Introduction to the Issue on Genomic and Proteomic Signal Processing

DURING the last several years, substantial progress has been made on developing high-throughput experimental techniques that produce large amounts of genomic and proteomic data pertaining to molecular activities in cells. Consequently, a great deal of research is being focused on addressing important problems in molecular biology by analyzing these data using mathematical and computational approaches. A problem of great interest is the development of methodologies that extract important information from genomic and proteomic data and use this information to build models of molecular biological systems. Computational analysis of these models can produce a wealth of knowledge about the structure, function, properties and control of living systems, which can be ultimately used to develop novel diagnostic tools, effective drug design, and therapeutic approaches for treating human diseases.

It is increasingly recognized that signal processing research may have an enormous impact on extracting, processing, and analyzing genomic and proteomic information and on using this information to build models of molecular biological systems. As a result, genomic and proteomic signal processing is a rapidly growing field within the signal processing community. Current activities in this area are gaining popularity in many universities and research labs worldwide. Many investigators are eager to expose themselves to basic research issues in molecular biology and understand how modern signal processing approaches can be used to address some of these issues.

From a signal processing perspective, genomic and proteomic data can be viewed as noisy (continuous or discrete) signals that convey valuable information about molecular structure and activities in cells. However, the high dimensionality, variability, and complexity of such data pose enormous challenges. Therefore, there is an immediate need to develop novel signal processing methodologies that effectively deal with these challenges. It is the purpose of this Special Issue to present the latest work in the field and demonstrate the enormous potential of signal processing in addressing problems of molecular biology.

In this Special Issue, we feature recent advances in the area of genomic and proteomic signal processing and highlight emerging challenges and applications. The articles herein bring to light the increasing role of genomic and proteomic signal processing in biology and medicine. Most importantly, the fifteen articles selected for this Special Issue present significant contributions in four primary areas: microarray analysis, sequence analysis, structure identification, and regulatory networks.

Microarray Analysis: The articles in this area provide new directions for the analysis of microarray data. The article by Meyer *et al.*, entitled "Information-theoretic feature selection in microarray data using variable complementarity" proposes a filtering approach for feature selection in microarray data based on mutual information. The authors show that the problem of feature selection can be reduced to a well-known quadratic optimization problem.

The article by Parvaresh *et al.*, entitled "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays" examines the possibility of measuring differentially expressed genes using microarrays with a number of properly designed probes that is much smaller than that of probes in conventional microarrays.

In the article "Modeling and estimation for real-time microarrays," Vikalo *et al.* propose a new method for measuring differentially expressed genes, called real-time microarray. The proposed method provides higher signal-to-noise ratio and broader assay detection dynamic range.

Sequence Analysis: The articles in this subtopic study the deterministic and statistical properties of genomic sequences. In the article "Feature selection for self-supervised classification with applications to microarray and sequence data," Kung and Mak examine several learning strategies pertaining to feature selection. The authors apply their method to time-course microarray data as well as sequence data.

The article by Akhtar *et al.*, entitled "Signal processing in sequence analysis: advances in eukaryotic gene prediction," brings a new view for advancing digital signal processing techniques applied to genomic sequences. The article reviews and evaluates the wealth of methodologies for symbolic to numeric transformation of the original DNA sequence in connection to DSP methods for gene and exon prediction problems on large data sets.

The article by Liang and Wang, entitled "A deterministic sequential Monte-Carlo method for haplotype inference," deals with the problem of detecting nucleotide alterations in the DNA of a species. The authors employ a statistical approach using a deterministic sequential Monte-Carlo sampling method.

The article by Arora *et al.*, entitled "Latent periodicities in genome sequences," presents a new approach for finding repetitions and regularities in DNA sequences. The specific type of regularity addressed in the article is latent periodicity, a less obvious periodic phenomenon when compared to the well-studied homologous or eroded periodicities.

In the article "A DSP approach for finding the codon bias in DNA sequences," Tuqan and Rushdi provide a mechanism to explain the period-3 component in the Voss-based short-time discrete Fourier transform. They establish a relationship between the detection of the period-3 component and nucleotide bias in the codon structure.

The article by Bouaynaya and Schonfeld, entitled "Non-stationary analysis of coding and non-coding regions in nucleotide

sequences," presents a new view in the study of long-range correlations of DNA sequences. With the appropriate signal processing tools, new light can be shed on the behavior of coding versus non-coding regions of DNA.

