Bioinformatics in China: A Personal Perspective

Liping Wei1*, Jun Yu2*

1 Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing, People’s Republic of China, 2 CAS Key Laboratory in Genome Sciences and Information, Beijing Institute for Genomics, Chinese Academy of Sciences, Beijing, People’s Republic of China

In this personal perspective, we recall the history of bioinformatics and computational biology in China, review current research and education, and discuss future prospects and challenges. The field of bioinformatics in China has grown significantly in the past decade despite a delayed and patchy start at the end of the 1980s by a few scientists from other disciplines, most noticeably physics and mathematics, where China’s traditional strength has been. In the late 1990s and early 2000s, rapid expansion of the field was fueled by the Internet boom and genomics boom worldwide and in China. Today bioinformatics research in China is characterized by a great variety of biological questions addressed and the close collaborative efforts between computational scientists and biologists, with a full spectrum of focuses ranging from database building and algorithm development to hypothesis generation and biological discoveries. Although challenges remain, the future of bioinformatics in China is promising thanks to advances in both computing infrastructure and experimental biology research, a steady increase of governmental funding, and most importantly a critical mass of bioinformatics scientists consisting of not only converts from other disciplines but also formally trained overseas returnees and a new generation of domestically trained bioinformatics Ph.D.s.

Introduction

The field of bioinformatics has enjoyed significant growth in China. A rough yet useful indicator of the field is the number of bioinformatics and computational biology publications from China indexed in PubMed at NCBI. As shown in Figure 1A, this number has been increasing significantly over the past decade. The number of all publications from China indexed in PubMed has also been increasing (Figure 1B), but if we plot the percentage of bioinformatics publications from China among all PubMed publications from China, we observe that the percentage itself has been increasing rapidly (Figure 1C), indicating that the contribution of bioinformatics research is becoming more significant within the life sciences in China. Comparing it with the situation worldwide, we observe that the number of bioinformatics publications from China is growing faster than the number of bioinformatics publications worldwide (Figure 1A versus 1D). Additionally, the number of PubMed publications from China has also been growing faster than the total number of PubMed publications (Figure 1B versus 1E). Furthermore, we observe that, very interestingly, for each year starting from 2003, the percentage of bioinformatics publications among all PubMed publications is greater in China than worldwide (Figure 1C versus 1F), indicating that bioinformatics as a field within the life sciences is enjoying a faster growth rate in China than elsewhere in the world.

These numbers and trends, although only rough estimates, are fascinating and deserve a closer look. In this article, we (see Box I for Authors’ Biographies) take a historical, as well as a horizontal, survey of bioinformatics in China. Due to the nature of personal perspectives and space limitations, what we present is by no means exhaustive, but merely suggestive.

The Early Years

China had a late and rather sluggish start in bioinformatics, largely due to lack of institutional support for bioinformatics and governmental funding during the early years. It was not until 1996 that the first Center for Bioinformatics in China was established within the College of Life Sciences at Peking University and the first government funding dedicated to bioinformatics was established as part of the National High Technology Development Program by the Ministry of Science and Technology (MOST). Despite the difficulties, starting from the end of the 1980s bioinformatics research was pioneered by a few scientists from other disciplines, most noticeably physics and mathematics where China’s traditional strength has been, applying theoretical frameworks and analytical tools from their original specialty to study biological questions.

A great example of these early scientists was Bailin Hao, who, trained in the former Soviet Union, was at the time already an accomplished theoretical physicist. Faced with computable and predictable characteristics of biological systems, he first studied protein structures and then devoted himself to the analysis of DNA and protein sequences. In visualizing very long DNA sequences, including the complete genomes of several bacteria and yeast and segments of human genome, his group observed fractal-like patterns [1]; subsequently they proposed the description of a genome using statistics of K-strings (by counting a series of K-tuples in a linear sequence of symbols), enabling a new paradigm for phylogenetic analysis without sequence alignments [2]. Chunting Zhang, another established physicist, also started his bioinformatics research at the end of the 1980s working on protein structure prediction [3]. His group later created the concept of Z-curve to display DNA composition dynamics in geometric spaces [4,5]. Z-curve has been successfully applied to genome sequence analyses such as prediction of coding genes in yeast genome [6], isochores in human genome [7], and replication origins in archaean genomes [8]. Runsheng Chen, with a background in biophysics, was involved in the early phases of Chinese genomics research and later pioneered small RNA...
Box 1. Authors’ Biographies

