

Supplementary Information: Interdisciplinary and transferable concepts in bioinformatics education: observations and approaches from a UK MSc course

Iain G. Johnston^{1,2,*}, Mark Slater³, Jean-Baptiste Cazier^{4,5}

Appendix A – QAA descriptors

Corresponding elements of designed module outcomes (Table 1) are given in parentheses.

The general QAA descriptors at this level involve (paraphrasing):

- Systematic understanding of knowledge and critical awareness of forefront of discipline [(b), (c), (e)]
- Comprehensive understanding of techniques [(b), (c), (d), (e)]
- Original applications of knowledge [(d)]
- Critically evaluate current research and methodologies [(e)]
- Systematically and creatively deal with complex issues [(d)]
- Self-direction and originality [entire module]
- Employment and personal skills [entire module]

On graduating with an honours degree in biosciences, graduates will have the following core knowledge, understanding and skills:

- experience and competence in a broad range of appropriate practical techniques and skills relevant to the biosciences including data collection, analysis and interpretation of those data, and testing of hypotheses and the ability to place the work in context and to suggest lines of further investigation [mainly (d)]
- the ability to explain biological phenomena at a variety of levels (from molecular to ecological systems) and how evolutionary theory is relevant to their area of study [(c), (e), (f)]
- the ability to plan, execute and present a piece of hypothesis-driven work within a supported framework in which qualities such as time management, problem solving, and independence are evident [entire module]
- the ability to access and evaluate bioscience information from a variety of sources and to communicate the principles both orally and in writing in a way that is organised and topical, and recognises the limits of current hypotheses [(b), (e), (f)]
- an appreciation of ethical issues and how they underpin professional integrity and standards [(e)]
- an appreciation of the impact on society of advances in the biosciences [(e)]
- the ability to record data accurately, and to carry out basic manipulation of data (including qualitative data and statistical analysis, when appropriate) [(c), (d)]
- an understanding of the use of bioinformatics approaches in the analysis of large datasets [entire module]

On graduating with an honours degree in biosciences, in addition to demonstrating the core knowledge, understanding and skills specified in the threshold standard, a graduate will be able to:

- plan, execute and present an independent piece of work, in which qualities such as time management, problem solving and independence are evident, as well as interpretation and critical awareness of the quality of evidence [individual assessment; also covered in individual projects in a parallel module]
- construct reasoned arguments to support their position on the ethical and social impact of advances in the biosciences [(e), (f)]
- demonstrate a secure and accurate understanding of the explanation of biological phenomena at a variety of levels (from molecular to ecological systems) and explain the relationship of evolutionary theory to their area of study [(b), (c), (e)]
- apply relevant advanced numerical skills to biological data [(d)]
- communicate science to peers and non-scientists [(f)]
- demonstrate well developed strategies for updating, maintaining and enhancing their knowledge of the biosciences, including cross-disciplinary awareness access bioscience databases and use appropriate selection

Category	Number
Theory/practice combination	13
Accessible and helpful instructors	7
Breaks and session structure	4
Other positive	5
More time / spreading content	8
Earlier release of material	5
Deeper coverage	3
Other for improvement	2

Table 1: **Summary of student responses by category.**

criteria to mine, manipulate and interpret data. [entire module]

Appendix B – Student feedback

General positive observations:

- ‘Learning by doing’ [Warhurst, 2008] explicitly praised (I.1.c) with many other positive notes about the combination of theory and practice (I.1.c; I.2; I.3.a; I.3.b; I.4; I.5; I.7.a; I.8.a; I.9; I.10; I.11; I.14.a; I.15).
- Step-by-step approach helps students from different backgrounds avoid feeling lost, increasing accessibility (I.1.a; UK PSF V1, V2).
- Well-spaced breaks (which aren’t strictly aligned with hourly schedule) praised for allowing consolidation and troubleshooting (I.6; I.7.b; I.13; I.16.c)
- Other: confidence building (I.1.b), accessible and helpful lecturers (I.3.a; I.6; I.10; I.16.a; I.16.b, III.3.b, III.11), online support (I.8.b), Python (I.14.b); course structure (I.16.d)

General notes for improvement:

- Earlier release of material and/or reading would be useful (II.1; II.13.b; III.3.a)
- Spreading out content and consolidating over multiple days would have helped (even without lecturer contact) (II.3; II.5; II.6; II.11)
- More time on content / ensure all content is completed (II.4; II.7; II.12; II.13.a; III.5)
- Other: lecturer delivery (II.2), deeper coverage and more content (II.9, II.10, II.16), smoother transition from introduction to practice (II.13.c)

Numerical labels below reference individual student responses.

