

INTRODUCTION

Open Access

# Summary of talks and papers at ISCB-Asia/SCCG 2012

Konstantin Tretyakov<sup>1†</sup>, Tatyana Goldberg<sup>2†</sup>, Victor X Jin<sup>3</sup>, Paul Horton<sup>4\*</sup>

From ISCB-Asia 2012

Shenzhen, China. 17-19 December 2012

## Abstract

The second ISCB-Asia conference of the International Society for Computational Biology took place December 17-19, 2012, in Shenzhen, China. The conference was co-hosted by BGI as the first Shenzhen Conference on Computational Genomics (SCCG).

45 talks were presented at ISCB-Asia/SCCG 2012. The topics covered included software tools, reproducible computing, next-generation sequencing data analysis, transcription and mRNA regulation, protein structure and function, cancer genomics and personalized medicine. Nine of the proceedings track talks are included as full papers in this supplement.

In this report we first give a short overview of the conference by listing some statistics and visualizing the talk abstracts as word clouds. Then we group the talks by topic and briefly summarize each one, providing references to related publications whenever possible. Finally, we close with a few comments on the success of this conference.

## Introduction

Following the success of the first ISCB-Asia, held jointly with APBioNET as InCoB/ISCB-Asia 2011 [1,2], ISCB-Asia/SCCG 2012 took place on December 17-19, 2012, in Shenzhen, China. This year BGI co-hosted ISCB-Asia as the first Shenzhen Conference on Computational Genomics (SCCG). ISCB-Asia/SCCG 2012 was immediately followed by the Asian Young Researchers Conference on Computational and Omics Biology (AYRCOB), also cohosted by BGI.

More than 146 people from more than 18 countries attended ISCB-Asia/SCCG 2012. The 45 conference talks included: 9 proceedings talks (selected from 26 submissions), 6 keynotes, 7 highlights, 3 technology track talks, 2 program chair-invited talks, and 4 special sessions (*Cancer genome informatics*, *Workflows and the cloud for reproducible bioinformatics*, *Computational statistics for modern biology*, *BGI special session*).

The talks were given by researchers from 16 countries, representing many of the leading centers of bioinformatics research worldwide, and the selection of topics was, in the opinion of the authors of this report, quite representative of modern-day trends in computational biology (Figures 1, 2).

This report briefly summarizes each talk given at the conference, grouped into six broad subject areas, ranging from data processing to statistics in modern biology.

## Data processing

As our community is struggling with the continuous deluge of “Big Data” [3], introduction of efficient tools and infrastructure for handling what are now petabyte-sized file collections, designing computational workflows and maintaining reproducible results becomes key for the future success of computational biology. Consequently, two sessions, organized by Scott Edmunds of *GigaScience*, were devoted to the topic of cloud-based tools, workflows and reproducible computing.

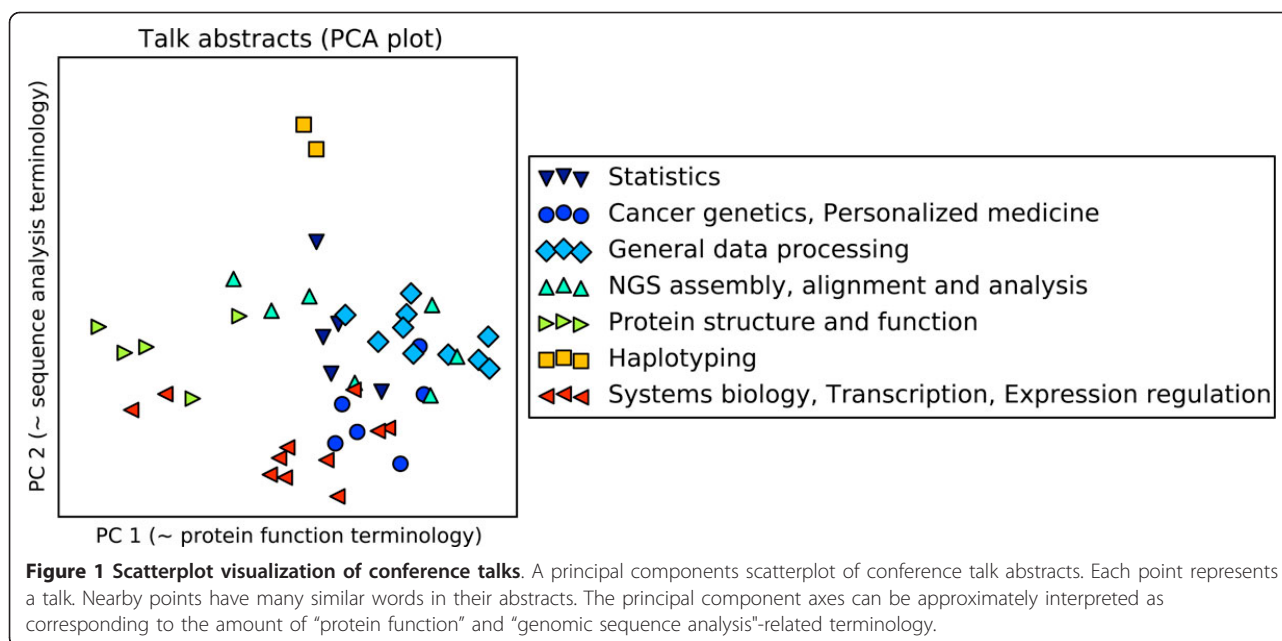
One focus area of ISCB-Asia/SCCG 2012 was workflow-management systems, with Galaxy [4-6] being one