Liping Wei is Professor and Director of the Center for Bioinformatics at Peking University and Associate Director of the National Laboratory of Protein Engineering and Plant Genetic Engineering in China. She received her undergraduate training in Electrical Engineering and Information Sciences from the University of Science and Technology of China. She holds a Master’s degree in Applied Mathematics from Brown University and a PhD in Medical Informatics (now called Biomedical Informatics) from Stanford University. She worked in the biotech industry for four years, first at Exelixis, Inc., and then at Nexus Genomics, Inc., while continuing to serve as a Consulting Assistant Professor of bioinformatics in the Department of Medicine at Stanford University. She moved back to China in early 2004 to resume an academic career in bioinformatics at Peking University. Her current research interests include the regulation by, and of, noncoding RNAs and antisense transcripts and the signaling and regulatory networks underlying neurobiological disorders such as autism and addiction.

Jun Yu is a professor in the Beijing Institute of Genomics, Chinese Academy of Sciences (CAS). Dr. Yu obtained a B.S. degree in biochemistry from Jilin University in 1983 and a PhD in biomedical sciences from New York University School of Medicine in 1990. He joined the University of Washington Genome Center in 1993 and attuned his primary research interests toward genomics and bioinformatics. He started to work in China in 1998 and has led many major genome projects there, such as the International Human Genome Project (the Chinese effort), the Superhybrid Rice Genome Project, and the Silkworm Genome Project. His current research interests include comparative genome analysis, transcriptome modeling, human genetic diversity, and model organisms for phenotypic plasticity. He has published more than 120 scientific papers and won numerous academic awards, such as Scientific Leader of the Year, 2002 (Scientific American), 100-Talent Plan (CAS, 2002–2005), and China–US Biology Examination and Application (CUSBEA, 1983).

The Internet Boom and the Genomic Boom

Two significant developments in the second half of the 1990s proved critical to the expansion of bioinformatics in China. They were the Internet boom and the genomic boom. It was not until 1994 that full TCP/IP Internet connection between China and the rest of the world was established. At that time, the only established connection was an existing link, a dial-up X.25 connection, between the Institute of High-Energy Physics (IHEP) in Beijing and the Stanford Linear Accelerator Center (SLAC) at Stanford University. On May 17, 1994, the official connection to FTX-West was announced, and the U.S.-based Energy Sciences Network (ESnet) agreed to carry China IP traffic. Despite the late start, China caught up quickly with the worldwide Internet boom and established fast and broad Internet connections which were critical for biological data exchange.

At about the same time, the genomic boom in the late 1990s to early 2000s exemplified by the Human Genome Project [49] started to generate huge and exponentially growing amounts of data. After earlier efforts such as sequencing of full-length cDNAs [50], Chinese scientists officially participated in the Human Genome Project in 1999 and sequenced 1% of the human genome [9,49]. Since then the Chinese genome centers have independently sequenced several large genomes including those of rice (*Oryza sativa* L. ssp. indica) [51], silkworm (Bombyx mori) [52], and numerous microbial genomes such as that of *Thermoanaerobacter tengcongensis*, a rod-shaped, gram-negative, anaerobic eubacterium isolated from a freshwater hot spring in Tengchong, China [53].

Data is a major driving force for bioinformatics—it demands new methods for data storage and analyses and motivated the computational discovery of biological patterns. In all genomic projects, bioinformatics teams played significant roles in data analysis. Databases were created to store the genomic data, for instance BGI-RIS [54] and SilkDB [55] for rice and silkworm genomic sequence data and related biological information, and ChickVD [56] for chicken genomic sequence variations. Methods and software packages were developed to analyze the huge amount of genomic data, including for instance BGF, an ab initio gene prediction method developed for the rice genome based on Hidden Markov Model (HMM) and dynamic programming [57], a non-supervised gene prediction algorithm for bacterial and archaeal genomes [58], and a method for genome comparison using Gene Ontology (GO) with statistical testing [59].

Present Day: A Diversified Landscape

Since the end of the 1990s, bioinformatics in China has experienced significant growth fueled by not only the aforementioned Internet and genomic boom but also by official institutional support, steadily increasing government funding, and very importantly, a growing number of Chinese scientists returning to China after formal overseas training and research experience in bioinformatics and genomics. Today the field is characterized by a great variety of biological questions addressed and the close collaborations between computational scientists and bench biologists, with a full spectrum of focuses ranging from providing resources and services, developing databases and algorithms, to generating hypotheses and making biological discoveries.

