

Design, delivery and development of a modern and translational genomics model for bioscience students at The University of York

Andrew S Mason ¹

Department of Biology and York Biomedical Research Institute, The University of York, UK

Coding and data science skills have become a core component of bioscience undergraduate degrees in the majority of UK universities, reflecting the importance of these skills in research, industry and wider graduate employment. These skills are crucial for the analysis and interpretation of 'omics' based data, where large datasets are used to explore biological processes at scale. Most undergraduate degrees do not provide a direct link between these coding skills and the analysis of raw sequencing data, leaving students unprepared or unaware of the bioinformatic skills required for a research career.

To develop these skills in York undergraduates, a new, optional stage 3 module 'Genomics' was designed after a review of programmes at other Russell Group universities, to deliver modern and translational genomics research-led teaching. This module was first delivered in semester 1 of the 2023/24 academic year and this report represents a critical review of the design and delivery of the module to facilitate its continued development.

Genomics students had already demonstrated good data acumen through first- and second-year courses. Now, by building on existing R skills and learning the Linux command line, students processed, analysed and interpreted prokaryotic and eukaryotic raw sequencing data to address novel research questions. Students were exposed to cutting edge genomics research and techniques through lectures and asynchronously-delivered videos. Broadly distributed, but often exceptional results highlighted the propensity of this new module to challenge the students and to develop and assess research-specific skills.

Overall, the first delivery of Genomics was highly successful, but this review has identified four achievable proposals to improve teaching practice in this module.

genomics; bioinformatics; data science; biology; coding; omics

Word count: 3955

¹ Email: andrew.mason@york.ac.uk

Introduction to the teaching of genomics and bioinformatics

Until the human genome project, computational biology (or “bioinformatics”) was relatively niche, with most biological research focused on traditional laboratory or field studies. However, with the advent of high throughput and increasingly inexpensive technologies for reading the sequence of DNA and RNA within cells and tissues (“genomics”), there has become an increasing need for data science skills for all biologists (Marx, 2013; Leonelli, 2019).

Initially the scale of the data was small, so browser-based tools became the main choice for researchers, with many resources such as NCBI’s BLAST (Altschul *et al.*, 1990), the UCSC (Kent *et al.*, 2002) and Ensembl (Harrison *et al.*, 2023) genome browsers, and the EMBOSS suite (Rice, Longden and Bleasby, 2000) still attracting thousands of users daily, and running their own series of tutorials and webinars. Consequently, these tools have also become the main exposure to bioinformatics in undergraduate education, as they have clear user interfaces, are robust to changes in university computer architecture, and require very little updating of material (Attwood *et al.*, 2019; Brown, 2016; Suciu, Aydin and Chen, 2015). However, bioinformatics and genomics are rapidly changing fields, with technologies commonplace today (such as single cell RNA sequencing) which did not exist 10 years ago. Consequently, our teaching practices must constantly update as well.

Another challenge to this browser-based dogma has been the sheer scale of data now available to biologists, making programming an essential skill (Leonelli, 2019; Mariano *et al.*, 2019), as even high-powered user interfaces such as Galaxy (Afgan *et al.*, 2016) struggle to cope. Whilst this has become more common in bioscience undergraduate degrees, until recently it has often been limited to postgraduate level, even though many PhD and industry roles now request these skills from graduates (Gatherer, 2020). A common challenge at all levels has been resistance to learn these new skills as not being traditional biology (Carey and Papin, 2018). A key aim of computational teaching to biologists must therefore be contextualising these skills within answerable (and interesting) biological questions, requiring biologists (and not just computer scientists or university IT) to deliver these materials.

The issue then becomes how to make that leap, as informatics has a steep learning curve and many datasets are large and complex, requiring substantial computational power to analyse. Often students are provided with model or heavily subsampled datasets, or simply fully processed data where analysis is more about plotting than statistics. This falls short of the skills biologists need to work with their own large datasets, and fails to expose bioinformaticians-in-training to research projects which might spark their interests in postgraduate work. This requires design and development of bespoke genomics modules.

Genomics and bioinformatics undergraduate teaching at The University of York

All undergraduate students associated with the degree of Biological Sciences (BSc/MBiol) have core R coding teaching throughout stages 1 and 2. With ‘semesterisation’, these modules have

been serialised into “Becoming a Bioscientist (BABS)” modules 1 to 4, where data science is taught alongside traditional laboratory skills. Many teaching staff naturally introduced genomics data into their teaching, though it was not until semester 2 of stage 2 (“BABS4”, formerly “Big data biology” taken by current stage 3 and 4 students) that over 80% of students are explicitly exposed to bioinformatic analysis of genomics data.

Until the 2023/24 academic year, Big data biology was the last specific bioinformatics module available to students. Given the predominance of genomics and bioinformatics, this was reviewed to introduce a new “Genomics” module to stage 3 bioscience undergraduates.

