Available online at www.sciencedirect.com
ScienceDirect
J. Genet. Genomics 35 (2008) 577−584
JOURNAL OF GENETICS AND GENOMICS
www.jgenetgenomics.org

Developmental systems biology flourishing on new technologies

Jing-Dong J. Han [a, *], Yi Liu [a, b, 1], Huiling Xue [a, 1], Kai Xia [a, 1], Hong Yu [a, 1], Shanshan Zhu [a, 1], Zhang Chen [a], Wei Zhang [a], Zheng Huang [a], Chunyu Jin [a], Bo Xian [a], Jing Li [a], Lei Hou [a], Yixing Han [a], Chaoqun Niu [a], Timothy C. Alcon [a, c]

[a] *Chinese Academy of Sciences Key Laboratory of Molecular Developmental Biology*, *Center for Molecular Systems Biology*, *Institute of Genetics and Developmental Biology*, *Chinese Academy of Sciences*, *Beijing 100101*, *China*
[b] *Key Laboratory of Machine Perception* (*Ministry of Education*), *Peking University*, *Beijing 100871*, *China*
[c] *Bioinformatics and Computational Biology Graduate Program*, *Iowa State University*, *Ames*, *IA 50011-1040*, *USA*

Received for publication 13 August 2008; revised 4 September 2008; accepted 5 September 2008

Abstract

Organism development is a systems level process. It has benefited greatly from the recent technological advances in the field of systems biology. DNA microarray, phenome, interactome and transcriptome mapping, the new generation of deep sequencing technologies, and faster and better computational and modeling approaches have opened new frontiers for both systems biologists and developmental biologists to reexamine the old developmental biology questions, such as pattern formation, and to tackle new problems, such as stem cell reprogramming. As showcased in the International Developmental Systems Biology Symposium organized by Chinese Academy of Sciences, developmental systems biology is flourishing in many perspectives, from the evolution of developmental systems, to the underlying genetic and molecular pathways and networks, to the genomic, epigenomic and noncoding levels, to the computational analysis and modeling. We believe that the field will continue to reap rewards into the future with these new approaches.

Keywords: developmental systems biology; new technologies; symposium

Introduction

The CAS International Symposium on Developmental Systems Biology (May 18−20, 2008, Beijing) was organized by the Institute of Genetics and Developmental Biology and the Institute of Biophysics, the Chinese Academy of Sciences . It was supported by the Chinese Academy of Sciences (CAS), the National Science Foundation of China (NSFC), our partner the CAS Key Laboratory of Molecular Developmental Biology, and commercial sponsors, in particular IBM.

The symposium covered a wide range of topics on systems and computational biology with an emphasis on organism development. The cutting-edge research and technology for Developmental Systems Biology presented by many speakers fascinated the participants. Here we summarize the talks into five topics and wish to share the excitement with our readers.

Evolution

Evolutionary selection acts at different levels in an organism. DNA evidence for evolution includes mutations and nucleotide similarities among species, while recent studies find that gene expression is also tuned during evolution to control biological processes. Its effects can even be observed in complex gene networks at the systems level.

Philipp Khaitovich (CAS-MPG Partner Institute for Computational Biology, China) presented an unpublished work on transcriptome evolution in primates. In development, timing is crucial, and evolution often works by altering the timing of various developmental processes. In

* Corresponding author. Tel: +86-10-6484 5843; Fax:+ 86-10-6484 5797.
E-mail address: jdhan@genetics.ac.cn
[1] These authors contributed equally to this report.

primate evolution, such a change can be observed in the timing of sexual maturation, which is obviously delayed in human relative to other primates. By analyzing mRNA expression in the brains of humans, chimpanzees and rhesus macaques, his group finds that delayed maturation is also evident at the gene expression level.

Ralph Greenspan's group (the Neurosciences Institute, USA) illustrated the flexibility of gene networks under different selection strategies (van Swinderen and Greenspan, 2005) in the evolution of fly behavior (Dierick and Greenspan, 2006; Weber et al., 2008). He demonstrated that gene interactions are rewired under different backgrounds in fruitfly. Their findings have implications for the role of degeneracy and complexity in setting gene network states, which in turn may offer insights into evolutionary mechanisms.

Co-evolution refers to reciprocal evolutionary change between interacting partners. It plays a critical role in driving interacting components to evolve. MicroRNAs may be the most active class of genes in the genome in evolutionary terms.