Providing biological data resources and services is still an important part of bioinformatics. A good example is the Shanghai Center for Bioinformation Technology (http://www.scbit.org/), led by
Yixue Li and founded in 2002. It has served as the central repository for biological data generated throughout Shanghai and neighboring areas. It has also provided extensive public online bioinformatics resources. Furthermore, many new primary databases with raw data and value-added secondary databases were created by Chinese scientists (Table 1), and many locally developed methods were implemented and incorporated into Web servers (Table 2). These databases and tools address a broad range of biological questions and are open to the entire international scientific community, and are enjoyed by a large worldwide user base.

Independently or in collaboration with bench scientists, bioinformatics scientists have played significant roles in important biological discoveries. In 2003, China and many other countries suffered from the severe acute respiratory syndrome (SARS) epidemic. A team of bioinformaticians quickly joined the Chinese SARS Molecular Epidemiology Consortium, led by Guo-Ping Zhao, to sequence and analyze SARS coronavirus genomic sequences.
Table 1. Examples of biological databases developed and maintained in China.

| Category                  | Name             | Main content                                                                 | URL                               | Reference |
|---------------------------|------------------|------------------------------------------------------------------------------|-----------------------------------|-----------|
| Genome resources          | BGI-RIS          | An integrated information resource and comparative analysis workbench for rice genomics | http://rise.genomics.org.cn/     | [54]      |
|                           | SilkDB           | A knowledgebase for silkworm biology and genomics                            | http://silkworm.genomics.org.cn  | [55]      |
|                           | ChickVD          | A sequence variation database for the chicken genome                         | http://chicken.genomics.org.cn   | [56]      |
|                           | The Z curve database | A graphic representation of genome sequences                                | http://tubic.tju.edu.cn/zcurve/   | [148]     |
|                           | MED-Start        | Predicted translation initiation sites in microbial genomes                  | http://ctb.pku.edu.cn/MED_Start.htm | [100]     |
|                           | ProTISA          | A comprehensive resource for translation initiation site annotation in prokaryotic genomes | http://mech.ctb.pku.edu.cn/protisa | [149]     |
|                           | DoriC            | A database of oriC regions in bacterial genomes                              | http://tubic.tju.edu.cn/doric/    | [150]     |
| Transcription regulation resources | DRTF          | A database of rice transcription factors                                     | http://drtf.cbi.pku.edu.cn/      | [97]      |
|                           | DATF             | A database of Arabidopsis transcription factors                             | http://daf.cbi.pku.edu.cn        | [98]      |
|                           | NATsDB           | A database of natural antisense transcripts identified in ten genomes        | http://natsdb.cbi.pku.edu.cn/     | [151]     |
|                           | NONCODE          | An integrated knowledge database of noncoding RNAs                          | http://noncode.bioinfo.org.cn    | [13,152]  |
|                           | NPInter          | A database of noncoding RNA and protein interaction.                         | http://bioinfo.ibp.ac.cn/NPInter  | [14]      |
|                           | ATID             | A collection of publicly available alternative translational initiation events | http://bioinfo.au.tsinghua.edu.cn/atie/ | [153]     |
|                           | dbRES            | A database for annotated RNA editing sites                                   | http://bioinfo.au.tsinghua.edu.cn/dbRES/ | [154]     |
|                           | PASDB            | A collection of genes reported to be alternatively spliced in plants, spanning 44 plant species | http://pasdb.genomics.org.cn      | [155]     |
| Protein resources         | MPSS             | An integrated database of protein annotations                               | http://www.scbit.org/mpss/       | [156]     |
|                           | SPD              | A secreted protein database                                                  | http://spd.cbi.pku.edu.cn        | [157]     |
|                           | SynDB            | A synapse protein database based on a synapse ontology                       | http://syndb.cbi.pku.edu.cn      | [158]     |
|                           | DBSub-Loc        | Database of protein subcellular localizations                               | http://www.bioinfo.tsinghua.edu.cn/dbsubloc.html | [45]      |
|                           | dbNEI            | A database for neuro-endocrine-immune interaction.                          | http://bioinfo.au.tsinghua.edu.cn/dbNEIweb/ | [159]     |
|                           | SPIDer           | Saccharomyces protein–protein interaction database                          | http://cmb.bnu.edu.cn/SPIDer/index.html | [130]     |
|                           | InterDom         | A database of putative interacting protein domains for validating predicted protein interactions and complexes | http://InterDom.lit.org.sg       | [160]     |