Scope of research

This critical review covers:

- i) strategies to co-develop the Genomics module by the author and the module organiser (MO), Dr Daniel Jeffares;
- ii) an assessment of the students attracted to the Genomics model as a feature of their data acumen;
- iii) a critique of the delivery and assessment of the module incorporating student feedback and teaching team reflection; and
- iv) identification of achievable proposals to develop and refine the Genomics module ready for second delivery in semester 1 of 2024/25.

The overall aim of the review is therefore to improve the quality of teaching practice for this new and exciting module.

Methodology

Bioscience undergraduate exposure to coding and omics teaching in the Russell Group

The Russell Group contains 24 Higher Education Institutions in the UK. Publicly available 2023/24 course and module catalogues for Bioscience related undergraduate programmes were reviewed online during June, July and August 2023. The London School of Economics and Political Science was excluded due to the absence of a Bioscience undergraduate degree.

Programmes were reviewed for their earliest compulsory or optional delivery of: i) coding, and ii) omics-based teaching material. A clear distinction was drawn between “genomics” and “genome structure/organisation”. Taught coding language and computational tools applied in the genomics modules were recorded in all cases.

Development and delivery of the stage 3 Genomics module at The University of York

The concept of a Genomics module was approved for first delivery in stage 3 semester 1 of the 2023/24 academic year in Spring 2023. The teaching team was the module organiser (MO; Dr Dan Jeffares, Biology) and the author of this report, who worked collaboratively to design the specifics of the module. Module design and content creation was fine-tuned through Summer 2023 ready for delivery from September.

Module design focused on delivery of cutting edge, up to date and practical genomics teaching, rather than a “history of genomics” or purely theoretical material. The module assumed some experience of coding and handling large biological datasets, as covered in stages 1 and 2 by Biology students. No ‘wet lab’ components were to be included. Module design integrated features from the review of genomics teaching in the Russell Group, with particular inspiration from The University of Edinburgh “Genomes and Genomics” module.

Review of the 2023/24 first Genomics module cohort assessment results

The Genomics assessment marking was completed in February 2024. Student surnames were used to query stage 1 and stage 2 module scores (after resit if applicable) for each student using the University of York eVision portal. Average results were calculated for each stage, and the results from the Biology-specific data-focused modules in stage 1 (compulsory; Laboratory & Professional Skills for Bioscientists - BIO00017C; “17C”) and stage 2 (optional; Big data biology - BIO00047I; “47I”) were extracted and also averaged. Cohort-wide Big data biology module results were also available from 2022/23 (the matched cohort).

Throughout the analysis, the presence of missing values required individual students to be removed from particular statistical tests or plots. Missing values were due to: i) transferring to Biology after stage 1 (2); ii) completion of an exchange programme during stage 2 (1); iii) focus on biochemistry modules so did not complete 17C or 47I (3); and iv) biology students not taking the optional 47I module in stage 2 (9). Furthermore, the stage 1 and stage 2 results

for 3 enrolled BSc Natural Sciences students were not available, so these were removed from all analyses, as was data from a visiting year abroad student.

Following this initial curation, student names and unique identifiers were removed, the data rows shuffled, and then anonymous identifiers assigned following the format: uoy01.

Responsive and reflective module review by teaching staff

Student feedback was collected informally during the module by the MO in weekly 'office hours' sessions, and by teaching staff at lectures and workshops. Teaching staff met regularly to discuss any issues with delivery of the new module, responding by altering future material where relevant. Teaching staff met at the end of the module to critique all aspects of the module, and discussed student-submitted feedback (below) with pertinent themes and actions summarised here.

Critical review of student-submitted feedback

Students submitted anonymised feedback through the official departmental end of semester feedback forms. Following return of the assessments, some students submitted additional anonymised feedback through a Google form set up by the stage 3 student representative, who then shared these comments with the teaching staff. Feedback from both submissions were collated and summarised with relevant themes presented here.

Statistical analysis and plotting

All quantitative data were processed and analysed using Python v3.10.7, using the pandas v1.5.0, NumPy v1.23.4, SciPy v1.9.1, Matplotlib v3.6.0 and Seaborn v0.12.0 libraries. Data were visualised and checked for normality before choosing parametric or non-parametric statistical tests. $p < 0.05$ was the accepted threshold for statistical significance throughout.

All analysis has been recorded and summarised through a jupyter notebook available at: <https://gist.github.com/asmasonomics/a92503549e4be37b4c0c18506adb9c0b>.

Results

The University of York compares favourably for coding and omics teaching in the Russell Group

47.8% of Russell Group universities, including York, provide compulsory coding modules in at least one language to their undergraduate Bioscience students within the first two years of study (Figure 1A). The predominant coding language was R (16/18 universities across compulsory and optional modules, including at York). Coding teaching was often packaged into “essential skills” modules along with traditional laboratory work, rather than in specific statistics, data science or bioinformatics modules. The University of York’s strategy is now consistent with these other universities, as the stage 2 Big data biology (471) module has been incorporated within “Becoming a Bioscientist 4”, which includes laboratory practicals.