I. What aspects of the teaching best helped your learning?

- I.1.a. A step by step approach to introducing new software and the fundamentals of programming, meant students from a non-computing background could follow along without feeling too lost.
- I.1.b. Learning command line then Python and finally applying it to R really helped build confidence with using different software/languages.
- I.1.c. Learning by “doing”. Given the opportunity to work out a pipeline before being shown solutions helped think creatively about problem solving.
- I.2. They showed the examples and asked to do some practices which is a good way to understand what I have learnt
- I.3.a. Lecturers been very understanding and helpful when teaching us. Exercises provided very helpful at getting us to practise.
- I.3.b. Practising better than learning by theory.
- I.4. The examples and interactive aspect of the lectures.
- I.5. Letting us do the activities in the [programming] sessions to reinforce our learning
- I.6. This module has been highly relevant as an introduction to the course, however I have found it very full on with a lot of new information in a short space of time. [...] I found the lecturers to be very helpful and willing to give breaks every 45mins with a chance to ask questions – very helpful.
- I.7.a. Working through the code as a class when learning Python was highly effective.
- I.7.b. Frequent short breaks after learning new concepts made it easier to consolidate learning.
- I.8.a. Many practical examples
- I.8.b. The information found on Canvas
- I.9. Practising different aspects of the code that we are learning

- I.10. Time for exercises where the lecturer was able to solve our questions individually
 - I.11. The practical sessions were the most helpful. However, the theoretical parts were also very useful but a little intense
 - I.13. Frequent breaks for questions & troubleshooting
 - I.14.a. Practical mixed with theory (really good).
 - I.14.b. Python lectures were good though I found them too slow
 - I.15. Lots of helpful worked examples so we can apply things we have learnt, right away
 - I.16.a. Very enthusiastic and knowledgeable
 - I.16.b. Engaging at all times
 - I.16.c. Good well spaced breaks
 - I.16.d. Good amount of content
- II. What could have been done differently to help you learn more effectively?
- II.1. Earlier release of pre-course material! It would really help if applicants could have a chance to look [at] command line, R, Python before the course started.
 - II.2. The speed of speaker is a little bit quickly for me to follow
 - II.3. Spreading out the teaching content over the whole 2 weeks – we had 4 days off which could've been used for applying what we have learnt with help available if we needed.
 - II.4. If there was more time to go through the introductions.
 - II.5. Provide a tutorial a day after learning a coding language. This will help us understand and remember concepts, more than having just a lecture followed with [...]?
 - II.6. [I wonder if it might be useful to have 1 day of lectures followed by one day of self-study, spread more over the first 2 weeks so that there is time to make sense of each topic straight after being taught it.]
 - II.7.a. Ensure all scheduled content is completed.
 - II.7.b. It is slightly overwhelming knowing that some key concepts were not [covered] due to time constraints.
 - II.9. Some deeper explanations of some of the "R" code with examples
 - II.10. More examples for the codes on the lecture handouts
 - II.11. Split out the lectures, in order to some parts would be more understood
 - II.12. More time for practice
 - II.13.a. More time on each system – we ran out of time with each one.
 - II.13.b. Some recommended reading to introduce concepts could cut down time spent just listening about them instead of practice.
 - II.13.c. Generally, I felt there was a steep step up between the introduction to concepts & using them in practice
 - II.16. More on Python and self created functions
- III. Any other comments you wish to make:
- III.1. Really liked the xkcd comics!
 - III.3.a. We were only sent the reading 1 week before the course started – it had lots to read on it, so would have been much better to get this earlier so we would have time to do more.
 - III.3.b. Lecturers so far have been great :D
 - III.5. For lecturers to have more time to teach as all 3 coding languages weren't completed due to running out of time.
 - III.11. Mr Johnston was very useful and with good vibes

Appendix C – Course content details

In this appendix section we give further details, references, and reflections on the specific content taught and discussed in the different parts of the module outlined in the main text.