\* Correspondence: horton-p@aist.go.jp

† Contributed equally

<sup>4</sup>Computational Biology Research Center, AIST, 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan

Full list of author information is available at the end of the article



of the key platforms. *Genomic Data Submission and Analytical Platform (GDSAP)* [7] was presented by Tin-Lap Lee (Chinese University Hong Kong), as a CBIIT-led effort to provide a unified, Galaxy-based online toolkit for biomedical scientists. IRRI-Galaxy is a similar effort from the International Rice Research Institute (talk by Ramil Mauleon). Mohamed Abouelhoda (Nile and Cairo University) presented Tavaxy [8], another cloud-based system, which focuses on letting users combine and run workflows designed in both Galaxy and Taverna [9]. Finally, two commercial cloud-based data analysis systems were presented at the conference: *ClusterTech Life-science Analysis Suite (CLASS)* (tech talk by Ping Chung Ng) and BGI's *EasyGenomics* [10] (talk by Xu Xing).

While workflow systems allow researchers to efficiently design data analysis pipelines, they are not always successful at ensuring long-term reproducibility. More often than not, running the same workflow a year later would not yield the same results. This issue is addressed by the *Wf4Ever* project [11] presented by Marco Roos (Leiden University Medical Center), in which the idea of a *workflow* is generalized to the concept of a *research object*.

Reproducibility, interoperability and automation of workflows can be facilitated by the use of universally recognized metadata formats. The *ISA* metadata framework [12] aims to provide a set of such formats, standards and tools (talk by Eamonn Maguire, University of Oxford).

The issue of preserving privacy in the analysis of omics data is a growing concern [13]. In a technical talk, Kana Shimizu (AIST) described a clever application of additive homomorphic encryption which allows a

database of chemical compounds to be interrogated for the presence or absence of records similar to a query compound - without revealing the query itself.

Finally, as detailed in article S8 of this supplement [14], Konstantin Tretyakov (University of Tartu) presented a new file fingerprinting method enabling fast synchronization of large biological data repositories within the cloud and between data centers.