Compulsory teaching of genomics was rare across Russell Group universities (8.7%; Figure 1B), but was a common optional module in later years. Most universities which did not provide specific coding teaching also did not provide genomics teaching. Detailed review of genomics teaching strategies showed that in most cases students were provided with partially processed data or used graphical user interfaces (GUIs) such as Galaxy, rather than working with raw data at the Linux command line. Consequently, students were left unprepared to work with ‘real’ genomics data.

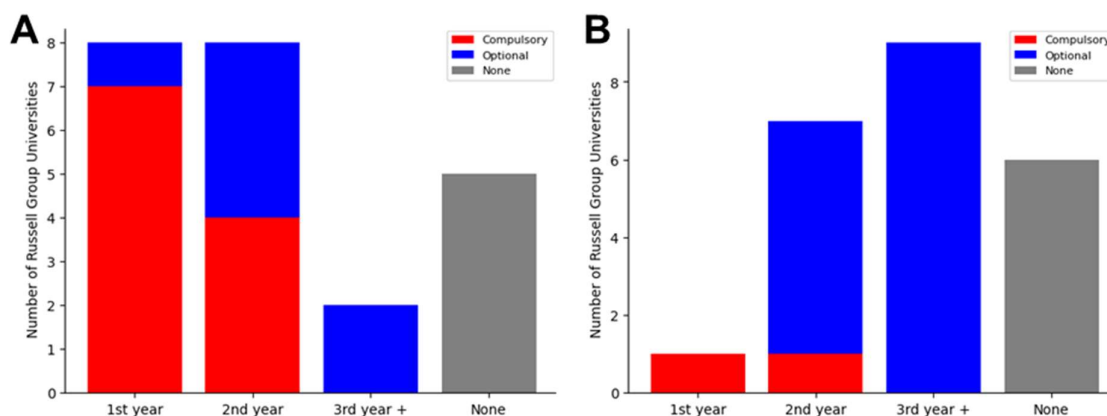


Figure 1: First exposure of coding and genomics modules to Bioscience undergraduates. Coding modules are typically first delivered in stages 1 and 2, and are often compulsory (A), whereas genomics modules are commonly optional (B). Data sourced from the 2023/24 publicly available module catalogues.

Design of a modern and translational Genomics module

The aim of this module was to provide real practical experience of working with raw data in genomics and to provide students with the background knowledge of theory and bioinformatic techniques to interpret such analyses. Most stage 3 students at York have the benefit of two years of compulsory R teaching. True bioinformatic analysis of genomic data requires experience working with Linux, and York has available managed compute to facilitate

this experience for Bioscience students. Furthermore, development of diverse data science skills, as well as an under-considered appreciation of ethical issues in genomics (with broader impacts to public health for all), and dissemination of results are all highly transferrable skills for graduates moving to careers outside bioscience and academia (Table 1).

Table 1: Genomics module learning outcomes and associated teaching. Delivery schematic presented in Figure 2.

Learning objective	Matched teaching/assessment
1. Comprehend and critique different genomic techniques	Lectures, independent study videos
2. Explain applications of genomics across biology	Lectures, group poster session, workshops
3. Design experiments that use genomic technology to solve a problem	Workshops, summative project report
4. Conduct analysis of genomic data using a variety of computational tools	Workshops, summative project report
5. Determine novel biological insight from newly generated sequencing data	Workshops, summative project report
6. Critically analyse research papers in genomics	Weekly set paper, summative project report
7. Explain ethical aspects of genomics	Ethics workshop

To facilitate the practical elements in the workshops, theoretical content was ‘front loaded’ into the module supported by independent study videos (Figure 2). Genomics courses at other universities, notably The University of Edinburgh, had almost double the contact hours than is standard at York. Videos were used to deliver important material outside the scope of lectures and in a format more easily digestible when the content was very different to that experienced previously. However, there was no clear metric for assessing student engagement with these videos.

Group poster work was initially proposed as a summative assessment (c.20%), formally assessing engagement with the lecture material while the report assessed practical skills and interpretation. This was rejected by the department due to timing of the session in the semester, and concerns over having multiple assessments.

Workshops 3, 4, 5 and 7 were designed as standalone elements which could individually form the basis of the summative assessment, with students expected to go beyond the workshop material to access higher grades. Two further workshops were used to i) guide students through the assessment, and ii) provide additional time and support for completing project work without set material. The ethical workshop was the greatest unknown, as students needed to engage in group and consider diverse, multifaceted viewpoints on ethical questions in genomics. The author had conducted preliminary work through stage 2 tutorials, but it was unclear how successful this would be for a full class, particularly towards the end of the semester with material not directly relevant for the assessment.

