

## Bioinformatics Applications in Biological and Clinical Studies

Mao JH and Weier HUG\*

Department of Cancer & DNA Damage Responses, Life Sciences Division, University of California - LBNL, Berkeley, USA

\*Corresponding author: Heinz-Ulli G Weier, Department of Cancer & DNA Damage Responses, Life Sciences Division, E.O. Lawrence Berkeley National Lab, 1 Cyclotron Road, MS 977, Berkeley, CA 94720, USA, Tel: (001)510-486-5363; Fax: (001)510-486-5343; E-mail: ugweier@lbl.gov

Rec date: Feb 24, 2014, Acc date: Feb 27, 2014, Pub date: Feb 28, 2014

Copyright: © 2014 Mao JH, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Editorial

Welcome to the first Issue of Volume 5 of the Journal of Data Mining in Genomics and Proteomics (JDMGP). As in previous publications [1,2], articles compiled in the present issue of JDMGP are not centered on a sole topic. Instead, the Publisher, the Editorial Board and Reviewers have succeeded in assembling a compendium of nine articles that cover a wide range of bioinformatics applications ranging from bioinformatics strategies and challenges in proteome analysis to structural biology and the analysis coding or non-coding DNA expression under various conditions to map construction, ranking of genomes and strategies for sequence alignments.

In contrast to previous issues of JDMGP which placed a strong emphasis on Research Articles, JDMGP Volume 5, Issue 1, opens by presenting a set of three timely, succinctly written editorials introducing the reader to present strategies and challenges in applying bioinformatics tools in proteomics.

The first editorial article by D. Josic entitled 'Strategies for Complete Proteomic Analysis of Hydrophobic Proteins in Complex Biological Samples - Hyde-and Seek' [3] discusses shortfalls in commonly performed analyses of proteins in complex mixtures, namely the incomplete extraction of hydrophobic proteins prior to mass spec analysis. Since highly abundant hydrophilic proteins often mask the less abundant hydrophobic moieties, new strategies for proteomic analysis had to be developed and are discussed in this paper. Recent developments in mass spectrometry, sample preparation and separation are the focus of the contribution by B. Willard and colleagues 'Current Bioinformatics Challenges in Proteome Dynamics using Heavy Water-based Metabolic Labeling' [4]. This elegantly and concisely written paper describes the present challenges of metabolome analysis using heavy-water based labeling and freely available software for advanced proteome dynamics studies recently developed by the authors at the University of Texas. Very much in line with the two previous papers, the editorial by S.S. Natarajan entitled 'Analysis of Soybean Seed Proteins Using Proteomics' discusses approaches and challenges in the field of agricultural proteome analysis stressing the importance of the proper protein extraction method from soybean seeds for subsequent analysis by 2 dimensional-polyacrylamide gel electrophoresis (PAGE) [5]. While these three editorials present fine descriptions of protein extraction methods, analytic strategies and challenges, future contributions to the JDMGP might place a stronger emphasis on the role of pre-analytical parameters [6].

The following research article by K. Bamdad and co-workers 'Conformational Changes of a Chemically Modified HRP: Formation of a Molten Globule like Structure at pH 5' [7] discusses the impact of chemical modifications and pH changes on the structure and enzymatic activity of horseradish peroxidase (HRP). This article nicely

bridges the proteomics research discussed in the preceding three editorials to two contributions that focus on gene expression. In their interesting paper 'Measuring Inequalities in Gene Co-expression Networks of HIV-1 Infection Using the Lorenz Curve and Gini Coefficient' C. Ma and colleagues applied mathematical modeling to describe relations between variables with regard to gene co-expression in HIV-1 infected cells using a factor or coefficient that has previously been used in economics to describe income inequalities [8].

Much wet lab work has been invested in the study described by H. Yuwh et al. [9] called 'To Study the Effects of Shenqifuzheng Injection (SFI) on Micro-RNAs in Human Dendritic Cells' [9]. Here, the authors demonstrate the treating monocyte-derived human dendritic cells with one particular traditional Chinese medicine (SFI) did not show consistent upregulation in the expression of three relevant micro RNAs. Increased expression was only found in one cell lines, but not in peripheral blood cells from seven donors, thus raising questions about the importance of a previously described in vitro model.

The last three papers in this issue change the focus from proteomics and gene expression to the assembly of maps for mammalian genomes with goal to organize interactive properties of long non-coding RNAs, investigate amino acid alignment derived from multiple 3-dimensional structures and an approach to rank genomes to find associations that reveal ranks in regards to various properties.