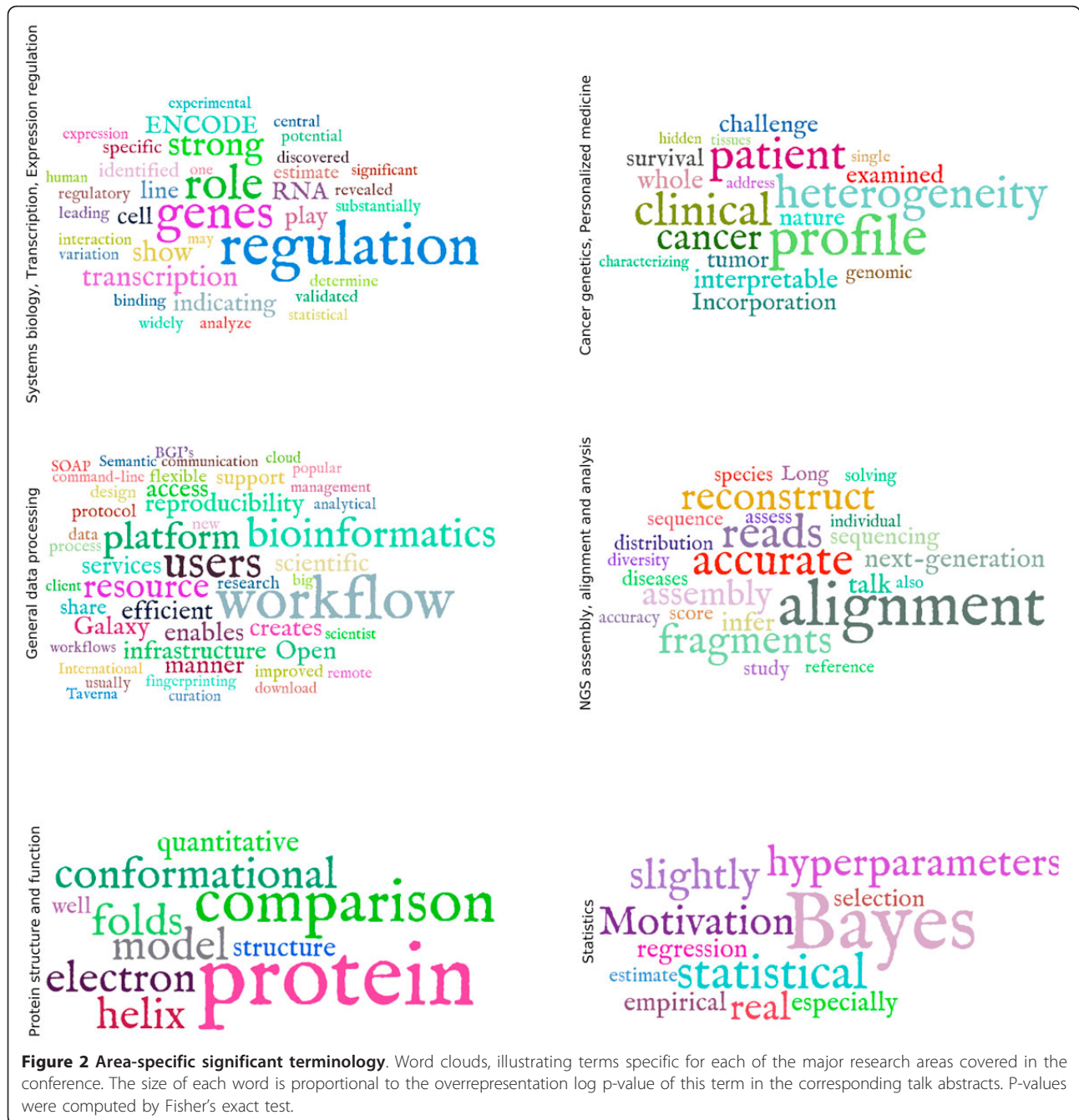
### Sequence and NGS data analysis

Genomic sequence analysis techniques such as sequence alignment have long been considered a mature field and one of the cornerstones of computational biology. However, the rapid development of next-generation sequencing (NGS) technologies in the last decade continues to raise new unexpected challenges.

In meta-genomics, the genomes of all the organisms found in an environmental sample (e.g. soil) are pooled and sequenced together. The resulting genomic diversity severely complicates the task of NGS read assembly and introduces the related problem of *binning* reads by their source, where source is one or more closely related species. Francis Chin (The University of Hong Kong) presented Meta-IDBA [15] as a tool for solving both problems.

*De novo* assembly of RNA-seq (transcriptome) data also requires specialized algorithms and tools. Dongxiao Zhu (Wayne State University) presented SPATA [16], an RNA-seq assembly method based on a divide-and-conquer strategy.

Genome-wide methylation is often measured by applying NGS to bisulfite-converted DNA [17], in which unmethylated cytosines are converted to thymines. This



**Figure 2 Area-specific significant terminology.** Word clouds, illustrating terms specific for each of the major research areas covered in the conference. The size of each word is proportional to the overrepresentation log p-value of this term in the corresponding talk abstracts. P-values were computed by Fisher's exact test.

conversion allows the position of methylated cytosines to be inferred by traditional sequence alignment procedures as described in a highlights talk by Martin Frith [18] (AIST).

An interesting method for performing phylogenetic analysis on NGS data without the need to assemble reads was presented in a highlights talk by Urmila Kulkarni-Kale [19] of the University of Pune.

Articles S6 and S7 in this supplement apply machine learning techniques to the problem of automatic detection

and classification of non-coding RNAs. Kun Sun (The Chinese University of Hong Kong) presented iSeeRNA [20], a novel tool applying Support Vector Machine (SVM) classifiers to detect long intergenic non-coding RNAs (lincRNAs) from transcriptome sequence data. Meanwhile, Mark Menor (University of Hawaii) presented a state of the art multiclass classifier McRUM [21], evaluated on a dataset of small non-coding RNAs (small ncRNAs).