Introductory philosophy: information and life

The first 'diversity-addressing' foundational content required non-biologists to meet core biological concepts and biologists to meet key informational concepts. This was addressed by first providing a set of basic biology lectures with an 'information and life' flavour, and next providing an entry-level set of coupled lectures and exercises on core technologies. The learning objectives for this section were simply for students to be exposed to an information-centric picture of biology, and to the core types of information in molecular biology and our ways of measuring

them.

The introductory lectures started with the pitch that ‘all biology is computational biology’ [Markowitz, 2017] – that is, modern biology requires computational methods, and ‘computational biology’ will soon become as redundant as ‘pipette biology’. Some classical high-profile bioinformatic studies, with simple core concepts and strikingly visual outputs, were summarised (for example Novembre *et al.*’s finding that European genetic structure closely aligns with physical geography [Novembre *et al.*, 2008]). Aligning with this focus on computation and information, life was presented as the reproduction and maintenance of *information* in the face of entropic disorder. The central dogma of molecular biology was covered from an informational perspective – following the transfer of information between different molecules to facilitate function in the cell, through information storage and use (the central dogma), replication (evolution), and processing (signalling and regulation). Pursuing ‘life as information’ further, the Kolmogorov complexity associated with life’s information content [Edmonds, 1995] was heuristically illustrated with an exercise using gzip to compress photographs of a rich jungle ecosystem (1.4MB compressed to 1.2MB) versus a dead moonscape (1.4MB compressed to 644KB); and the human genome (779MB compressed to 676MB) to explicitly link ‘biological’ with the more familiar ‘computational’ information. The idea that Kolmogorov complexity can describe evolved ‘algorithms’ that give rise to apparently more complex outputs was probed further using L-studio to simulate plant growth and form [Prusinkiewicz *et al.*, 1999].

A group exercise followed where students from biological and non-biological backgrounds were paired, asked to choose one of a list of biological topics, and after a short period for preparation and research, give a minute’s presentation on informational aspects of their chosen concept. This was expanded into an overview of some examples of topics of bioinformatic study (transcriptomics, proteomics, metabolomics, etc) and the information storage, replication, and/or processing involved in each case. Generally, we found students from biological backgrounds very receptive to ideas of structuring quantifying information. Several viewed the (semi-)quantitative formalisation of an idea, that they had informally considered through their training, appealing and intuitive. Contributions from their quantitative partners helped to reinforce this picture and provoked some interesting discussions comparing life and human computation. Finally, an aside on artificial life [Levy, 1993], Von Neumann machines [Von Neumann *et al.*, 1966], cellular automata [Conway, 1970], and the Tierra / Avida environment [Ray, 1991, Ofria and Wilke, 2004, Adami *et al.*, 2000] was presented as an informal and visually appealing example of the relationship between information and (biological or artificial) life.

Teaching computational core concepts with virtual machines

Command line. The command line section involved a joint student-instructor exploration of Unix file structure (students and instructor used the command line to navigate and explore together a preconstructed example file system) and the Unix philosophy, explored together with the instructor by using grep, sed, awk, and other tools to construct pipelines to analyse preconstructed data. Short single-line exercises accompanied lecture slides on the theory of these commands¹. Following the joint lecture-exercises delivery, more involved practical exercises were explored, including working in small groups (2-4) to build analysis pipelines for specific bioinformatics problems motivated by the instructor’s research (for example, identification of G-quadruplex-forming regions in DNA sequences [Kikin *et al.*, 2006]). The connection with ongoing and current research was emphasised to illustrate the cutting-edge and practical nature of these modes of working [Brown, 2016, St Clair and Visick, 2010]. The instructor worked in parallel with the class via the VM environment, solving the multistage problem step-by-step with suggestions from the class groups.

Python. In order to teach Python to the students, it was decided to focus on giving them a good grounding in the fundamentals of general programming, then moving on to common concepts before introducing more complex applications. We considered alternatives, including concentrating on ‘teaching by example’ by providing complete code blocks that would cover the main use cases the students would face. However, we believed that this approach could lead to students developing a narrower ability, not only to develop their own code but also to understand and modify third party code. Given that bioinformatics often uses several languages and development paradigms, and indeed that later course content also develops skills and practise in R, we also considered an introduction to general programming techniques as an appropriately transferable introduction.

