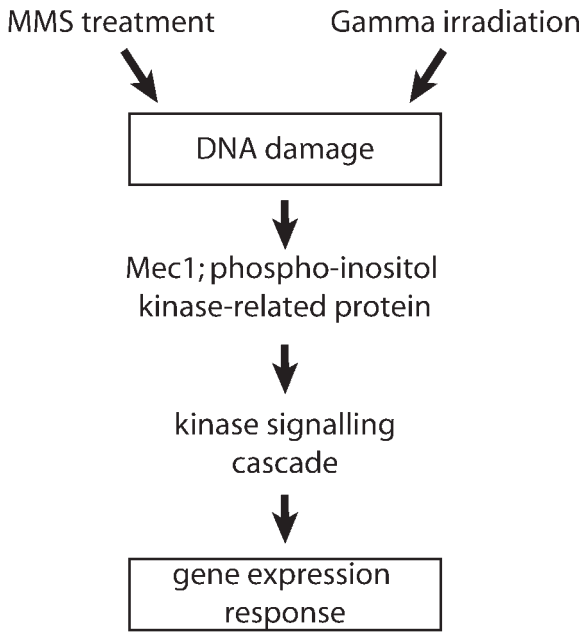

FIG. 1. **A model for the yeast DNA damage response**. MMS, methyl methanesulfonate, based upon an overview presented in [5].

RESULTS

Students were asked to prepare their results in the form of a research paper, with the results grouped under four defined headings to guide their definition and testing of research hypotheses. They were provided with guidelines as to the style, length and content of their report, that served to both mimic the requirements of scientific journals for publication, and to provide a transparent framework using which they could be assessed. Since students were free to choose genes and hypotheses to investigate within the workshop series, the content of each report reflected those individual choices.

Example Student Data

Here we have presented some of the results obtained by one of the students registered for an MSc Cell and Molecular Systems Biology, for which this computer workshop was a compulsory element (a full laboratory report is presented in Supporting Information S5).

Correlation of the Responses to MMS and Gamma Radiation Exposure—Students were asked to correlate the responses to MMS treatment and gamma irradiation as both causes DNA damage, and might be expected to elicit similar, overlapping transcriptional responses. Students plotted ln (induction ratio) of the gamma radiation and MMS transcriptional responses. An example graph

[1]The Abreviations used are: SGD, *Saccharomyces* Genome Database; MMS, methyl methanesulfonate.

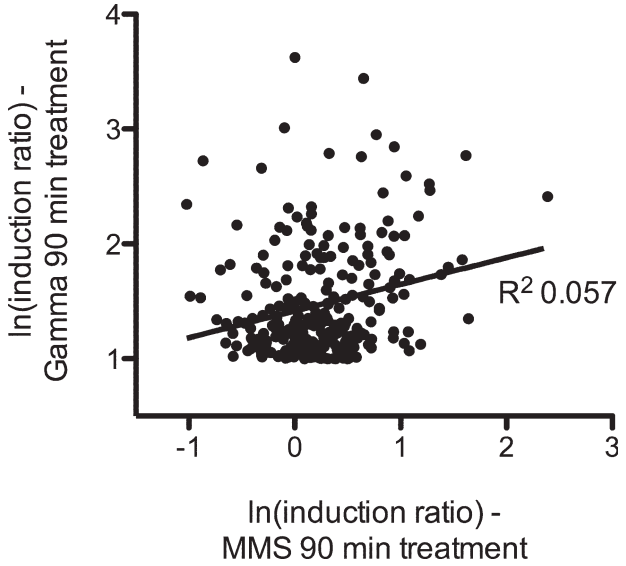

FIG. 2. **Scatter plot of ln(induction ratios) for *S. cerevisiae* gamma (γ) radiation induced genes (*x*-axis) versus ln (induction ratios) for MMS-induced genes (*y*-axis) in wild type cells, 90 minutes after exposure to each treatment.** Source data was taken from [5].

produced by a student is shown in Fig. 2, showing that those genes showing greater than 2-fold induction following 90 minutes MMS treatment are for the most part unresponsive to gamma radiation treatment. This simple correlation demonstrated to the students that cellular responses to DNA damage agents were likely to be complex and agent-specific.