	LECTURES	WORKSHOPS	IND STUDY	ASSESSMENT
W1	L1. Module outline L2. Principles in Genomics		DataCamp Linux Video 1: Linux I Paper reading	
W2	L3. Genome size & style L4. Small genomes		DataCamp Linux (cont.) Video 2: Linux II Paper reading	
W3	L5. Large genomes L6. Transcriptomics		DataCamp Linux (cont.) Video 3: Linux III Paper reading	
W4		Workshop 2: Linux IV	Video 4: DNA sequencing Paper reading	Workshop 1: Posters Q&A
W5	L7. Cancer genomics	Workshop 3: Assembly	Video 5: Assembly Paper reading	Group poster work
W6	L8. Population genomics	Workshop 4: RNAseq	Video 6: Read mapping	Group poster work
W7			Video 7: Genome annotation	Formative group poster presentations
W8		Workshop 5: Gene finding	Video 8: Genome consortia	Workshop 6: Project guidance
W9		Workshop 7: Population genomics	Video 9: Genome browsers	Analysis & report writing
W10		Workshop 8: Consolidation & catch up		Analysis & report writing
W11		Workshop 9: Ethics of Genomics		Report Drop-In Guidance Analysis & report writing

Figure 2: Genomics module plan. Delivery across the semester was largely in two sections, with genomics content through lectures in the first half, practical skills through workshops in the second half, and instructional video material throughout to support learning outcomes. Formative group assessment (W7) focused on application of lecture material, and the summative assessed report required extension of material and skills developed in one or more of workshops 3, 4, 5 or 7.

Students taking Genomics were already data-focused learners

One working assumption was that the majority of Genomics students would have completed Big data biology (47I) in second year. This was the case (32/49; 65.31%), but represented a lower proportion than was expected. The 2022/23 47I cohort numbered 185 students, with a very broad distribution of results around a mean value typical for stage 2 modules of 63.72% (Figure 3). The 32 students who went on to do Genomics achieved significantly higher in 47I compared with those who did not (MWU=3701.0; $p=9.70 \times 10^{-5}$), averaging 71.03% (Figure 3).

These eventual Genomics students appeared to develop particular propensity for the data-focused module during second year. Whilst the 17C results in first year were well correlated with stage 1 averages ($p=0.69$; Figure 4A), 17C grades were not significantly different from them (MWU=295.0; $p=0.19$). However, by stage 2, these students were performing

significantly better in 47I than their second averages (MWU=56.0; $p=2.84 \times 10^{-5}$), particularly for those students averaging a 2.1 or 2.2 (Figure 4B).

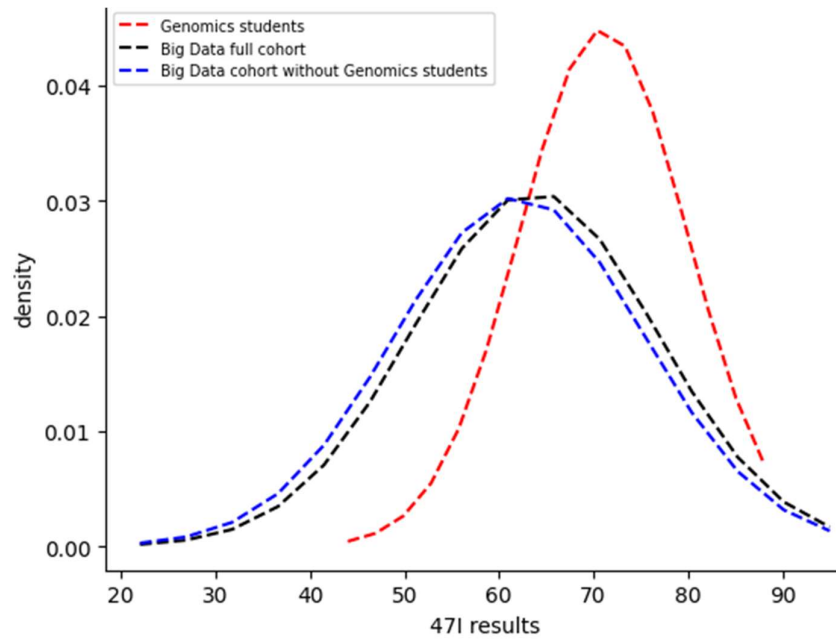


Figure 3: Students taking Genomics were higher than average achievers in Big data biology (47I). Dotted lines represent cumulative normal probability density functions from equal 15-bin histograms from each indicated set of data. Full Big data biology cohort: $\mu=63.72$, $\sigma=12.97$; Genomics students only: $\mu=71.03$, $\sigma=8.89$; Big data biology cohort without Genomics students: $\mu=62.18$, $\sigma=13.16$.

