SIGBioinformatics Record

Newsletter of the SIGBioinformatics
ACM Special Interest Group

Volume 3, Issue 1, January 2013 ISSN 2159-121

SIGBioinformatics Organization

Chair

Aidong Zhang
State University of New York at Buffalo (UB)

Vice-Chair

Gultekin Ozsoyoglu
Case Western Reserve University

Secretary-Treasurer

Armin R. Mikler
University of North Texas

Newsletter Editors

Pietro Hiram Guzzi
University of Catanzaro

Young-Rae Cho
Baylor University

Board of Directors

Vasant Honavar
Iowa State University

Sun Kim
Seoul National University

Vipin Kumar
University of Minnesota

Yi Pan
Georgia State University

Marie-France Sagot
INRIA

Cathy Wu
University of Delaware

Director of Conferences

Vasant Honavar
Iowa State University

Awards Committee

Umit V. Catalyurek (Co-Chair)
Ohio State University

Wei Wang (Co-Chair)
University of California at Los Angeles

Constantin Aliferis
NYU School of Medicine

Jonathan Dennis
Florida State University

Tony Hu
Drexel University

Tamer Kahveci
University of Florida,

Parthasarathy Srinivasan
Ohio State University

Joe Zhang
University of Southern Mississippi

Zhongming Zhao,
Vanderbilt University Medical Center

Information Director

Mohammed Zaki
Rensselaer Polytechnic Institute

Associate Information Directors

Guo-zheng Li
Tongji University

Li Liao
University of Delaware

Yunlong Liu
Indiana-Purdue University at Indianapolis (IUPUI)

Liaison to Industry and Other Societies

Sanjay Ranka
University of Florida

European liaison

Mario Cannataro,
University of Catanzaro,

Asian Liaison

Guo-zheng Li
Tongji University

Women in Bioinformatics Outreach Chair

May Wang
Georgia Institute of Technology

Community Outreach Chair

Arup K Ghosh
University of Central Florida

Advisory Board

Pierre Baldi (Co-Chair)
University of California-Irvine

Mark Borodovsky (Co-Chair)
Georgia Institute of Technology

Jeff Bennetzen
University of Georgia

Volker Brendel
Iowa State University

Robert Cottingham
Oak Ridge National Laboratory

Dmitrij Frishman
Technische Universität München

Mark Gerstein
Yale University,

Nick V. Grishin
University of Texas, Southwestern Medical Center

Minoru Kanehisa
University of Tokyo

Nikos Kyrpides
DOE-Joint Genome Institute (JGI)

Meral Ozsoyoglu
Case Western Reserve University

Sir Richard John Roberts
New England Biolabs

Dong Xu
University of Missouri-Columbia

Notice to Contributing Authors to SIG Newsletters

By submitting your article for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- to publish in print on condition of acceptance by the editor
- to digitize and post your article in the electronic version of this publication
- to include the article in the ACM Digital Library and in any Digital Library related services
- to allow users to make a personal copy of the article for noncommercial, educational or research purposes

However, as a contributing author, you retain copyright to your article and ACM will refer requests for republication directly to you.

Table of Contents

1. Editor-in-Chief's Notes

The SIGBioinformatics Record presents some changes that reflect the discussion made on the last BCB Conference. This issue presents a novel structure based on following sections:

- *Contributed Articles*: It contains a paper by Hu et al. which discusses novel trends of research in microbiome analysis.
- *Reports*: It contains two workshop reports. The first paper by Bellazzi et al. presents main results on a workshop on Biobanks held in Pavia on November 2012. The second paper by Romano et al. discusses experiences from the NETTAB workshop series.

We also thank the work by Amarda Shehu that is now responsible for a novel section "Profiles" that will be devoted to the presentation of biographies.

Finally the next ACM BCB conference is presented.

Pietro Hiram Guzzi,
Newsletter Editor-in-Chief

2. Notes from Chair

SIG Bioinformatics Status:

As of January 1, 2013, we have 430 members of the society, which includes those new members who were just transferred from the SIG Health Informatics (SIGHIT). As you may know, the ACM SIG Governing Board Executive Committee has decided not to move ahead with the final chartering of the Special Interest Group on Health Informatics (SIGHIT). All members of SIGHIT were offered membership in SIGBioinformatics. SIGHIT members who already hold membership in SIGBioinformatics will be provided with a 1 year extension. If requested, transfer of membership to another SIG will be considered on a case-by-case basis.

We are a much bigger community now. To better represent our community, the SIGBioinformatics Executive Committee has unanimously approved to change our name to SIGBCB (Bioinformatics, Computational Biology, and Biomedical Informatics), which was first proposed at the BCB Conference in Orlando, Oct, 2012. A formal request will be submitted in the coming weeks.

Special Program 1 – Women in Bioinformatics

In 2011, we started a special program, called Women in Bioinformatics, sponsored by the National Science Foundation during the BCB2011 conference. The purpose of this program is to encourage female students to get involved in the bioinformatics research. We featured one keynote speech by Nancy Cox, from University of Chicago and a special panel moderated by the panel chair, Cathy Wu, from the University of Delaware. We were able to support 21 female and minority students to participate in the BCB2011 conference; many of them would not have been able to attend otherwise.

We continued this program in BCB2012 meeting. A keynote speech was delivered by Martha L. Bulyk and a “Women in Bioinformatics” Forum was held on Oct. 9, which featured a set of presentations of both female faculty members and PhD students. There was a large audience. Many good questions were raised.

We have now appointed the “Women in Bioinformatics Outreach Chair” in the SIG Bioinformatics organization, Dr. May Wang from Georgia Tech and Emory University. Some students who participated in the “Women in Bioinformatics” have volunteered to help. These students will help the chair to gather information relevant to Women in Bioinformatics, and the chair will facilitate information exchange for Women in Bioinformatics such as a quarterly news update or a website for discussion.

Special Program 2 – PhD Students Forum

A PhD students Forum was held in BCB2012 which featured the poster presentations of 23 PhD students from various universities. This forum provided a platform for PhD students to network, practice presentation skills and exchange their ideas and research experiences.

Special Program 3 – Knowledge Repository

As part of our knowledge repository, we have started our PhD Dissertation Abstract Repository. An initiative to create a PhD dissertation abstract repository has been proposed and started by Dr. Armin Mikler. This repository will host PhD dissertation abstracts collected from PhD students and will be available for public access.

SIG Bioinformatics Awards

The SIG Bioinformatics award committee has proposed the following plan for establishing new SIGBioinformatics awards:

- *Best Paper* and *Best Student Paper Awards* – through our annual BCB conference, which is the flagship conference of our SIG, we plan to give two awards: Best Paper and Best Student Paper. The award selection process will follow the formal review process of the conference. The PC Chair(s) of the conference, in consultation with their Vice Chairs, will create a short list of papers based the peer reviews of the papers. Awards committee will review these short lists and make a selection. Selection will be announced during the BCB conference. We are in the process of finalizing the proposal to ACM and planning to give the first award during BCB 2013.
- *Best PhD Dissertation Award* – Starting from 2014, by leveraging the “Special Program 3 – PhD Dissertation Abstract Repository” mentioned above, awards committee will review the PhD dissertations submitted to repository throughout the year, and will select best dissertation.
- Starting from 2015 we plan to start two new awards 1) *Outstanding Service Award,* that will awarded to an individual for their service to SIGBioinformatics community, 2) *Research Achievement Award* to an individual for a body of work presented at the conference 5 years (or more) earlier and had the most impact, measured by using quantifiable metrics like citations, software usage etc.