The Python course was taught in a single intensive block over one and a half days and was principally derived from a C++ course taught by one of us (MS) to postgraduate physics students. It started by covering fundamentals such as variables, objects, and operators, before moving on to other general programming concepts such as loops, conditionals and functions. There was a through-line for the majority of the course of writing an implementation of

¹Outside of scheduled contact time, students were encouraged to play ‘Regex Golf’ (<http://regex.alf.nu/>) to gamify an introduction to this rather technical concept.

the Caesar Cipher which allowed progressively more involved exercises to be set. The aim of this structure was to give the students as much hands-on programming time as possible during the course and thus develop their code-based problem solving skills, without yet requiring the additional time overhead of solving harder, 'real-world' problems. This approach also fitted well with the text and file manipulation they were most likely to use Python for, at least in the short term. Consistent with the overarching active learning and mixed lecture-practical philosophy, an effort was made to not lecture for more than 20-30mins before providing another hands-on exercise. In addition, where appropriate, both the Python interpreter and scripting were used in order to give the students practice in both.

Overall, student reception to the course was positive (Appendix B), and even the students that started off the course with no programming knowledge had grasped the basic concepts by the end. Their uptake of basic programming concepts was demonstrated by their familiarity with loops, conditionals, data structures and so on in later R practise, underlining the 'diversity-addressing' results of this approach. The basic course structure was retained through the years it has been run; however, we have recently decided to move away from the Caesar Cipher example and employ a more bioinformatics-related 'gene manipulation' example. This task will have a similar level of difficulty but provide a more direct connection to biological data.

This is one part of the wider module where more time could be readily invested in more detailed and/or thorough coverage of core concepts. Going forward, it was thought that outcomes would be even better by spreading this content out over a longer period, thus allowing students time for reflection on the material, and the setting of 'homework' to which feedback could be given. These ideas were also raised in student feedback (Appendix B). However, as a necessary precursor to other content in the course, there is a logistic tension between time spent here and time on other more specific content later. Given the restrictions of the broader course timetable, we currently retain this short, intensive introduction format, with a view to more advanced and specific concepts being covered in subsequent modules as required.

R. The concepts met in the Python introduction were next translated into an introduction to the R statistical environment [R Core Team, 2015]. Given their importance in bioinformatics, particular attention was focused on the variety of data types available in R. Complementing the previous Python coverage, exercises in program control and input/output were completed in groups of 2-4 as before. A brief introduction to implementing simple hypothesis testing in R was demonstrated, but more statistical details were withheld for the future module on that topic. Finally, an interactive visualisation practical was held, first using R core capacity and then the ggplot2 library [Wickham, 2009]. Students were very engaged in these visualisation practicals, enjoying exploring the styling options afforded by R and particularly ggplot2, and working in an open-ended manner to compete and co-operate in producing beautiful visualisations of data.

Git. Due to time constraints, an introduction to version control and Git was presented in a light-touch format. Again working jointly with the instructor on the command line, students forked an existing repository from GitHub (www.github.com), proposed changes and submitted pull requests, which were accepted by the instructor in real time to demonstrate both sides of the process.

Structuring and logistics of investigation, and scientific communication

In addition to the topics in the main text, some informal discussion of Belbin's team roles [Belbin, 2012], academic personalities, and good practice in small team dynamics [Salazar et al., 2012] was held. In particular, a study summarising the meeting dynamics of innovative teams was explored [Salazar, 2018]. Here, team reflection (at the start and end of a meeting), promotive voice (at the start of a discussion), perspective seeking (as ideas develop) and managing interdisciplinary connections were the recipe for innovative and productive meeting. The idea here was not to insist upon employing ideas from management studies, but to make students aware of some existing ideas about beneficial combinations of skills, approaches, and collaboration dynamics. The plasticity of team roles – that a given researcher may be called upon (or find it desirable) to change roles throughout a project – was highlighted in conjunction with the expected plasticity of interdisciplinary scientific research projects above.

