Reviewer’s report

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

Version: 3 Date: 5 September 2014

Reviewer: Lars Kaderali

Reviewer’s report:

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.

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.

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.

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.

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.

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.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I declare that I have no competing interests