Another central topic of NGS analysis is fast and accurate inference of genotypes. Articles S2 and S5 presented

novel algorithms for haplotype reconstruction. Fei Deng (University of Hong Kong) presented a dynamic programming algorithm and heuristic speed-up for the minimum error correction formulation of this problem [22]. In a more probabilistic approach, Hirotsuka Matsumoto (University of Tokyo) presented a mixture model based reconstruction method (MixSIH) and a new quality metric (MC) [23]. Meanwhile, in the special session on cloud computing, Junwen Wang (EasyGenomics, BGI) presented FaSD [24], a cloud based solution for fast and accurate detection of SNPs (Single Nucleotide Polymorphisms) from sequence data.

Finally, Ge Nong (Sun Yat-Sen University) gave an invited talk on efficient (linear-time) construction algorithms for suffix arrays [25], a data structure underlying many NGS applications.

### **Systems biology, transcription and expression regulation**

Elucidation of the mechanisms involved in gene expression, mRNA regulation and protein-protein interactions is among the central goals of contemporary biological research. Indeed, five out of the six conference keynote talks were largely related to this area of research. This field has recently gained a significant boost thanks to the completion of the ENCODE project [26], whose press releases attracted significant media attention – in particular the claim that more than 80% of the human genome “works as a kind of control panel packed with genetic dials” [27], when previous conventional wisdom was that more than 95% of human DNA might be nonfunctional “junk”.

The true significance of the often quoted 80% figure has, however, been widely debated [28]. According to the keynote speaker Philip Green (University of Washington), “junk DNA hasn’t gone anywhere and is here to stay”. In one of the most memorable moments of ISCB-Asia/SCCG 2012, Prof. Green eloquently challenged transcription (*i.e.* interaction with RNA-polymerase) as sufficient evidence of function, pointing out that although 100% of the genome interacts with DNA-polymerase (replication) we do not conclude that 100% of the genome is functional.

Nevertheless, no one doubts that transcription regulation is a vital and complex process, whose details remain unclear. A keynote by Piero Carninci (RIKEN) described studies using DeepCAGE [29], which enables highly sensitive detection of transcription start sites via NGS sequencing of 5' capped mRNAs. In a recent study, Carninci's group has used DeepCAGE to discover specific patterns of expression of retrotransposon elements in different cell compartments.

Methylation is one of the key mechanisms of mammalian gene regulation. In a keynote by Takashi Ito (University of Tokyo), a novel whole-genome bisulfite sequencing method was presented, which, in contrast to previous

technologies, can be applied to minuscule quantities of material (only a few thousand cells) without the need for global PCR amplification.

Two talks focused on transcription factors and their binding. A keynote by Zhiping Weng (University of Massachusetts Medical School) introduced Factorbook [30], a *de novo* motif discovery analysis performed on ENCODE data; and Stephen Kwok-Wing Tsui (The Chinese University of Hong Kong) presented a method of discovering Protein-DNA binding sites using association rule mining [31].

miRNA was another topic addressed by multiple talks. As described in article S3 of this supplement [32], Toyofumi Fujiwara (INTEC) introduced a novel method for predicting miRNA targets, based on the hypothesis that promoters of miRNAs and their targets tend to share (predicted) cis-elements [32]. In a highlights talk, Michal Linial (The Hebrew University of Jerusalem) discussed a novel algorithm, miRror2.0, which enables the discovery of combinatorial regulation of transcription by several miRNAs [33].

Several talks introduced advances in the analysis of gene expression data. As detailed in article S9 [34], Koki Tsuyuzaki (Tokyo University of Science) proposed a novel information criterion based technique for detecting differentially expressed genes. Luonan Chen (Shanghai Institute for Biological Sciences) discussed the analysis of expression of the genes in the quiescence signaling pathways of *S. cerevisiae*. Finally, a keynote by Eric Xing (Carnegie Mellon University) presented TREEGL [35], an approach based on  $\ell_1$ -penalized linear regression, capable of reconstructing the evolution of a gene network from relatively scarce gene expression data.

We close this section by mentioning the highlights talk given by Mark Ragan (The University of Queensland), who reported an analysis of human transcriptomic data to elucidate the effect that alternative splicing has on the protein-protein interaction network, via inclusion or omission of certain interaction domains [36].

### **Protein structure and function**

The structure of a protein provides crucial insights into its functional role in a cell. Although the first protein structures were determined more than half a century ago, the structures of many proteins remained unsolved and the mechanism of protein folding is not yet fully understood.