Best paper and best student paper awards will be continuous and there will be at least one recipient each year. However, we envision that Best PhD Dissertation Award, Outstanding Service Award and Research Achievement Award may not be awarded every year.

SIG Outreach Activities

We now have appointed a “European liaison for SIG Bioinformatics,” Dr. Mario Cannataro from University "Magna Græcia" of Catanzaro, Italy. The main activities Mario proposes are to create a network between the SIG Bioinformatics and the main European Bioinformatics

societies and institutions. The Main aim of this network is to enable exchange of ideas and experiences (e.g., by cross referencing web portals, by allowing the conduction of surveys among society members, and by providing mutual advertising). Moreover, plans are under way to provide mutual discounts to members in the registration fee of conferences.

We have also appointed an “Asian liaison for SIG Bioinformatics,” Dr. Guo-Zheng Li from Tongji University, China. Dr. Li will help us to reach out to Asian communities to get involved in SIG Bioinformatics activities. Dr. Li will create and enhance the relation between SIG Bioinformatics and the main Chinese, Japanese and Korea Bioinformatics societies and institutions’ with the aim of providing mutual recognition that may be useful in many ways, e.g., enabling discounts on the society subscription fee or in the registration fee of conferences organized by such societies, and especially enabling exchange of ideas and experiences, e.g., by cross referencing web portals, by allowing the conduction of surveys among society members, and by providing mutual advertising (such as bioinformatician recruitment, and/or by reviewing books/software tools).

We have also appointed the “Community Outreach Chair” in the SIG Bioinformatics organization, Mr. Arup Ghosh from University of Central Florida. Mr. Ghosh will help SIG Bioinformatics to reach out to local ACM Chapters for them to be involved in SIG Bioinformatics activities. A Bioinformatics job board will be edited in the newsletter and a directory of Bioinformatics PhD programs will be maintained in the SIG Bioinformatics website

3. Contributed Articles

Exploratory analysis of human microbiome by linear and nonlinear methods

Xingpeng Jiang, Xiaohua Hu

Drexel University

xpjiang@drexel.edu, xh29@drexel.edu

Weiwei Xu

Wuhan University

xh29@drexel.edu

Tingting He

Central China Normal University

tthe@mail.ccnu.edu.cn

ABSTRACT

There is a growing recognition that the human microbiome - microbes living in intimate association with us - forms a vital part of our biology, and plays an important role in both health and sickness. A huge amount of data are being generated about these communities, much through metagenomics methods, which sequence DNA without directly identifying which organisms they come from. These data pose a tremendous opportunity for understanding, and a tremendous computational and theoretical challenge. In this paper, we compare several linear and nonlinear methods to explore human microbiome.

Keywords

Human microbiome, nonnegative matrix factorization, principle component analysis, multidimensional scaling analysis, t-distributed stochastic neighbor embedding

1. INTRODUCTION

Recently several classic methods have been employed extensively to visualize and explore the complex structure embedded in metagenomic profiles which summarize the abundance of functional or taxonomic categorizations in metagenomic sequences. These data-driven methods are usually linear methods such as principal component analysis (PCA), nonnegative matrix factorization (NMF), classical multidimensional scaling (MDS, also called principle coordinate analysis, PCoA).

A metagenomic profile typically has hundreds of metabolic pathways, thousands of species, tens of thousands of protein families. After reducing the dimension, metagenomic profiles are usually represented by several "components" which may facilitate biological interpretation and discovery. For example, PCA has been used frequently in metagenomic profiles which summarized the abundance of functional or taxonomic categorizations of metagenome sequences (Qin, Li et al. 2010). Another method - MDS which is based on the dissimilarities of data instead of similarity in PCA has been adopted as a standard technology for visualizing the taxonomic relationships in microbial communities (Kolehmainen, Kuvaja-Mikkonen et al. 1984). Recently a non-negative matrix factorization (NMF) framework has been used in analyzing metagenomic profiles to gain a different and complementary perspective on relationships between functions, environment, and biogeography of global ocean (Jiang, Langille et al. 2012).

An enterotype is a classification of living organisms based on its bacteriological ecosystem in the gut microbiome (Arumugam, Raes et al. 2011; Siezen and Kleerebezem 2011). The discovery indicates that the three human enterotypes which are not dictated by age, gender, body weight, or national divisions. Another study indicates that long-term diet influences enterotype (Wu, Chen et al. 2011). It is also interesting to see that Chimpanzees and humans have similar enterotypes and the same study found that the enterotype of an individual chimpanzee varied over time (Moeller, Degnan et al. 2012). Type 1 is characterized by high levels of Bacteroides, type 2 has few Bacteroides but Prevotella are common, and type 3 has high levels of Ruminococcus. Although PCA and MDS have been used in these studies to discover enterotypes in human microbiome, it is still interesting to see if other linear and nonlinear methods can identify novel signals or patterns in human microbiome.

In this paper, we compare several dimensional reduction and visualization methods to see the organization of human gut microbiomes. These methods include PCA, MDS, NMF, Isomap and t-distributed stochastic neighbor embedding (t-SNE). See Table 1 for a summary of these methods.

Table 1. Methods summary

Methods	Abbreviation	Linearity	References
Principle component analysis	PCA	Yes	(Pearson, 1901)
Nonnegative matrix factorization	NMF	Yes	(Lee and Seung 1999)
Classic Multidimensional analysis	MDS	Yes	(Carroll and Arabie 1980)
Isometric Feature Mapping	Isomap	No	(Tenenbaum, de Silva et al. 2000)
t-distributed stochastic neighbor embedding	t-SNE	No	(Matten and Hinton, 2008)

2. MOTIVATION

The vast majority of methods employed in current metagenomics analysis are under the hypothesis that structures and relationships in a microbial community are linear. However, the interactions among microbiota are most likely nonlinear instead of linear and the mathematical space of microbiota is most like in manifold instead of Euclidean space. The deep investigation of manifold learning method in reducing the high-dimension metagenomic profile can discover novel and important structure of functional or taxonomic components which are not discovered

by previous linear methods such as PCA and NMF. For this purpose, we introduce Isomap - a manifold learning method to visualize nonlinear structure of human microbiome (Tenenbaum, de Silva et al. 2000; Balasubramanian and Schwartz 2002). We can visualize and explore these structures using only several components which are the intrinsic dimensions discovered by manifold learning. We also employ the t-distributed stochastic neighbor embedding (Jamieson, Giger et al. 2010) to see if we can discover nonlinear structure in human microbiome.

3. METHODS

3.1 Principle component analysis

PCA was invented in 1901 by Karl Pearson (Pearson, 1901). Now it is the mostly used tool in exploratory data analysis and dimension reduction. PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated principal components (Zhou, Chou et al. 2006). In PCA, the first principal component corresponding the largest eigenvalue of the covariance matrix can explain the largest possible variance in the dataset (that is, accounts for as much of the variability in the data as much as possible), and each succeeding component in turn can explain the highest variance possible under the constraint that it be orthogonal to preceding components.

