

Statistical and bioinformatics applications in biomedical omics research

From the initiation of human genome project in 1990, and announcement of its completion in 2003, to the publications of first drafts of human proteome in Nature on May 29, 2014, these omics research projects have not only revolutionized the scientific research and clinical practices, but also have profound influence on agriculture, renewable energy development, biotechnology, and many other disciplines. Rapid advancements in high-throughput molecular technologies in the past several decades have enabled the identification and profiling of high-dimensional omics data, including DNA/RNA sequencing, RNA expression, methylation, DNA copy number variation, metabolomics, proteomics, and post-translational modification data. It is by no means a trivial task to analyze these Big Data generated by different platforms since there are various complex computational and analytical issues associated with them. The field of Biostatistics and Bioinformatics has played an essential role in solving computational issues and making sense of these data so that the research findings could lead to better understanding of human health, improve disease diagnosis, and provide opportunities for personalized treatment for complex human diseases such as cancer.

To equip readers with most updated analytical approaches for their potential use in cancer research, this special issue covers a wide range of modern statistical methodologies and bioinformatics tools. Hsu *et al.* review sparse principal component analysis, a powerful statistical tool for dimension reduction and feature extraction. A simulation study and an analysis on gene signature in a lung cancer dataset are used to illustrate the advantage of this approach. It is followed by a review of artificial neural networks, a powerful machine learning method, which deals with nonlinearity, by Oustimov and Vu. As datasets from multiple omics platforms become available, the analytical complexity also increases. Chalise and his coauthors provide a systematic review of existing integrative clustering methods for identifying novel molecular subtypes of a disease. Another computational challenge is how to properly incorporate temporal information in an analysis. Koestler and his colleagues propose a method, time-course recursively partitioned mixture model (TC-RPMM), a modified version of the RPMM, for clustering subjects based on

temporal profiles of gene expression using a mixture of mixed effects models. Modeling temporal information from a completely different angle, Qi and Voit uses a dynamic model of purine metabolism as the simulation system and metabolomics data as the input data to infer potential critical components of the uncontrolled cellular growth in colorectal cancer. Teer reviews new challenges we are facing due to the technological developments in massively-parallel sequencing and provides a comprehensive view of modern bioinformatics tools for sequence analyses. Guo *et al.* have evaluated several batch removal techniques for analyzing microRNA-Seq data and discussed their effectiveness. Protein kinases, the proteins that add a phosphate group to the substrate proteins during phosphorylation events, have become one of the largest groups of ‘druggable’ targets in cancer therapeutics in recent years. Chen and Eschrich review existing bioinformatics resources and statistical approaches for phosphorylation network inference, and their connection to therapeutics. Finally, Chen *et al.* offer a broad review of statistical issues for biomarker adaptive designs in clinical trials including adaptive signature design, interaction tests to identify predictive biomarkers, multiple testing, study power, and model validation.

We sincerely thank all contributors, authors and the editorial office for supporting this special issue with review or research articles on statistical approaches and bioinformatics tools for omics research.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Translational Cancer Research* for the series “Statistical and Bioinformatics Applications in Biomedical Omics Research”. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.3978/j.issn.2218-676X.2014.06.02>). The series

“Statistical and Bioinformatics Applications in Biomedical Omics Research” was commissioned by the editorial office without any funding or sponsorship. YAC served as the unpaid Guest Editor of the series. DTC served as the unpaid Guest Editor of the series and serves as an unpaid editorial board member of *Translational Cancer Research*. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are

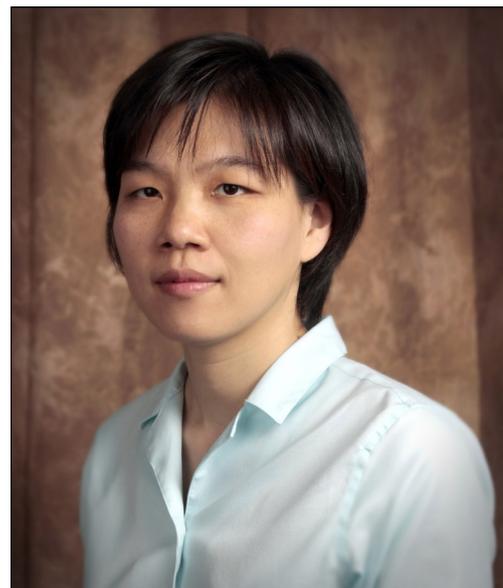
appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Dung-Tsa Chen & Yian Ann Chen



Dung-Tsa Chen, Senior Member. *Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, 12902 Magnolia Drive Tampa, FL 33612, USA. (Email: dung-tsa.chen@moffitt.org.)*



Yian Ann Chen, Assistant Member. *Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, 12902 Magnolia Drive Tampa, FL 33612, USA. (Email: Ann.Chen@moffitt.org.)*

Submitted: 31 May 2014; Accepted for publication: 3 June 2014

doi: 10.3978/j.issn.2218-676X.2014.06.02

View this article at: <http://dx.doi.org/10.3978/j.issn.2218-676X.2014.06.02>

Cite this article as: Chen DT, Chen YA. Statistical and bioinformatics applications in biomedical omics research. *Transl Cancer Res* 2014;3(3):180-181. doi: 10.3978/j.issn.2218-676X.2014.06.02