doi:10.1371/journal.pcbi.1000020.t001

derived from the early, middle, and late phases of the SARS epidemic as well as viral sequences from palm civets. Their work uncovered the molecular evolution of the SARS coronavirus and the viral invasion from animal to human [60,61]. Another group proposed a mathematical model to estimate the evolution rate of the SARS coronavirus genome (0.16 base/day) and the time of the last common ancestor of the sequenced SARS strains (August or September of 2002) [62]. To identify potential anti-viral drug targets, proteomic and bioinformatic methods were used to investigate key SARS viral proteins including structural proteins [63], spike protein [64], and 3C-like proteinase [30]. Genomic packaging signals, which may be used to design antisense RNA and RNA interference (RNAi) drugs treating SARS, were predicted by comparative genomics methods [65]. The three-dimensional structure of 3C-like proteinase was constructed by homology modeling, based on which virtual screening of chemical databases was performed to search for potential inhibitors [31]. Chinese SARS patients were statistically studied to elucidate the association of symptom combinations with different outcome and therapeutic effects [66] and the association between mannose-binding lectin gene polymorphisms and susceptibility to SARS virus infection [67].

The large Chinese population provided invaluable resources for genetic studies. Patient and control groups were genotyped to establish associations between genetic variations and susceptibility to diseases such as esophageal squamous cell carcinoma [68], hepatocellular carcinoma in a particularly high-risk region of China [69], hypertension [70], and coronary heart disease [71]. A link between an Asian-enriched SNP in human sialidase and severe adverse reactions to the anti-viral drug Tamiflu (oseltamivir carboxylate) was proposed and confirmed by
bioinformatic methods and enzymatic assays [72]. A number of interesting works studied the migration history of Chinese and East Asian populations by the sequencing and phylogenetic analysis of Y-chromosomes [73–77]. Recently the Chinese HapMap Consortium participated in the international HapMap project to genotype the Chinese Han population [78]. As more and more genetic data were generated, new algorithms continued to be developed, for instance a Java-based method to analyze linkage disequilibrium [79], software for haplotype block partition and InSNPs selection [80], a method

| Table 2. Examples of Web servers and software packages developed and maintained in China. |
|---|---|---|---|---|
| Category | Name | Main Function | URL | Reference |
| Genome analysis tools | BGF | Prediction of genes in the rice genome | http://tlife.fudan.edu.cn/bgf/ | [57] |
| | CVTree | A phylogenetic tree reconstruction tool based on whole genome sequences without alignment | http://cvtree.cbi.pku.edu.cn | [161] |
| | ZCURVE | A system for recognizing protein coding genes in bacterial and archaeal genomes | http://tubic.tju.edu.cn/Zcurve_B/ | [5] |
| | Zplotter online | A program to draw and manipulate the Z curve online based on a user's input DNA sequence | http://tubic.tju.edu.cn/zcurve/ | [148] |
| | GS-Finder | A program to find bacterial gene start sites with a self-training method | http://tubic.tju.edu.cn/GS-Finder/ | [162] |
| | GC-Profile | A Web-based tool for visualizing and analyzing the variation of GC content in genomic sequences | http://tubic.tju.edu.cn/GC-Profile/ | [163] |
| | FGF | A Web tool for Fishing Gene Family in a whole genome database | http://fgf.genomics.org.cn/ | [164] |
| | BPhyOG | An interactive server for genome-wide inference of bacterial phylogenies based on overlapping genes | http://cmb.bnu.edu.cn/BPhyOG/ | [165] |
| | SAPRED | Using new structural and sequence attributes and Support Vector Machine to predict possible disease association of single amino acid polymorphism | http://sapred.cbi.pku.edu.cn/ | [82] |
| Expression regulation analysis tools | GBA | EST-based digital gene expression profiling | http://gba.cbi.pku.edu.cn | [166] |
| | CEAS | An online server to analyze transcription factor binding sites based on CHIP-chip data | http://ceas.cbi.pku.edu.cn | [99] |
| | OTFBS | Over-represented Transcription Factor Binding Site Prediction Tool | http://www.bioinfo.tsinghua.edu.cn/%7Ezhengjsh/OTFBS/index.html | [48] |
| | SVAP | Identification and expression analysis of alternatively spliced isoforms | http://svap.cbi.pku.edu.cn | |
| | CPC | Prediction of the protein-coding potential of transcripts using sequence features and support vector machine | http://cpc.cbi.pku.edu.cn | [107] |
| | RDfolder | A Web server for prediction of RNA secondary structure | http://rna.cbi.pku.edu.cn | [108] |
| | miRAS | A data processing system for miRNA expression profiling study | http://e-science.tsinghua.edu.cn/miras/ | [106] |
| | MiPred | Classification of real and pseudo microRNA precursors using random forest prediction model with combined features | http://www.bioinf.seu.edu.cn/miRNA/ | [103] |
| | RFRCD-siRNA | Improved design of siRNAs by random forest regression model coupled with database searching | http://www.bioinf.seu.edu.cn/siRNA/index.htm | [167] |
| Protein analysis tools | EasyGO | Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species | http://bioinformatics.caau.edu.cn/easygo/ | [168] |
| | CTKPred | An SVM-based method for the prediction and classification of the cytokine superfamily | http://www.bioinfo.tsinghua.edu.cn/%7Ehengl/CTKPred/index.html | [169] |
| | GNBSL | A new integrative system to predict the subcellular location for Gram-negative bacteria proteins | http://166.111.24.5/webtools/GNBSL/index.htm | [127] |
| | MeMo | A Web tool for prediction of protein methylation modifications | http://www.bioinfo.tsinghua.edu.cn/~tigerchen/MeMo/contact.html | [170] |
| | KOBAS | A Web-based platform for pathway identification | http://kobas.cbi.pku.edu.cn | [171] |
| | IntNetDB | An integrated protein–protein interaction network database generated by a probabilistic model | http://hanlab.genetics.ac.cn/IntNetDB.htm | [172] |
| Platforms | BOD | A customizable bioinformatics on demand system accommodating multiple steps and parallel tasks | http://e-science.tsinghua.edu.cn/bod/index.jsp | [173] |
| | ABCGrid | Application for Bioinformatics Computing Grid | http://abcgrid.cbi.pku.edu.cn | [174] |
to detect human recombination hotspots based on a multiple-hotspot model and an approximate log-likelihood ratio test [81], and an SVM-based method that used new sequence and structure features to predict the possible disease association of non-synonymous SNPs with increased accuracy [82].

