State of the Art on: Integrative analysis of transcriptional, mutational and DNA structural profiles in ovarian cancer of chemotherapy sensitive vs. resistant patients

Sara Sansone, sara.sansone@mail.polimi.it

1. Introduction to the research topic

Genomic computing is a new science focused on understanding the functioning of the genome in order to make fundamental discoveries in biology and medicine. The challenge is to answer to relevant questions for biological and clinical research, e.g. how cancer arises and develops, how driving mutations occur, how much complex disorders such as cancer depend on environmental factors or genetic predisposition; and then use individual genomic information for personalized/precision medicine. In particular, computational genomics refers to the use of computational and statistical analysis to decipher biology from genome sequences and related data, including both DNA and RNA sequence as well as other "post-genomic" data. Thus, it might be considered as a subset of bioinformatics and computational biology, but with a focus on using whole genomes rather than individual genes. This work use genomic computing to make an integrative analysis of one specific disease, i.e. ovarian cancer.

The research topic is mentioned in very prestigious journals, such as Nature, BMC Medical Genomics, PNAS (Proceedings of the National Accademy of Science of the United States of America).
First published in 1869, Nature is the world's leading multidisciplinary science journal. It publishes most significant discoveries—findings that advance knowledge and address some of the greatest challenges that we face as a society today. On the other hand, BMC Medical Genomics is an open access journal publishing original peer-reviewed research articles in all aspects of functional genomics, genome structure, genome-scale population genetics, epigenomics, proteomics, systems analysis, and pharmacogenomics in relation to human health and disease. Finally, PNAS one of the world's most-cited and comprehensive multidisciplinary scientific journals, publishing more than 3,200 research papers annually.
Ovarian cancer and its chemoresistance are also discussed during important conferences, like the meetings of the International Ovarian Cancer Consortium (IOCC), the World Cancer Congress (WCC) and the Global Summit on Oncology and Cancer. In particular, the WCC is a recognized international conference which encourages effective knowledge transfer and best practices exchange amongst over 3,500 cancer control and public health experts from 150 countries. On the other hand, the Global Summit on Oncology and Cancer hosts over 600+ Conferences, 1200+ Symposiums and 1200+ Workshops on Medical, Pharma, Engineering, Science, Technology and Business.

1.1. Preliminaries

In order to understand the research topic and to deal with it, it is necessary to work with some technological tools used in the field of genomic computing. A description of them will follow.
First of all, the data need to be extracted from The Cancer Genome Atlas (TCGA). It is a public funded project that aims to catalogue and discover major cancer-causing genomic alterations to create a comprehensive "atlas" of cancer genomic profiles. So far, TCGA researchers have analysed large cohorts of over 30 human tumours through large-scale genome sequencing and integrated multi-dimensional analyses. Studies of individual cancer types, as well as comprehensive pan-cancer analyses have extended current knowledge of tumorigenesis. A major goal of

the project was to provide publicly available datasets to help improve diagnostic methods, treatment standards, and finally to prevent cancer.

Secondly, in order to retrieve the data contained in TCGA it is used GenoMetric Query Language (GMQL), which operates upon aligned genomic data in a variety of data formats; it provides parallel computation in the cloud, thereby supporting queries over thousands of samples, such as the ones provided by ENCODE and TCGA consortia. The language's name indicates its ability to compute massive operations on genomic regions, which take into account region relative positions and distances.

Finally, there is the necessity to exploit some tool able to find the relevant regions of the genome, with respect to Copy Number Alteration (CNA). The state of the art on this is the use of GISTIC 2.0.23 (Genomic Identification of Significant Targets in Cancer) [2]. Indeed, the GISTIC module identifies regions of the genome that are significantly amplified or deleted across a set of samples. Each aberration is assigned a G-score that considers the amplitude of the aberration as well as the frequency of its occurrence across samples. False Discovery Rate q-values are then calculated for the aberrant regions, and regions with q-values below a user-defined threshold are considered significant.

