Author's response to reviews

Title: Computational Cancer Biology: Education is the natural key to many locks

Authors:

Frank Emmert-Streib (v@bio-complexity.com)
Shu-Dong Zhang (s.zhang@qub.ac.uk)
Peter Hamilton (p.hamilton@qub.ac.uk)

Version: 4
Date: 18 September 2014

Author's response to reviews: see over
General Response to all reviewers: First of all we would like to thank all reviewers for their thoughtful and constructive comments, which helped us to significantly improve the paper.

CC1: As a general comment we would like to remark that we submitted this paper for the paper category 'debate'. This information may not have been accessible to the reviewers and could have given the misleading impression that our paper is a 'report', which would follow a different goal and structure. In order to make this also clear to the readers of the paper, we added a remark that our paper ‘intends to indicate a discussion’ of this important topic rather than clamming to provide a final solution to it.

CC2: The comments of the three reviewers are interconnected with each other for this reason, we added references whenever there are overlapping comments/meanings.

Reviewer 1:

The authors' premise is that 'computational cancer biology is not yet fully recognized as a coequal field in oncology, leading to a delay in its maturation and, as an immediate consequence, an under-exploration of high-throughput data in translational research'.

To support this premise, the authors argue that:
- There is little funding for computational cancer biology
- Nature and Science do not publish computational cancer biology papers
- oncology as a field [...] is mainly concerned with wet lab work and clinical trials. They argue that this lack of recognition for computational biology is due to lack of understanding of computational approaches, which can be remedied by education. They mention that this is supported by a poster competition they organized after students had received training in computational biology among other disciplines. The best poster was a poster promoting computational biology, thus suggesting that the students understood the importance of that field even if they were only briefly exposed to it.

Overall, I am sympathetic to the author's point and I do agree that more education in computational and quantitative biology would help address the recognition problem that computational biology is facing.

On the other hand, I also think the arguments are rather superficial, simplistic and one-sided, and not very well supported by the data.

Q1: First the authors don't support their claims very well. For example 12 training Fellowships awarded by the MRC for Bioinformatics, Biostatistics and Methodology
compared to 92 in all fields would appear to be a lot. The poster competition is anecdotal.

A1: We agree with the reviewer that the numbers can be seen differently (the glass is half-full or half-empty, depending on your perspective). On the other hand, from the perspective we set at the beginning of our paper (post-genomic era and high-throughput data) every modern experiment generates a data set that is challenging to analyze. In such a context, one would assume a much larger appreciation by funding agencies, e.g., reflected in more balanced numbers of fellows and a discussion/presentation of their research strategy.

See also the comments from reviewer 2 in Q4.

Q2: Second, the absence of computational cancer biology papers in Nature and Science is not necessarily reflective of a bias against computational biology. More likely, it is due to the fact that few high quality computational biology papers are submitted to these journals.

A2: We think that a problem is that a quantification of this is difficult due to the fact that neither nature nor science publish a list of rejected papers so one could trace them to see where they finally end up.

On the other hand, it would be interesting to create such a list, e.g., in the form of a web side where author's of rejected nature and science papers can add this information so everyone can assess if such papers have been rejected rightly or wrongly. We would be willing to discuss this possibility with the reviewer on a collaborative base because this sounds interesting.

We personally do not think that there is a lack of good papers in computational cancer biology that would be publishable in Nature or Science. However, this is our subjective opinion.

Q3: Third, the authors do not properly acknowledge that the field of computational biology is littered with papers that have no relevance whatsoever to the rest of the biological field - and accordingly are never cited by papers in that field.

A3: This is true. Unfortunately, this holds for every other field as well, including biology, medicine and clinical sciences (and also physics, chemistry, statistics etc) because the number of citations follows a power law.

_EPL (Europhysics Letters) Volume 61 Number 4_
**Measuring preferential attachment in evolving networks**
Furthermore, it is interesting to note that the number of citations of papers is not necessarily a reflection of their quality, see http://iopscience.iop.org/0295-5075/86/6/68001 for an interesting discussion.

