Highlights from the 5(th) International Society for
Computational Biology Student Council Symposium at the
17(th) Annual International Conference on Intelligent
Introduction

**Highlights from the 5th International Society for Computational Biology Student Council Symposium at the 17th Annual International Conference on Intelligent Systems for Molecular Biology and the 8th European Conference on Computational Biology**

Thomas Abeel*1,2, Jeroen de Ridder3,4 and Lucia Peixoto5

Address: 1Department of Plant Systems Biology, VIB, Ghent University, Gent, Belgium, 2Broad Institute of MIT and Harvard, Cambridge, MA, USA, 3Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands, 4Bioinformatics and Statistics, Dept. of Molecular Biology, Netherlands Cancer Institute, Amsterdam, The Netherlands and 5Department of Biology and PENN Genome Frontiers Institute, University of Pennsylvania, Philadelphia, PA, USA

Email: Thomas Abeel* - thomas@abeel.be

* Corresponding author

**Abstract**

This meeting report gives an overview of the keynote lectures and a selection of the student presentations at the 5th International Society for Computational Biology Student Council Symposium at the 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and the 8th European Conference on Computational Biology (ECCB). The symposium was held in Stockholm, Sweden from June 27 to July 2, 2009. We also report on other Student Council events that were organized at ISMB/ECCB 2009.

**Introduction**

The Student Council (SC) of the International Society for Computational Biology (ISCB) is a world-spanning organization for students in the broad field of computational biology and bioinformatics.

The major aims of the Student Council are to organize events and facilitate networking opportunities for students. The main contribution of the SC is to nurture soft skills, such as working in a team, organizational and networking skills, to complement the normal academic program. Since its inception, the Student Council has organized an annual student symposium for the benefit of the student community. This year the fifth Student Council Symposium was held in conjunction with the joint ISMB and ECCB conferences on June 27th. Over 120 delegates took part in this first anniversary edition of the Student Council Symposium. The symposium featured three keynote lectures, a research partner session with two speakers, a tutorial, a panel discussion, nine student presentations and a poster session.

This year’s keynotes were delivered by three very well-established and acclaimed scientists. The symposium started with a keynote lecture by Peer Bork (EMBL Heidelberg, Germany) on the topic of “Integration of heterogeneous data in Biology: from chemicals to ecosystems”. The afternoon session was initiated by Michal Linial (The Hebrew University of Jerusalem, Jerusalem, Israel) with a presentation titled “In search of overlooked functions: hidden connections among proteins, toxins and viruses”. The day was concluded by Overton Prize winner Trey
The fourteen abstracts can be grouped in four very broad categories: (i) functional genomics and gene expression, (ii) protein structure and function, (iii) evolution and population biology and (iv) next-generation sequencing analysis and applications.

1. Functional genomics and gene expression
Genome-wide sequencing has undoubtedly revolutionized the way we do biology. Nowadays, thousands of complete genomes are available, and many more will be available within the next decade. The challenges of genome assembly and annotation have largely been conquered. However, the functional output of those genomes is not yet completely understood, and is currently the focus of research. One of those functional outputs is gene expression, which is the first step towards the generation of a phenotype. During the symposium several studies have taken advantage of the capabilities of genome-wide expression technologies and the availability of such data that has been accumulated in the past decade to answer relevant biological questions. For example, Szczepińska et al. [1] use gene expression data to explore relations between gene expression and genomic context. They identify and functionally annotate distant genomic clusters within co-expression clusters. The number of annotated clusters was significantly higher than random. Koeva et al. [2] combine shared differential expression with homology to define "core stemness mechanisms" in mouse stem cell types. Co-expression analysis has become an essential part of the data exploration phase when dealing with genome-wide expression data. Although several algorithms are available for this purpose, new and more efficient ones are constantly being developed. One new algorithm, designed by Schulteiss et al. [3], uses a newly developed SVN-kernel to identify regulatory modules. Exploring functional aspects of the genome does not need to involve generating new data, since a vast amount of data is publicly available. This allows us to search for sets of related experiments with biological relevance. Caldas et al. [4] describe an algorithm to query the data based on the actual measurements instead of the textual annotations. Such a huge amount of data requires specialized visualization to be interpretable. Standard heat maps and profile plots cannot handle hundreds of thousands of samples. This is where Space Maps come in, a visualization technique developed by Gehlenborg et al. [5]. Expressed genes are by no means the only functional output of genomes. The role of non-coding RNAs in the regulation of cell processes cannot be ignored. Moreover, their function can actually be used to uncover the functions of genes, by using high-throughput siRNA screens. These screens can identify genes that underlie similar phenotypes by the clustering of their screen profiles. Samusik et al. [6] describe how genes that show the same phenotypes are more likely to interact with each other and they developed a method of cross-validation of siRNA screening data using protein-protein interactions.