1.2. Research topic

Ovarian cancer is the deadliest gynecologic malignancy, with a 5-year survival rate of approximately 47%, a percentage that has remained constant over the past two decades. Most ovarian cancers are epithelial in origin and their treatment prioritizes surgery and cytoreduction followed by cytotoxic platinum and taxane chemotherapy. While most tumors will initially respond to this treatment, recurrence is likely to occur within a median of 16 months for patients who have the disease at an advanced stage. Thus, this research project is concerned about stage III and IV of ovarian cancer and in particular, high-grade serous ovarian adenocarcinoma[1](HGS-OC). It is a rapidly growing carcinoma believed to have tubal origin with a high chromosomal instability; its peculiarity stands in the relapse timing of the patients affected by it. Indeed, patients can be recognized and differentiated into three classes, according to the time elapsed from the end of the first line therapy to relapse:

- relapse within 6 months since the end of treatment: *resistant*;

- relapse after 12 months since the end of treatment: *sensitive*;

- relapse after 36 months since the end of treatment: *sensitive long term*.

HGS-OC generally responds to platinum-based chemotherapy, but 80% of the patients relapse within 18 months from the diagnosis and progressively becomes resistant to treatment, up to becoming incurable: less than 20% of the patients survive after five years from the initial diagnosis.

For this reason, new treatment options separate from traditional chemotherapy, which consider achievements in understanding of the pathophysiology of ovarian cancer are needed to improve outcomes. Moreover, it is crucial to find a mechanism that allows to identify and discriminate resistant and sensitive patients, at the time of diagnosis. Hence, this study involves the analysis of resistance to chemotherapy in ovarian cancer patients, based on their transcriptional, mutational, and DNA structural profiles, in particular of the molecular differences between patients that are sensitive to therapy compared to patients that are resistant, using both in-house data and data sets from the TCGA (The Cancer Genome Atlas).

The ultimate aim is the identification of a molecular signature that could be used to predict the response to therapy (sensitive / resistant) at the time of diagnosis, starting from the Copy Number Alteration (CNA) profiles of the patients.

---

[1]Adenocarcinoma: malignant epithelial tumor that originates specifically from cells of the glandular epithelium

2. Main related works

2.1. Classification of the main related works

Studies have been done on chemo-resistance, either considering the problem in general or analyzing it with respect to ovarian cancer. All these works can be classified considering the kind of analysis made and its final goal. Indeed, the analysis might be done considering gene expression data, CNA data, microRNA data and so on, both separately or using some of them in combination. The goal instead might be to either predict chemoresistance in a patient analyzing its genome or try to understand what is the biological phenomenon that led to it.

In particular, there are studies based on gene expression alone with the aim of classifying cell line chemosensitivity. Other studies are concerned with the possibility to find relevant mutated genes and pathways of genes that are drivers for ovarian cancer and responsible of the chemo-resistance.

The latters [4] led to the identification of the mutation of TP53 in almost all tumor and of other nine genes, which mutations are statistically recurrent, and their results provide opportunities for new therapeutic treatment. Overall, these discoveries set the stage for approaches to the treatment of HGS-OvCa in which aberrant genes or networks are detected and targeted with therapies selected to be effective against these specific aberrations.

Another study [3] suggested that integrated analysis on DNA methylation and gene expression may allow for the identification of new therapeutic targets and/or biomarkers prognostic of disease response.

Thus, all the researches made so far on chemoresistance in ovarian cancer do not individuate a molecular signature that could be used to predict the response to therapy, which is what is needed to decide the right treatment at the time of diagnosis. They just tried to find biomarkers or therapeutic targets. Also, the precise mechanism(s) underlying the development of platinum resistance in late-stage ovarian cancer patients currently remains unknown.

2.2. Brief description of the main related works

The Cancer Genome Atlas researchers measured comprehensively genomic and epigenomic abnormalities on clinically annotated HGS-OvCa samples to identify molecular abnormalities that influence pathophysiology, affect outcome and constitute therapeutic targets [4]. They identified 10 recurrent mutated genes, 113 significant focal DNA copy number aberrations, and promoter methylation events involving 168 genes, thus providing a large-scale integrative view of the aberrations in HGS-OvCa.