Ethics in bioinformatics

The comparison of ideas between students and instructor in the main text set the tone for the future coverage of ethics topics. Students and instructor took part in guided but free discussion of several ethical topics in research, (from core scientific practice to bioinformatic-specific issues) through the lens of the individual researcher (Fig. 2D). These first included the role research plays in society as a whole, from the broad ('what should science achieve?', 'who should pay for it?') to the specific ('how can we justify the carbon costs associated with large-scale

computational biology?’ and the ‘10/90’ gap, where only 10% of health research funding is spent addressing 90% of the world’s health problems [Schüklenk and Kleinsmidt, 2006]).

Following the previous examples of scientific philosophy, we briefly touched upon different ideas of ‘doing good’, from the principles of biomedical ethics [Beauchamp et al., 2001] to Bentham’s ‘greatest happiness of the greatest good’, and Pearce’s associated ‘hedonistic imperative’ of the abolition of suffering through the use of biotechnology [Pearce, 1995]. Considerations for individual researchers choosing a research project or direction were discussed, including the individual’s personal motivations, the likelihood that the ‘advertised’ benefits of the research will be achieved, and the balance of contributions and costs associated with the project. Students were encouraged to consider which scientific disciplines and individual scientists had saved the most lives in history; subsequently we explored the thought-provoking website ‘Science Heroes’ (<http://www.scienceheroes.com>) which attempts to estimate this latter quantity, leading to class conversation on the inherent ‘goodness’ of particular research topics, the moral compass behind research intentions, and decoupling research from researcher.

We next discussed whether the issues that often motivate bioinformatic studies are best addressed through fundamental science, public health, or other policy interventions. Several examples were used to crystallise this discussion, including the topic of plant science and food security. Substantial scientific and bioinformatic effort is spent characterising, at the molecular level, crop plant traits that may improve productivity and improve food security. However, a third of the world’s food is simply lost or wasted every year [fao, 2019], eclipsing the gains from many molecular insights. Students were encouraged to discuss whether societal resource was best spent addressing these logistic issues, more scientifically advanced biotechnology, or both. We found that most students had not previously connected a scientific project to a real economic cost to society (for example, grant funding from a research council). When presented with costs of some real research grants, and the overall budget of UK research councils, they were able to compare research and policy solutions on a more even footing.

Focusing increasingly on bioinformatics, students were asked to populate a list of what they considered ethical topics particularly pertinent to the field of bioinformatics. Students initially had difficulty in compiling this list, perhaps suggesting that they had rarely considered ethical concerns in the field previously. ‘Designer babies’ was often among the first ideas to appear [Amer, 2017], and ‘personal data’ appeared in several suggestions. Correspondingly, the next point of discussion was from the Exome Aggregation Consortium (ExAC) study, highlighting the possible overselling of ‘pathological’ variants discovered in early bioinformatics investigation [Lek et al., 2016] and catalysing a discussion on functional validation of bioinformatic results [Editorial, 2016].

We next focused on the structure, acquisition, and use of bioinformatic data. We discussed a 2016 study highlighting the traditional ethnic focus of GWAS on participants of white European ancestry [Popejoy and Fullerton, 2016]. This was striking news for our students, working in classes of rather international background, and set the stage for a discussion of how global the benefits of GWAS are, and should aim to be [Bustamante et al., 2011]. Following this point, the heterogeneous worldwide impact and benefits of bioinformatics research was next discussed, with particular reference to ‘benefit sharing’, defined by Ref. [Schroeder, 2007] as ‘...the action of giving a portion of advantages/profits derived from the use of human genetic resources to the resource providers in order to achieve justice in exchange...’. A critical ethical and legal analysis of north-south benefit sharing in genetics research [Schüklenk and Kleinsmidt, 2006] was discussed, along with a critique debating whether and how developing countries will benefit from their participation in genetics research [Ndebele and Musesengwa, 2008].

We next covered a current issue in the ethics of bioinformatics – the re-identifiability and the subsequent potential for misuse of personal data [Homer et al., 2008, Greenbaum et al., 2011]. Students had generally met ideas of anonymising and pooling genetic data, but were surprised at the ease with which individuals could be re-identified under many of these circumstances [Homer et al., 2008]. A provocative article on this topic about 23AndMe’s use of genomic data [Seife, 2013] was read in depth. This article raised the points of the company’s permissive terms and conditions allowing the use of personal genomic information for marketing, and the fact that any family member using such a service is automatically providing a large amount of the genetic information they share with other family members to that service.