In his keynote address, Gunnar von Heijne (Stockholm University) described quantitative analyses of the energetics and kinetics of membrane protein assembly *in vivo* that lead to a better understanding of the mechanisms of transmembrane protein assembly, and improvement of topology and structure-prediction methods. Many of these prediction methods sometimes confuse helical regions in

membrane proteins with signal peptides in classically secreted proteins. In a highlights talk, Henrik Nielsen (Technical University of Denmark) described SignalP 4.0 [37], a new version of that popular software for predicting signal peptides. In addition to having a new architecture, SignalP now explicitly discriminates between signal peptides and transmembrane regions.

Superficially dissimilar protein sequences can adapt similar 3D structures and biological functions. Thus protein structure comparison methods are indispensable. Article S1 by Xiuzhen Huang et al. (Arkansas State University) describes ePC, an accurate and fast algorithm that is able to compare whole structures as well as specific substructures [38]. As detailed in article S4 [39], Prasad Gajula (Indian Agricultural Statistics Research Institute) presented a molecular dynamics simulation of the protein Vinculin and showed that the simulation is highly consistent with local mobility as determined experimentally by electron paramagnetic resonance (EPR) spectroscopy.

Two novel methods dealing with multimeric protein complexes were introduced by Daisuke Kihara (Purdue University): Multi-LZerD [40] for modeling the structure of protein complexes and EMLZ-erD [41] for fitting them into electron microscopy maps.

### **Cancer genomics and personalized medicine**

The special session on cancer research organized by Chen-Hsiang Yeang (Academia Sinica) focused on techniques for the integrative analysis of heterogeneous cancer data. Yinyin Yuan (Institute of Cancer Research) presented a quantitative image-based approach [42] that, in combination with molecular assays, is able to uncover new knowledge about breast tumor biology and predict patient survival. Robert Beckman (Daiichi Sankyo) introduced a computational model of targeted cancer therapy incorporating genetic evolutionary dynamics and single-cell heterogeneity [43]. He reported that in a large virtual clinical trial of cancer patients the model may lead to improved outcomes compared with the current personalized medicine approach. Shihua Zhang (Chinese Academy of Sciences) developed a method for the systematic analysis of multi-dimensional genomics data [44]. Using DNA methylation, gene expression and microRNA expression data of ovarian cancer samples, he showed that the method is able to uncover biologically relevant patterns of complex gene regulatory systems. Biaoyang Lin (Zhejiang-California International Nanosystems Institute) presented MAPS [45], a massive parallel sequencing method, to conduct a global survey of Hepatitis B virus (HBV) in HBV-related hepatocellular carcinomas (HCCs). The outcome of the survey contributed significantly to the understanding of the mechanism of HCC development.

In a technical track presentation, Frank Schacherer (BIOBASE GmbH) described Genome Trax and the Human Genome Mutation Database (HGMD) [46], two human-curated annotation sources for identifying functionally relevant variants in the human genome and understanding their effects in a medical context. HERRA and HaploShare [47] are two other sources for linking genetic variants with underlying disease phenotypes; presented by Wanling Yang (The University of Hong Kong), who also stressed the importance of making disease associations from exome or whole genome sequencing data easily interpretable for clinicians.

A tool for non-invasive prenatal diagnosis, FetalQuant [48], was presented in a highlights talk by Hao Sun (The Chinese University of Hong Kong). This tool estimates fractional fetal DNA concentration directly from massively parallel sequencing on DNA in maternal plasma, eliminating the need for prior genotype information.

### **Statistics in modern biology**

Recent advances in DNA chip technology and the discovery of thousands of SNPs in genome-sequencing projects motivated genomic selection using high-density markers. However, the increasing number of available biomarkers presents both computational and statistical challenges. The special session on Statistics in Modern Biology organized by Dabao Zhang (Purdue University) addressed both issues.

Vitara Pungpapong (Chulalongkorn University) proposed a fast and accurate algorithm for biomarker selection, implementing an empirical Bayes method for variable selection in regression models [49]. POCRE, the penalized orthogonal-components regression method [50], is another predictor for genomic selection, presented by Min Zhang (Purdue University). It outperforms its competitor BayesB [51] in both time and performance at estimating breeding values. Yuanhui Xiao (Georgia State University) presented analysis of retinal pigment epithelium flatmount images.

