

Meeting report

Open Access

Proceedings of the Seventh Annual UT-ORNL-KBRIN Bioinformatics Summit 2008

Eric C Rouchka*¹, Julia Krushkal² and Daniel Goldowitz³

Address: ¹Department of Computer Engineering and Computer Science, University of Louisville, 123 JB Speed Building, Louisville, KY 40292, USA, ²Department of Preventive Medicine and Center of Genomics and Bioinformatics, University of Tennessee Health Science Center, 66 N. Pauline Street, Suite 633, Memphis, TN 38163, USA and ³Child and Family Research Institute, Center for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Room 2026, 850 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada

Email: Eric C Rouchka* - eric.rouchka@louisville.edu; Julia Krushkal - jkrushka@utmem.edu; Daniel Goldowitz - dang@cmmt.ubc.ca

* Corresponding author

from UT-ORNL-KBRIN Bioinformatics Summit 2008
Cadiz, KY, USA. 28–30 March 2008

Published: 8 July 2008

BMC Bioinformatics 2008, 9(Suppl 7):11 doi:10.1186/1471-2105-9-S7-11

This article is available from: <http://www.biomedcentral.com/1471-2105/9/S7/11>

© 2008 Rouchka et al; licensee BioMed Central Ltd.

The University of Tennessee (UT), Oak Ridge National Laboratory (ORNL), and Kentucky Bioinformatics Network (KBRIN) have extensive research and educational ties in bioinformatics and have a long standing tradition of holding annual joint bioinformatics regional summits that bring together researchers, educators, and students interested in bioinformatics. These summits provide unique opportunities for enhancing collaborative links and integration of multidisciplinary research efforts and have resulted in numerous new collaborative projects in bioinformatics research and education. The Seventh Annual UT-ORNL-KBRIN Bioinformatics Summit was held at Lake Barkley State Park in Cadiz, Kentucky on March 28–30, 2008. A total of 174 participants registered for the summit, with 81 from various Tennessee institutions, and 78 from Kentucky institutions. Eighty-two registrants were faculty, while 45 were students, 33 staff, and 7 post-docs with the remainder as invited speakers.

The conference program included three days of presentations, the first devoted to the National Center for Biotechnology Information (NCBI) short courses, a discussion of the revolutionary changes in the concept of a gene in light of recent data generated by the ENCODE project and other advances in molecular biology and genome analysis [1], and a presentation on educational opportunities at UT-Knoxville and ORNL. The last day and a half was dedicated to scientific presentations divided into three ple-

nary sessions: Pathways to Prediction; Biomedical Informatics; and Regulatory Analysis as well as a keynote speech given by Jeremy Smith of ORNL and a bioinformatics education workshop led by A. Malcolm Campbell of Davidson College.

NCBI short courses

Wayne Matten (NCBI) presented two of the NCBI mini-course modules, "Entrez Gene Quick Start" and "Correlating Disease Genes and Phenotypes." Both of these mini-courses were 2.5 hours in length with a one and a half hour lecture overview, followed by a one-hour interactive problem-based session. The first module concentrated on using NCBI's resources for searching for gene-based information based on a RefSeq record, including sequence information such as reference gene, isoforms, single nucleotide polymorphisms and associated phenotypes, homologs, and protein structure. The second mini-course focused on associating a diseased gene with a corresponding change in phenotype using a number of NCBI resources [2].

Pathways to prediction

Nitin Baliga (Institute for Systems Biology) started the first plenary session with an exciting presentation titled "A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell." What followed was a discussion of systems biology approaches incorporating transcrip-

tion, translation, and interactions to derive dynamic temporal relationships that allow for organisms to thrive in extreme conditions. The main focus of this session was the molecular mechanisms of response of Archea to extreme temperature, radiation, pH levels, and high salinity and the wealth of systems biology methods and approaches to investigate this response [3-6].