**Q4:** Fourth, it is unclear why the authors suggest that biologists should receive quantitative training (meaning of a p-value etc) - but do not seem to suggest that quantitative and engineering folks should receive biological training. Training programs that emphasize biological problem finding and sound problem solving based on quantitative methodologies would likely help in that regard.

A4: This is a good point relating to an asymmetry of our view. This asymmetry comes from the perspective we assume in this paper, by starting from the beginnings of oncology in 1975. At that time (and in the following 1.5 decades) none (only very few people) would consider ‘computational methods’ important in this area. Instead, oncology started as an experimental field. Now, 3 decades later, we think such a view point is no longer valid. That means if one wants to be an ‘oncologist’ limiting the skill set to experimental techniques is not enough because every experiment generates data and high-throughput data are challenging to analyze.

On the other hand, if one want to be a (computational) oncologist analyzing data only, this is perfectly fine because, as noted by reviewer 2 (see Q4), the availability of public databases allows to obtain hundreds to thousands of data sets for analysis to generate novel hypothesis.

To summarize, if computational biologists should or should not receive training in experimental skills is an interesting topic for itself that is somewhat different from our main argument (computational cancer biology should be a co-equal field). A discussion of this should be provided in a separate paper in order to avoid a confusion/dilution of arguments.

**Q5:** Fifth, it is unclear that education is the only remedy to this problem. While education may help, many other initiatives would too. For example, funding agencies could start enforcing a requirement for computational biology expertise on grants that propose significant data generation. This is already happening for statistical considerations with statisticians sitting in many study sections. As another example, other fields such as physics have had a long tradition of quantitative research and large-scale multidisciplinary collaborations (especially experimental physics) - one would think there would be something to be learnt from these fields.
A5: There is some important point underrepresented and that is we suggest education as a ‘naturally induced’ factor (see section title). In contrast, the reviewer mentions an ‘enforced’ process, which is possible too. However, we think that enforcing rules can generate a hostile environment that can counteract the initial purpose.

Nevertheless, we agree with the reviewer by (1) changing the title of our paper (adding ‘natural’) and (2) adding a section about the responsibilities of funding agencies in this respect, which should indeed not remain unmentioned.

Reviewer 2:
In their manuscript Emmert-Streib and coauthors claim that oncology benefits tremendously from high-throughput data, but that “computational oncology” is not seen as a co-equal field in oncology, leading to a delay in its maturation and under-exploration of high-throughput data. The authors argue that over time this can be resolved by training of scientists, and discuss how such training can be achieved at the local university level.

The manuscript is more an opinion piece than a scientific research report, and many of the assumptions and claims made can and will lead to discussion. There are only very few “hard facts” presented in the manuscript, hence pro- and counter-arguments for points made should be given and discussed in the manuscript. Overall I agree with the authors that computational cancer biology is a crucial research field that holds big promises for novel and personalized treatment approaches for cancer. I also agree that training is a crucial element to promote and ultimately harvest its potential. I do not agree to all conclusions drawn in the manuscript however, and feel that important aspects are missing and / or need to be detailed.

A - Clarification: We agree with the assessment of the reviewer and for this reason the paper has been submitted as a ‘debate’ paper and not as a ‘report’ because it is more argumentative in nature than expected from a report. See also CC1 above.

Q1: Most strikingly, while the manuscript speaks about the importance of “training”, it remains relatively vague with respect to who should receive what training, and with what aims. As the authors point out as well, computational oncology is a very interdisciplinary field, and researchers working in this area typically have very diverse, and often interdisciplinary backgrounds. Training needs to be adapted to the individual background a person has, and computer scientists entering the field need very different training than biology students, and again entirely different training is required for people with a clinical background. I am firmly convinced that the field needs this influx of people with different backgrounds, but hence we need
to establish not one training program for everyone but an array of different training methods for people with different backgrounds.

A1: Please see A2.