On the other hand, an integrated analysis of DNA methylation and gene expression revealed signaling pathways related to platinum resistance in ovarian cancer. This study has been done treating clonally derived, drug-sensitive A2780 epithelial ovarian cancer cells with increasing concentrations of cisplatin [3] .

Also, it has been implemented an algorithm for classification of cell line chemosensitivity based on gene expression profiles alone [5]. Using oligonucleotide microarrays, the expression levels of 6,817 genes were measured in a panel of 60 human cancer cell lines for which the chemosensitivity profiles of thousands of chemical compounds have been determined. The study tried to determine whether the gene expression signatures of untreated cells were sufficient for the prediction of chemosensitivity. The final conclusion is that at least for a subset of compounds genomic approaches to chemosensitivity prediction are feasible.

Nevertheless, none of this works tried to use the Copy Number Altered regions of the genome to implement a classifier able to identify chemo-resistance in ovarian cancer. Moreover, even if comprehensive analysis of the genomic profiles of patients affected by this disease have been carried out, none of them led to a solution to the problem of predicting the resistance at the time of diagnosis.

For what concern the possibility to use the CNA data, different published studies make use of GISTIC 2.0 in order to analyze a set of patients, having the same disease, and extrapolate regions of the genomes with relevant CNA. An example is the research made by TCGA, which has already been mentioned above [4]. Also, Ducie et al. [1] used GISTIC for their molecular analysis of HGS-OvCa. In particular, they exploited this tool to identify the most significant focal DNA Somatic CNAs from TCGA and try to validate their thesis about the possible common biologic origin of HGS-OvCa with or without STIC associated lesions.

However, these works did not use the results obtained from GISTIC as a way to identify features for building a classifier, which is how it will be used in this research.

2.3. Discussion

Ovarian cancer has a 5-year survival rate of approximately 47%. Most deaths (around 70%) are of patients presenting with advanced-stage, high-grade serous ovarian cancer (HGS-OvCa). The standard treatment is aggressive surgery followed by platinum–taxane chemotherapy. After therapy, platinum-resistant cancer recurs in approximately 25% of patients within six months, and the overall five-year survival probability is 31%.
Studies have been done in order to understand the cause of this chemoresistance and to find aberrant genes or networks that can be targeted by specific therapies, which might be more effective than the standard ones. In particular, all kind of genomic data have been used and also some laboratory analysis have been done.
However, there is still no algorithmic tool able to predict the response to therapy of a patient, starting from her genomic profile. The existence of relevant regions of CNA and recurrent mutated genes might in fact be used in order to build a classifier able to achieve this goal. Thus, an accurate analysis of the data and about the feasibility of this proposal needs to be carried out. Indeed, it is crucial to first understand how far this research can go and then make the effort to achieve the best possible result in order to help biologists and doctors finding the right treatment for different ovarian cancer patients.

References

[1] Ducie, J., Dao, F., Considine, M., Olvera, N., Shaw, P. A., Kurman, R. J., Shih, I.-M., Soslow, R. A., Cope, L., and Levine, D. A. Molecular analysis of high-grade serous ovarian carcinoma with and without associated serous tubal intra-epithelial carcinoma. *Nature Communications 8*, 1 (2017), 990.

[2] Gad Getz, Rameen Beroukhim, C. M. S. S., and Dobson, J. Genomic identification of significant targets in cancer. GISTIC Technical Report Release 2.0.23, Broad Institute, March 2017.

[3] Li, M., Balch, C., Montgomery, J. S., Jeong, M., Chung, J. H., Yan, P., Huang, T. H., Kim, S., and Nephew, K. P. Integrated analysis of dna methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer. *BMC Medical Genomics 2*, 1 (Jun 2009), 34.

[4] Network, T. C. G. A. R., and Bell, e. A. Integrated genomic analyses of ovarian carcinoma. *Nature 474* (06 2011), 609 EP –.

[5] Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J., Scherf, U., Lee, J. K., Reinhold, W. O., Weinstein, J. N., Mesirov, J. P., Lander, E. S., and Golub, T. R. Chemosensitivity prediction by transcriptional profiling. *Proceedings of the National Academy of Sciences 98*, 19 (2001), 10787–10792.