Journal of Biomedicine and Biotechnology • 2005:2 (2005) 63–64 • DOI: 10.1155/JBB.2005.63

EDITORIAL

Data Mining in Genomics and Proteomics

Halima Bensmail[1] and Abdelali Haoudi[2]

[1] *Department of Statistics, The University of Tennessee, Knoxville, TN 37996-0532, USA*
[2] *Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk, VA 23507-1696, USA*

There is no doubt that both computational biology and bioinformatics, and the interface of computer science and biology in general, are central to the future of biological research. The disciplines span a process that begins with data collection, analysis, classification, and integration, and ends with interpretation, modeling, visualization, and prediction. Data mining plays a role in the middle of this process. Overall, the focus is on identifying opportunities and developing computational solutions (including algorithms, models, tools, and databases) that can be used for experimental design, data analysis and interpretation, and hypothesis generation.

Data mining is the search for hidden trends within large sets of data. Data mining approaches are needed at all levels of genomics and proteomics analyses. These studies can provide a wealth of information and rapidly generate large quantities of data from the analysis of biological specimens from healthy and diseased tissues. The high dimensionality of data generated from these studies will require the development of improved bioinformatics and computational biology tools for efficient and accurate data analyses.

This issue of the Journal of Biomedicine and Biotechnology consists of seventeen papers that describe different applications of data mining to both genomics and proteomics studies in yeast, and plant and human cells and tissues. Papers by Bensmail et al, Ghosh and Chinnaiyan, and Mao et al present different classification and clustering approaches for disease biomarkers discovery. Genomics and proteomics studies have shown great promises and have been applied to studies aiming at generating expression profiles and elucidating expression networks in different organisms as shown in the papers by Samsa et al, Mungur et al, Liu et al, Baldwin et al, and Joy et al. Data mining in genomics and proteomics studies reveals new regulatory pathways and mechanisms in different health and disease conditions as presented by Wren and Garner, and provides comparative sequence analysis approaches as presented by Gambin and Otto and Gao et al. Those studies have also provided approaches for subcellular localization of proteins suggesting that such approaches can produce an objective systematics for protein location and provide an important starting point for discovering sequence motifs that determine localization as presented by Chen and Murphy. Chen et al studied the performance of five nonparameteric tests to select genes and proved that the popular F test does not perform well on gene expression data since the heterogeneity behavior assumption is the most dominant in the gene expression data. Corder et al explored a statistical approach called grade of membership (GOM) and proved that brain hypoperfusion contributes to dementia, possibly to Alzheimer's disease (AD) pathogenesis, and raises the possibility that the APOE ε^4 allele contributes directly to heart value and myocardial damage. Hand and Heard present in their review article various tools for finding relevant subgroups in gene expression data. Alkharouf et al conduct an OLAP cube (online analytical processing) to mine a time series experiment designed to identify genes associated with resistance of soybean to the soybean cyst nematode, which is a devastating pest of soybean. Brylinski et al created a sequence-to-structure library based on the complete PDB database. Then an early-stage folding conformation and information entropy were used for structure analysis and classification.

Whilst postgenomic science is producing vast data torrents, it is well known that data do not equal knowledge and so the extraction of the most meaningful parts of these data is key to the generation of useful new knowledge. More sophisticated data mining strategies are needed for mining such high-dimensional data to generate useful relationships, rules, and predictions.

Halima Bensmail
Abdelali Haoudi

Correspondence and reprint requests to Abdelali Haoudi, Eastern Virginia Medical School, Department of Microbiology and Molecular Cell Biology, Norfolk, VA 23507-1696, USA, E-mail: haoudia@evms.edu

Halima Bensmail received her PhD degree jointly from Pierre & Marie Curie University, Paris, France, and the University of Washington, Seattle, in 1996 in statistics and mathematical modelling. She then joined the University of Washington for a Researcher position for a period of two years, then worked as a Consultant and Associate Researcher at the Fred Hutchinson Cancer Research Center. She joined the University of Leiden for a period of three years. She joined the University of Tennessee for a Statistics Assistant Professor position at the Department of Statistics in 2000. Dr. Bensmail is also an Associate Editor for the Journal of Biomedicine and Biotechnology and a Reviewer at the NIH and NSF. She has established numerous collaborations both within the academia (EVMS, Georgia State University, University of California, Oak Ridge Laboratory) and with the private sector (HIRST Company for Hedge Fund Strategy Benchmarks, Federal Bank of Atlanta). She has advised numerous PhD and Master's students and cochaired many conferences particularly on data mining. She is a member of several scientific organizations and has received numerous scientific and teaching awards.

Abdelali Haoudi received his PhD degree in cellular and molecular genetics jointly from Pierre & Marie Curie University and Orsay University in Paris, France. He then joined the National Institutes of Health (NIEHS, NIH) for a period of four years after winning the competitive and prestigious NIH Fogarty International Award. Dr. Haoudi then joined the faculty in the Department of Microbiology and Molecular Cell Biology at Eastern Virginia Medical School in Norfolk, Va, in 2001. Dr. Haoudi is primarily interested in the elucidation of the molecular basis of cancer including cell cycle checkpoints, DNA repair and apoptosis, in addition to the development of cancer gene therapeutic strategies. Dr. Haoudi is also the codirector of the Cancer Biology and Virology Focal Group. He has founded the Journal of Biomedicine and Biotechnology (http://www.hindawi.com/journals/jbb) and is also the Founder and President of the International Council of Biomedicine and Biotechnology (http://www.i-council-biomed-biotech.org). He is a member of several scientific organizations and has received numerous scientific awards.