Q2: Also the aim of the training needs to be specified more clearly. What is the goal of training? Should every biology student receive sufficient training to enable them to apply standard-methods for high-throughput data analysis themselves, or even so much that they are able to develop own algorithms for complex problems that cannot be solved with standard methods, or should they simply receive training to be able to discuss with a computational biologist who carries out the actual data analysis? Similarly, to what level should training in computational oncology be included in the curriculum of medical students and medical doctors? It remains very unclear what the authors suggest here, but in my opinion this is a critical piece of information when it comes to actually implementing training programs at the local level as the authors propose.

A2: The reviewer gives a very good example in A5 mentioning CASyM. This demonstrates the difficulty of the problem making it necessary to fund an initiative to ‘develop an implementation strategy (road map) for Systems Medicine’ rather than providing them at the start of their project. We think that there are different steps and each has its own difficulties in implementation. In the present paper we ‘debate’ the first step of this process, namely the recognition of ‘computational cancer biology’ as a coequal field. After there is a broad acceptance of this, further steps need to be discussed and these are indeed specific rather than generic for the needs of different institutions.

Q3: On a further note, even the definition of the field “computational oncology” is not very clear, and correspondingly, it remains also very unclear what concrete training requirements the authors see. The manuscript mainly motivates the use of computational methods with high-throughput data, presumably mostly from deep sequencing. Methods directly associated with this would be, for example, tools for raw data processing and data normalization, identification of differentially expressed transcripts / mutations / differential methylation and chromosomal aberrations, and then methods to correlate these with cancer phenotypes. The next level then are network- and pathway-based analysis approaches, clustering and classification methods, machine learning tools, etc., and it gets more complicated from there the further we move away from “raw data” to “interpretation”. However, technology is going to change, and it then needs to be discussed if training should focus on a specific technology, or rather on underlying general algorithmic and methodological approaches. What do the authors propose here? Are you suggesting training e.g. on how to use a mapping tool and a specific tool for differential gene expression analysis, or teaching more general concepts of data processing and data analysis and machine learning algorithms? For example, should a biology or medical student learn what supervised machine learning is, and what broad group
of methods are widely in use in that field and how they can be applied to high-throughput data? To take this even further – what about mathematical approaches to computationally model cancer-related pathways and simulate e.g. treatment effects on these pathways – would the authors also include this in training to some level? I guess my point is - it becomes very tricky when it comes to defining concrete training goals and identifying training needs – but I think this is a crucial requirement for a publication that essentially claims that training is required for computational oncology.

A3: We think that in order for computational cancer biology to receive a broad recognition it is first of all important to provide a comprehensive education on all levels
- summer internships (summer research program or summer school for BSc, MSc and PhD students)
- undergraduate training (BSc course)
- postgraduate training (MSc course or PhD program)
- postdoctoral training (seminars)
- mentoring of colleagues (seminars, conferences, workshops)

The question regarding the proficiency is a next step the requires specific rather than generic decisions. See also Q1 and Q2 for related points.

Importantly, we think a key for succeeding is a fair and competent introduction to computational biology because this enables a seemingly transition between different proficiency levels.

Q4: I found it somewhat surprising that the authors feel that computational oncology should be seen as a field co-equal to other fields in cancer research, such as other subareas of oncology (the authors mention fields specialized on specific cancer subtypes as examples). I think there is a fundamental difference here, and that difference is that while breast cancer can indeed be seen as a specialization in oncology, “computational oncology” is fundamentally different as it is much more method oriented and not focused on specific biological subgroups (such as breast vs. lung tumours) – even more so if the authors by “computational oncology” mainly refer to high throughput data analysis. This immediately then raises the question why it should even be desirable to establish “computational oncology” as a co-equal subfield in oncology, instead of as a complementary research field – similar as medical genetics would certainly not be seen as a coequal subfield in oncology, comparable to a subfield focusing on breast cancers. I similarly see “biometry and statistics” also rather as an independent research field than a subfield of oncology or any other medical field. I feel that the respective paragraph at least needs to be reworded, and am not convinced that computational oncology really has reason to complain at this time that it is seen as a “auxiliary” field. In fact, an interesting transformation is ongoing, and many of us working in the field are suddenly in a position where bioinformatics and computational biology is no longer used for testing of hypothesis that were
generated in a “wet-lab” group, but due to “big data” is more and more coming into a position where the computational part is also generating the hypothesis in the first place and defining the research question – in fact making the “wet” sciences the auxiliary sciences that merely generate data that is then analyzed and interpreted elsewhere. This obviously has fundamental consequences also for collaborative work in oncology in general, and will also change authorship order on ensuing publications – and is in fact merely driven by the amount of data all of a sudden available, that cannot be interpreted anymore without computational methods.