Chung-I Wu's group (University of Chicago, USA) finds that only a very small fraction of stem-loop structures that have emerged through evolution are actually functional and preserved, likely driven by selection (Lu et al., 2008a, 2008b). They find that the adaptation of a new miRNA to the transcriptome is a long process that involves many changes in miRNAs themselves. They also examine how miRNAs and their targets co-evolve and co-adapt during evolution. They hypothesize that co-adaptation may also be important for genetic buffering circuits, such as feedback and feed-forward loops, established by miRNAs and their targets to stabilize the transcriptome output. They carry out transgenic experiments on evolving miRNAs across species in *Drosophila* and find that the predicted targets of *Drosophila melanogaster* (Dm) miRNA Dm310 tends to be expressed at a normal or elevated level when Dm310 is over-expressed in Dm. However, the target expression is repressed when the orthologous miRNA cluster in *Drosophila pseudoobscura* (Dp) Dp310 is over-expressed in Dm. They propose that Dp310 (but not Dm310) might disrupt the feedback or feed-forward loops in Dm and that co-adaptation between miRNAs and their targets may be important to maintain the stability of the transcriptome.

Taijiao Jiang (Institute of Biophysics, CAS, China) presented a method based on a nucleotide co-occurrence network to understand human influenza evolution (Du et al., 2008). This network model effectively captures the antigenic evolutionary patterns of H3N2 virus and identifies the genetic basis for human influenza epidemics. They find that many human influenza virus genomic mutations are not random, but instead participate in highly correlated networks of cooperative mutations and that correlated amino acid substitutions are preferentially located in the known antigenic regions of the viral hemagglutinin (HA). They also find a specific correlated amino acid substitution that may account for a recent human-avian infectivity shift.

Evolutionary events can also be used to predict gene/protein interactions. **Zhirong Sun** (Tsinghua University, China) described a method for inferring functional associations between proteins from evolutionary events (such as gene loss, gene gain and horizontal transfer) during speciation (Zhou et al., 2006). The profiles of evolutionary events are constructed from phylogenetic profiles and a species tree. They find that this method has better prediction performance than one simply based on phylogenetic profiles (presence or absence of genes in different species). The two methods can complement each other in predicting protein-protein interactions (PPIs).

Developmental pathways and networks

Pernille Rorth (Temasek Life Science Laboratory, Singapore) presented a fascinating work on cell migration and guidance mechanisms in *Drosophila*. They dissect signaling events in border cells, a cluster of about eight cells which perform a spatially and temporally controlled migration during *Drosophila* oogenesis. They find that a cell cluster uses two fundamentally different types of guidance signaling. One is already found in single-cell migrations, while a second mode revealed by their study is specific for the group of cells which harbor spatial information and act as a collective (Bianco et al., 2007; Rorth, 2007). Understanding such collective migratory behavior will be of great importance to cancer research, since cancer cells may metastasize in groups rather than as single cells.

Nicholas Baker (Albert Einstein College of Medicine, USA) showed a good example of cell fate regulation in a dynamic model of *Drosophila* eye development. He and his collaborators do a quantitative analysis focusing on the activator-inhibitor system responsible for the regular spacing of the R8 photoreceptors that define the eye's ommatidial pattern. A novel, non-Turing mechanism is found where R8 induction is determined by an intrinsically dynamic process involving long-range activation and short-range inhibition with the existing R8s acting as a template. This model predicts that R8 cells are defined before the appearance of the full group of proneural cells, and not selected by interactions between these cells. The model therefore provides insight into the selection of other neural cells by lateral inhibition.

By high-throughput RNAi screen in *C. elegans*, **Hui Ge's group** (Whitehead Institute for Biomedical Research, USA) focuses on the gene pleiotropic effect. They define pleiotropic genes by their "phenotypic signatures" and find that pleiotropy occurs extensively among genes involved

in early embryogenesis. In addition, they hypothesize that these genes are organized into partially overlapping functional modules, with the pleiotropic genes as connectors between these modules (Zou et al., 2008).

Stem cells are defined by the ability to self-renew and differentiate into mature somatic cell types. **Sheng Zhong's laboratory** (University of Illinois at Urbana-Champaign, USA) studied the conserved as well as distinct regulatory network components in human embryonle stem (hES) and mouse embryonic stem (mES) cells. They generated a time-course microarray dataset for differentiating mES cells and compared it with several differentiation datasets of hES cells. They demonstrated that a species-specific regulatory module is critical to maintain pluripotent phenotype in mES cells. They reported that two target genes, which are activated in mES cells by this regulatory module, are activated in hES cells by other transcription factors. These data offer an example that the transcription network is rewired but the same transcriptional output is maintained (Xie et al., 2008).