### **Conclusions**

We subjectively conclude that ISCB-Asia/SCCG 2012 was without doubt a successful event. At least three continents (Asia, Europe & North America) and all age groups were well represented, but there was naturally a strong contingent of Asians and young people, which we believe bodes well for the future of computational biology in Asia. Indeed, The Asian Young Researchers Conference on Computational and Omics Biology (AYRCOB), also co-hosted by BGI, immediately followed ISCB-Asia/SCCG 2012 in Shenzhen and effectively served as an extension of ISCB-Asia/SCCG 2012 for student participants. Finally we note that the 23rd Annual Genome Informatics Workshop held in Tainan,

Taiwan during the week previous to ISCB-Asia/SCCG 2012 was also highly successful, enjoying a record number of paper submissions.

Among other memorable aspects of ISCB-Asia/SCCG 2012 were the social events organized after the long days of presentations and poster sessions. The quiet get-togethers in the nearby Chinese pubs and the organized excursion to BGI, followed by a dinner, all were great opportunities for the speakers and other participants of the meeting to come together in an informal environment and participate in relaxed and open discussions. The rather compact size of the conference added an exceptional degree of friendliness. Established researchers seemed to be much more accessible to graduate students than it is usually the case in larger conferences.

#### Authors' contributions

KT, TG contributed equally to writing the manuscript. KT prepared the figures. PH edited an initial draft of the manuscript. All authors read and approved the final manuscript.

#### Competing interests

KT is a coauthor of article S8 of this supplement and mentioned in this report. KT and TG gave talks at the conference. VJ was program committee chair. PH was conference chair.

#### Acknowledgements

We would like to acknowledge BGI for co-hosting the conference and the conference sponsors BIOBASE, CLUSTERTECH and AIST for financial support. Huinan Hao and the rest of the BGI staff for excellent ground support during the conference. The Kingkey Palace hotel staff for their cooperation and delicious lunches. The ISCB management: in particular B.J. Morrison for her time and care, Janet Kelso for her encouragement, and Burkhard Rost, Rheinhard Schneider and Michal Linial, who traveled great distance to support the conference. We thank all members of the conference advisory committee, steering committee, and program committee and the anonymous volunteers who served to review proceedings and highlights track submissions. We thank the keynote speakers, who took time out of their busy schedules to give six excellent talks. Last but not least, we thank Catherine Wells and the BMC editors for their help in publishing this supplement.

#### Declarations

The publication costs for this article were funded by general research funds of the National Institute of Advanced Industrial Science and Technology (AIST), Japan.

This article will be published as the introduction to *BMC Genomics* Volume 14 Supplement 2, 2013: Selected articles from ISCB-Asia 2012. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/14/S2>.

#### Author details

<sup>1</sup>Institute of Computer Science, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia. <sup>2</sup>Faculty of Informatics, Department for Bioinformatics and Computational Biology, Technical University Munich, Boltzmannstrasse 3, Garching 85748, Germany. <sup>3</sup>Department of Biomedical Informatics, The Ohio State University, 460 W 12th Ave., 212 BRT, Columbus, OH 43210, USA. <sup>4</sup>Computational Biology Research Center, AIST, 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan.

Published: 29 April 2013

#### References

- Schönbach C, Tan TW, Kelso J, Rost B, Nathan S, Ranganathan S: **InCoB celebrates its tenth anniversary as first joint conference with ISCB-Asia.** *BMC Genomics* 2011, **12**(Suppl 3):S1.