The increasing number of completely and partially sequenced genomes have enabled the study of the evolution of genomes and gene families. Of particular interest to many Chinese scientists is the study of the evolution of plants, largely due to the country’s longstanding strength in plant biology and partly due to its dedication to the sequencing of the rice genome. Two groups had studied the collinearity of duplicated genes along rice chromosomes and discovered that the rice genome had undergone an ancient whole-genome duplication 70 million years ago and a recent segmental duplication involving Chromosomes 11 and 12 five million years ago [83,84]. Their findings settled previous disputes over whether rice was an ancient aneuploid or an ancient polyploid, bringing the two contradictory sides to agreement in the form of a graciously co-authored Commentary [85]. In addition to whole genomes, the evolution of many important plant gene families was also studied in detail, such as transcription factor families [86] and enzymes [87]. A group studied the origin of new genes in rice and found a surprisingly high number of retrogenes as well as chimeric genes originated by retroposition, suggesting that retroposition is an important mechanism that governs gene evolution in rice and other grass species [88]. Another group designed a theoretical model based on the molecular clock to test the hypotheses of hybridization as an evolutionary mechanism [89]. Several groups had proposed interesting theories about the genetic code (amino acid codons and stop codons) [36,37,40,90,91].

When and where genes in a genome are transcribed and translated and how this expression of genes and proteins is regulated are important biological questions that had interested numerous scientists for decades. A first step to study this important question is to measure and compare the expression level of mRNAs. Several Chinese groups had developed algorithms for microarray experiments including designing probes [92], processing raw data by an integrated pipeline [93], detecting differentially expressed genes by relative entropy [94], identifying clusters of co-expressed genes by Hidden Markov Models [47], and finding disease-related genes by an ensemble decision approach [95]. Transcription factors bind to upstream regions of genes to regulate their transcription. About 2,000 transcription factors (~10% of the genomes) were identified in each of the three sequenced plant genomes, Arabidopsis, rice, and populus, by combining similarity and motif-based approaches [96–98]. A new method was developed to identify potential transcription factors binding sites in the upstream regions of genes by searching for over-represented cis-elements with Position Weight Matrix-based similarity scores [48]. Another method uncovered the patterns of conservation, genomic distribution, and co-factor binding of transcription factor binding sites given ChIP-chip tiling array data [99]. A new four-component statistical model was proposed to improve accuracy of the identification of translation initiation sites in microbial genomes [100]. Alternative splicing adds great variety to the proteome. A new method, named Splicing Variant Analysis Platform (SVAP, http://svap.cbi.pku.edu.cn), could identify alternatively spliced isoforms from EST data ten times faster than other existing methods. Increased prediction accuracy of splice sites was achieved by using quadratic discriminant analysis with diversity measure or by introducing a competition mechanism of splice sites selection [22,101]. The impact of very short alternative splicing on protein structures and functions was studied [21]. An interesting work identified 2,695 newly evolved exons in rodents and calculated the new exon origination rate at about $2.71 \times 10^{-5}$ per gene per million years; they suggested that most new exons might originate through “exonization” of intronic sequences and appear to be alternative exons that are expressed at low levels [102].