The paper by Zhen et al. [10] 'Variant Map Construction to Detect Symmetric Properties of Genomes on 2D Distributions' [10] is quite timely by taking into account the roles of non-coding RNAs, many of which have been described just recently [11,12] and using an approach termed 'Variant Map Construction (VMC)'. The contribution by A. Bolshoy and coworkers 'Ranking of Prokaryotic Genomes Based on Maximization of Sortedness of Gene Lengths' [13,14] focuses on robust algorithms to group homologous genes into defined categories and defining the most expeditious alignment algorithm. Finally, in the paper 'Units: Universal True SDSA (Structure-Dependent Sequence Alignment)' S. Fey and G. Wyckoff describe a novel algorithm to align full protein sequences and to utilize the alignment to calculate improved structural quality assessment scores.

In summary, the current issue of JDMGP (Vol. 5, Issue 1) contains an exciting collection of nine research articles from labs that work at the cutting edge of bioinformatics applications in the life sciences. The articles describe ongoing work in a broader field of research focus mostly on the interests of the teams of authors. Present efforts at the publishing house are underway to publish a further volume with additional contributions on 'Bioinformatics for High Throughput Sequencing' [1] before the end of the 2014. Please see the JDMGP's Special Issue web site for the timeline and further information.

## Acknowledgements

This work was supported in parts by NIH grants CA168345 and CA116481 carried out at the Ernest Orlando Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231. We also wish to express our gratitude to the many peer reviewers, who devoted time and effort to select the very best contribution for this issue of the Journal.

## Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

## Conflict of Interest

The author declares no conflict of interest.

## References

1. Weier HU (2013) Bioinformatics for High Throughput Sequencing. *J Data Mining Genomics Proteomics* 4: e108.
2. Weier HU (2013) New Algorithms for Genome Characterization, Epigenetic Profiling Analysis, Data Mining and Population-based Studies. *J Data Mining Genomics Proteomics* 4: e109.
3. Josic D (2014) Strategies for Complete Proteomic Analysis of Hydrophobic Proteins in Complex Biological Samples – Hyde-and Seek. *J Data Mining Genomics Proteomics* 5: e111.
4. Willard B, Sadygov RG, Kasumov T (2014) Current Bioinformatics Challenges in Proteome Dynamics using Heavy Water-based Metabolic Labeling. *J Data Mining Genomics Proteomics* 5: e112.
5. Natarajan SS (2014) Analysis of Soybean Seed Proteins Using Proteomics. *J Data Mining Genomics Proteomics* 5:e113.
6. Moore HM, Kelly AB, Jewell SD, McShane LM, Clark DP, et al. (2011) Biospecimen Reporting for Improved Study Quality (BRISQ). *J Proteome Research* 10: 3429-3438.
7. Bamdad K, Bijan Ranjbar B, Naderimanesh H, Sadeghi M (2014) Conformational Changes of a Chemically Modified HRP: Formation of a Molten Globule like Structure at pH 5. *J Data Mining Genomics Proteomics* 5: 147.
8. Chuang Ma, Sheng-He Huang, Yanhong Zhou (2014) Measuring Inequalities in Gene Co-expression Networks of HIV-1 Infection Using the Lorenz Curve and Gini Coefficient. *J Data Mining Genomics Proteomics* 5: 148.
9. Yuwh H, Sze WH, Yip DMY, Cho SP, Boost WCS, et al. (2014) To Study the Effects of Shenqifuzheng Injection (SFI) on Micro-RNAs in Human Dendritic Cells. *J Data Mining Genomics Proteomics* 5:149.
10. Zheng J, Zhang W, Luo J, Zhou W, Liesaputra V (2014) Variant Map Construction to Detect Symmetric Properties of Genomes on 2D Distributions. *J Data Mining Genomics Proteomics* 5: 150.
11. Ren S, Peng Z, Mao JH, Yu Y, Yin C et al. (2012) RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Res* 22: 806-821.
12. Weier HUG (2012) Rags to Riches: Re-viewing the Tumor-specific Expression of Long Noncoding DNA and Satellite DNA Repeats in Humans in the Era of Next Generation Sequencing. *J Data Mining in Genomics Proteomics* 3: e101.
13. Bolshoy A, Salih B, Cohen I, Tatarinova T (2014) Ranking of Prokaryotic Genomes Based on Maximization of Sortedness of Gene Lengths. *J Data Mining Genomics Proteomics* 5: 151.
14. Foy S, Wyckoff G (2014) Units: Universal True SDSA (Structure-Dependent Sequence Alignment). *J Data Mining Genomics Proteomics* 5: 152.