3.2 Classic Multidimensional analysis

Multidimensional scaling (MDS) is a family of statistical methods often used in data visualization for exploring similarities or dissimilarities in data (Carroll and Arabie 1980). MDS algorithm starts with a similarity matrix of items (samples or species), then assigns a location to each item in a K-dimensional space, where K is specified by the investigator. For sufficiently small K, the resulting locations may be displayed in a graph or 3D visualization. The typical MDS procedure takes an input matrix giving dissimilarities between pairs of items and outputs a coordinate matrix whose configuration minimizes a loss function. Several loss functions can be used here, in classic MDS we used the loss function “strain” (Schmidt 1972).

3.3 Isometric Feature Mapping

We will also investigate the potential of a widely utilized global manifold learning approach for nonlinear dimensionality reduction: Isometric Feature Mapping (Isomap) (Tenenbaum, de Silva et al. 2000). Isomap is a global geometric framework for nonlinear dimensionality reduction which is built on classical Multidimensional Scaling (MDS). Although MDS is developed to preserve the Euclidean distance in a low dimensional space, Isomap seeks to preserve the intrinsic geometry of the data as captured in the geodesic distances in a manifold between all pairs of data points.

3.4 Nonnegative matrix factorization

Non-negative matrix factorization (NMF) provides an exciting alternative to traditional dimensional-reduction method (Lee and Seung 1999). In NMF, samples are represented by non-negative combinations of canonical components. The structure and patterns found by NMF is thus often very different from and more intuitive to interpret than that of more traditional eigenvector-based methods, such as PCA. By constructing samples as positive combinations, NMF also has the

potential to “disentangle” components which often overlap to create particular community samples. NMF approximates a data matrix X with non-negative entries as the product of two non-negative matrices W and H (Lee and Seung 1999). In a metagenomic example, X typically has p rows corresponding to metabolic pathways, and s columns corresponding to environmental samples; the entries would then represent the amount of evidence for a certain kind of pathway in a certain sample [often a number of DNA reads (Gianoulis, Raes et al. 2009)]. The matrix W is $p \times k$ whereas the matrix H is $k \times s$. Hence, each column of W has one entry for each of the p metabolic pathways; we can thus think of W as a collection of k “samples components”, where k is the “degree” of the factorization. In this interpretation the s columns of H give each of the s environmental samples as linear combinations of these canonical samples. In the dual interpretation, the k rows of H are “pathways components” and the p rows of W give the observed pathways as linear combinations of them.

3.5 t-SNE

We introduce a nonlinear dimensionally reduction technique to visualize metagenomic profiles that “embeds” high-dimensional metagenomic samples into low-dimensional space. This technique is a variation of the Stochastic Neighbor embedding (SNE) that was proposed by Hinton and Roweis in 2002, where conditional probability is adapted to describe item similarities in the high-dimensional Euclidean distances between data points. t-SNE, based on the same idea, is developed to be easier for optimizing the objective function and to solve the “crowding problem” (Maaten and Hinton, 2008). In addition, it is also showed that t-SNE has the potential to be applied to large data sets by adopting random walks on neighborhood graphs. The performance of t-SNE is demonstrated on a wide variety of data sets and compared with many other visualization techniques (Maaten and Hinton, 2008).

4. RESULTS

NMF and MDS can be considered as linear methods because they are based on Euclidean distance or operations. Isomap is based on data manifold and t-SNE considers the distance of data distributions (normal distribution or t-student distribution), thus they can be thought as nonlinear methods. Although early results from PCA indicate that there are three enterotyples, we have not found the three clusters in sample visualization from all methods.

MDS has the tendency to put healthy microbiomes together (Figure 1), although here is an overlapping between healthy and disease microbiomes. The results showed that NMF has clear linear results (Figure 2), it also seems that there is a tendency to discriminate healthy microbiome and disease microbiome. The sample size is not enough to support the hypothesis. If this is true, once we apply NMF on a large-size dataset, we may expect to find biomarkers indicating disease status of human gut microbiome.

For nonlinear methods, we found that Isomap has better visualization results than t-SNE (Figure 3 and 4). For t-SNE, although we found that heathy microbiomes are clustered together, but there are no clustering structure among them. Isomap cannot only identify robust enterotypes analogous to previous results, but also have clear clustering structure. Furthermore, most of the healthy microbiome are located at the upper-left corner of the visualization (blue points in Figure 3). From Isomap, it also seems that obese and elderly microbiomes are similar to each other, and there are not any difference between them from the results. Additionally there are three branch in the result of

Isomap and it seems that each branch is corresponding to a enterotype. Overall, we suggest that nonlinear manifold learning methods can discover the enterotype structure well.