The majority of students correctly identified that there was little correlation between those genes induced by MMS and those induced by gamma radiation. On a practical level this exercise did however reveal that in a typical class there was a wide range of Excel competence and familiarity, and some students undoubtedly struggled with the implementation of a basic scatterplot using the Excel graphing function. Therefore, care needs to be taken to explain to students how to correctly correlate datasets.

Identification of Genes and Their GO Categories Induced by MMS and Gamma Radiation Treatment—Students then examined the microarray data using the MeV viewer, for example, to compare the MMS and gamma radiation responses of specific subsets of genes with a GO category of DNA repair. In a separate exercise, the transcriptional responses of all 5,500 yeast genes to MMS and gamma radiation was examined. Using this approach, students were able to visually identify clusters of genes that were up-regulated in response to both MMS and gamma radiation. An example of a student-generated result of this type is shown (Fig. 3). Then, using FuncAssociate web server, over-represented categories within such a group of genes could be identified. In this case, the student correctly identified GO:9266 "response to temperature stimulus" as the most highly over-represented functional category within this group. This result is counter-intuitive (albeit correct), since the student might expect GO categories such as "DNA repair" (GO:6281) or "response to DNA damage stimulus" (GO:6974) to be returned by such analysis. For this reason analysis of the yeast DNA damage microarray dataset can productively stimulate class discussion.

Regulation of MMS and Gamma Radiation Responses by the MEC1 Gene—Students were asked to test the importance of the Mec1 pathway in the DNA damage response by comparing DNA damage responses in the *mec1* mutant with the same treatment in a wild-type cell. Students again performed correlation analyses of genes with a ln(induction ratio) of either >1, or <-1 90mins after exposure to MMS in wild type and *mec1* mutant cells. A student-generated example of such a scatter plot is shown in Fig. 4.

Orthologues of DNA Damage Response Genes, and Their Significance in Human Genetic Disease—Students then used a variety of bioinformatic tools to identify homologues of DNA damage repair genes in other organisms including humans. Students were then able to link some of these genes to human diseases using the OMIM database. Students were thus able to link functional studies in model organisms such as *S. cerevisiae* to gene function, and genetic disease in humans. Students were free to select their own genes with which to carry out orthologue analysis. An example of this is given in the student report (Supporting Information S5).

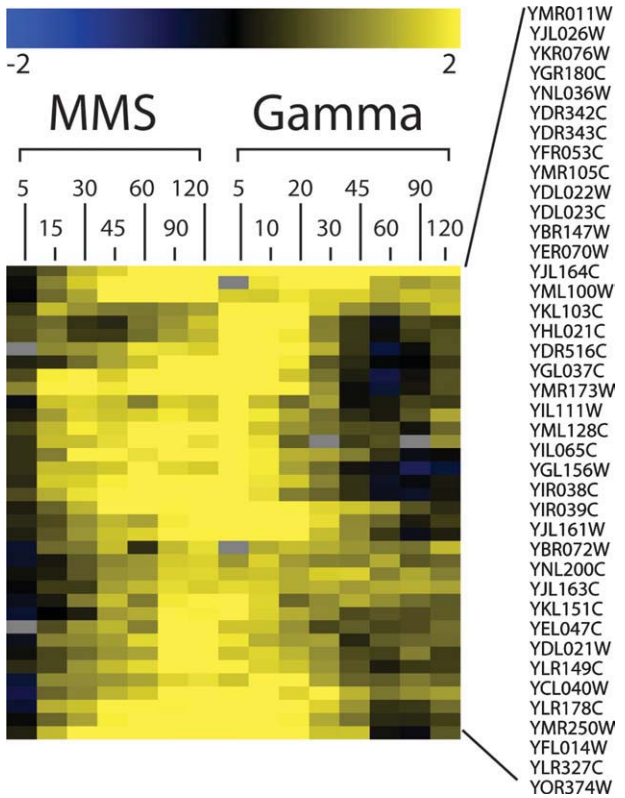

FIG. 3. **Microarray time-course profile of *S. cerevisiae* gene cluster that is induced by both MMS treatment and gamma (γ) radiation exposure, generated by microarray viewer MeV and subject to hierarchical clustering of genes with similar induction profiles (time points in minutes).** Induced genes are represented in yellow, repressed genes in blue. Source data was taken from [5].