Appendix D – Online resources

Artificial life:

- Video of emergence of self-reproducing ‘evolooop’ automata <https://www.youtube.com/watch?v=5IpK5meen28>
- Prime number calculation by automata in Game of Life <https://www.youtube.com/watch?v=68nEX5CEmZE>

Regex Golf:

- <http://regex.alf.nu/>

Science Heroes:

- <http://www.scienceheroes.com>

xkcd comics:

- 'Biology is largely solved' <https://xkcd.com/1605/>
- Regular expressions <https://xkcd.com/208/>
- tar <https://xkcd.com/1168/>
- p-values <https://xkcd.com/882/>
- Linux cheat sheet <https://store.xkcd.com/products/linux-cheat-shirt>
- Frequentist versus Bayesian statistics <https://xkcd.com/1132/>
- Dubious correlations <https://xkcd.com/1725/>

Readability:

- Readability Analyzer <https://datayze.com/readability-analyzer.php>

Science, ethics, and society:

- Causes of death and disability http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/
- Health and health spending <https://ourworldindata.org/health-meta/>
- UK cancer spending <https://www.statista.com/chart/2825/uk-cancer-spending/>
- Public health and smoking https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm and nutrition <http://www.who.int/mediacentre/factsheets/malnutrition/en/>
- Food waste <http://thebrazilbusiness.com/article/food-waste-sector-in-brazil>
- FAO Food Loss and Waste <http://www.fao.org/food-loss-and-food-waste/en/>
- Coverage of ExAC study findings <https://www.nature.com/news/rethink-the-links-between-genes-and-disease-1.20771>
- 23AndMe data use <https://www.scientificamerican.com/article/23andme-is-terrifying-but-not-for-the-reason/>

Appendix E – Reflections on student performance in interdisciplinary group projects

Encouragingly, the groups invariably delivered mature and well-presented research projects, with both written and oral presentations incorporating the ideas from their parallel work on communication and critique. Group write-ups critiqued existing work and provided a well-argued motivating framework for the research question. Results were generally presented in a clear and rhetorically compelling way, although there were several instances where different styling choices could have made deliverables clearer (from small axis labels to different choices of distributional plots). Conclusions were sensible, with little 'overselling' of results to fit a narrative, and discussions often drew on broader literature to frame the outputs and motivate future work. Overall, the group reports read encouragingly like scientific manuscripts ready for submission.

Group oral presentations were also of a high standard. Slides were well designed and the sometimes challenging task of allocating different parts of the delivery to different group members was generally well managed. There was heterogeneity between groups in the amount of time dedicated to motivating and 'setting up' the research question; a common feedback point was that in presenting to an interdisciplinary audience, investing several slides in the broader picture behind the research can help quickly bring everyone up to speed. Q&A was also generally positive, with student responses to aligned technical questions generally confident and well-reasoned, and with good responses (including some honest answers of 'I'm not sure') to more testing broader questions.

Rather by contrast, the individual written abstracts were more heterogeneous in quality. Technical abstracts were sometimes rather verbose, reading more like the introduction to a paper. Lay abstracts were sometimes very similar copies of the technical version, although there were many more thoughtful instances where well-chosen vocabulary and phrasing was used to simplify the message (Fig. 2). As fitting these individual abstracts to a given audience appears to be a more challenging task, we intend in future to make explicit reference to readability scores like the Flesch-Kincaid grade level [Kincaid et al., 1975] to give a quantitative handle on the complexity of written text.

References

[fao, 2019] (2019). Fao: Food loss and food waste.

[Adami et al., 2000] Adami, C., Ofria, C., and Collier, T. C. (2000). Evolution of biological complexity. *Proceedings of the National Academy of Sciences*, 97(9):4463–4468.