More recently, several Chinese bioinformatics groups had made progress in studying the regulation of transcription and translation by microRNAs, noncoding RNAs, and natural antisense transcripts. New computational methods were developed to identify precursor and mature microRNAs by sequence features, hairpin structural features, and cross-species conservations [18,20,103,104]. MicroRNAs in Chlamydomonas reinhardtii, a unicellular green alga, were identified in reads obtained by the highly parallel 454 pyrosequencing technology; the results suggested that the miRNA pathway may be an ancient mechanism of gene regulation that evolved prior to the emergence of multicellularity [105]. To study the expression profile of microRNAs, two methods were developed using data from EST sequencing and SAGE-based total RNA clones, respectively [19,106]. Longer non-coding RNAs were also studied extensively. Hundreds of novel noncoding RNAs were identified in C. elegans first by a cloning strategy and then by a whole-genome tiling array technology [11,12]. The expression of intron-encoded noncoding RNAs was profiled using a custom-designed microarray combining noncoding RNAs and their host genes, and many noncoding RNAs were found to be independently transcribed with ncRNA-specific promoter elements [10]. To facilitate future studies, new databases were created to catalogue noncoding RNAs and noncoding RNA–protein interactions [13,14], and new algorithms were developed to accurately predict the protein-coding potential of a given transcript and the secondary structure of an RNA [107,108]. Many RNAs function through an antisense mechanism. Several Chinese groups have identified cis- and trans-natural antisense transcripts at the whole-genome scale and found that they are highly abundant and have interesting features of function, expression, and evolution [109–111].

Research at the protein level is gaining increasing momentum thanks in part to China’s recognition of protein sciences as one of the major research areas to promote and support in the next 20 years. Progress continued to be made in the prediction of protein secondary and tertiary structures [46,112–116]. An interesting new method achieved rapid multiple alignment of protein three-dimensional structures by the use of conformational letters, which were defined as discretized states of 3D segmental structural states [117]. Several effective scoring functions and algorithms were developed for the docking of protein with small compounds [28,29,33,34]. In the field of proteomics, China joined the Human Proteome Organization (HUPO) with a special focus on the study of the human liver proteome, led by Fuchu He [118,119]. The fetal and adult liver proteomes were extensively profiled using mass spectrometry technologies [120–122]. Driven by the need to identify peptides and proteins from mass spectrometry data, new algorithms were developed with increased accuracy [15,16,123–126]. Computational prediction of subcellular localizations of proteins is an important problem in proteomics, and several good methods had been developed in China using a variety of classification methods such as Support Vector Machine [44,45], fuzzy k-NN method [23], and an integrated meta-approach [127]. Another method predicted the protein submitochondria locations.
by hybridizing pseudo-amino acid composition with physicochemical features of segmented sequence [24]. Protein–protein interactions were under intense study. They were predicted from sequence features [120], Gene Ontology annotations [129,130], and integration of 27 heterogeneous genomic, proteomic, and functional annotation datasets [27]. The protein–protein interfaces were studied in detail, such as the potential of mean force used in the ranking of the binding energies of different protein–protein complexes [25] and the hydrogen bond, hydrophobic and vDW interactions used to estimate the individual contribution of each interfacial residue to the binding [26]. A recent paper reported the design of nonnatural protein–protein interaction pairs by key residues grafting [27]. The topology of the protein–protein interaction network in yeast was analyzed using a spectral method derived from graph theory to uncover hidden topological structures such as quasi-cliques and quasi-bipartites [131], and visualized as a clustering tree [17]. Metabolic pathways were identified using the KEGG Othology as an alternative controlled vocabulary [132].