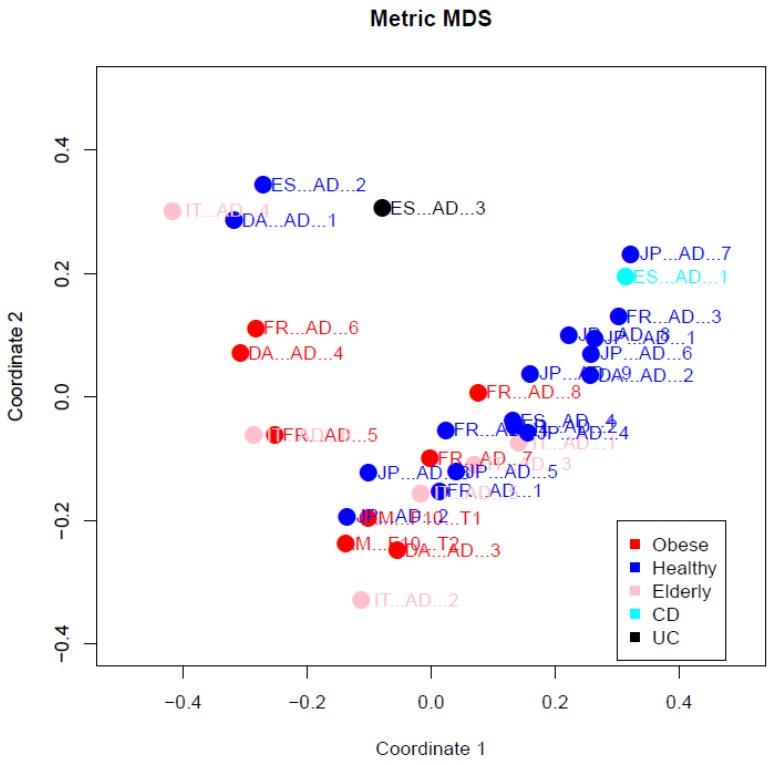

Figure 1. Visualization of human microbiome by MDS

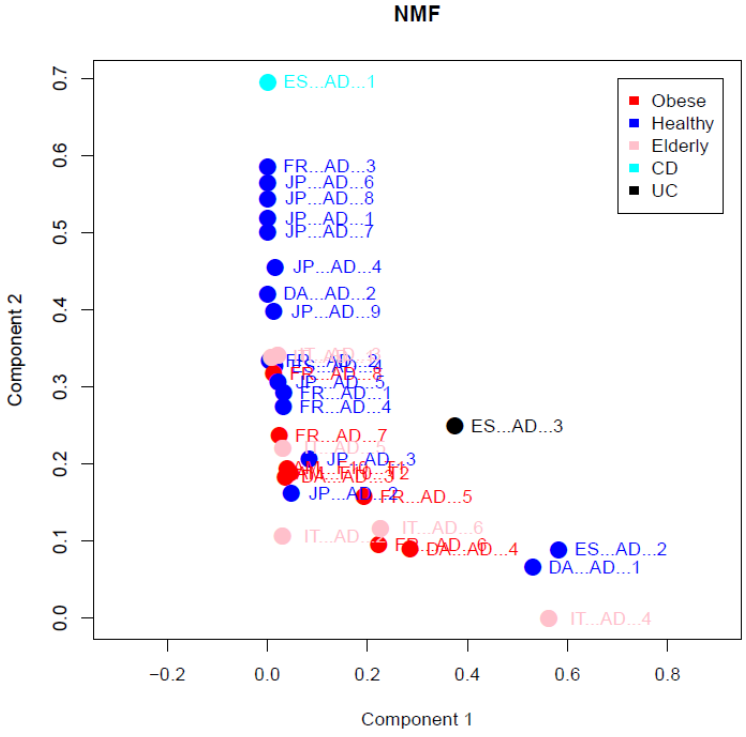

Figure 2. Visualization of human microbiome by NMF

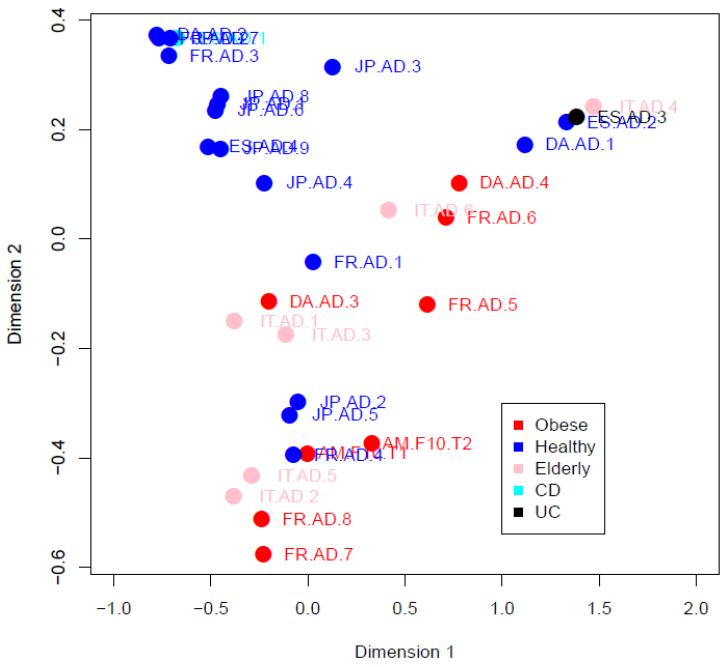

Figure 3. Visualization of human microbiome by Isomap

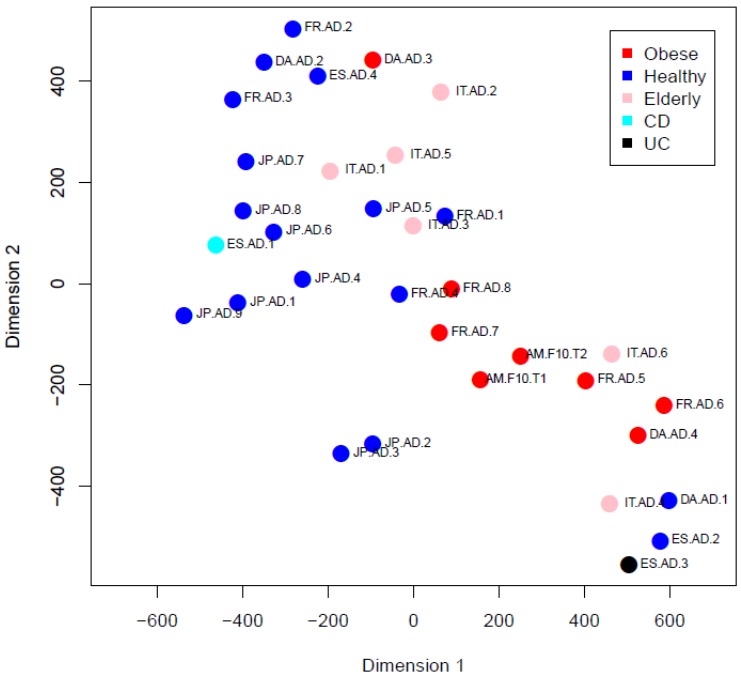

Figure 4. Visualization of human microbiome by t-SNE

Early results (Arumugam, Raes et al. 2011) indicate that ES-AD-2, ES-AD-3, IT-AD-1, DA-AD-1, FR-AD-6, IT-AD-4 formed an enterotype corresponding to a particular function. However, we found the last two microbiomes (FR-AD-6 and IT-AD-4) are not always close to the other four microbiomes in our results. For example, these four samples formed a clear cluster in the result of MDS, but be apart away from FR-AD-6 and IT-AD4. The two samples with CD sample are not

clustered to disease samples indicating that it is a stable structure.

5. CONCLUSION

Data visualization is an important step before any further statistical analysis. Although more and more studies in metagenomic are produced recently, there are less a systematically investigation of visualization methods for metagenomic samples. We compared several visualization methods, either linear or nonlinear, and found that the clustering patterns of human gut microbiomes from different methods are similar. We also found that healthy people may have different gut microbiome structure to disease gut especially in the results of Isomap. This indicates that nonlinear methods may be useful in discovering novel patterns in human microbiome. In future, more advanced nonlinear data representation methods may be adapted for exploring human gut microbiome.

6. ACKNOWLEDGEMENT

This work was supported in part by NSF CCF 0905291, NSF CCF 1049864, NSF IIP 1160960, NSFC 90920005, NSFC 61170189 and China National 12-5 plan 2012BAK24B01.

7. REFERENCES

[1] Arumugam, M., J. Raes, et al. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346): 174-180.

[2] Balasubramanian, M. and E. L. Schwartz (2002). The isomap algorithm and topological stability. *Science*, 295(5552): 7.

[3] Carroll, J. D. and P. Arabie (1980). Multidimensional scaling. *Annu Rev Psychol*, 31: 607-649.

[4] Gianoulis, T. A., J. Raes, et al. (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc Natl Acad Sci USA*, 106(5): 1374-1379.

[5] Jamieson, A. R., M. L. Giger, et al. (2010). Exploring nonlinear feature space dimension reduction and data representation in breast Cadx with Laplacian eigenmaps and t-SNE. *Med Phys*, 37(1): 339-351.

[6] Jiang, X., M. G. Langille, et al. (2012). Functional biogeography of ocean microbes revealed through non-negative matrix factorization. *PLoS One*, 7(9): e43866.

[7] Kolehmainen, E., A. Kuvaja-Mikkonen, et al. (1984). Comparison of different myelin isolation methods by use of nonequilibrium pH gradient gel electrophoresis. *Neurochem Res*, 9(9): 1253-1265.

[8] Lee, D. D. and H. S. Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788-791.

[9] van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9 (2008): 2579-2605

[10] Moeller, A. H., P. H. Degnan, et al. (2012). Chimpanzees and humans harbour compositionally similar gut enterotypes. *Nat Commun*, 3: 1179.

[11] Qin, J., R. Li, et al. (2010). A human gut microbial gene catalogue established by metagenomic

sequencing. *Nature*, 464(7285): 59-65.

[12] Schmidt, C. F. (1972). Multidimensional scaling analysis of the printed media's explanations of the riots of the summer of 1967. *J Pers Soc Psychol*, 24(1): 59-67.

[13] Siezen, R. J. and M. Kleerebezem (2011). The human gut microbiome: are we our enterotypes? *Microb Biotechnol*, 4(5): 550-553.