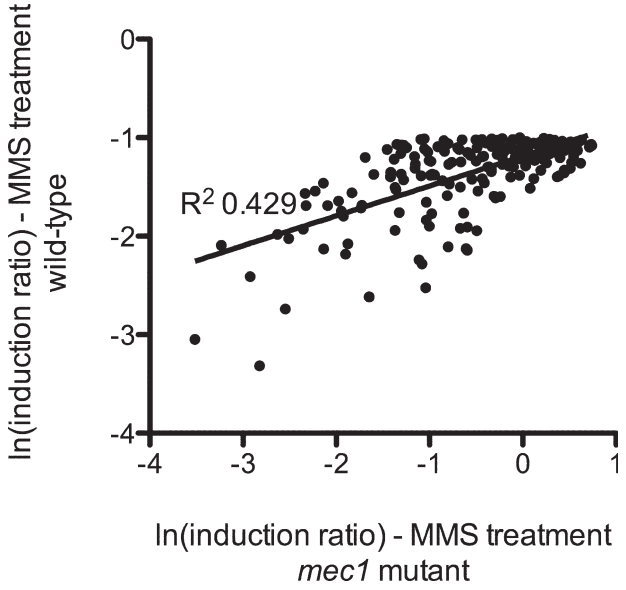

FIG. 4. **Scatter plot of ln(induction ratios) of those *S. cerevisiae* genes highly repressed in a wild-type cell in response to MMS treatment for 90 minutes (*x*-axis), versus the ln(induction ratios) for the same genes in a *mec1* mutant subject to identical treatment (*x*-axis).** Source data was taken from [5].

Evaluation of Student Learning and Educational Impact

Student learning and critical thinking skills were assessed using the practical report that the students wrote. There is also a final course examination, which requires the students to integrate their learning from both practical class and lecture course and provides an additional evaluation opportunity. Students completed standard student course feedback forms upon completion of the workshops. Overall the students' reaction to this exercise (ranked on a scale of 5 down to 1) was extremely positive. 52.6% of students ranked the workshops as 5 (totally effective for their learning), and 26.3% scored the workshops 4 (Fig. 5). On a voluntary basis, students were also asked a series of questions through a supplementary questionnaire designed to assess the educational impact of these workshops (42% of the class responded to the questionnaire in Supporting Information S6). This included six questions in which students had to rank their understanding and abilities of the specific tasks within the practical workshops, before and after the workshops had taken place.

All students who participated in this questionnaire believed there to be a gain in understanding and ability for all of the tasks performed during the workshops (Fig. 5).. Students ranked their perceived learning, understanding and ability on a scale of 1 (none) to 5 (complete) for each question. The majority of students (between 62 and 88% depending on the task the question related to) ranked their understanding or ability to perform a task as 1 or 2 before the workshops, and at 4 or 5 (between 50 and 96% for each question) after participating in the workshops. Free text responses indicated that students were pleased to be given the opportunity to try out bioinformatic analyses for themselves. Negative comments relating to these workshops centred on the issue that they sometimes had to wait for help from teaching staff, especially during the first two classes, highlighting the requirement for software problem resolution during the practicals.

Written reports were double marked using the University of Aberdeen's common assessment scale (CAS) in accordance with the University of Aberdeen's academic standards and quality assurance procedures. A range of abilities was recorded within the class of 30 students, with all students meeting or exceeding the minimum requirements to pass.

DISCUSSION

Student Responses

Overall the students responded very positively to these workshops. All students engaged with the tasks during the classes, many were happy to work independently, others worked in small groups. The final reports were well laid out, often bound and color printed, indicating that all the participants had taken care and effort in the generation of the reports and that they had fully engaged with the teaching.

Evaluation of Learning

It is clear from the student responses, both through the whole course evaluation and through the evaluation questionnaire for the practical workshops, that all the students questioned perceived that they had learnt a range of bioinformatic skills during the workshops. Student responses were generally very positive, and the majority of participants indicated that they had learnt what they expected to learn, and saw the addition of the practical workshops to the course as a very beneficial element, that aided their learning experience significantly. Instructors were able to have positive discussions with students throughout the course of the four workshops and build up relationships with the students. Students