- [Amer, 2017] Amer, S. (2017). Ethical concerns regarding the use of bioinformatics and computational genomics. In *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*, pages 58–62. The Steering Committee of The World Congress in Computer Science, Computer .
- [Beauchamp et al., 2001] Beauchamp, T. L., Childress, J. F., et al. (2001). *Principles of biomedical ethics*. Oxford University Press, USA.
- [Belbin, 2012] Belbin, R. M. (2012). *Team roles at work*. Routledge.
- [Brown, 2016] Brown, J. A. (2016). Evaluating the effectiveness of a practical inquiry-based learning bioinformatics module on undergraduate student engagement and applied skills. *Biochemistry and molecular biology education*, 44(3):304–313.
- [Bustamante et al., 2011] Bustamante, C. D., Francisco, M., and Burchard, E. G. (2011). Genomics for the world. *Nature*, 475(7355):163.
- [Conway, 1970] Conway, J. (1970). The game of life. *Scientific American*, 223(4):4.
- [Editorial, 2016] Editorial, N. (2016). Rethink the links between genes and disease.
- [Edmonds, 1995] Edmonds, B. (1995). What is complexity?-the philosophy of complexity per se with application to some examples in evolution. In *The evolution of complexity*. Kluwer, Dordrecht.
- [Greenbaum et al., 2011] Greenbaum, D., Sboner, A., Mu, X. J., and Gerstein, M. (2011). Genomics and privacy: implications of the new reality of closed data for the field. *PLoS Computational Biology*, 7(12):e1002278.
- [Homer et al., 2008] Homer, N., Szlinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., and Craig, D. W. (2008). Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4(8):e1000167.
- [Kikin et al., 2006] Kikin, O., D'Antonio, L., and Bagga, P. S. (2006). Qgrs mapper: a web-based server for predicting g-quadruplexes in nucleotide sequences. *Nucleic acids research*, 34(suppl.2):W676–W682.
- [Kincaid et al., 1975] Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Research Branch Report 8-75, Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN*.
- [Lek et al., 2016] Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285.
- [Levy, 1993] Levy, S. (1993). *Artificial life: A report from the frontier where computers meet biology*. Random House Inc.
- [Markowitz, 2017] Markowitz, F. (2017). All biology is computational biology. *PLoS biology*, 15(3):e2002050.
- [Ndebele and Musesengwa, 2008] Ndebele, P. and Musesengwa, R. (2008). Will developing countries benefit from their participation in genetics research? *Malawi Medical Journal*, 20(2):67–69.
- [Novembre et al., 2008] Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., et al. (2008). Genes mirror geography within europe. *Nature*, 456(7218):98.
- [Ofria and Wilke, 2004] Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial life*, 10(2):191–229.
- [Pearce, 1995] Pearce, D. (1995). *Hedonistic imperative*. David Pearce.
- [Popejoy and Fullerton, 2016] Popejoy, A. B. and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature News*, 538(7624):161.

- [Prusinkiewicz et al., 1999] Prusinkiewicz, P., Karwowski, R., Měch, R., and Hanan, J. (1999). L-studio/cpfg: a software system for modeling plants. In *International Workshop on Applications of Graph Transformations with Industrial Relevance*, pages 457–464. Springer.
- [R Core Team, 2015] R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [Ray, 1991] Ray, T. S. (1991). An approach to the synthesis of life. *Artificial life II*, pages 371–408.
- [Salazar, 2018] Salazar, M. (2018). Brainstorming for innovation in team science.
- [Salazar et al., 2012] Salazar, M. R., Lant, T. K., Fiore, S. M., and Salas, E. (2012). Facilitating innovation in diverse science teams through integrative capacity. *Small Group Research*, 43(5):527–558.
- [Schroeder, 2007] Schroeder, D. (2007). Benefit sharing: its time for a definition. *Journal of medical ethics*, 33(4):205–209.
- [Schüklenk and Kleinsmidt, 2006] Schüklenk, U. and Kleinsmidt, A. (2006). North–south benefit sharing arrangements in bioprospecting and genetic research: a critical ethical and legal analysis. *Developing World Bioethics*, 6(3):122–134.
- [Seife, 2013] Seife, C. (2013). 23andme is terrifying, but not for the reasons the fda thinks. *Scientific American*, 27(11).
- [St Clair and Visick, 2010] St Clair, C. and Visick, J. (2010). *Exploring Bioinformatics: A Project-based Approach*. Jones and Bartlett Publishers. USA.
- [Von Neumann et al., 1966] Von Neumann, J., Burks, A. W., et al. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14.
- [Warhurst, 2008] Warhurst, R. P. (2008). cigars on the flight-deck: new lecturers participatory learning within workplace communities of practice. *Studies in higher education*, 33(4):453–467.
- [Wickham, 2009] Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.