By integrating protein-protein interaction (PPI) data and gene expression profiles during fruitfly and human brain aging, **Jing-Dong Jackie Han's group** (Institute of Genetics and Developmental Biology, CAS, China) finds two pairs of transcriptionally anti-correlated modules associated with two temporal switches of cellular functions, "proliferation to differentiation" and "reductive metabolic to oxidative metabolic". Network analysis and RNAi experiments in *C. elegans* demonstrate that the genes connecting different modules in PPI networks seem to preferentially affect network stability as well as aging and/or longevity (Xue et al., 2007).

Using multiple types of network data, including expression profiling, PPIs, genetic interaction and phenotypic profiling in human and other model organisms, **Marc Vidal's group** (CCSB, Dana-Farber Cancer Institute, USA) presented a network perspective on breast cancer. Novel genes are found to be associated with breast cancer by their network modeling strategy, and one of the novel predicted genes, HMMR, is experimentally validated (Pujana et al., 2007). They also use a drug-target network to characterize the relationships between drug targets and disease genes, and open up novel insights into diseases and strategies for selecting drug target genes (Yildirim et al., 2007). They recently developed a method to evaluate node removal versus edge removal on the behaviors and properties of networks. Node perturbations (such as null mutations) and edge perturbations (such as point mutations in a PPI domain) are introduced based on human disease-related gene mutations. They have also revisited gene annotations in yeast. Training a naïve Bayes predictor with a combination of different "omics" data, they are able to predict and experimentally validate that many open reading frames (ORFs) previously thought to be "spurious" are actually actively transcribed (Li et al., 2008b).

Finding causal genes for complex diseases is a big challenge. **Michael Zhang** (Cold Spring Harbor Laboratories, USA & Tsinghua University, China) presented an improved method, CIPHER, to predict disease candidate genes (Wu et al., 2008). They first derive phenotypic descriptions of gene mutations based on literature mining, then use an integrated network, consisting of a phenotype network, a gene-phenotype network and a protein network, to construct a regression model to score the consistency of the phenotypic description and disease description. Based on the concordance across different networks, they rank and predict genes that might be associated with a disease based on their confidence scores. CIPHER outperforms many algorithms proposed previously in the precision and recall of prediction, but relies little on known disease genes of the same phenotype. It is also effective in predicting disease genes for phenotypes without any known genetic origins.

An important aim of computational biology is to predict protein function. **Peer Bork's group** (EMBL-Heidelberg, Germany) uses several approaches to achieve this goal: homolog-based function prediction (Kensche et al., 2008) and protein-network based function prediction. For the second approach, he presented a drug-target protein network based on side-effect similarities between drugs, under the assumption that two drugs with similar side-effects may have the same drug targets. Unexpectedly, the network contains 261 side-effect-driven drug-drug relations formed by chemically dissimilar drugs. They experimentally test 20 of these. Thirteen drug-target relations are indicated by *in vitro* binding assays and 9 of these are confirmed in cell assays. This study leads to a new usage of phenotypic information to infer molecular interactions between marketed drugs and their protein targets (Campillos et al., 2008; Kuhn et al., 2008). They also analyze temporal and spatial characteristics of biological networks and find different levels of biological regulation, such as dynamic changes of gene expression in a time-series microarray experiment, using a newly developed tool, KEGG Atlas (Okuda et al., 2008).

Nicholas Luscombe (EMBL-EBI, UK) presented a genome-scale analysis of how regulatory feedback controls the metabolic system in *E. coli*, and examines how two models (enzyme concentration or enzyme activity) are deployed throughout the system. Rather than using a currently popular metabolic-related analysis, he presented an analysis based on the metabolic regulatory network of *E. coli* dataset from EcoCyc and RegulonDB. The total of 878 interactions integrated from the datasets, which exist in more than 80% of the pathways among more than 30% of the metabolic molecules, are gathered into two groups, "indirect" and "direct" interactions, corresponding to the enzyme concentration model and the enzyme activity model, respectively. Through analysis of time-scales, specificity of feedback, location of regulated reactions,

feedback metabolites, global organization, and overlap regulations, they find that "metabolism is regulated on a genomic scale in which direct and indirect interactions selectively control catabolism and anabolism by coordinating time-scales, specificities and concentrations".