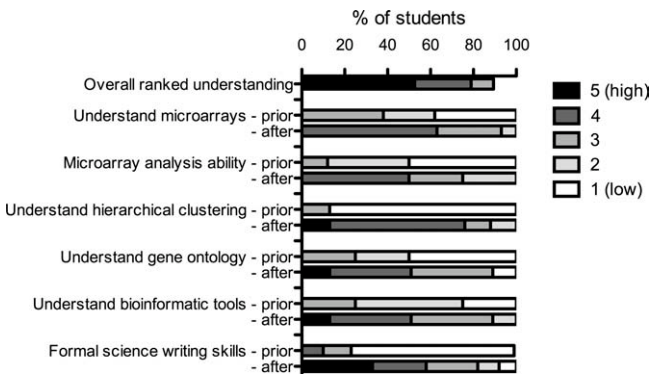

FIG. 5. **Bar chart of student feedback responses**. Students were asked to rank their level of understanding on a scale of 5 (high) to 1 (low) for the workshop series overall, then in response to more targeted questions, where students ranked their understanding in a particular subject area before, and after, the workshop delivery. Percentages of students responding at each point on the scale are reported.

were given the opportunity to ask questions and to relate what they were learning to the wider context of genomic and bioinformatic research tools. This dynamic interaction would not have been possible through a didactic lecture based delivery style.

Student feedback consistently highlighted that a key positive feature of the workshop series was the ability to learn basic bioinformatic techniques in class, but then to have the intellectual freedom to explore the large and complex microarray dataset in a manner of the student's choosing while preparing the report for assessment. Anecdotally, through discussion with students, the instructors learned that one principal frustration students feel with many class practicals, both "wet" and computer based is their canalized nature. The expected answer is preordained, and there is little latitude for true experimentation. The advantage of using a microarray dataset in this regard is its size and complexity. There are new discoveries to be made, and new inferences to be drawn from every dataset, particularly since functional genomics knowledge is ever-expanding. Furthermore, unexpected answers can emerge from microarray experiments, which further engage students' interest. In this regard, the DNA damage microarray experiment used in this workshop is intriguing in that many of the genes that are induced by gamma radiation or MMS treatment do not sit nicely within a DNA damage repair GO category, but rather are core components of a heat-shock/general stress response [5]. Such biological surprises can provide fertile ground for class discussion,

Practical Issues and Improvements for the Future

Each workshop lasted two hours and was carried out in a computer classroom, with two instructors present to facilitate learning. However, particularly for the first two workshops when students were first assimilating the aims and methods of the workshops, we found that instructor time was in heavy demand, and future iterations of the workshops will engage more demonstrating staff. This issue was highlighted in student feedback. During the course of the practical workshops, a wide range of background knowledge and familiarity with basic and specialized computer programs was noted throughout the group, with some students very familiar and confident in the use of most of the programs, and others unable to perform basic data manipulation in Microsoft Excel. Additionally, some students struggled to complete the tasks within the 2 hours while others finished early. The main student pitfalls were misinterpretation of the data at an early stage of the workshop series; this appeared to be related to a lack of knowledge of, or confidence with, Microsoft Excel. Therefore, in the future, we propose to deliver, to those students that need it, basic Excel familiarization training. Running this additional training before the start of the practical course will allow the workshop time to be used to more fully explore bioinformatic techniques and microarray data, rather than to try to individually help students with Excel.

The workshops are not expensive to run, and easy to administer practically, since all that is required is a computer classroom; all software (with the exception of Microscoft Excel) is open-source and hence freely available. Additionally, students are able to carry out the practical workshops themselves at a remote location as long as they have access to a computer and an internet connection.

CONCLUDING REMARKS

Microarray Data as a Tool for Research Skills Development in Students

We have designed a series of practical workshops that can be used at Masters level or advanced undergraduate level, that use real research data to teach genomic and bioinformatic skills to students. Published microarray data is an ideal tool for student data analysis workshops. Students get a flavor of real research skills, can attach a relevance and meaning to what they are doing and are free to explore large datasets in a self-directed manner. By using data published from a model organism such as *S. cerevisiae* a wealth of freely available supporting genomic resources can be used to complement the microarray analysis. All the online bioinformatic programs used in these workshops are freely available open source software. Students appreciate the opportunity to work with real research data and learn key problem solving skills that will be of use if they pursue research careers in a wide variety of bioscience areas. Material and explanations introduced in a lecture-based setting can be explained in context and reinforced, and students also get the benefit of learning how to present research findings in publication style format. Therefore microarray data, especially that generated from model organisms such as *S. cerevisiae*, is an ideal research playground for advanced undergraduate or graduate students wishing to learn real genomic and bioinformatic skills.