Biomedical informatics

Dan Masys (Vanderbilt University) led the second plenary session on Biomedical Informatics. Dr Masys, along with Paul Harris of Vanderbilt, presented an overview of the Clinical and Translational Science Awards (CTSA) program at NIH. Vanderbilt's CTSA program, the Vanderbilt Institute for Clinical and Translational Research (VICTR), was discussed. In order to efficiently process the large influx of patient information and tissue samples into the Vanderbilt DNA Databank, VICTR has implemented a research portal, StarBRITE. StarBRITE offers clinical researchers a number of tools for conducting patient-based research, including regulatory support, patient recruitment support, data management, and clinical systems integration for electronic medical record (EMR) support. The VICTR Informatics Core is involved in the consortium developing the REDCap (Research Electronic Data Capture) <http://www.iwg-online.org/projects/redcap/index.php> data collection project as a portion of the EMR functionality of StarBRITE.

Mikael Benson (Göteborg University, Sweden) was an additional plenary speaker for this session. He discussed detection of markers for personalized medicine specifically targeted for allergies, including locating epigenetic markers using hay fever as a disease model. Potential markers could emerge from systems biology analysis of decomposed transcriptional networks using DNA microarrays [7-9]. Genes identified in these networks are further analyzed for polymorphisms, which may eventually be used as diagnostic markers.

Regulatory analysis

The final plenary session, Regulatory Analysis, was led by Ziv Bar-Joseph (Carnegie Mellon University) with his platform presentation "Reconstructing Dynamic Regulatory Networks in Multiple Species." Dr. Bar-Joseph discussed current limitations of common approaches to analysis of time series gene expression data, including dealing with populations of cells instead of single cells and the use of static data for the study of dynamic time course responses. In order to better address these limitations and to incorporate temporal features of such data, he presented an approach known as STEM [10] that aligns time course expression data to pre-determined sets of expression profiles, allowing for comparisons of significant profiles found across time course experiments. To better group

together expression patterns of genes, a hidden Markov model approach for bifurcating profiles of dynamic regulatory events (DREM) [11] was presented, with evidence of experimental studies from yeast [12] and *E. coli* [13].

David Galas, (Institute for Systems Biology and Battelle Memorial), was the second speaker for the Regulatory Analysis session providing an intriguing closing talk in his presentation "Grappling with Complexity: Key Problems in Computational Biology." Dr. Galas discussed the current and future issues in computational biology, particularly in relationship to the massive explosion of available genome information and technical breakthroughs such as NextGen sequencing [14]. The importance of these issues is dictated by the fact that available biological data are growing exponentially rather than linearly, thus outpacing the rate at which the data can be handled by the growth of processor speed as dictated by Moore's Law [15]. He discussed how the Institute for Systems Biology is approaching this issue by looking at a thorough understanding of subsets of data in order to aid in P4 (Predictive, Personalized, Preventative, and Participatory) Medicine, particularly through their innovative Personalized Genome and Blood Organ Fingerprint projects.

Keynote speaker

Jeremy Smith, director of the ORNL Center for Molecular Physics, was the keynote speaker for the summit, giving a talk titled "Computer Simulation of Molecular Machines." Dr. Smith discussed numerous examples of how physical principles drive various processes at a molecular level, including protein folding [16] and muscular contraction. He also discussed the computational resources available at ORNL for bio-computing, including the 119-teraflop Cray XT3/XT4 Jaguar supercomputer, which currently ranks as the second most powerful supercomputer in the world and will provide a significant boost to computing power of Tennessee bioinformaticians and biology researchers.

Bioinformatics education presentation and workshop

Dr. Cynthia Petersen, the director of UT/ORNL Graduate School of Genome Science and Technology, provided an exciting brief overview of new educational initiatives in bioinformatics aiming to raise computationally-enabled bioscientists. She discussed the new SCALE-IT (Scalable computing and leading-edge innovative technologies) program, the availability of TeraGrid portals for biology, and educational opportunities for undergraduate and graduate students.