Non-coding RNA

Stephen Cohen (Temasek Life Sciences Laboratory, Singapore) presented recent work on microRNA mutants in *Drosophila*, which provides insights into the functions of miRNAs in the brain and in the control of metabolism (Karres et al., 2007). They propose that a developmental switch can be accomplished by a change in transcription or a miRNA-mediated change in post-transcriptional gene expression. MicroRNA-mediated post-transcriptional gene regulation is thought to contribute to robustness, in part through noise reduction. They find that miRNA targets are often expressed at a very low level, possibly indistinguishable from noise, in miRNA-expressed cells. In such cases, the job of the miRNA is to keep target genes at a low expression level or to turn them off completely. The function of miRNA is very diverse: some miRNAs act as components of regulatory feedback loops; some are modulators that ensure robustness or set a threshold for switch activation; another set of miRNAs switch off target gene expression. Failure to regulate targets may have severe consequences or subtle effects, depending on the type of the miRNA-target relationships. These studies raise questions about the numbers of biologically important targets. His group also developed a new method to purify duplexes of miRNAs bound to their targets *in vivo*.

Yongqing Zhang (Institute of Genetics and Developmental Biology, CAS, China) presented his group's findings that FMRP, a *Drosophila* protein closely related to the human fragile X syndrome protein, appears in the RNA-induced silencing complex RISC and has a similar mutant phenotype to that of piRNA pathway piwi mutants, and it also has specific interactions with PIWI subfamily proteins PIWI and Aub. Based on these results, he suggests that a possible relationship between the non-coding piRNA pathway and FMRP might help reveal the mechanism of FMRP absence-caused mental retardation.

Genomics and epigenetics

Jun Wang (Shenzhen BGI, China) described their new findings based on the complete sequencing of a Han Chinese genome using the new generation of sequencing technology, the Solexa massive parallel signature sequencing, on unique molecular arrays. It is the first time that an Asian genome has been sequenced. As a representative Asian genome, it provides a better understanding of human evolution when compared with European genomes. Technically, it also provides a successful case study of applying the new generation short fragments (~35 mer) sequencing for resequencing and assembling the human genome based on a reference genome.

The human body is composed of a large number of cell types, each defined by a specific gene expression profile. *Cis*-regulatory elements, such as promoters near the transcriptional start site, enhancers distant from the transcriptional start site and insulators in the intergenic regions, control gene expression by associating with specific transcription factors, many of which modify local chromatin structures. While each class of *cis*-regulatory elements may contribute to cell-type dependent gene expression, previous studies have mainly focused on the role of promoters as a driving force behind tissue-specific and differential expression, partly due to the lack of knowledge of the long range regulatory elements. To better understand the mechanisms of cell-type specific gene expression, **Bing Ren's group** (University of California, San Diego, USA) performs ChIP-on-chip experiments to localize the genomic binding sites of general transcription factors, active chromatin modifications, and the insulator binding protein CTCF in the human genome in five cell types. They find that promoters are characterized by a high level of H3K4me3 and a relatively low level of H3K4me1, whereas the enhancers are the opposite. Using these characteristics, they further computationally predict enhancer and promoter sites in the human genome in Hela cells (Heintzman et al., 2007). Through examining the localization pattern of the insulator-binding protein CTCF, they identify a consensus binding site for CTCF and find that their locations remain largely invariant across various cell types (Kim et al., 2005; Kim et al., 2007). They also find that predicted enhancers are enriched near upregulated genes after CTCF depletion, suggesting a role of CTCF in blocking enhancer activities. Their recent unpublished ChIP-chip results on histone modification and CTCF profiles in proliferating and differentiated human embryonic stem cells reveal a switch of acetylation to trimethylations on H3K27 and a change in the majority of enhancers upon hESC differentiation, and that the enhancers with histone modifications correlate with cell-specific gene expression.