A4: We feel that maybe the term ‘co-equal’ has been misinterpreted. By this term we mean that two fields are equally respected and important. Assuming this meaning, we think that our views are not far apart, except the ‘auxiliary field’ argument. For this reason, we submitted our paper under the section ‘debate’ because other views than ours are possible and welcome. Yet we consider our views valid and looking forward to the reactions of the community.

Q5: I strongly disagree with a statement made in the paper that we as scientists cannot influence political decisions and funding programs. Many of us are indeed active in grant committees and even strategic committees of funding agencies, and are in positions where we can and do influence research calls of funding bodies and funding decisions, and I do think that it is critical to use these channels to promote computational biology approaches in research. I am not at all convinced that just “training” is going to be enough, as the authors suggest. Furthermore, what about research infrastructure (access to suitable computational infrastructure, methods and data), strategies for translation of computational research results into industrial and clinical application, etc.? I fully agree with the authors that computational approaches are critical and will transform the field of oncology and medicine. However, ultimately, broad recognition will come with success of computational cancer research, and this success is measured in terms of new drugs and new treatment approaches resulting from computational oncology research. On a final note, there are some initiatives at the national and international levels that concern themselves with strategies to promote computational biology / systems medicine / computational oncology. The European Network “CASyM – Coordinating Systems Medicine across Europe” (https://www.casym.eu/) is one such initiative, and there is extensive discussion within CASyM on the development of multidisciplinary training concepts for systems medicine. The present manuscript should discuss and include concepts developed within this and other similar initiatives.

A5: This comment and Q5 of reviewer 1 are related. We agree with both reviewers and added for this reason an additional section ‘Responsibility of funding agencies’.

Reviewer 3:
The debate article addresses the role of computational cancer biology and comes to the conclusion that additional education in the field will be the key to more progress in the future.

1. Does the debate present a novel argument, or a novel insight into existing work? Computational Cancer Biology is a novel and growing field that as the authors state has not gotten enough attention. This debate might increase the visibility and acceptance of the field - there are not many discussions of the kind.

2. Does the debate address an important problem of interest to a broad biomedical audience? The discussion will be of interest to most of the audience of the journal.

3. Is the piece well argued and referenced? The piece is well argued. However: the authors already assume that the reader already accepts and knows about the importance of the field. Additional arguments might help to convince the critical reader. What are the achievements of computational cancer biology this far? Name some success stories? What is the potential advancement that can be gained by putting more emphasis on the field? One could cite articles like:

   Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.

   - A colorectal cancer classification system that associates cellular phenotype and responses to therapy.

The authors come to the conclusion that a better education is the key? Are there other models possible? How is the current interaction between Computer Scientist or Statistician with the biologist? How can this be improved? Is it feasible the one person can have enough insights in both fields or does it need better interaction?

A3: This is a good point. We added a discussion about this to the paper and included also citations to such accomplishments. See also Q5 (reviewer 1).

4. Has the author used logical arguments and sound reasoning? Yes.

5. Is the piece written well enough for publication? Yes.
Minor Essential Revisions:
More citations are needed to demonstrate or prove the points.

A MER: We added more citations, see also A3.