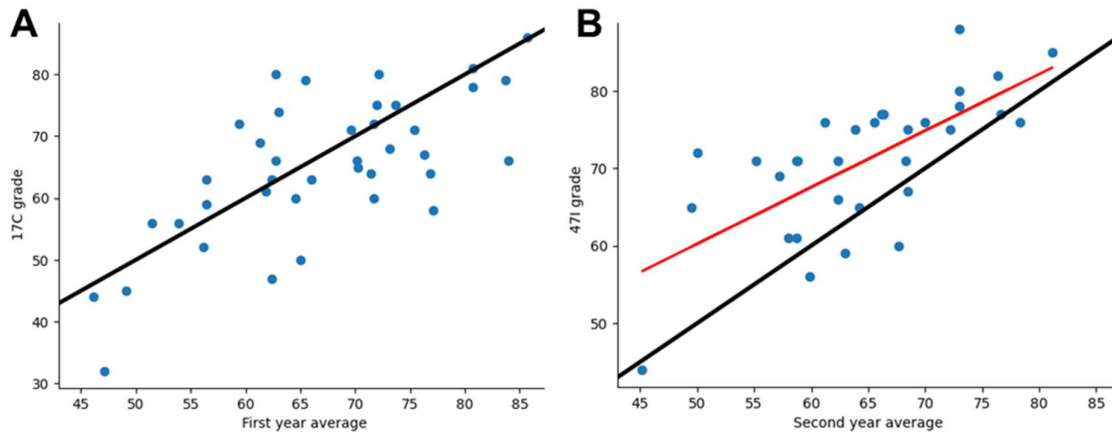
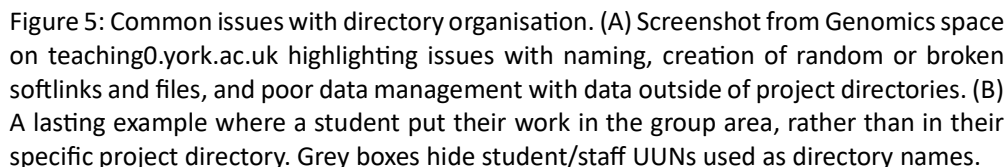


Figure 4: Genomics students' Big data biology results reveal data acumen exceeding their second year averages. (A) 17C data-focused module in stage 1 correlates well with first year average ($p=0.69$; $p=9.90 \times 10^{-7}$) with scores highly congruent on black $x=y$ line (MWU=295.0; $p=0.19$). (B) 47I also correlates well with 2nd year average ($p=0.69$; $p=1.21 \times 10^{-5}$), but 81.25% of students achieved higher than their 2nd year average (MWU=56.0; $p=2.84 \times 10^{-5}$), with the effect greatest for students with lower averages (red regression line indicated).

Linux was introduced to the students through DataCamp, a large, online and asynchronous training suite with videos and live coding checked by the software portal. Students responded favourably to this and most were happy to work on the university Linux servers from workshop 2 (week 4; Figure 2). Some students struggled with the absence of a GUI and lacked clear directory structures and file organisation; features also observed in first and second year data teaching (Figure 5). This often created problems and frustrations, including the erroneous complaint by some that “the code doesn’t work”. This required careful management from teaching staff to ensure students remained confident to try things and fail again, learning more each time. The majority of students were working confidently on Linux by the end of the course, though confidence performing fully unguided bioinformatic analysis was not tested.



The Genomics summative project report required biological interpretation and technological skill; bringing together the twin facets of genomics in research. The 2000-word report included a suggested 1000-word introduction, engaging with primary literature to build a narrative for the presented results. Students were then assessed on their development and personalisation of workshop material, presentation of data, critique and discussion of results, and coding documentation. The class mean was 66.11%, with a median of 65%. There was a broad range of results (43%-89%), with over a third gaining a first class (Figure 6A).

10

multiple linear regression ($F=824.3$; $DF_n=2$; $DF_d=30$; $p=6.06 \times 10^{-27}$), second year average was not itself a significant variable ($p=0.085$), but the 47I result was ($p=0.031$).

Considering all enrolled students against their second year averages (which was strongly correlated with performance in Genomics; $p=0.48$), this elevation of scores was not consistent, with 11 students scoring between 5 and 20 marks lower than their second year average, and 11 scoring the same above their average (Figure 6C, 6D). However, this was not due to lower grades from those students who did not complete 47I, with no significant difference between their results ($MWU=172.5$; $p=0.62$). Instead, these data are congruent with students being tested on different skillsets to most modules, with a particular focus on data acumen. Consistent with this, but somewhat anecdotally, both students who had completed years in industry performed far better than their second year averages indicated.