[14] Tenenbaum, J. B., V. de Silva, et al. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): 2319-2323.

[15] Wu, G. D., J. Chen, et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334(6052): 105-108.

[16] Zhou, X., J. Chou, et al. (2006). Protein structure similarity from Principle Component Correlation analysis. *BMC Bioinformatics*, 7: 40.

4. Reports

IT solutions for integrating clinical and molecular data to support biomedical research: from Biobanks to Knowledge Repositories

Riccardo Bellazzi[1], Arianna Dagliati[1], Flavio Licciulli[2], Paolo Romano[3], Luigi Varesio[4] and Mario Cannataro[5]

[1]Università degli Studi di Pavia, Pavia, Italy

[2]Istituto di Tecnologie Biomediche, CNR Bari, Italy

[3]IRCCS San Martino IST, Genova, Italy

[4]Istituto Giovanna Gaslini, Genova, Italy

[5]Università Magna Graecia di Catanzaro, Italy

RATIONALE

Biobanks, which are organized collections of biological samples and associated data, are increasingly recognized as a fundamental infrastructure for biomedical research. Biobanks allow to perform powerful molecular medicine studies and to share samples and data between research centers. During the last few years, the emergence of personalized medicine and the advent of new techniques of genome sequencing have made biobanks a crucial element to investigate existing hypothesis and to generate new ones.

However, the heterogeneity of technical specifications and the regulatory issues could represent a barrier for the effective collaboration between medical centres, and restrict the access to samples and data. An essential step to properly exploit the biobanks potentials is to develop valid methods to optimize the management of their data.

This paper summarizes the talks and discussions held during the first Workshop on "IT solutions for integrating clinical and molecular data to support biomedical research: from biobanks to knowledge repositories", held in Pavia on November 23rd, 2012. During the workshop, the experiences of Italian and European scientific centres of excellence has been presented and compared. The main goal of the workshop was to analyze the current opportunities for biobanks management methodologies' harmonization and sharing.

The main purpose of the symposium has been to find out and outline possible frameworks for exchanging data, procedures and knowledge as well as to reuse clinical data. The discussion focused on topics of clinical importance - such as high-throughput molecular diagnostics – but also on current regulatory issues- such as the legal aspects of biobanks accreditation.

WORKSHOP GOALS

The main goals of the workshop were:

[1] describe and compare experiences of national and international research centers;

[2] evaluate current opportunities to harmonize the technological solutions that deal with biobank data management;

[3] define strategies and potential projects to share data and information between Italian biobanks, taking into account regulatory and technological aspects.

The workshop comprised both invited talks and a panel discussion. The rest of the paper summarizes invited talks and reports results of the panel.

SPEAKER AND TOPICS

Riccardo Bellazzi, Università degli Studi di Pavia

Dr. Bellazzi described the essential steps needed to transform Biobanks databases into knowledge repositories, including in particular the need of reporting and maintaining the results of the samples analysis in a centralized database.

Jan Talmon - Maastricht University

Dr. Talmon described the PSI biobank infrastructure of the Academic Medical Centers in the Netherlands. In particular, the String of Pearls Initiative (in Dutch “Parelsnoer Initiatief” PSI) involves the eight academic medical centers in the Netherlands and is building a large collection of data and samples in relevant medical domains [1]. Dr. Talmon reported the main feature of the PSI model, as an infrastructure for multi-center prospective biobanks, able to support collection of data and materials. He also outlined the PSI frame of reference and the PSI Process.

Maria Grazia Daidone - Istituto Nazionale Tumori, Milano

Dr. Daidone discussed aspects related to the management, regulatory issues and access procedures of biobanks. She gave some main recommendations on the specimen collection and planning, on quality assurance and she provided information about policy issues. She focused on important aspects of biobank building and management such as "what to collect, "how to do it", "who can do it", "why can do it" [2]. Details have been reported on the ethical aspects and confidentiality and data protection. The criteria for the identification of a biobank have been also discussed.

Luigi Varesio - Istituto Gaslini, Genova

Dr. Varesio talked about the issues related to the integration in Biobank databases of Tissue&Omics results as an added value for clinic and research. He provided as case studies: BIT-NB Biobank in Europe, the Neuroblastoma Research Consortium (NRC) and the European Network for Cancer Research in Children and Adolescents (ENCCA). In particular he discussed the aspect of a centralized service to perform standardized analysis (mRNA profiling, miRNA profiling, CGH arrays, SNP arrays, methylation profiles), so that all resulting data are integrated in one on-line accessible database for all partners. The GRID was presented as a suitable infrastructure and a possible solution for the computational aspects related to data analysis of biobanks [3].

Daniele Segagni and Alberto Zambelli - Fondazione Salvatore Maugeri, Pavia

Dr. Zambelli and Ing. Segagni described the Onco-I2b2 and Cardio-i2b2 Projects at the Fondazione Salvatore Maugeri of Pavia. Both projects are based on the i2b2 (Informatics for Integrating Biology and the Bedside) software [4], an IT infrastructure that provides an open-source data warehouse enabling fast querying and retrieval of phenotype and molecular data (https://www.i2b2.org/). Onco-i2b2, in particular, provides a complete framework for integrating data coming from the hospital information system and the information collected by a biobank for tissue and blood storage.

Flavio Licciulli - Istituto di Tecnologie Biomediche CNR, Bari

Dr. Licciulli described the experience of Apulia's centers in integrating oncological Biobanks and biomedical databases, with particular reference to the Apulia Tumor Biobank (BioBOP) initiative. The technological aspects of Biobank and biomolecular data integration, as well as the solution called GaianDB [5], a lightweight Dynamic Distributed Federated Database (DDFD), have been discussed. GaianDB is based on Apache Derby, an open source relational database implemented in Java (http://db.apache.org/derby/).

Mario Cannataro - Università Magna Graecia, Catanzaro

Dr. Cannataro reported the experiences at University of Catanzaro on microarray data analysis for translational research and pharmacogenomics. A number of tools and solutions have been discussed, including micro-CS [6] for preprocessing and annotation of gene expression data, DMET-Analyzer for preprocessing and analysis of DMET (Drug Metabolism Enzymes and Transporters) SNP (Single Nucleotide Polymorphism) genotyping data [7]. Finally some applications in Medical Research have been presented.

PANEL DISCUSSION - Luciano Milanesi (Istituto di Tecnologie Biomediche CNR, Milano), Paolo Romano (San Martino IST, Genova) and Riccardo Bellazzi (Università degli Studi di Pavia)

The meeting was closed by a panel discussion on the "Technological infrastructures for research data sharing". The panelists talked about some interesting solutions for data and information sharing, comprising the open-source platform SHRINE (Shared Health Research Information Network), from Harvard Medical School, the use of Wiki systems for biobanks for easy reporting and annotation of samples analysis results, and the opportunities related to the BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) European project (http://www.bbmri.eu/).

REFERENCES

[1]. Jan L. Talmon, Maurits G. Ros, and Dink A. Legemate,) PSI: The Dutch Academic Infrastructure for shared biobanks for translational research, Summit on Translational Bioinformatics 2008; Volume 2008, pp. 110–114, Proceedings AMIA Joint Summits on Traslational Science.