Malcolm Campbell (Davidson College) presented a hands-on guide to teaching bioinformatics education at an undergraduate level through his workshop "Functional

Genomics Cafeteria Style." His workshop was divided into Part I: Introducing Genomics and Bioinformatics Early; Part II: Meshing Mathematics with Biology; and Part III: A Fun Way to Integrate Biology, Math, CS, and Engineering. In Part I, he introduced the notion of teaching microarrays at the undergraduate level, through wet lab simulations [17] and creation of synthetic microarrays with known expression levels [18]. Part II focused on the use of MAGIC Tool for statistical analysis of microarrays [19,20]. Part III focused on facilitating interdisciplinary research at the undergraduate level by breaking down departmental boundaries and encouraging creativity through the International Genetically Engineered Machine Competition (iGEM).

Additional discussions of the role of bioinformaticians in student education and K-12 educational outreach were held throughout the summit.

Poster session

A total of 44 posters were presented during a 3-hour poster session on Saturday afternoon. These poster abstracts, a number of which are published in this supplement, were grouped in the topics of Bioimaging, Bioinformatics Infrastructure, Bioinformatics of Health and Disease, Comparative Genomics, Databases, Functional Genomics, Gene Regulation, Genomics, Machine Learning and Algorithms, Microarrays, Ontologies and Text Mining, Proteomics, Structure and Function Prediction, and Systems Biology.

Seven of the poster abstracts were selected for inclusion in the Summit program as short 15-minute platform presentations. Those posters presented as platform presentations were "Impact of sequence variants on the genetic analysis of expression" (Daniel Ciobanu); "Evaluation of pooled allelotyping versus individual genotyping for genome-wide association analysis of complex disease" (Siddharth Pratap); "Using spacings to infer regions of loss-of-heterozygosity from paired genotype array data" (Stanley Pounds); "Integrated bioinformatics platform for differential proteomics" (Xiang Zhang); "Towards ultimate quantification – recent advances in statistical analysis of real-time PCR data with linear models" (Joshua Yuan); "Geometric databases of protein structures" (Di Wu); and "Extracting putative gene networks from microarray data using graph theoretical algorithms" (Sudhir Naswa).

Future plans

The 2009 Bioinformatics Summit will return to Fall Creek Falls State Park, Tennessee in the Spring of 2009. Potential focus areas for next year may include translational informatics, epigenetics, and current technological advances for generating biological data.

Acknowledgements

We would like to thank the additional Conference Program Committee members Nigel Cooper (University of Louisville), Jack Dongarra (University of Tennessee), Ramin Homayouni (University of Memphis), Mike Langston (University of Tennessee), Terry Mark-Major (University of Tennessee-Memphis), Dan Masys (Vanderbilt University), Terry Moore (University of Tennessee), Jay Snoddy (Vanderbilt University), Chuck Staben (University of Kentucky), Arnold Stromberg (University of Kentucky), Bruce Whitehead (University of Tennessee Space Institute) and Rob Williams (University of Tennessee-Memphis) for putting together a well received scientific program. In addition, we wish to thank Bethany Coates, Stephanie Dearing, Terry Mark-Major, and Michelle Padgett for all of their efforts in putting together all of the details that allowed for the meeting to proceed. We would also like to thank Kerry Allen and the delightful staff of Lake Barkley State Park, who helped to make our stay as pleasant and smooth as possible. Funding for the UT-ORNL-KBRIN Summit is provided in part by The Kentucky Biomedical Research Infrastructure Network (KBRIN), The University of Tennessee Center for Information Technology Research, the University of Tennessee Molecular Resource Center, The UT-ORNL Science Alliance, The University of Tennessee Health Science Campus Research Center of Genomics and Bioinformatics, and NIH grants P20RR16481 and R13LM009315.