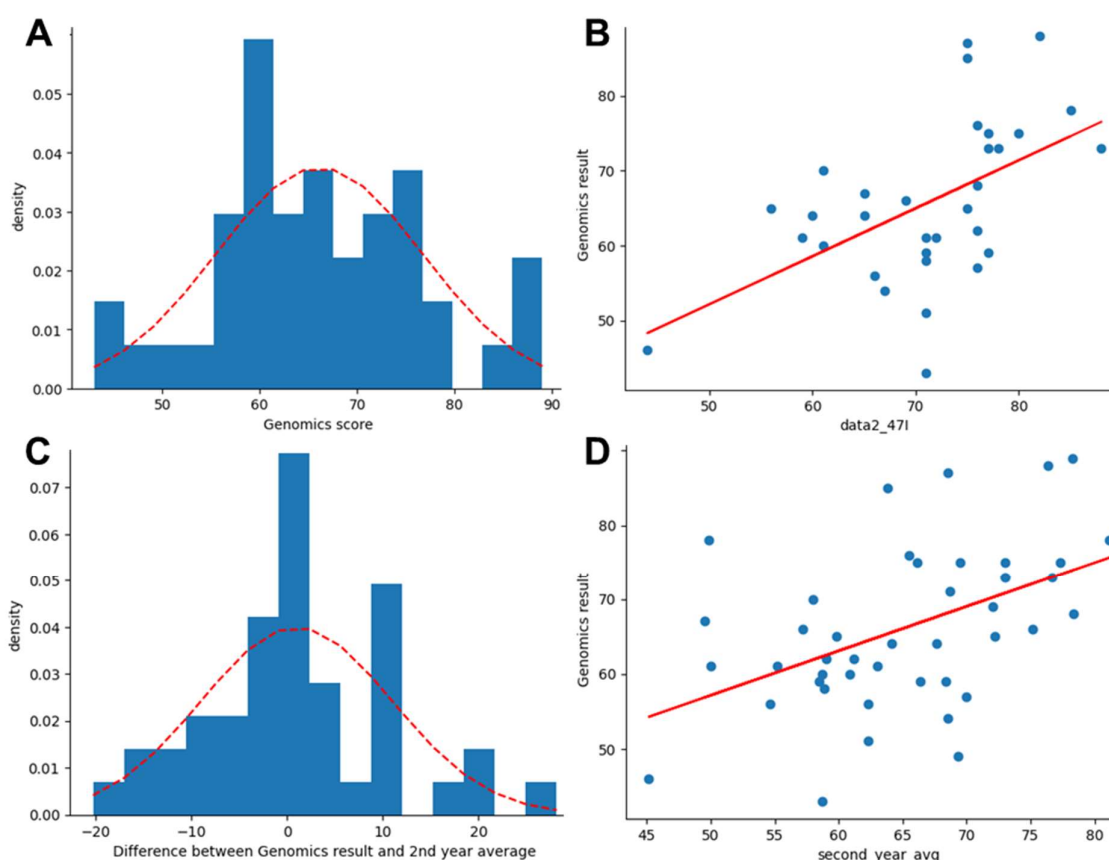


Figure 6: Genomics assessment correlated with second year performance. (A) Summative assessment produced 46 point range with no fails and 36.36% gaining a first class mark: $\mu=66.11$, $\sigma=10.66$, $\tilde{x}=65.0$. (B) Big data biology scores correlated positively with Genomics ($p=0.53$; $p=1.62 \times 10^{-3}$), with students typically improving on Big data ($t=3.15$; $p=3.60 \times 10^{-3}$). (C) Genomics scores produced highly divergent results compared with second year averages. (D) Good correlation maintained ($p=0.48$; $p=1.04 \times 10^{-3}$), but divergent results rendered paired measures non-significant (paired $t=-0.65$; $p=0.52$). Dotted red lines represent cumulative normal probability density functions from equal 15-bin histograms from each indicated set of data. Solid red lines indicate linear regression lines.

Student feedback was positive overall

30.61% of students responded to the departmental feedback survey (average 3rd year response rate was 29%), giving an average overall rating of 4.0 for Genomics, consistent with the 3rd year average of 3.9 (Figure 7A). Students perceived that the workload was comparable to other modules, tending to higher (Figure 7B). The knowledge, helpfulness and personable nature of the teaching staff was highlighted consistently, as was the relevance of the topics and use of Linux. Suggestions for improvements consistently mentioned the robustness of workshop material and the availability of material in advance through the VLE. Students really enjoyed the group poster work and the poster session, but many felt frustrated that this was formative assessment only.