Interesting progress had been made in the systems biology studies of biological networks. The global dynamical properties and stabilities of the networks had been studied. In particular, the yeast cell-cycle network and the Drosophila segment polarity network appeared to be highly robust [133,134]. The protein interaction networks of Saccharomyces cerevisiae, Caenorhabditis elegans, and Drosophila melanogaster were found to have a scale-free and high-degree clustering nature with a small-world property with similar diameter at 4–5 [135]. Key proteins, subgraphs, and modules were identified and analyzed in the signal transduction networks [136], transcriptional regulation networks in yeast [137], and protein–protein interaction networks during fruitfly and human brain aging [138,139]. Simulation analysis of the energy metabolism network in mammalian myocardia revealed that the systemic states of metabolic networks did not always remain optimal, but might become suboptimal when a transient perturbation occurs [140]. Finally, the dynamic properties of the arachidonic acid metabolic network which includes several targets for anti-inflammatory drugs was analyzed using ordinary differential equations, and the flux balance in the network was found to be important for efficient and safe drug design [141].

The bioinformatics discipline has grown too broad to be reviewed comprehensively in any one article. Many works cannot be covered here because of the space limitations. Just to give two examples of interesting areas of bioinformatics research that we did not adequately review here: first, a small group of researchers in China have made progress in computational neurobiology, including proposing a computational model as a neurodecoder based on synchronous oscillation in the visual cortex [142] and a simulation study on the Ca2+-independent but voltage-dependent excytosis and endocytosis in dorsal root ganglion neurons [143]; second, many groups have developed databases and methods of imaging, such as the digital image datasets of Chinese Visible Human (a male and a female) [144], a high-resolution anatomical rat atlas [145], and methods for optical molecular imaging [146].

Education: A New Generation of Locally Trained Bioinformatics

Bioinformatics talent in China is reaching a critical mass. In addition to local bioinformatics scientists who converted from other disciplines and scientists who returned to China after formal training overseas, two new talent pools are forming. First, more and more scientists from other countries, especially European countries, are drawn to work in China by the improved research and funding environment. A good example is the Max Planck–Chinese Academy of Science Partners Institute in Computational Biology in Shanghai which employs a number of European scientists and plays key roles in facilitating international collaborations.

Second, a growing significant talent pool that has emerged in the past few years is the new generation of young bioinformatics scientists trained in local bioinformatics degree programs. Take the Center for Bioinformatics at Peking University as an example. By summer 2007, it had graduated 24 PhDs and four Master’s degrees in bioinformatics. Although a slight majority of the graduates (15 out of 28, or 54%) still chose to pursue postdoctoral positions overseas, the remaining had decided to stay in China, with nine working in academia and four in industry. Many other Chinese universities are now offering bioinformatics degree programs at the PhD, Master’s, and even Bachelor’s levels (see list in Table 3). In addition to formal degree programs, many university courses and ad hoc workshops have been offered to train bench biologists on how to use bioinformatic databases and tools.

The Future: Promises and Challenges

Today bioinformatics research in China still lags behind the best in the world. However, it is catching up quickly, and several positive factors suggest a bright future. First, the aforementioned critical mass of bioinformaticians will be a driving force for the future development of bioinformatics in China. Second, life science research at large is gaining momentum in China and progressing quickly. Collaborating closely with talented local and international biologists, bioinformatics scientists will have access to new and numerous biological data to analyze, important biological questions to solve, and bench experimental laboratories capable of validating their predictions and hypotheses. Third, China’s total funding for scientific research has been steadily increasing in the past five years, reaching 1.40% of her GDP in 2006 and a total annual R&D expenditure of 300 billion Yuan (or U$38.5 billion) [147]. The recent national goal for R&D growth has been set to an incremental annual increase of 20%. Last but not least, China is investing heavily in its computing infrastructure, including the China Grid initiated by the Ministry of Education and the China National Grid (CNGrid) initiated by Chinese Academy of Sciences, both supported largely by the Ministry of Science and Technology. Bioinformatics scientists benefit as welcomed major users of these computing infrastructures.

Despite the promising future, current challenges remain. Scientists returning from overseas often face a salary reduction, reverse culture shock, and a different funding system and application process. China lacks the equivalent of the Biomedical Information Science and Technology Initiative Consortium (BISTIC) launched at the NIH in 2000. BISTIC consists of senior-level representatives from each of the NIH institutes and centers plus representatives of other federal agencies interested in bioinformatics. It plays important roles in coordinating sustained funding and other support for bioinformatics. No similar initiatives exist in China yet. There is also no official professional bioinformatics society in China. More noticeable perhaps is the lack of a Chinese national center for bioinformation, the equivalent of the US National Center for Biotechnology Information (NCBI) that is the central repository and information hub for biological data and tools. As China is generating a large and growing amount of biological data, a Chinese national center with a dedicated
budget and staff could more effectively collect and store the local data and exchange it with the international scientific community. Journals and other publications in Chinese could also be collected locally and shared internationally. A Chinese national center could also offer better and faster online bioinformatics service to the large user base in China and neighboring Euro-Asian countries.