In the article entitled "Efficient updating of biological sequence analyses," Hong and Tewfik deal with the important task of updating an appropriately obtained DNA match of a DNA sequence, when a newer version of that sequence becomes available. They propose a novel solution to this problem, which employs a hidden Markov model and a Viterbi recursion.

Structure Identification: The articles in this category investigate the detection and identification of the structure of genes and proteins. The article by Ramachandran and Antoniou, entitled "Identification of hot-spot locations in proteins using digital filters," uses digital filters to locate hot-spots in proteins by finding the characteristic frequency of the amino-acid sequence of a protein.

The article by Savir and Tlusty, entitled "Optimal design of a molecular recognizer: molecular recognition as a Bayesian signal detection problem," studies the ability of molecules to recognize target molecules. The authors show that the structure of the optimal molecular recognizer changes dynamically with the structural properties of the molecules and interaction energies between them.

In the article "Fast structural alignment of RNAs by optimizing the adjoining order of profile-csHMMs," Yoon and Vaidyanathan propose a novel approach for speeding up the SCA algorithm used with profile-cs hidden Markov models to provide significant speed up in RNA alignment.

Regulatory Networks: The articles in this group explore the relation and control of interacting genes. The article by Faryabi *et al.*, entitled "Optimal intervention in asynchronous genetic regulatory networks," develops methods for optimal intervention strategies in regulatory networks. The objective is to ensure that asynchronous genetic regulatory networks spend very little time in undesirable states.

In the article "Intrinsically multivariate predictive genes," Martins *et al.* focus on identification of predicting genes based on a set of genes such that only the full set can be used to predict the target. This situation arises when a gene exerts tight regulation over multiple uncorrelated pathways.

We believe that the articles published in this Special Issue provide an overview of the state-of-the-art in the field of genomic and proteomic signal processing. We also hope that the Special Issue will accelerate growth of this exciting field of research and help in disseminating some important biological issues to the signal processing community. We are grateful to the authors for their outstanding contributions and to the reviewers for their thorough evaluations. We finally extend our gratitude to the Editor-in-Chief, Dr. Lee Swindlehurst, for his kind support and guidance.

Dan Schonfeld, *Guest Editor*
Department of Electrical & Computer Engineering
University of Illinois at Chicago
Chicago, IL 60607-7053
dans@uic.edu

John Goutsias, *Guest Editor*
Center for Imaging Science
The Johns Hopkins University
Baltimore, MD 21218-2686
goutsias@jhu.edu

Ilya Shmulevich, *Guest Editor*
Institute for Systems Biology
Seattle, WA 98103-8904
ishmulevich@systemsbiology.org

Ioan Tabus, *Guest Editor*
Institute of Signal Processing
Tampere University of Technology
Tampere, Finland
ioan.tabus@tut.fi

Ahmed H. Tewfik, *Guest Editor*
Department of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455
tewfik@umn.edu

Dan Schonfeld (SM'05) received the B.S. degree in electrical engineering and computer science from the University of California at Berkeley in 1986 and the M.S. and Ph.D. degrees in electrical and computer engineering from the Johns Hopkins University, Baltimore, MD, in 1988 and 1990, respectively.

In 1990, he joined the University of Illinois at Chicago, where he is currently a Professor in the Department of Electrical and Computer Engineering. He has authored over 120 technical papers in various journals and conferences. His current research interests are in multi-dimensional signal processing, image and video analysis, computer vision, and genomic signal processing.

Dr. Schonfeld was co-author of papers that won the Best Student Paper Awards in Visual Communication and Image Processing 2006 and IEEE International Conference on Image Processing 2006 and 2007. He is currently serving as Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING as well for the IEEE TRANSACTIONS ON IMAGE PROCESSING.

John Goutsias (SM'94) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981 and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, CA, in 1982 and 1986, respectively.

In 1986, he joined the Department of Electrical and Computer Engineering at The Johns Hopkins University, Baltimore, MD, where he is currently a Professor of Electrical and Computer Engineering, a Whitaker Biomedical Engineering Professor, and a Professor of Applied Mathematics and Statistics. His research interests include computational systems biology, modeling and analysis of biochemical reaction systems, bioinformatics, genomic signal processing, and image processing and analysis.