2. Protein structure and function
Proteins are the building blocks of life. Ultimately, the expression of genes leads to the expression of proteins. However, computationally speaking, proteins are much more complex than DNA, since their function is extremely dependent on its tridimensional structure, which varies from protein to protein.
Predicting how a protein will fold from its linear sequence of amino acids is in itself a complex problem and it is even more challenging for transmembrane proteins. Tran et al. [7] implemented a new algorithm that can successfully deal with super-secondary structures in beta-barrel transmembrane proteins. Once folds and structures have been determined, structural alignment allows extending them to closely related homologs. Although several methods exist today to achieve such a task, they are mostly heuristic. Wohlers et al. [8] developed PAUL, a non-heuristic approach that can outperform existing methods. There are several factors that influence variability in structural alignments. Pirovano et al. [9] determined that certain types of folds (helices and coils) are usually observed in inconsistent alignment regions. Since this observation is at odds with currently used alignment strategies, more care should be taken in the development of new algorithms. Protein structures, more than the sequences themselves, capture the function of a protein. Hence, they are better suited to identify functional homologs. Combining structure searches with advanced HMMs is a much more reliable way to search for conserved functions than simple sequence similarity searches. Petrossian et al. [10] uses this strategy to predict novel methyltransferases that are also shown to bind the appropriate substrate.

3. Evolution and population biology
If proteins determine phenotype, evolution determines which of those phenotypes are likely to be seen by natural selection. Selection itself depends on diversity. Functional divergence or selection for new function is the hallmark of successful adaptation. Hence, the identification of proteins under functional divergence is of broad interest. Williams et al. [11] present a fast new method for detecting these changes on the whole-genome level across a complex phylogenetic tree. They also apply the method successfully to the evolution of pathogenicity in divergent bacterial lineages. The diversity that selection operates on can also be the source of disease susceptibility among individuals. Genome-wide association studies aim to find the association between genetic markers, usually SNPs, and disease. Lee et al. [12] address the challenge of selecting representative SNPs for supporting a disease-gene association. They demonstrate superior performance when using a Pareto optimality based approach.

4. Next-generation sequencing analysis and applications
Next-generation sequencing technology is enabling massive production of high-quality data with a variety of applications. It will undoubtedly revolutionize the limits of what is possible and affordable in experimental biology. Applications of next-generation sequencing include de novo sequencing, re-sequencing and sequencing of RNAs for applications formerly limited to microarrays. Many platforms (Illumina Genome Analyzer, Applied Biosystems SOLiD, Helicos HeliScope) are currently able to produce "ultra-short" paired reads of lengths starting at 25 nt. It is still an open question whether genome re-sequencing is feasible with ultra-short paired reads. Chikhi et al. [13] show that re-sequencing requires significantly (48.3%) shorter paired reads to produce results comparable to unpaired reads. High-throughput sequencing technologies open exciting new approaches to transcriptome profiling (RNA-Seq). Bohnert et al. [14] developed a new technique to accurately infer the underlying transcript abundances based on linear programming, which is also a powerful tool to reveal and quantify novel (alternative) transcripts.

Conclusion
In total we received 101 abstract submissions. Each abstract was rigorously reviewed by at least 3 reviewers. From the 101 abstracts, we could only select 8 for oral presentation, due to time limitations on the program. One abstract was rejected and the authors of the remaining 92 abstracts were invited to present their poster, which 64 of them accepted. The presented research covers a broad range of topics including microarrays, proteomics, genetics and genomics. Fourteen outstanding abstracts have been selected to compile this supplement.

Thanks to our generous sponsors, we were able to hand out nine travel fellowships to help cover expenses for students to attend ISMB/ECCB and the Student Council Symposium. Two of the travel fellowships were given to excellent students from developing nations to give them the opportunity to present their work and to meet potential collaborators.

A new initiative introduced this year was the Student Council Career Central (SCCC) in form of a lounge in the exhibit area and seminars. This initiative is an endeavor by the Student Council to address the needs of recruiters and job seekers in the field of computational biology. SCCC provides a platform for research institutes and companies to attract some of the best talents in computational biology from the Student Council community and attendees at ISCB conferences all over the world. During the ISMB/ECCB 2009 conference SCCC offered one-on-one resume critique sessions sponsored by FASEB with Dr. Clifford S. Mintz, who helped students to improve their CVs and resumes. Seminars about various career options available at the European Bioinformatics Institute and non-academic career paths were also offered as a part of SCCC this year.

The future
The 6th Student Council Symposium will be held in Boston, MA, USA in conjunction with ISMB 2010 in July 2010. Further information about the Student Council and
the events we organize is available on our website: http://www.iscbsc.org.

Competing Interests
The authors declare that they have no competing interests.

Authors contribution
TA, JdR and LP drafted the manuscript.

Acknowledgements
We would like to thank ISCB Executive Officer BJ Morrison McKay, ISCB Administrative Support Suzi Smith and ISMB conference organizer Steven Lead for logistic support and invaluable advice in organizing the symposium.

We are indebted to all speakers for their time and effort. In particular we should mention Peer Bork, Michal Linial and Trey Ideker, our keynote speakers. We would also like to thank the ISCB board of directors for their enthusiastic support and invaluable advice in organizing the symposium.

We would like to thank ISCB Executive Officer BJ Morrison McKay, ISCB Administrative Support Suzi Smith and ISMB conference organizer Steven Lead for logistic support and invaluable advice in organizing the symposium.

We are grateful to Oxford University Press for sponsoring the best poster and best presentation awards.

This article has been published as part of BMC Bioinformatics Volume 10 Supplement 14, 2009: Highlights from the Fifth International Society for Computational Biology (ISCB) Student Council Symposium. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/10?issue=S13.

References