David Gifford (Massachusetts Institute of Technology, USA) and his collaborators have developed a system to study motor neuron differentiation from ES cells, where mouse ES cells are differentiated *in vitro* to form mature spinal motor neurons. Using this system, they find that *Hox* genes regulating this developmental course undergo distinct temporal changes in the patterns of chromatin marks and gene expression that are related to the exogenous signals retinoic acid (RA) and sonic hedgehog (Shh). Three types of data are gathered for ES cells in the process of becoming motor neurons: chromatin changes, transcription and the targets of transcription factor Olig2,

which is a master regulator in spinal motor neuron development. Chromatin modifications of Hox clusters, especially H3K27me3 and H3K4me3, change during this process. They also find that H3K27me3 domains and H3K4me3 domains at the Hox A cluster are not spread along the chromatin, but localize close together. The master regulator Olig2 is expressed at day 4 after mouse ES cells are induced to differentiate, and Olig2 binding sites are found primarily on H3K27me3 free domains on the same day. In addition, a lag of chromatin marks compared with gene expression is observed, including the clearance of H3K27me3. H3K4me3 is found to be correlated with GC content in the genome. Another study mapping the binding sites of highly conserved transcription factors in mouse and human using ChIP-chip experiments has revealed large variations in binding sites between the two species (Odom et al., 2007).

Histones are characterized by numerous post-translational modifications that influence gene transcriptional regulation. Histone acetylations are generally associated with gene activation, whereas histone methylations could be associated with either gene activation or repression through modifying different sites. However, due to the lack of global distribution data in higher eukaryotic organisms, it remains to be determined to what extent gene-specific combinatorial patterns of histone modifications exist. **Keji Zhao's group** (National Institute of Health, USA) performs a genome-wide analysis of 38 modifications including both acetylation and methylation in human $CD4^+$ T cells using ChIP (chromatin immunoprecipitation) with specific methylation or acetylation antibodies followed by Solexa sequencing ("ChIP-seq"). By analyzing the genome-wide distribution patterns of histone methylations and acetylations around the transcriptional start sites of all the genes together with their expression profiles, they find that all of the examined acetylations positively correlated with gene expression, which is consistent with their involvement in transcriptional activation. Their analysis also indicates that different acetylations may target different regions of genes, which is consistent with previous studies that specific histone acetyl transferases can associate with diverse regions of genes. To identify the combinatorial patterns of histone modifications, they examine all possible combinations of 18 acetylations, 19 methylations and H2A.Z. They find that only a small fraction of all possible combinations of modifications exist at promoter regions, among which only a few are prevalent. They also examine the modification patterns in the enhancer regions and also find that only a small number of patterns are prevalent. Their results reveal that a limited number of patterns are associated with promoters and enhancers. In particular, they identify a common modification module consisting of 17 modifications (H2A.Z, H2BK5ac, H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K18ac, H3K27ac, H3K36ac, H4K5ac, H4K8ac and H4K91ac), which is considered as a modification "backbone", at 3,090 promoters. They also test the robustness of this backbone through perturbations and principal component analysis. These modifications tend to colocalize in the genome and are correlated with each other at an individual nucleosome level. Their analysis of putative enhancers reveals various patterns of histone modifications. Their data also suggest that although the genes associated with these prevalent histone modification patterns tend to have high expression levels, the histone modifications themselves do not uniquely determine expression but may function cooperatively to prepare chromatin for transcriptional activation (Barski et al., 2007; Roh et al., 2007; Schones et al., 2008; Wang et al., 2008).

Results in sperm precursor cells of fruitfly from **Xin Chen's group** (Johns Hopkins University, USA) confirmed that high levels of H3K4 trimethylation and Pol II binding at transcription start sites label highly expressed genes.

Based on recent progress, **Jingde Zhu** (Shanghai Jiaotong University, China) presented a comprehensive review of current technical advances in epigenetic studies.

Using available high-resolution maps for many histone modifications on the human genome ("ChIP-seq" data), **Jing-Dong Han's group** built a Bayesian network to infer the logical relationships among histone modifications and gene expression. The network not only confirms known relationships, but also finds some new relationships, demonstrating network-based methods to be a promising new approach to deciphering the complex "histone code" (Yu et al., 2008).

To determine whether a significant abundance bias of a particular protein exists between non-diabetic and diabetic cohorts, **Jiarui Wu's group** (Shanghai Institutes for Biological Sciences, CAS, China) developed a computational strategy called LSPAD which stands for Localized Statistics of Protein Abundance Distribution. With this strategy, they uncover protein markers of high, medium and low abundance in human blood that are associated with diabetes, and the involvement of the ficolin-related complement system in type II diabetes.

Fuchu He's group (Fudan University & the Academy of Military Medical Sciences, China) shared their results on fetal liver proteomics. They identify 328 proteins with different expression in four fetal liver development stages. Based on the proteins' function categories, they discussed the relationship of hepatogenesis and development of fetal liver.