Dr. Goutsias served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1991–1993), the IEEE TRANSACTIONS ON IMAGE PROCESSING (1995–1997), the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (2002–2006), and the *Journal of Visual Communication and Image Representation* (1993–2005). He also served as a Guest Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING Special Issue on Genomic Signal Processing (2006). He is currently a co-editor for the *Journal of Mathematical Imaging and Vision* and an Associate Editor for the EURASIP *Journal on Signal Processing and Bioinformatics*.
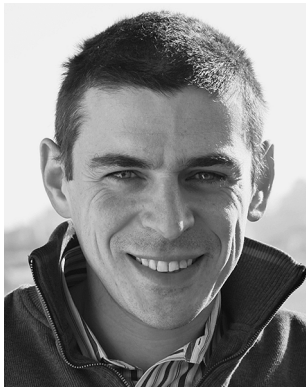
Ilya Shmulevich (SM'04) received the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1997.

In 1997–1998, he was a Postdoctoral Researcher at the Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands, and at the National Research Institute for Mathematics and Computer Science, University of Amsterdam, The Netherlands, where he studied computational models of music perception and recognition. In 1998–2000, he was a Senior Researcher at the Tampere International Center for Signal Processing, Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland. From 2001–2005, he was an Assistant Professor at the Cancer Genomics Laboratory, Department of Pathology, The University of Texas M. D. Anderson Cancer Center, and an Adjunct Professor in the Department of Statistics, Rice University, Houston, TX. Presently, he is an Associate Professor at the Institute for Systems Biology, Seattle, WA, and an Affiliate Professor in the Departments of Bioengineering and Electrical Engineering at the University of Washington. His research interests include systems biology, nonlinear signal and image processing, and computational learning theory.

Dr. Shmulevich is an Associate Editor of the EURASIP *Journal on Bioinformatics and Systems Biology*.

Ioan Tabus (SM'99) received the M.S. degree in electrical engineering in 1982 and the Ph.D. degree in 1993 from the "Politehnica" University of Bucharest, Bucharest, Romania, and the Ph.D. degree (Hons.) from Tampere University of Technology (TUT), Tampere, Finland, in 1995.

He was a Teaching Assistant, Lecturer, and Associate Professor with the Department of Control and Computers, "Politehnica" University of Bucharest, Bucharest, Hungary, between 1984 and 1995. From 1996 to 1999, he was a Senior Researcher, and since January 2000, a Professor at the Institute of Signal Processing at TUT. He is co-author of more than 170 publications in the fields of signal processing, signal compression, and genomic signal processing.

Dr. Tabus was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING between 2002 and 2005. He is a member of the TC Bio Image and Signal Processing of the IEEE Signal Processing Society. He is currently the Editor-in-Chief of the EURASIP *Journal on Bioinformatics and Systems Biology* and an Associate Editor of the journal *Signal Processing*. Dr. Tabus was co-recipient of the 1991 "Traian Vuia" Award of the Romanian Academy, co-recipient of the NSIP 2001 Best Paper Award, and co-recipient of the Norsig 2004 Best Paper Award.

Ahmed H. Tewfik (F'96) received the B.Sc. degree from Cairo University, Cairo, Egypt, in 1982 and the M.Sc., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1984, 1985, and 1987, respectively.

He was with Alphatech, Inc., Burlington, MA, during 1987. He is the E. F. Johnson Professor of Electronic Communications with the Department of Electrical Engineering, University of Minnesota. He served as a Consultant to several companies in Minnesota, California, and New York in the areas of industrial measurements, multimedia protection and distribution, wireless communications, medical instrumentation, and gigabit Ethernet. Between 1997 and 2001, he was the President and CEO of Cognicity, an entertainment marketing software tools publisher that he co-founded. His current research interests are in genomics and proteomics, wearable sensors for patients at cardiac risk or with traumatic brain injury, programmable wireless networks, and brain computing interfaces.

Dr. Tewfik was a Distinguished Lecturer of the IEEE Signal Processing Society during 1997–1999. He is the recipient of several awards, including the IEEE Third Millennium Award in 2000. He was elected to the Board of Governors of the IEEE Signal Processing Society in 2005.