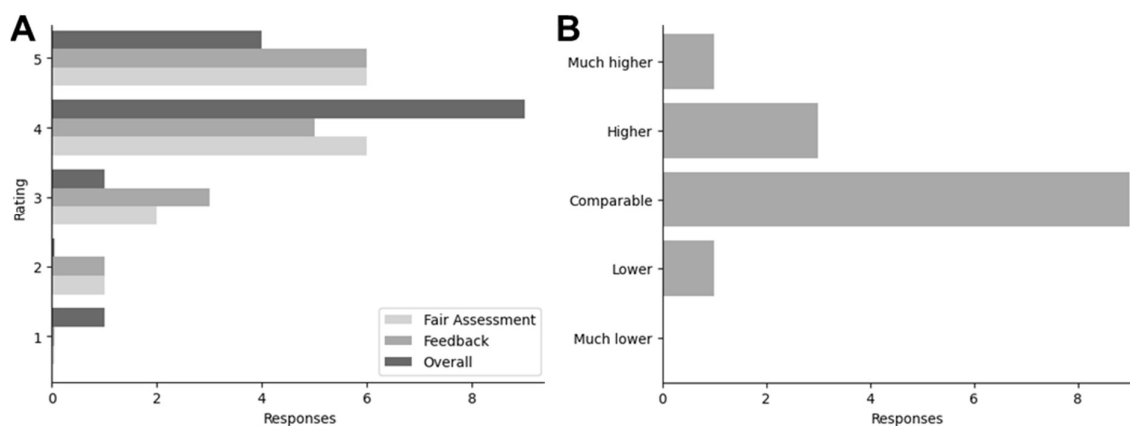


Figure 7: Student feedback across all categories was predominantly positive. Data were collected as part of the Department of Biology feedback forms (n=15; 30.61% of cohort). (A) Module ratings out of 5. (B) Perception of workload relative to other stage 3 modules.

Additional, more critical, feedback came from some students after assessment scores and feedback were released. This likely reflected the highly differential marks observed across the cohort compared to second year averages (Figure 6C), and the somewhat limited feedback given per marking rubric section, rather than as a single and longer overall comment. Assessment marking was completed through Turnitin for the first time for both markers, with students receiving a rubric with reduced section names lacking the explanation of marking criteria. A particular issue here was that many students received low marks for the “Biological Insight and Technical Skill” section (maximum score 15) as this was intended as the “excellence criteria”, but this was not clear to students.

The MO analysed marks for consistency between chosen topic and maker (finding no differences; data not shown), and released cohort-wide feedback and an expanded rubric to help students interpret their marks better. Students were especially keen to use feedback from the Genomics project report to directly inform and improve their capstone project writing.

Discussion

Timeliness and relevance of the Genomics module

The provision of coding and genomics-based teaching to Bioscience undergraduates at York is now above the average student experience from a Russell Group university. The adoption of Linux into the Genomics curriculum is a real advance beyond York's peers, as students are now exposed to authentic data analysis, rather than a focus on analysing processed data with R or Python, or simply completing browser-based problem solving.

Adoption of Linux does come with challenges. Students have a new coding language to learn and many have not been exposed to 'proper' command line coding without a GUI. Even in a system as intuitive (and familiar) as R Studio, many students do not appreciate file and directory structures and this is exacerbated without the familiar methods for visualising. DataCamp (or another guided system where a virtual environment is used to practice coding where answers are automatically checked) was well received by the students and provided a good foundation for the workshops. Another challenge is that whilst students can work with raw data, they can't feasibly work with all types of raw data due to limitations on server RAM and cores, and the time large jobs take to run, potentially days. Students would not be able to identify mutations in a Eukaryotic genome, for example, an area of interest to any biomedical students looking towards careers in cancer research, an area of research strength at York. Working with unicellular organism data (workshops 3, 5 and 7) is often more tractable for teaching, but may be of limited appeal to students with diverse interests.

This diversity is particularly relevant when considering the translational role of the Genomics module. There is clear data science relevance for graduates across multiple sectors, but this module also facilitates development of an academic career. Outside the final year capstone projects, Genomics is one of the few modules where students are able to ask and test their own research questions, often crucial for inspiring undergraduates to undertake postgraduate degrees. Indeed, multiple students have already used their assessment for interviews this Spring. A crucial requirement, therefore, of any Genomics module is that it remains current, creating a continuing work burden for teaching staff. Students really appreciated that in lectures 5, 6 and 7 the majority of mentioned papers were from the 2020s, including work published just weeks before lecture delivery. Students embraced this timeliness and novelty, and the responses and assessments of many students were inspiring.

Review of the Genomics assessment and potential indicators for success

The assessment was divisive, however, with many, often high-achieving, students struggling with the combined requirements of analysing and interpreting complex data and then contextualising it within the wider literature; often areas students had never come across before. The broad spread of grades without fails (Figure 6A) was encouraging to see, as students were clearly challenged (to excel for over a third of students), but sufficiently

supported to pass the module in all cases. A fair criticism of the module was the assessment feedback (already addressed for future years), which in many cases was insufficient for students to develop their writing and analysis skills. Students and teaching staff were also keen to include the group posters in the summative assessment. It is possible this may favour students where the data and research acumen are less well developed.