- Ranganathan S, Schönbach C, Kelso J, Rost B, Nathan S, Tan TW: **Towards big data science in the decade ahead from ten years of InCoB and the 1st ISCB-Asia Joint Conference.** *BMC Bioinformatics* 2011, **12**(Suppl 13):S1.
- Graham-Rowe D, Goldston D, Doctorow C, Waldrop M, Lynch C, Frankel F, Reid R, Nelson S, Howe D, Rhee SY, et al: **Big data: science in the petabyte era.** *Nature* 2008, **455**:1-50.
- Giardine B, Riemer C, Hardison RC, Burhans R, Eltniski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A: **Galaxy: a platform for interactive large-scale genome analysis.** *Genome Res* 2005, **15**(10):1451-1455.
- Goecks J, Nekrutenko A, Taylor J, Team G: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**(8):R86.
- Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J: **Galaxy: a web-based genome analysis tool for experimentalists.** *Curr Protoc Mol Biol* 2010, Chapter 19:Unit 19.10-21.
- CBIIT-Galaxy. [<http://galaxy.cbiit.cuhk.edu.hk/>].
- Abouelhoda M, Issa SA, Ghanem M: **Tavaxy: integrating Taverna and Galaxy workflows with cloud computing support.** *BMC Bioinformatics* 2012, **13**:77.
- Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T: **Taverna: a tool for building and running workflows of services.** *Nucleic Acids Res* 2006, **34**(Web Server):W729-W732.
- EasyGenomics. [<https://www.easygenomics.com/>].
- Belhajjame K, Corcho O, Garijo D, Zhao J, Missier P, Newman D, Palma R, Bechhofer S, Garcia-Cuesta E, Gómez-Pérez J, Klyne G, Page K, Roos M, Ruiz J, Soiland-Reyes S, Verdes-Montenegro L, Roure DD, Goble C: **Workflow-centric research objects: a first class citizen in the scholarly discourse.** *Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web* 2012.
- Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, Begley K, Booth T, Bougueleret L, Burns G, Chapman B, Clark T, Coleman LA, Copeland J, Das S, de Daruvar A, de Matos P, Dix I, Edmunds S, Evelo CT, Forster MJ, Gaudet P, Gilbert J, Goble C, Griffin JL, Jacob D, Kleijns J, Harland L, Haug K, Hermjakob H, Sui SJH, Laederach A, Liang S, Marshall S, McGrath A, Merrill E, Reilly D, Roux M, Shamu CE, Shang CA, Steinbeck C, Trefethen A, Williams-Jones B, Wolstencroft K, Xenarios I, Hide W: **Toward interoperable bioscience data.** *Nat Genet* 2012, **44**(2):121-126.
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y: **Identifying personal genomes by surname inference.** *Science* 2013, **339**(6117):321-324.
- Tretyakov K, Laur S, Smant G, Vilo J, Prins P: **Fast probabilistic file fingerprinting for big data.** *BMC Genomics* 2013, **14**(Suppl 2):S8.
- Peng Y, Leung HCM, Yiu SM, Chin FYL: **Meta-DBA: a de Novo assembler for metagenomic data.** *Bioinformatics* 2011, **27**(13):i94-101.
- Zhao Z, Nguyen T, Deng N, Johnson K, Zhu D: **SPATA: A Seeding and Patching Algorithm for de novo Transcriptome Assembly.** *Bioinformatics & Biomedicine Workshops, IEEE International Conference* 2011.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci USA* 1992, **89**(5):1827-1831.
- Frith MC, Mori R, Asai K: **A mostly traditional approach improves alignment of bisulfite-converted DNA.** *Nucleic Acids Res* 2012, **40**(13):e100.
- Kolekar P, Kale M, Kulkarni-Kale U: **Alignment-free distance measure based on return time distribution for sequence analysis: applications to clustering, molecular phylogeny and subtyping.** *Mol Phylogenet Evol* 2012, **65**(2):510-522.
- Sun K, Chen X, Jiang P, Song X, Wang H, Sun H: **iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data.** *BMC Genomics* 2013, **14**(Suppl 2):S7.
- Menor M, Baek K, Poisson G: **Multiclass relevance units machine: benchmark evaluation and application to small ncRNA discovery.** *BMC Genomics* 2013, **14**(Suppl 2):S6.
- Deng F, Cui W, Wang L: **A highly accurate heuristic algorithm for the haplotype assembly problem.** *BMC Genomics* 2013, **14**(Suppl 2):S2.
- Matsumoto H, Kiryu H: **MixSIH: a mixture model for single individual haplotyping.** *BMC Genomics* 2013, **14**(Suppl 2):S5.
- Xu F, Wang W, Wang P, Li MJ, Sham PC, Wang J: **A fast and accurate SNP detection algorithm for next-generation sequencing data.** *Nat Commun* 2012, **3**:1258.

