Genomics Proteomics Bioinformatics 11 (2013) 253–256

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

MEETING REPORT

Promote Connections of Young Computational Biologists in China

Shihua Zhang [1,*], **Xiu-Jie Wang** [2,*]

[1] *National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*
[2] *Center for Molecular Systems Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*

Received 21 June 2013; revised 23 June 2013; accepted 1 July 2013
Available online 5 July 2013

Background

The first Youth Investigator Forum on Interdisciplinary Research of Mathematics, Computer Science and Biological Science (YMCB) took place on May 18–19, 2013 in Beijing, China. The conference was hosted by the National Center for Mathematics and Interdisciplinary Sciences (NCMIS), Chinese Academy of Sciences (CAS), and was initiated by Dr. Xiu-Jie Wang (Institute of Genetics and Developmental Biology, CAS) and Dr. Shihua Zhang (Academy of Mathematics and Systems Science, CAS). This conference was launched to nurture the next generation of computational biologists in China. More than 300 people from all over the country attended this conference and 21 talks and 40 posters were presented at this meeting. The topics covered next-generation sequencing data analysis, transcription and mRNA regulation, microRNA and lncRNA regulation, network biology, cancer genomics, epigenetics and personalized medicine. The conference was opened by Prof. **Shihua Zhang** with a warm welcome to the participants and an introduction to the goal of the conference. Then Prof. **Lei Guo**, **Zhirong Sun** and **Xiu-Jie Wang** gave brief remarks respectively to wish a successful conference. In this report we give a short introduction about the talks and briefly recapitulate each one of them, providing references to related publications whenever possible. We summarize this report with a few comments on the success of this conference.

Topics

In the morning session on the first day (May 18), four speakers presented their work on epigenetics, cancer biology and proteome science. Histone modifications have been shown to play critical roles in classic epigenetic regulation. Deciphering how newly-deposited histones acquire these modifications during and after DNA replication is vital for understanding the mechanisms underlying epigenetic inheritance. Attempting to tackle this important question, Prof. **Bing Zhu** and his team employed combinatorial approaches through integrating biochemistry, quantitative mass spectrometry and high-throughput sequencing technology. He highlighted some of their recent progresses along this direction [1–5]. 5-Methylcytosine (5mC) is a major epigenetic modification, which is often referred to as ''the fifth nucleotide''. However, we only have limited knowledge about how offspring inherit the DNA methylome from their parents. Prof. **Jiang Liu** reported their recent important work published in *Cell* journal [6]. Nine single-base resolution DNA methylomes were generated for zebrafish at various early developmental stages including gametes and early embryos. They found that compared to sperm,

* Corresponding authors.
E-mail: zsh@amss.ac.cn (Zhang S), xjwang@genetics.ac.cn (Wang XJ).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2013.07.001

methylome from oocytes is significantly hypo-methylated. Maternal DNA methylation pattern is maintained until the 16-cell stage; surprisingly, the paternal DNA methylation pattern is maintained throughout early embryogenesis. They thus concluded that "*besides DNA sequences, sperm DNA methylome is also inherited in zebrafish early embryos*". Prof. **Lijian Hui** reported his long-term interest in understanding the mechanisms underlying tumorigenesis of normal cells [7]. Using hepatocytes as an experimental system, his lab explores cell lineage conversion with particular focus on inducing non-hepatic cells to form functional hepatocyte-like cells. By doing so, they try to address the essential scientific question, that is, how do cells maintain their identity through preventing the conversion of terminally differentiated cells to other cell types, including transdifferentiation to different lineages and transformation to tumor cells? Finding answers to this question would facilitate the development of novel therapies for liver diseases and cancers. Chemical cross-linking combined with mass spectrometry (CXMS) is a powerful tool for analyses of protein structures and protein–protein interactions. Prof. **Meng-Qiu Dong** and her team have developed a CXMS workflow, which combines a variety of readily available cross-linkers, high-resolution mass spectrometry and a data analysis software program called pLink [8]. Their method was validated with proteins of known structures and further tested on protein complexes, crude immunoprecipitates and whole-cell lysates. In her talk, she highlighted several key challenges and stated that using CXMS to improve protein structure prediction is a task that requires further development of MS approaches, better understanding of the meaning of CXMS data with respect to protein dynamics and dedicated computational tools.