References

1. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al.: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799-816.
2. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvermin V, Church DM, Dicuccio M, Edgar R, Federhen S, et al.: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2008, **36**:D13-D21.
3. Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, Kottemann MC, Hood L, Diruggiero J: **Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1.** *Genome Res* 2004, **14**:1025-1035.
4. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsen V, Shannon P, Johnson MH, Bare JC, et al.: **A predictive model for transcriptional control of physiology in a free living cell.** *Cell* 2007, **131**:1354-1365.
5. Kaur A, Pan M, Meislin M, Facciotti MT, El-Gewely R, Baliga NS: **A systems view of haloarchaeal strategies to withstand stress from transition metals.** *Genome Res* 2006, **16**:841-854.
6. Whitehead K, Kish A, Pan M, Kaur A, Reiss DJ, King N, Hohmann L, Diruggiero J, Baliga NS: **An integrated systems approach for understanding cellular responses to gamma radiation.** *Mol Syst Biol* 2006, **2**:47.
7. Benson M, Carlsson L, Guillot G, Jernas M, Langston MA, Rudemo M, Andersson B: **A network-based analysis of allergen-challenged CD4+ T cells from patients with allergic rhinitis.** *Genes Immun* 2006, **7**:514-521.
8. Benson M, Langston MA, Adner M, Andersson B, Torinsson-Naluai A, Cardell LO: **A network-based analysis of the late-phase reaction of the skin.** *J Allergy Clin Immunol* 2006, **118**:220-225.
9. Benson M, Cardell LO, Hohmann S, Jirstrand M, Langston M, Mobini R, Nerman O: **[Systems biology can radically change health care. Basis for individualized prediction, prevention and treatment].** *Lakartidningen* 2007, **104**:3037-3041.
10. Ernst J, Bar-Joseph Z: **STEM: a tool for the analysis of short time series gene expression data.** *BMC Bioinformatics* 2006, **7**:191.
11. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z: **Reconstructing dynamic regulatory maps.** *Mol Syst Biol* 2007, **3**:74.
12. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al.: **Transcriptional regulatory networks in *Saccharomyces cerevisiae*.** *Science* 2002, **298**:799-804.
13. Ernst J, Beg QK, Kay KA, Balazsi G, Oltvai ZN, Bar-Joseph Z: **A Semi-Supervised Method for Predicting Transcription Factor-Gene Interactions in *Escherichia coli*.** *PLoS Comput Biol* 2008, **4**:e1000044.

14. Mardis ER: **The impact of next-generation sequencing technology on genetics.** *Trends Genet* 2008, **24**:133-141.
15. Moore GE: **Cramming more components onto integrated circuits.** *Electronics* 1965, **38**.
16. Daidone I, Ulmschneider MB, Di NA, Amadei A, Smith JC: **Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding.** *Proc Natl Acad Sci USA* 2007, **104**:15230-15235.
17. Campbell AM, Zanta CA, Heyer LJ, Kittinger B, Gabric KM, Adler L, Schulz B: **DNA microarray wet lab simulation brings genomics into the high school curriculum.** *CBE Life Sci Educ* 2006, **5**:332-339.
18. Campbell AM, Hatfield WT, Heyer LJ: **Make microarray data with known ratios.** *CBE Life Sci Educ* 2007, **6**:196-197.
19. Campbell AM, Ledbetter ML, Hoopes LL, Eckdahl TT, Heyer LJ, Rosenwald A, Fowlks E, Tonidandel S, Bucholtz B, Gottfried G: **Genome Consortium for Active Teaching: meeting the goals of BIO2010.** *CBE Life Sci Educ* 2007, **6**:109-118.
20. Heyer LJ, Moskowitz DZ, Abele JA, Karnik P, Choi D, Campbell AM, Oldham EE, Akin BK: **MAGIC Tool: integrated microarray data analysis.** *Bioinformatics* 2005, **21**:2114-2115.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