[2]. Angelo Paradiso, Maria Grazia Daidone, Peter Riegman, Introduction to Cancer Biobanking: Why, Where, How? Biopreservation and Biobanking. June 2011, 9(2): 139-140. Special Issue on Interactive Biobanking in Cancer and Trends in the New Milleniums (Angelo Paradiso, Maria Grazia Daidone, and Peter Riegman Guest Editors).

[3]. Massimiliano Izzo, Andrea Schenone, Sara Barzaghi, Fabiola Blengio, Marco M Fato, Luigi Varesio, A Grid-enabled web platform for integrated digital biobanking in paediatrics, EMBnet Journal, Vol 18. 2012. Proceedings of the NETTAB 2012 Workshop on “Integrated Bio-Search” 14-16 November 2012, Como, Italy.

[4]. Murphy SN, Mendis ME, Berkowitz DA, Kohane I, Chueh H. Integration of clinical and genetic data in the i2b2 architecture. AMIA Annu Symp Proc. 2006:1040. PMID:17238659.

[5]. Graham Bent, Patrick Dantressangle, David Vyvyan, Abbe Mowshowitz and Valia Mitsou, A Dynamic Distributed Federated Database, Proceedings of the Annual Conference of ITA (International Technology Alliance), Imperial College, London, 2008.

[6]. Pietro Hiram Guzzi, Mario Cannataro. mu-CS: An extension of the TM4 platform to manage Affymetrix binary data. BMC Bioinformatics 11: 315 (2010).

[7]. Pietro Hiram Guzzi, Giuseppe Agapito, Maria Teresa Di Martino, Mariamena Arbitrio, Pierfrancesco Tassone, Pierosandro Tagliaferri, Mario Cannataro. DMET-Analyzer: automatic analysis of Affymetrix DMET Data. BMC Bioinformatics 13: 258 (2012).

The ongoing evolution of ICT for bioinformatics through twelve years of NETTAB workshops

Paolo Romano[1], Nicola Cannata[2]

[1] IRCCS AOU San Martino IST, Genoa, Italy

[2] School of Science and Technology, University of Camerino, Camerino, Italy

Introduction

Bioinformatics is a relatively young discipline. Its actual start could be identified in 1980, when the EMBL Nucleotide Sequence Data Library was first established. At that time, needs and resources, as well as technologies, were infinitely lower than today. Since then we have assisted to the creation of some essential algorithms and software that are now the basis for every molecular data analysis, to the explosion of data availability that has been generated by the genome projects and by the related advent of new high-throughput technologies, and to the incredible changes that emerged from the initial network leading to the current highly interoperable, semantic aware, social and collaboratively based Internet systems.

In the meantime, bioinformatics has been able to leverage from the best tools that Information and Communication Technologies (ICTs) have produced and has developed to a technology savvy status where databases adhere to most recent standards and include semantic metadata, software are distributed in the network and actively interoperate for the carrying out of the most frequent and required tasks, automated data analysis procedures can be carried out on the network, and bioinformaticians all around the world collaborate on-line for the development of new information resources, knowledge bases, software tools and platforms.

This has been possible by a continuous monitoring of technological developments in ICT research domains, by a careful analysis of the possible impact of new technologies and by a swift adoption of most promising tools and platforms. A contribution in this activity was also given by the series of workshops NETTAB that started in 2001 and since then have introduced some of the most interesting ICT technologies to a large community of biologists and bioinformaticians.

NETTAB Workshops

NETTAB Workshops are a series of International meetings on “Network Tools and Applications in Biology” held annually in Italy (1). They are aimed at introducing participants to the most promising among those innovative Information and Communication Technologies (ICTs) that are being applied to the biomedical domain. They provide a unique forum for bringing together physicians, biologists, bioinformaticians and computer scientists with researchers expert in novel ICT approaches able to address emerging application challenges in the life sciences.

Workshops include many focused sessions which are devoted to involved technologies, tools, systems and platforms, and early applications. Keynote lectures introduce the sessions' topics, and are followed by presentations selected from among the submitted contributions by peer review. Discussion is a key factor, both within sessions and in a Panel Discussion which is usually aimed at introducing and discussing both challenges that could be faced by the adoption of the new

technology and perspectives in its adoption. Tutorials and poster sessions complete the agenda of the NETTAB workshops.

Because of the continuous technological evolution, the workshop is focused each year on a different technology or domain. Since 2001, many different focus themes were discussed. These included, e.g., Standardization for data integration (Genoa, 2001), Multi agent systems (Bologna, 2002), Scientific workflows (Naples, 2005), Grid and Web Services (Santa Margherita di Pula, 2006), The Semantic Web (Pisa, 2007), Collaborative research and development (Catania, 2009), Biological wikis (Naples, 2010), and Integrated Bio-Search (Como, 2012).

Many well known keynote speakers have presented their research activity at NETTAB workshops, including: Carole Goble, Thure Etzold, Rolf Apweiler, Martin Bishop, Mike Taussig, Susanna Assunta Sansone Francis Ouellette, Martin Senger, Gary Bader, Olivier Bodenreider, Robert Stevens Andrey Rzhetsky, Alexander Kel, Alex Bateman, Tim Clark, Yves Lussier, Eric Neumann, Barend Mons, and Erik Bongcam-Rudloff.

Various Supplements or Special Issues were published by peer-review journals, including BMC Bioinformatics and Briefings in Bioinformatics, with best papers presented at NETTAB workshops (2-7).

During last years, NETTAB workshops have received support by some European Bioinformatics Societies, including the Bioinformatics Italian Society (BITS), the Swiss Institute of Bioinformatics (SIB-ISB), the Netherlands Bioinformatics Center (NBIC), and the EMBnet organization.

NETTAB 2013: Semantic, Social and Mobile applications for Bioinformatics

The NETTAB 2013 workshop will be held in Venice Lido on next October 16-18, 2013. The focus theme will be "Semantic, Social and Mobile applications for Bioinformatics and Biomedical laboratories" and it may be able to attract quite a big number of researchers interested in these innovative research and application domains.

Human beings are social animals and ICT have recently permeated human society of newer forms and ways of participation in social activities. In the Internet, the hype has shifted from Web2.0 to Social Media; in science too, this development is evident. Beside facilitating communication and making easier the sharing of information, social platforms and technologies are enhancing learning, problem solving and crowdsourcing. In biology, and especially in the “-omic” disciplines, we already rely on a wide diffusion of social tools and applications, e.g for distributed annotations, Wiki knowledge bases, documentation and productivity.

On the other hand, access to the Internet is nowadays increasingly happening through mobile devices. Mobile Internet access is expected to soon overtake access from standard PCs and workstations. Moreover, mobile phones are expected to become the main personal computing devices. Smartphones and tablets represent the most practical computing devices in biomedical laboratories and they actually are the ideal companions for always on the move scientists. While we can observe a widespread diffusion of health and lifestyle mobile applications and a rapid adoption of mobile solutions in medicine and healthcare we cannot say the same for life sciences and bioinformatics.

Semantic methodologies and technologies are instead well established in “-omic” projects. It can even be proudly observed that the bioinformatics community was an early adopter of Semantic Web technologies.

In the NETTAB 2013 workshop, mobile and/or social and/or semantic solutions for bioinformatics and laboratory informatics problems will be explored. It is our opinion that a savvy combination of these three technologies could greatly enhance the research outcome of life scientists and markedly simplify the workflows in biomedical laboratories.