In the afternoon session on the first day (May 18), six speakers presented their work. Expression quantitative trait loci (eQTL) studies have generated large amounts of data in different organisms. However, the role of epistasis in the joint regulation of multiple genes has not been explored, largely due to the computational complexity issue. Prof. **Minghua Deng** and his team proposed a computationally-feasible approach to identify pairs of chromosomal regions that interact to regulate co-expression patterns of gene pairs based on a bivariate model, the covariance matrix of which depends on the joint genotypes at the candidate loci [9]. They also proposed a filtering process to reduce the computational burden. Prof. **ChengGang Zhang** delivered a survey talk about how data analysis technique and interdisciplinary research face the application demand in the translational research era with lots of interesting case studies. Exome sequencing has been widely used in detecting nonsynonymous single nucleotide polymorphisms (nsSNPs) associated with human inherited diseases. Prof. **Rui Jiang** presented a bioinformatics method called SNP prioritization via the integration of genomic data (SPRING) to predict pathogenic nsSNPs for a given query about disease in exome sequencing studies. Prof. **Yuling Jiao** presented his studies on the auxin movement of leaf polarity formation and highlighted the important roles of mathematical simulation in their studies. In the following talk, Prof. **Ling-Yun Wu** reported recent progresses on network comparison, which is an important topic in network biology [10]. Particularly, he presented a new network alignment method based on a machine learning model and showed their promising results. With the rapid expansion of deep sequencing technologies, plenty of individual genomes were sequenced, offering both serious challenges and great opportunities for bioinformaticians. Combining multiple level genomic data with novel algorithms, Prof. **Ge Gao** and his team systematically screened more than 950 human genomes and, surprisingly, identified multiple unknown functional elements/units among the whole genome.

In the morning session on the second day (May 19), five speakers presented their studies on regulation of microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Prof. **Xiu-Jie Wang** presented the recent work in her lab [11]. They developed a computational method and systematically identified intergenic or noncoding gene-originated endogenous miRNA target mimic (eTM) for 20 conserved miRNAs in Arabidopsis and rice. The predicted miRNA binding sites were well conserved among eTMs of the same miRNA, whereas sequences outside of the binding sites varied a lot. The effectiveness of eTMs for three other miRNAs was also confirmed by transient agroinfiltration assay. They also developed a web-based plant small RNA meta-analysis toolbox, PsRobot, to meet the increasing need for batch analysis of small RNAs (sRNAs) in plants with the development of high-throughput sequencing technology [12]. In the following talk, Prof. **Jian Lu** reported the impact of miRNA regulation on variation in human gene expression [13]. By examining variations in gene expression patterns in human populations and between human and other primate species, they found that during primate evolution, miRNAs have stabilized expression of a small number of target genes. They integrated expression data with genotypes determined in the HapMap3 and the 1000 Genomes Projects and revealed that the increased variation in expression of miRNA target genes is mainly attributed to variations in expression of miRNAs, genetic variants in miRNA loci and mutations in miRNA target sites. Accumulating evidence demonstrates that lncRNAs play many important roles in diverse biological processes. Therefore, there is a great demand for functional annotation of available lncRNAs. Prof. **Yi Zhao** presented the recent work of his lab, trying to tackle this issue for the first time using a global network-based strategy. They integrated gene expression data and protein interaction data to develop lncRNA global function predictor (lnc-GFP) based on a bi-colored network to predict potential functions of lncRNAs on a large scale [14]. Prof. **Qinghua Cui** summarized the work of his lab on the function of miRNAs as well as their association with diseases [15]. Prof. **Xiaowo Wang** presented his recent studies on quantitative models on cell-specific miRNA regulation.

In the afternoon session on the second day (May 19), Prof. **Yi Xing** presented his talk on the regulatory variation in expression of mRNA isoforms in human populations [16]. Alternative splicing is the process by which exons from precursor mRNA transcripts are differentially included during splicing, resulting in different mature mRNA isoforms from a single gene locus. Common genetic variations affecting splicing regulation can confer differential alternative splicing between human individuals and, as a consequence, impact expression level or function of proteins. Prof. Yi Xing and his team have developed GLiMMPS, a robust tool for the detection of genetic variations in alternative splicing from RNA-seq data. As population-scale RNA-seq studies become increasingly affordable and popular, GLiMMPS provides a useful tool for elucidating the genetic architecture

of alternative splicing variations in humans and model organisms. Prof. **Wenfeng Qian** and his team investigated the cellular efficiency in protein translation, which is an important fitness determinant in rapidly-growing organisms [17]. They estimated the *in vivo* translational speeds of all sense codons from the budding yeast *Saccharomyces cerevisiae* and found that preferentially-used codons are not translated faster than those unpreferred. CCCTC-binding factor (CTCF) is a key factor in regulating gene expression, DNA loop formation and maintaining high order 3D chromatin architecture. CTCF has been hypothesized to exhibit distinct properties through binding various DNA sequence motifs by different combinations of the zinc fingers. To better understand the relationship between the binding patterns and its versatile properties, Prof. **Zhihua Zhang** presented his motif pipeline to reanalyze the ENCODE data of CTCF ChIP-seq experiments in 14 cell types and study functional distinctive CTCF bindings. Prof. **Yong Wang** talked about how to identify key biomarkers from high-dimensional biological data using optimization models [18]. They proposed an efficient approach to model the disease complexity by ellipsoids and look for a set of heterogeneous biomarkers. The barcode of a genome is formulated from its composition. Prof. **Fengfeng Zhou** demonstrated this interesting fractal feature of a genome, together with potential biological applications in his talk. Prof. **Yong Zhang** presented a talk on new gene evolution in the light of omics [19]. Regardless of the under-characterization of lineage-specific new genes in molecular biology, integrative functional genomics demonstrated their prevalence and functional importance. Specifically, they found a high flux of new gene addition to the evolving genome, which was originated via DNA-level duplication, retroposition and *de novo* origination. Then further found that new genes were very important for the evolution of reproduction and development, as well as several other interesting findings.