25. Nong G, Zhang S, Chan WH: **Two Efficient Algorithms for Linear Time Suffix Array Construction.** *Computers, IEEE Transactions on* 2011, **60**(10):1471-1484.
26. Maher B: **ENCODE: The human encyclopaedia.** *Nature* 2012, **489**(7414):46-48.
27. Conner S: **Scientists debunk 'junk DNA' theory to reveal vast majority of human genes perform a vital function.** *The Independent* 2012 [http://www.independent.co.uk/news/science/scientists-debunk-junk-dna-theory-to-reveal-vast-majority-of-human-genes-perform-a-vital-function-8106777.html].
28. McKie R: **Scientists attacked over claim that 'junk DNA' is vital to life.** *The Observer* 2013 [http://www.guardian.co.uk/science/2013/feb/24/scientists-attacked-over-junk-dna-claim].
29. Kurosawa J, Nishiyori H, Hayashizaki Y: **Deep cap analysis of gene expression.** *Methods Mol Biol* 2011, **687**:147-163.
30. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, Rando OJ, Birney E, Myers RM, Noble WS, Snyder M, Weng Z: **Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors.** *Genome Res* 2012, **22**(9):1798-1812.
31. Leung KS, Wong KC, Chan TM, Wong MH, Lee KH, Lau CK, Tsui SKW: **Discovering protein-DNA binding sequence patterns using association rule mining.** *Nucleic Acids Res* 2010, **38**(19):6324-6337.
32. Fujiwara T, Yada T: **miRNA-target prediction based on transcriptional regulation.** *BMC Genomics* 2013, **14**(Suppl 2):S3.
33. Balaga O, Friedman Y, Linial M: **Toward a combinatorial nature of microRNA regulation in human cells.** *Nucleic Acids Res* 2012, **40**(19):9404-9416.
34. Tsuyuzaki K, Tominaga D, Kwon Y, Miyazaki S: **Two-way AIC: Detection of Differentially Expressed Genes from Large Scale Microarray Meta-Dataset.** *BMC Genomics* 2013, **14**(Suppl 2):S9.
35. Parikh AP, Wu W, Curtis RE, Xing EP: **TREEGL: reverse engineering tree-evolving gene networks underlying developing biological lineages.** *Bioinformatics* 2011, **27**(13):i196-i204.
36. Davis MJ, Shin CJ, Jing N, Ragan MA: **Rewiring the dynamic interactome.** *Mol Biosyst* 2012, **8**(8):2054-66.
37. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**(10):785-786.
38. Ashby C, Johnson D, Walker K, Kanj IA, Xia G, Huang X: **New enumeration algorithm for protein structure comparison and classification.** *BMC Genomics* 2013, **14**(Suppl 2):S1.
39. Gajula MP, Vogel K, Rai A, Dietrich F, Steinhoff H: **How far *in-silico* computing meets real experiments. A study on the structure and dynamics of spin labeled vinculin tail protein by molecular dynamics simulations and EPR spectroscopy.** *BMC Genomics* 2013, **14**(Suppl 2):S4.
40. Esquivel-Rodríguez J, Yang YD, Kihara D: **Multi-LZerD: multiple protein docking for asymmetric complexes.** *Proteins* 2012, **80**(7):1818-1833.
41. Esquivel-Rodríguez J, Kihara D: **Fitting multimeric protein complexes into electron microscopy maps using 3D Zernike descriptors.** *J Phys Chem B* 2012, **116**(23):6854-6861.
42. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bard-well H, Johnson N, Doyle S, Turashvili G, Provenzano E, Aparicio S, Caldas C, Markowitz F: **Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling.** *Sci Transl Med* 2012, **4**(157):157ra143.
43. Beckman RA, Schemmann GS, Yeang CH: **Impact of genetic dynamics and single-cell heterogeneity on development of nonstandard personalized medicine strategies for cancer.** *Proc Natl Acad Sci USA* 2012, **109**(36):14586-14591.
44. Zhang S, Liu CC, Li W, Shen H, Laird PW, Zhou XJ: **Discovery of multi-dimensional modules by integrative analysis of cancer genomic data.** *Nucleic Acids Res* 2012, **40**(19):9379-9391.
45. Ding D, Lou X, Hua D, Yu W, Li L, Wang J, Gao F, Zhao N, Ren G, Li L, Lin B: **Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach.** *PLoS Genet* 2012, **8**(12):e1003065.
46. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeyasinghe S, Krawczak M, Cooper DN: **Human Gene Mutation Database (HGMD): 2003 update.** *Hum Mutat* 2003, **21**(6):577-581.
47. **HKU Lab Software Downloads.** [http://paed.hku.hk/uploadarea/yangwl/html/software.html].
48. Jiang P, Chan KCA, Liao GJW, Zheng YWL, Leung TY, Chiu RWK, Lo YMD, Sun H: **FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma.** *Bioinformatics* 2012, **28**(22):2883-2890.
49. Pungpapong V: **Empirical Bayes Variable Selection for High-Dimensional Regression.** *PhD thesis* Purdue University; 2012.
50. Pungpapong V, Muir WM, Li X, Zhang D, Zhang M: **A fast and efficient approach for genomic selection with high-density markers.** *G3 (Bethesda)* 2012, **2**(10):1179-1184.
51. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**(4):1819-1829.

doi:10.1186/1471-2164-14-S2-11

**Cite this article as:** Tretyakov *et al.*: Summary of talks and papers at ISCB-Asia/SCCG 2012. *BMC Genomics* 2013 **14**(Suppl 2):11.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