The clear association of previous flair for data acumen and performance in Genomics was not unexpected, but it was curious that this outstripped general performance in stage 2 (Figure 4B, 6B, 6D). This potential predictor module should be reflected on carefully, as the teaching staff were consistent between modules, other stage 3 results were not considered when assessing performance in Genomics, and, crucially, Big data biology no longer exists as a standalone module. 47I has now been incorporated within “Becoming a Bioscientist 4” (BABS4), a skills module with combined and integrated laboratory and data components. It is interesting to note that the first year 17C (now “Becoming a Bioscientist 2”) which was already a mixed module, was not a significant predictor of performance in Genomics beyond the first-year average. Future monitoring will determine how well BABS4 performance tracks with Genomics in future years, and whether having a separate data-only module was more relevant for some York undergraduates.

Proposals for improving the Genomics module

On first delivery, the Genomics module appears to have been successful, with both teaching staff and students generally happy with the content and balance of the module. Students performed well and were exposed to research-led teaching of current topics relevant in bioscience research and industry. However, further development is needed. Integrating student feedback and discussions from the teaching staff critical review has outlined the following priorities for immediate development before delivery in September 2024.

- i) Lecture material on core ideas in genomics were overly reinforced. These concepts are crucial, but students were engaging with the material asynchronously and multiple times between lectures, reducing need for lecturer-led reinforcement. **Proposed action** to merge lectures 2 and 3 (Figure 2), providing space for an additional lecture on more advanced material.
- ii) All practical workshops had sections which did not run smoothly. **Proposed action** to refine material to ensure computational work is efficient, leaving time to discuss the biological interpretation and reasons for particular bioinformatic steps.
- iii) **Proposed action** to adopt the group poster work into summative assessment, perhaps through a reflective piece of writing to be submitted within semester assessment weeks.
- iv) Data analysed by students was recent, but tested, to ensure that students would have results in this first delivery period. **Proposed action** to fully address Learning Objective 5 (Table 1) by generating new data to facilitate truly research-led teaching.

Concluding remarks

The University of York already had above average provision of data science skills for bioscience undergraduates. The development of a new stage 3 Genomics module which incorporates theoretical and advanced practical components has provided novel opportunities for undergraduates to experience research outside their capstone projects and to develop highly transferable data skills applicable to a broad range of graduate careers.

This research project has facilitated an objective review of this new module, providing robust data to support its continued development. 4 key areas have been identified to improve content and assessment, and to enhance the prominence of research-led teaching. These proposals will directly improve delivery of the module in the 2024/25 academic year.

Acknowledgements

The author extends his sincere thanks to Dr Daniel Jeffares (Genomics MO) for his support, mentorship and encouragement through the design, delivery and review of the Genomics module. It is fantastic to work so closely with someone equally passionate about teaching genomics and bioinformatics to our undergraduates. Finally, thank you to the Genomics 2023 cohort for their patience and enthusiasm as we delivered this module for the first time.

Statement of ethics

This project represents an internal critical review of non-public domain data on student assessment at The University of York not for wider dissemination. Formal ethical approval was not required on the basis of full anonymisation of data already collected for this purpose under the Department of Biology's data use statements provided through the website and therefore accessible to all students.

References

Afgan, E. et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, 44 (W1), pp.W3–W10.

Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410.

Attwood, T. K. et al. (2019). A global perspective on evolving bioinformatics and data science training needs. *Briefings in bioinformatics*, 20 (2), pp.398–404.

Brown, J. A. L. (2016). Evaluating the effectiveness of a practical inquiry-based learning bioinformatics module on undergraduate student engagement and applied skills. *Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology*, 44 (3), pp.304–313.

Carey, M. A. and Papin, J. A. (2018). Ten simple rules for biologists learning to program. *PLoS computational biology*, 14 (1), p.e1005871.

Gatherer, D. (2020). Reflections on integrating bioinformatics into the undergraduate curriculum: The Lancaster experience. *Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology*, 48 (2), pp.118–127.

Harrison, P. W. et al. (2023). Ensembl 2024. *Nucleic acids research*. [Online]. Available at: doi:10.1093/nar/gkad1049.

Kent, W. J. et al. (2002). The human genome browser at UCSC. *Genome research*, 12 (6), pp.996–1006.

Leonelli, S. (2019). The challenges of big data biology. *eLife*, 8. [Online]. Available at: doi:10.7554/eLife.47381.

Mariano, D. et al. (2019). Introducing programming skills for life science students. *Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology*, 47 (3), pp.288–295.

Marx, V. (2013). Biology: The big challenges of big data. *Nature*, 498 (7453), pp.255–260.

Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in genetics: TIG*, 16 (6), pp.276–277.

Suciu, R. M., Aydin, E. and Chen, B. E. (2015). GeneDig: a web application for accessing genomic and bioinformatics knowledge. *BMC bioinformatics*, 16 (1), p.67.