Similar to the combinatorial effects of histone modifications and histone isoforms or the "histone code", different ribosomal subunit isoforms have recently been found to have distinct functions, hence the existence of a potential "ribosome code". **Frederick Roth's group** (Harvard Medical School, USA) demonstrated paralog-specific re-

quirements for the translation of localized mRNAs by studying *ASH1* mRNA in yeast, and that these effects are restricted to a distinct subset of duplicated ribosomal proteins. Further comprehensive transcriptional and phenotypic profiling of cells lacking specific ribosomal proteins reveals differences between other functional roles of ribosomal protein paralogs in addition to their effects on mRNA localization. They find that the ribosomal protein paralogs have differential requirements for assembly and localization in specific genetic backgrounds (Suzanne Komili, 2007). They also introduced a new method they developed, called "the green monster" method. This method uses multiple GFP tags instead of one GFP at a time to screen the multiple mutants. The fluorescence intensity of GFP rather than the GFP species will increase with the number of mutations. They demonstrate that when fluorescence intensity is measured quantitatively, this then avoids the limitations of total number of GFP species/colors allowed in traditional methods of testing multi-gene interactions.

Bioinformatics

Many researchers presented new or improved bioinformatics methods. These works can be broadly categorized into five themes.

Finding transcription factor binding sites and gene regulatory modules

Wen-Hsiung Li's group (University of Chicago, USA) presented a new method to detect gapped motifs and to search for related gene regulatory modules in yeast (Tsai et al., 2006; Chen et al., 2008). Here, a gapped motif is a recurring DNA sequence pattern which is well conserved at both ends except for the inner "gapped" position. The key idea of their approach is to find short, conserved motifs first. Then, gapped motifs are identified by repeatedly linking two of the short motifs which flank a degenerate position.

Machine learning approach to disease status prediction

Xuegong Zhang's Lab (Tsinghua University, China) investigates the problem of classifying the status of breast cancer disease based on global gene expression profiles (Lu et al., 2008c). To this end, the authors propose a new learning algorithm, R-SVM, which jointly performs classification and feature selection (Zhang et al., 2006). Their results show that some features of the disease could be well predicted by gene expression, and the computation analysis can provide new understanding to the underlying biology even when some features can not be predicted satisfactorily with only molecular features.

Bioinformatics approaches to protein structure and molecular interactions prediction

Amy Keating's group (Massachusetts Institute of Technology, USA) studies the sequence and structural properties of the alpha-helical coiled coil motif, which is responsible for the dimerization of bZIP proteins (Newman and Keating, 2003; Grigoryan and Keating, 2006). They first conducted experiments to measure the strength of interactions between human bZIP transcription factors, as well as the stabilities of some homo- and hetero-dimer bZIPs in solution. Based on these results, they trained an SVM-classifier to predict bZIP interactions and proposed a computational approach to design novel peptides that specifically bind to a human coiled coil protein.

Luhua Lai's group (Peking University, China) study the arachidonic acid metabolic pathway (which is related to human inflammation) as a test case for investigating multi-way drug-targets interaction (Yang et al., 2007). Through computationally simulating drug effects on the pathway, they find that it is crucial to block both the 5-LOX and the COX-2 pathways to inhibit the inflammatory mediators effectively. Combinatorial target control strategies are proposed based on this finding to guide better drug design.

Network-based bioinformatics

As mentioned above in the Evolution and Network sessions, **Zhirong Sun's group** introduced their work on predicting PPI networks based on "evolutionary scenario" (Zhou et al., 2006), **Michael Zhang** presented a network-based approach for predicting human disease genes based on integrated functional networks (Wu et al., 2008). The basic assumption is that functionally related genes will be more likely to co-evolve or be responsible for phenotypically similar diseases.

As published knowledge and information increases rapidly, the scientific literature has become the richest resource for scientific research. In order to effectively utilize this resource, **Wei Li's group** (Institute of Genetics and Developmental Biology, CAS, China) presented a method for predicting human PPI using text mining with a naive Bayes model. Specifically, Bayes' theorem is used to combine evidence from the correlation of gene expression with prior knowledge of PPIs. They also validated some of the predicted PPIs through yeast two-hybrid assays, identifying a putative novel subunit of BLOC-1 complexes involved in vesicle trafficking.