References

1. NETTAB workshops' web site: http://www.nettab.org/

2. Priami C, Merelli E, Gonzalez PP, Omicini A (Eds.) Transactions on Computational Systems Biology III. Lecture Notes in Computer Science, Vol. 3737, Springer-Verlag Berlin Heidelberg (2005) VII, 169 p. ISBN 978-3-540-30883-6

3. Milanesi L, Armano G, Breton V, Romano P (Eds). Special Section on Grid, Web Services, Software Agents, and Ontology Applications for Life Sciences. IEEE Transactions on NanoBioscience (2007) 6 (2):101-167.

4. Romano P, Schroeder M, Cannata N, Marangoni R (Eds). A Semantic Web for Bioinformatics: Goals, Tools, Systems, Applications. BMC Bioinformatics 2008, 9(Suppl 4):S1-S13

5. Milanesi L, Romano P (Eds). Bioinformatics Methods for Biomedical Complex System Applications. BMC Bioinformatics 2009, 10(Suppl 12):I1, S1-S17

6. Romano P, Giugno R, Pulvirenti A (Eds). Special Issue: Collaborative Bioinformatics and RNA Analysis. Briefings in Bioinformatics (2011) 12(6): 547-625

7. Bellazzi R, Romano P, Masseroli M, Murphy S, Shabo A (Eds). Research from the Eleventh International Workshop on Network Tools and Applications in Biology (NETTAB 2011). BMC Bioinformatics 2012, 13(Suppl 14):S1-S14

5. Profile

We are excited to introduce a new feature to our newsletter, a profile of successful computer scientists that have made exciting new discoveries or significant contributions to bioinformatics and computational biology. The purpose of the profile is to provide insight into the professional and personal life of these computer scientists and allow them to share interesting lessons or personal experiences during their academic journey with colleagues and students. It is our hope that you will enjoy this feature, and that these profiles of your colleagues or mentors will encourage you in your own academic journeys. We debut with a profile of Mona.

Amarda Shehu

Department of Computer Science

George Mason University,

Computer Scientist in Profile: **Mona Singh**

Contributor: Amarda Shehu[1,2] [amarda@gmu.edu]
Dept. of [1]Computer Science, [2]Bioengineering
George Mason University, Fairfax, VA 22030
Tel: 703-993-4135 Fax: 703-993-1710

Unassuming and soft spoken, you would have a hard time believing that the woman shaking your hand is Mona Singh, a Professor of Computer Science at Princeton University, recipient of the Presidential Early Career Award for Scientists and Engineer in 2001, prolific researcher in bioinformatics, Chairperson of the NIH MABS study section, organizer of various top bioinformatics conferences, such as ISMB and ACM BCB, and esteemed and respected by colleagues across different disciplines intersecting with bioinformatics and computational biology.

Mona's academic journey did not initially follow a clear path though her interest in science and math began early. She was first exposed to biology when working in an immunology lab in high school during the summers. However, Mona took her first computer science class as a senior in high school, when her 10th grade geometry teacher, Ms. Watkins, convinced her that she would really enjoy it. This led Mona to pursue a computer science degree at Harvard University, with a summer research internship at AT&T with David Johnson on graph algorithms. It was this experience that convinced her to pursue graduate studies.

As a graduate student at MIT in the Laboratory of Computer Science, Mona initially studied machine learning in the theory group. Professor Bonnie Berger, who eventually became her PhD co-advisor at MIT, encouraged her to enter computational biology and introduced her to many interesting machine learning and algorithmic problems arising in molecular biology.

A postdoctoral fellowship in Professor Peter Kim's structural biology group at MIT and the Whitehead Institute was an important step in her transition from core computer science to computational biology. Peter Kim convinced Mona to join his group. "Joining Peter's group was one of the best decisions I have made professionally, and I learned an incredible amount there," she says. "While I think of myself as a core computer scientist, I think having this dual training has been very important for my science and my career. Further, I have been fortunate that both Bonnie and Peter were strong advocates for me professionally, and without their faith in me, I am not sure I would have had the confidence to pursue an academic career."

Peter said, "Mona, you're completely comfortable in the computer science world, but to really make a contribution in computational biology, you will have to become comfortable in the biology world, too."

Mona joined the Computer Science Department at Princeton in 1999. She recalls, "Computational biology was still a relatively new field at the time and very few computational biologists were being hired in Computer Science Departments." Luckily, Princeton was excited to promote interdisciplinary research. "When I started, I was largely developing data-driven methods for predicting aspects of protein structure and interactions. My work in this area has continued, and we now have significant efforts in developing algorithms for analyzing large-scale protein interaction networks."

These days, you will find Mona charting new territories in analyzing cellular networks. She is particularly interested in developing methods for predicting specificity in protein interactions and for uncovering how molecular interactions vary based on context, whether that is within a single organism, across organisms, or across individuals, particularly in the context of human disease. Mona is passionate not only about science, but about the role women can play in science. She has been

active in efforts to encourage women to pursue careers in science, recruiting woman PhD students and advising advised numerous undergraduate and graduate female students.

The intense academic career is interspersed with family time with her husband and their two sons (ages 6 and 9). The family goes on regular trips to Hawaii, the birth place of her husband, and to Birmingham, Alabama, where Mona spent her formative years. The family often enjoys nature walks, bike rides, and visiting with friends.

Family on election day, 2012.

Looking ahead, Mona sees the field continuing to expand as high-throughput experimental technologies, including sequencing (in many applications), mass spectrometry, and various types of imaging, will generate increasingly large and complex biological data sets that will drive novel computational paradigms.

6. Announcements

ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB) 2013

You are invited to attend ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB). ACM BCB is the main flagship conference of the ACM SIG Bioinformatics. ACM BCB 2013 is in its fourth year, building upon the success of ACM BCB 2010 in Niagara Falls, ACM BCB 2011 in Chicago and ACM BCB 2012 in Orlando. Each of the conferences had about 200 attendees.

The conference will provide a premier forum for interdisciplinary and multidisciplinary research encompassing disciplines of computer science, mathematics, statistics, biology, bioinformatics, and biomedicine.

ACM BCB 2013, Call for Papers

September 22-25, 2013
Washington D.C., USA

The ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB) is the flagship conference of the ACM SIGBCB. This is the conference's fourth year, building upon the success of the first three meetings in Niagara Falls, Chicago, and Orlando.

ACM BCB 2013 will be held in Washington D.C. from September 22-25, 2013. The conference offers a forum for premier interdisciplinary research linking computer science, mathematics, statistics, biology, bioinformatics, and biomedical informatics. The past two decades have led to a tremendous growth in the size and dimensionality of biological and biomedical data. This conference serves to showcase leading-edge research in processing, modeling and analyzing these datasets for a variety of applications.

ACM BCB 2013 welcomes original submissions that have not been published and that are not under review by another conference or journal. Examples of relevant topics include but are not limited to:

- Genomics and Evolution
- Protein and RNA Structure, Protein Function, and Proteomics
- Computational Systems Biology
- Next Generation Sequencing Data
- Medical Informatics and Translational Bioinformatics
- Cross-Cutting Computational Methods
- Bioinformatics Infrastructure
- Immunoinformatics and Computational Immunology
- Computational Epidemiology

- Biomedical Image Analysis
- Knowledge Representation and Inference
- Integration of Biomedical Data
- Databases, Knowledgebases & Ontologies
- Text Mining and Natural Language Processing

Submitted manuscripts should not exceed 10 pages in ACM template on 8.5 x 11 inch paper (see ACM templates). All submissions will be evaluated on their originality, technical soundness, significance, presentation, and interest to the conference attendees. Submission implies the willingness of at least one of the authors to register and present the work associated with the paper submitted. All submitted papers will be reviewed by ACM-BCB's technical program committee. All accepted papers of registered authors will be included in the proceedings published by ACM digital libraries. Selected papers will be invited to adapt their papers for journal publication. All accepted papers will be required to submit an online ACM Copyright Form. Details for electronic submission will be found on the conference web portal at http://www.cse.buffalo.edu/ACM-BCB2013/.