Other than these brilliant talks, 40 posters were also presented going along with coffee break time. These poster presentations greatly expanded and enriched the topics of this conference. More importantly, this arrangement enabled young graduate students to have a chance to present their studies, improve their social skills and make new friends. For example, Zhao et al. [20] presented their recent work about driver pathway identification in cancer, which has attracted the interest of several researchers on the conference and they promptly asked for the code and data used in this study.

Conclusion

The conference ends successfully and has received many positive feedbacks from the participants. It provided a good platform for young researchers to communicate and share their own findings. We hope it can also be helpful for the development of computational biology in China, especially for young investigators. Among other memorable aspects of this conference were the social events organized after the long days of presentations and poster sessions. The new format of reception enables the speakers and some participants of the meeting to have close interactions in a casual environment and participate in relaxed and open discussions. The rather young community of the conference added an exceptional degree of friendliness. The students also had more chances to interact with established young investigators.

Given the rapidly-increasing needs for interdisciplinary studies and cross-field collaboration, closer ties among young investigators will help to promote the relationships among different types of institutions. At least, participants of this conference have manifested this point, with their actions and willingness of exploring more possibilities of cooperation. Taken together, the Youth Investigator Forum on Interdisciplinary Research of different disciplines brings us closer. Many participants have expressed their will to attend such events in the future, and we will try our best.

Acknowledgements

We are grateful for the support from the National Center for Mathematics and Interdisciplinary Sciences, and the Youth Innovation Promotion Association, CAS.

References

[1] Huang C, Zhang Z, Xu M, Li Y, Li Z, Ma Y, et al. H3.3-h4 tetramer splitting events feature cell-type specific enhancers. PLoS Genet 2013;9:e1003558.
[2] Huang C, Xu M, Zhu B. Epigenetic inheritance mediated by histone lysine methylation: maintaining transcriptional states without the precise restoration of marks? Philos Trans R Soc Lond B Biol Sci 2013;368:20110332.
[3] Yuan W, Wu T, Fu H, Dai C, Wu H, Liu N, et al. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science 2012;337:971–5.
[4] Xu M, Chen S, Zhu B. Investigating the cell cycle-associated dynamics of histone modifications using quantitative mass spectrometry. Methods Enzymol 2012;512:29–55.
[5] Yang N, Wang W, Wang Y, Wang M, Zhao Q, Rao Z, et al. Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. Proc Natl Acad Sci U S A 2012;109:17954–9.
[6] Jiang L, Zhang J, Wang JJ, Wang L, Zhang L, Li G, et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. Cell 2013;153:773–84.
[7] Ji S, Zhang L, Hui L. Cell fate conversion: direct induction of hepatocyte-like cells from fibroblasts. J Cell Biochem 2013;114:256–65.
[8] Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, et al. Identification of cross-linked peptides from complex samples. Nat Methods 2012;9:904–6.
[9] Wang L, Zheng W, Zhao H, Deng M. Statistical analysis reveals co-expression patterns of many pairs of genes in yeast are jointly regulated by interacting loci. PLoS Genet 2013;9:e1003414.
[10] Huang Q, Wu LY, Zhang XS. An efficient network querying method based on conditional random fields. Bioinformatics 2011;27:3173–8.
[11] Wu HJ, Wang ZM, Wang M, Wang XJ. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. Plant Physiol 2013;161:1875–84.
[12] Wu HJ, Ma YK, Chen T, Wang M, Wang XJ. PsRobot: a web-based plant small RNA meta-analysis toolbox. Nucleic Acids Res 2012;40:W22–8.

[13] Lu J, Clark AG. Impact of microRNA regulation on variation in human gene expression. Genome Res 2012;22:1243–54.
[14] Guo X, Gao L, Liao Q, Xiao H, Ma X, Yang X, et al. Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. Nucleic Acids Res 2013;41:e35.
[15] Qiu C, Chen G, Cui Q. Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. Sci Rep 2012;2:318.
[16] Zhao K, Lu ZX, Park JW, Zhou Q, Xing Y. GLiMMPS: Robust statistical model for regulatory variation of alternative splicing using RNA-seq data. Genome Biol 2013; in press.
[17] Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet 2012;8:e1002603.
[18] Ren X, Wang Y, Chen L, Zhang XS, Jin Q. EllipsoidFN: a tool for identifying a heterogeneous set of cancer biomarkers based on gene expressions. Nucleic Acids Res 2013;41:e53.
[19] Zhang YE, Landback P, Vibranovski MD, Long M. Accelerated recruitment of new brain development genes into the human genome. PLoS Biol 2011;9:e1001179.
[20] Zhao J, Zhang S, Wu LY, Zhang XS. Efficient methods for identifying mutated driver pathways in cancer. Bioinformatics 2012;28:2940–7.