Easy-to-use databases

Databases are a *sine qua non* of bioinformatics research. **Xiujie Wang's group** (Institute of Genetics and Developmental Biology, CAS, China) introduced their Gene On-

tology Enrichment Analysis Software Toolkit (GOEAST) (Zheng and Wang, 2008). The tool contains data from up to 60 species. Using this database, they analyze the gene functions of differentially expressed genes between broiler and layer chicken skeletal muscles during different developmental stages, and find that these differentially expressed genes are related to muscle development and metabolism, and that they are enriched with miRNA targets, with fewer SNPs within the miRNA binding sites.

Liping Wei's group (Peking University, China) introduced the KOBAS toolkit they developed for KEGG pathway identification (Wu et al., 2006). Using this database, they find 18 statistically significant upstream or downstream signaling pathways involved in drug addiction. Five of those pathways are consistently enriched for some addictive drugs. They also find some feedback loops in these pathways, which might be helpful to explain why addiction is irreversible (Li et al., 2008a).

In addition to talks on research projects, a workshop on scientific English writing skills was given by **Michael Cusick** (Dana-Farber Cancer Institute, USA). He summarized common mistakes made by native and non-native English speakers alike, and clarified many misconceptions about writing a good paper on systems biology, with a major focus on how to convey a precise meaning with the least number of words. Many vivid examples made the audience laugh unstoppably, and also made an unforgettable impression of how one can easily write something ridiculous by not paying proper attention to good writing style. Overall, the variety and high quality of the given presentations made the meeting an unforgettable event to everyone present, but equally valuable was the opportunity for direct interaction with some of the greatest minds and most active scientists in the field of developmental systems biology.

Acknowledgements

We thank the speakers at the meeting for careful reading and editing the report. We acknowledge supports from the China National Science Foundation (NSF) (No. 30588001, 30620120433), National Basic Research Program of China (No. 2006CB910700) and funds from the Chinese Academy of Sciences to J.-D.J.H. and NSF fellowship to T.A.

References

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell **129:** 823–837.

Bianco, A., Poukkula, M., Cliffe, A., Mathieu, J., Luque, C.M., Fulga, T.A., and Rorth, P. (2007). Two distinct modes of guidance signalling during collective migration of border cells. Nature **448:** 362–365.

Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., and Bork, P. (2008). Drug target identification using side-effect similarity. Science **321:** 263–266.

Chen, C.Y., Tsai, H.K., Hsu, C.M., May Chen, M.J., Hung, H.G., Huang, G.T., and Li, W.H. (2008). Discovering gapped binding sites of yeast transcription factors. Proc Natl. Acad. Sci. USA **105:** 2527–2532.

Dierick, H.A., and Greenspan, R.J. (2006). Molecular analysis of flies selected for aggressive behavior. Nat. Genet. **38:** 1023–1031.

Du, X., Wang, Z., Wu, A., Song, L., Cao, Y., Hang, H., and Jiang, T. (2008). Networks of genomic co-occurrence capture characteristics of human influenza A (H3N2) evolution. Genome Res. **18:**178–187.

Grigoryan, G., and Keating, A.E. (2006). Structure-based prediction of bZIP partnering specificity. J. Mol. Biol. **355:** 1125–1142.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. **39:** 311–318.

Karres, J.S., Hilgers, V., Carrera, I., Treisman, J., and Cohen, S.M. (2007). The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in *Drosophila*. Cell **131:** 136–145.

Kensche, P.R., van Noort, V., Dutilh, B.E., and Huynen, M.A. (2008). Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. J. R. Soc. Interface **5:** 151–170.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature **436:** 876–880.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell **128:** 1231–1245.

Kuhn, M., Campillos, M., Gonzalez, P., Jensen, L.J., and Bork, P. (2008). Large-scale prediction of drug-target relationships. FEBS Lett. **582:** 1283–1290.

Li, C.Y., Mao, X., and Wei, L. (2008a). Genes and (common) pathways underlying drug addiction. PLoS Comput. Biol. **4:** e2.

Li, Q.R., Carvunis, A.R., Yu, H., Han, J.D.J., Zhong, Q., Simonis, N., Tam, S., Hao, T., Klitgord, N.J., Dupuy, D., Mou, D., Wapinski, I., Regev, A., Hill, D.E., Cusick, M.E., and Vidal, M. (2008b). Revisiting the *Saccharomyces cerevisiae* predicted ORFeome. Genome Res. **18:** 1294–1303.

Lu, J., Fu, Y., Kumar, S., Shen, Y., Zeng, K., Xu, A., Carthew, R., and Wu, C.I. (2008a). Adaptive evolution of newly emerged micro-RNA genes in *Drosophila*. Mol. Biol. Evol. **25:** 929–938.

Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., and Wu, C.I. (2008b). The birth and death of microRNA genes in *Drosophila*. Nat. Genet. **40:** 351–355.

Lu, X., Lu, X., Wang, Z.C., Iglehart, J.D., Zhang, X., and Richardson, A.L. (2008c). Predicting features of breast cancer with gene expression patterns. Breast Cancer Res. Treat. **108:**191–201.

Newman, J.R., and Keating, A.E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science **300:** 2097–2101.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat. Genet. **39:** 730–732.

Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., and Kanehisa, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res. **36:** W423–W426.

Pujana, M.A., Han, J.D., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., Rennert, G., Moreno, V., Kirchhoff, T., Gold, B., Assmann, V., Elshamy, W.M., Rual, J.F., Levine, D., Rozek, L.S., Gelman, R.S., Gunsalus, K.C., Greenberg, R.A., Sobhian, B., Bertin, N., Venkatesan, K., Ayivi-Guedehoussou, N., Sole, X., Hernandez, P., Lazaro, C., Nathanson, K.L., Weber, B.L., Cusick, M.E., Hill, D.E., Offit, K., Livingston, D.M., Gruber, S.B., Parvin,

J.D., and Vidal, M. (2007). Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat. Genet. **39:** 1338–1349.

Roh, T.Y., Wei, G., Farrell, C.M., and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res. **17:** 74–81.

Rorth, P. (2007). Collective guidance of collective cell migration. Trends Cell Biol. **17:** 575–579.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell **132:** 887–898.

Suzanne Komili, N.G.F., Frederick P. Roth, and Pamela A. Silver. (2007). Functional specificity among ribosomal proteins regulates gene expression. Cell **131:** 557–571.

Tsai, H.K., Huang, G.T., Chou, M.Y., Lu, H.H., and Li, W.H. (2006). Method for identifying transcription factor binding sites in yeast. Bioinformatics **22:** 1675–1681.

van Swinderen, B., and Greenspan, R.J. (2005). Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. Genetics **169:** 2151–2163.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. **40:** 897–903.

Weber, K.E., Greenspan, R.J., Chicoine, D.R., Fiorentino, K., Thomas, M.H., and Knight, T.L. (2008). Microarray analysis of replicate populations selected against a wing-shape correlation in *Drosophila melanogaster*. Genetics **178:** 1093–1108.

Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res. **34:** W720–W724.

Wu, X., Jiang, R., Zhang, M.Q., and Li, S. (2008). Network-based global inference of human disease genes. Mol. Syst. Biol. **4:** 189.

Xie, D., Cai, J., Chia, N.-Y., Ng, H.H., and Zhong, S. (2008). Cross-species de novo identification of cis-regulatory modules with GibbsModule: Application to gene regulation in embryonic stem cells. Genome Res. **18:**1325–1335.

Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S., Zhang, Z., Hou, L., Zhang, Q., Zhang, Y., and Han, J.D. (2007). A modular network model of aging. Mol. Syst. Biol. **3:** 147.

Yang, K., Ma, W., Liang, H., Ouyang, Q., Tang, C., and Lai, L. (2007). Dynamic simulations on the arachidonic acid metabolic network. PLoS Comput. Biol. **3:** e55.

Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L., and Vidal, M. (2007). Drug-target network. Nat. Biotechnol. **25:** 1119-1126.

Yu, H., Zhu, S., Zhou, B., Xue, H., and Han, J.D. (2008). Inferring causal relationships among different histone modifications and gene expression. Genome Res. **18:** 1314–1324.

Zhang, X., Lu, X., Shi, Q., Xu, X.Q., Leung, H.C., Harris, L.N., Iglehart, J.D., Miron, A., Liu, J.S., and Wong, W.H. (2006). Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinformatics **7:** 197.

Zheng, Q., and Wang, X.J. (2008). GOEAST: A web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res. **36:** W358–W363.

Zhou, Y., Wang, R., Li, L., Xia, X., and Sun, Z. (2006). Inferring functional linkages between proteins from evolutionary scenarios. J. Mol. Biol. **359:** 1150–1159.

Zou, L., Sriswasdi, S., Ross, B., Missiuro, P.V., Liu, J., and Ge, H. (2008). Systematic analysis of pleiotropy in *C. elegans* early embryogenesis. PLoS Comput. Biol. **4:** e1000003.