Important Dates:

Paper submissions due: April 22, 2013
Notifications sent to authors: June 21, 2013
Camera-ready papers due: July 15, 2013

Contributed papers will be assigned to one of the following tracks for review: Systems Biology; Sequencing and Sequence Analysis Methods; Functional Genomics; Gene Regulation and Transcription; Protein and RNA Structure; Biomedical Informatics; Data and Knowledge Bases, Text Mining, and Ontologies; Bioimage Analysis; Evolution, Population, and Comparative Genomics. However, submissions in any area related to the main conference topics are welcomed even if they do not fit cleanly into one of these tracks.

Organizing Committee

General Chairs
Sridhar Hannenhalli, University of Maryland
Cathy Wu, University of Delaware and Georgetown University

Program Co-chairs
Srinivas Aluru, Iowa State University
Donna Slonim, Tufts University

Workshop Chair
Umit Catalyurek, Ohio State University

Publicity Chair
Jianlin Cheng, University of Missouri - Columbia

Local Arrangement Chairs
Amarda Shehu, George Mason Unviersity
Liliana Floria, Johns Hopkins University

Tutorial Chair

Clare Bates Congdon, University of Southern Maine
Vasant Honavar, Iowa State University

Steering Committee

Aidong Zhang, State University of New York at Buffalo, Chair

Registration Chair

Preetam Ghosh, Virginia Commonwealth University

Proceedings Chair

Jing Gao, State University of New York at Buffalo

ACM BCB 2013, Call for Workshop Proposals

The ACM BCB 2013 (http://www.cse.buffalo.edu/ACM-BCB2013) provides a premier forum for interdisciplinary and multidisciplinary research encompassing disciplines of computer science, mathematics, statistics, biology, bioinformatics, and biomedicine. The ACM BCB 2013 organizing committee invites proposals for workshops to be held in conjunction with the conference. The purpose of a workshop is to provide participants with the opportunity to present and discuss novel research ideas on active and emerging topics in bioinformatics, computational biology and biomedical informatics.

Workshops that focus on a challenge problem, e.g., flexible docking, computational drug design, or emerging topics e.g., computational immunology and vaccinology, analysis of protein-RNA interactions, multi-scale modeling, are especially welcome. Please note that the conference does not have funding for invitation of workshop speakers. All workshop attendees are expected to register for the conference. Student participants in the workshops can apply for a limited number of student travel fellowships from the conference on a competitive basis.

The workshops organizers must be prepared to create a workshop web page to be linked to the main conference web page, announce the workshop and call for papers, gather submissions, conduct the review process and decide the final workshop program. The workshop proceedings will be part of ACM BCB 2013 proceedings. The logistics of the workshops will be coordinated with help from the ACM BCB-2013 organizers.

Workshop Proposals should be no more than 3 pages in length and must include the following:

- Description of the workshop abstract, objectives, goals, relevance and expected outcome
- Description of workshop topic and the associated research issues
- Description of the target audience
- One or more potential invited speakers
- Past workshops and other related recent workshops
- Preferred duration of the workshop (full day or half day)
- Contact information (address, email, and phone) for all organizers
- A designated contact person

The organizers may include 2 additional pages to provide the following additional information:

- A list of potential authors
- A list of potential attendees

Proposals should be submitted by email (with subject "ACM-BCB 2013 Workshop Proposal") to Umit V.Catalyurek (catalyurek.1@osu.edu).

Key dates for the ACM-BCB Workshops:

- Submission deadline: March 1, 2013
- Notification of Acceptance: March 15, 2013

ACM BCB 2013, Call for Tutorial Proposals

ACM BCB 2013 invites tutorials that address the interests of its varied audience of individuals interested in Bioinformatics, Computational Biology and Biomedical Informatics (BCB) including graduate students, researchers and educators from academia, and researchers and practitioners from industry and government.

We especially welcome proposals for tutorials that:

- Introduce a specific BCB topic, designed to make the topic (and the conference) more accessible to participants who are new to that topic.
- Provide a hands-on introduction to one or more databases, software tools, or other resources of broad interest to the conference participants.
- Provide a comprehensive review of the current state of the art in a specific BCB topic aimed at researchers and practitioners who are knowledgeable, but not necessarily experts in the topic.
- Present techniques from research fields e.g., machine learning,statistics, parallel computing, that are relevant to BCB research.
- Introduce new research problems, new application areas, or new or emerging technologies of relevance to BCB.

The tutorials will be held on September 22, 2013. We envision most tutorials to be 2 hours or 4 hours long although longer durations (in multiples of 2 hour slots) may be considered. Ideally, each tutorial must have more than one presenter and no more than three, preferably from different institutions, bringing different perspectives.

Tutorial proposals should not exceed 5 pages, using an 11 point font for the text, and should include:

- Tutorial title
- Names and affiliations of presenters
- Tutorial abstract (200 words maximum, suitable for inclusion on the conference website)
- Tutorial description, including the objectives of the tutorial, its relevance to ACM-BCB 2013, description of the intended audience and background assumed of the audience, sufficient detail regarding the scope of material to be covered and the depth at which it will be covered.
- Desired tutorial length (if there is flexibility regarding length, please specify the topics to be included for each length).
- Information about other venues in which tutorials on the same topic have been presented or planned to be presented, along with pointers to the relevant slides or other tutorial materials and a brief explanation of how the proposed tutorial differs from the other offerings.
- Brief professional biographies of presenters including their scientific and professional qualifications and experience (relevant research, teaching or tutorial presentation) and contact information.

Each proposal will be reviewed by the members of the ACM-BCB tutorial program committee, and ranked based on the significance of the proposed tutorial topic, overall quality of the proposal, the qualifications and experience of the presenters, and the tutorial's fit to the conference, and the number of tutorial slots and the space available.

Please email tutorial proposals to Tutorial Chairs:

Vasant Honavar, v.honavar@gmail.com

Clare Bates Congdon, congdon@gmail.com

To ensure full consideration, tutorial proposals must be received no later than March 15, 2013.

ACM SIGBioinformatics Record

Call for Contribution

Dear SIGBioinformatics members,

The aim of SIGBioinformatics Record is to report on the activities conducted by the community, and to publish early notes on current progress in bioinformatics and biomedical informatics areas. In addition, special articles will be published highlighting important developments in the relevant fields. An important place in our newsletter will be devoted to collect description of different research domains of the members. As Bioinformatics and Computational Biology have become a broad research area, SIGBioinformatics Record could be a mechanism to publicize problem domains. Nevertheless SIGBioinformatics Record could stimulate discussions and bridge different areas. Thus, we kindly invite you to outline your research area (description should not exceed two pages, following the attached template – latex template could be found on the Guidelines on the ACM Website). Moreover, you can send us a brief description of "student-activities", i.e., ongoing student projects and educational activities carried out.

Important Dates

- Publication dates: 30th January, 30th May, 30th September
- Due dates: 1st January, 1st May, 1st September.