Evaluation of the Doctoral Dissertation
submitted by Marcos Yébenes Mayordomo
at the University of Gdansk titled:

Integration as a solution: A multi-omic approach to cancer diseases

Evaluator: Catia Pesquita, Faculdade de Ciências, Universidade de Lisboa

Recommendation

I fully recommend the **Acceptance** of the submitted dissertation by Marcos Yébenes-Mayordomo.

Critical Appraisal

The dissertation investigates how integrating different molecular profiling technologies is key to provide a holistic view of the molecular phenotype of cancer cells. This central idea is explored in three complementary studies, each targeting a different type of cancer (Esophageal adenocarcinoma, Undifferentiated pleomorphic sarcoma, and Gorham-Stout syndrome).

The studies all have substantial experimental complexity and a high diversity of methods and techniques. The analysis of the obtained results across all three studies reveals interesting results which represent definite contributions to the understanding of the molecular landscape in cancer and tumor-like diseases. One of the studies has already been published in an international peer-reviewed journal.

The dissertation is written in clear language, although some more detail and clarity in some sections would have been an improvement. The dissertation is quite compact, and I would have appreciated that some aspects of the work had been made more explicit for the reader.

My critical appraisal is written from the point of view of a specialist in biological data integration. I will not focus as deeply on the wet lab aspects of the work.

Motivation and Problem definition

The Introduction chapter provides an overview of omics research in cancer. A substantial number of works are cited in part 1.2 however, the description of these related works is very high-level. It would have been beneficial to provide an overview of the state of the art of proteogenomics in cancer with a categorization of works and an understanding of their limitations. It would have brough more clarity to the contributions of the thesis if the experimental design of closely related works was also briefly presented, in a way that would highlight how the studies in the thesis address the challenges in proteogenomics analysis. It is not made clear in the text how the methodology followed in the three studies that comprise the dissertation differs from the state of the art.

A well-thought description of the challenges in proteogenomics is given in the Introduction, which I appreciate. However, in the description of the three studies there is no reference to how exactly these challenges were dealt with. Although the reader could potentially infer the answer in some cases, it is mostly left unclear.

I also find that although a two-class categorization of proteogenomics studies is given in the Introduction, neither study is clearly placed into one of these categories (although a reader can infer it).

Delving into the specific studies, each study is given a general objective, however, it is not clear why these were the selected studies and how they may be related to each other. It would have been interesting to have each study identified as an example of a type of application of proteogenomics. Again, although this can be inferred by the reader, having this clearly stated and presented would have made the whole work clearer.

Esophageal adenocarcinoma

The methodology and experimental design could be more clearly described. It is not easy to understand the full methodology, and how many samples were analyzed at each step. This is a complex experimental approach, combining a variety of techniques. A flowchart of the methodology would have been immensely helpful. We need to read through the whole methods section and part of the results to finally get a much clearer picture of the overall study at sections 3.3.5 and 3.3.6

The results found by this study are extremely interesting and identify candidate genes that would not be uncovered by single omics techniques. Figure 3.5 is particularly enlightening.

I greatly appreciated section 3.3.8 and the hypothesis that dysregulation between RNA and Protein levels are not due to mutations.

Undifferentiated pleomorphic sarcoma

The methodology and experimental design are clearer than in EAC. However, the specific goals of this work only become clear in the Conclusions section.

The concurrent loss of RB1 and p53 as a significant event in UPS is very interesting, and the partial testing of this hypothesis with the drug APR-246 was elucidating. This is a promising result for novel therapeutic targets for tumors lacking both genes, and

clearly points that the development of a dual knock-out model would be beneficial for therapeutic developments.

The high level of intra-tumor heterogeneity detected also supported another line of inquiry. This study demonstrated the potential of proteogenomics to filter out candidate genes based on protein quantification, with a reduction of over 85% of mutated gees as neoantigen sources.

Gorham-Stout disease

This study is present with considerably more clarity than the two previous ones. However, this is also a more straightforward study, focusing on a single patient. Its methodology is also less complex, without proteomics studies, and rather focusing on the genome, transcriptome and very targeted immunohistochemical analysis.

The conclusions of the study are very relevant for this rare disease, and I particularly appreciated the application of Gene Set Enrichment Analysis to detect relevant pathways and genes related to them.

Unfortunately, none of the supplementary figures, tables and files of the published paper are included in the dissertation document. The readability of the paper is also not ideal, with the PDF being insert as an image.

Conclusions

The Conclusions section would have been the perfect place to bring cohesion to the thesis and explore how the three studies afford a complementary view of multi-omics in cancer. However, it serves only as a summary of contributions. It is a very clear and well-written summary, and at times hints at some of the questions that I had while reading the dissertation.

The challenges introduced at the beginning of the document, namely the trade-off between wet lab and database approaches, the noise introduced by data integration and the high computational power needs, could have been delved into. The text briefly mentions data integration as an enabler of a study, but it does not really go into any detail about this challenge. There is an acknowledgement that data heterogeneity due to sample preparation can result in reduced sensitivity, but no clear indication on how this affected the studies. It would have been interesting to see a discussion on the influence of the number of samples on the study conclusions, especially when comparing the EAC and UPS studies. Sample numbers are quite different in some parts of the study. A discussion on the trade-offs between cost and analytical power would have been welcome.

In sum, the dissertation reports on a body of high-quality research that is more than sufficient to support the awarding of a PhD degree.

_____

Catia Pesquita
Assistant Professor and Principal Investigator
LASIGE
Departamento de Informática
Faculdade de Ciências
Universidade de Lisboa