Acknowledgments— L.G.B. gratefully acknowledges the University of Aberdeen for funding her to complete a Post-Graduate Certificate of Higher Education (Teaching and Learning), for which this study forms the central research project. The authors are also grateful to Mrs. Lee Picken (University of Aberdeen Graduate School) for compilation and analysis of the whole course evaluation reports. They also thank student Beth Marshall for examples of data referred to throughout the text, and for providing the student report submitted as Supporting Information S5. They again thank Russell Betney for assistance in instructing the class and Joy Perkins for critical reading of the manuscript.

REFERENCES

[1] B. Dalton, C. C Morocco, T. Tivnan, P. L. Rawson Mead (1997) Supported inquiry science: Teaching for conceptual change in urban and suburban science classrooms, *J. Learn. Disabil.* **30**, 670–684.
[2] R. Driver, B. Bell (1986) Student's thinking and learning of science: A constructivist view, *School Sci. Rev.* **67**, 443–456.
[3] R. Breitling (2010) What is systems biology? *Front. Physiol.* **1**, 1–5.
[4] T. Ideker (2004) Systems biology 1010-what you need to know, *Nat. Biotechnol.* **22**, 473–475.
[5] A. P. Gasch, M. Huang, S. Metzner, D. Botstein, S. J. Elledge, P. O. Brown (2001) Genomic expression responses to DNA damaging agents and the regulatory role of the yeast ATR homologue Mec1p, *Mol. Biol. Cell.* **12**, 2987–3003.
[6] S. J. Elledge (1996) Cell cycle checkpoints: Preventing an identity crisis, *Science* **274**, 1664–1672.
[7] W. A. Cliby, C. J. Roberts, K. A. Cimprich, C. M. Stringer, J. R. Lamb, S. L. Schreiber, S. H. Friend (1998) Overexpression of a

kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints, *EMBO J.* **17**, 159–169.

[8] J. A. Wright, K. S. Keegan, D. R. Herendeen, N. J. Bentley, A. M. Carr, M. F. Hoekstra, P. Concannon (1998) Protein kinase mutants of human ATR increase sensitivity to UV and ionizing radiation and abrogate cell cycle checkpoint control, *Proc. Natl. Acad. Sci. USA* **95**, 7445–7450.

[9] A. L. Saeed, N. K. Bhagabati, J. C. Braisted, W. Liang, V. Sharov, E. A. Howe, J. Jianwei, T. Mathangi, J. A. White, J. Quackenbush (2006) TM4 microarray software suite, *Methods Enzymol.* **411**, 134–193.

[11] The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology, *Nat. Genet.* **25**, 25–29.

[10] G. H. Berriz, J. E. Beaver, C. Cenik, M. Tasan, F. P. Roth (2009) Next generation software for functional trend analysis, *Bioinformatics* **25**, 3043–3044.

[12] K. R. Christie, E. L. Hong, J. M. Cherry (2009) Functional annotations for the *Saccharomyces cerevisiae* genome: The knows and the known unknowns, *Trends Microbiol.* **17**, 286–294.

[13] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**, 403–410.

[14] S. Hunter, R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, C. Yeats (2009) InterPro: The integrative protein signature database, *Nucleic Acids Res.* **37** (Database Issue), D224–D228.

[15] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson (2003) Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res.* **31**, 3497–3500.

[16] E. Talla, F. Tekaia, L. Brino, B. Dujon (2003) A novel design of whole-genome microarray probes for *Saccharomyces cerevisiae* which minimizes cross-hybridization. *BMC Genomics.* **4**: 38.

[17] M. D. Piper, P. Daran-Lapujade, C. Bro, B. Regenberg, S. Knudsen, J. Nielsen, J. T. Pronk (2002) Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**: 37001–37008.

[18] X. Tang, J. E. Coffey, A. Elby, D. M. Levin (2010) The scientific method and scientific inquiry: Tensions in teaching and learning, *Sci. Educ.* **94**, 29–47.

[19] B. Potts (1994) Strategies for teaching critical thinking, *Pract. Assess. Res. Eval.* **4**, 1–4.

[20] B. Brownstein (2001) Collaboration: The foundation of learning in the future. *Education* **122**, 240.

[21] A. H. Johnstone (1997). And some fell on good ground, *Univ. Chem. Educ.* **8**, 11–13.