## Concluding Remarks

Bioinformatics in China has come a long way since the early years. It has benefited from developments in the rest of the world and has made its own distinct contributions. As bioinformatics in China continues to grow, we expect to see an increasing exchange of data, tools, and talent between China and other countries.

We believe the future is bright for bioinformatics in China and worldwide.

## Acknowledgments

We thank Chuan-Yun Li and Xiaomo Li for help with compiling the references and tables. We thank Professors Chun-Ting Zhang, Ying Xu, Xiaole Liu, Heping Cheng, Manyuan Long, Louis Tao, Xiaocheng Gu, Jingchu Luo, and the editors for insightful suggestions.

### Table 3. Examples of bioinformatics training programs in China.

| City       | University/Academy                                      | Affiliation                                                                 | Degree     |
|------------|----------------------------------------------------------|----------------------------------------------------------------------------|------------|
| Beijing    | Peking University                                       | Center for Bioinformatics, College of Life Sciences; Center for Theoretical Biology | PhD        |
| Beijing    | Tsinghua University                                     | Department of Biological Sciences and Biotechnology, Institute of Bioinformatics; Department of Automation | PhD        |
| Beijing    | Chinese Academy of Sciences                             | Beijing Institute of Genomics; Center of Systems Biology, Institute of Biophysics; Center of Molecular Systems Biology, Institute of Genetics and Developmental Biology | PhD        |
| Beijing    | China Agricultural University                           | College of Biological Sciences                                             | PhD, Master's |
| Beijing    | China Pharmaceutical University                         | School of Life Science and Technology                                      | PhD, Master's |
| Beijing    | Beijing Normal University                               | College of Life Sciences, Laboratory of Computational Molecular Biology    | Master's   |
| Beijing    | The Academy of Military Medical Science                 | Institute of Basic Medical Science                                         | Master's   |
| Baoding    | Hebei University                                         | College of Life Science                                                    | Master's   |
| Chengdu    | Sichuan University                                      | School of Life Sciences                                                    | PhD, Master's |
| Chengdu    | University of Electronic Science and Technology of China | School of Life Science and Technology                                      | PhD, Master's |
| Chongqing  | Chongqing University of Post and Telecommunications     | School of Bioinformatics                                                   | Bachelor's |
| Guangzhou  | Sun Yat-Sen University                                   | School of Life Sciences                                                    | PhD, Master's |
| Hangzhou   | Zhejiang University                                     | James D. Watson Institute of Genome Sciences, College of Life Sciences     | PhD, Master's |
| Harbin     | Harbin Medical University                               | Department of Bioinformatics                                               | Master's   |
| Hefei      | University of Science and Technology of China            | School of Life Sciences                                                    | PhD, Master's |
| Kunming    | Yunnan University                                        | School of Life Sciences                                                    | PhD, Master's |
| Lanzhou    | Lanzhou University                                       | School of Life Sciences                                                    | PhD, Master's |
| Nanjing    | Nanjing University                                       | School of Life Science                                                    | PhD        |
| Nanjing    | Southeast University                                     | State Key Laboratory of Bioelectronics, School of Biological Science & Medical Engineering | PhD, Master's |
| Shanghai   | Fudan University                                         | Institute of Biodiversity Science, School of Life Sciences                 | PhD, Master's |
| Shanghai   | Shanghai Institute for Biological Sciences               | Key Laboratory of Systems Biology                                          | PhD        |
| Shanghai   | Shanghai Jiao Tong University                            | Department of Biomedical Engineering, College of Life Science and Biotechnology | PhD, Master's |
| Shanghai   | Shanghai University                                      | School of Life Sciences                                                    | Master's   |
| Shanghai   | East China Normal University                             | School of Life Sciences                                                    | Master's   |
| Shanghai   | Tongji University                                        | School of Life Science and Technology                                      | PhD, Master's, Bachelor's |
| Tianjin    | Nankai University                                        | College of Life Sciences                                                   | PhD, Master's |
| Tianjin    | Tianjin University                                       | Tianjin University Bioinformatics Centre, School of Science                 | PhD, Master's |
| Xiamen     | Xiamen University                                        | Department of Chemistry                                                    | Master's   |
| Wuhan      | Huazhong University of Science and Technology            | School of Life Science and Technology                                      | PhD, Master's |
| Yangling   | Northwest A&F university                                 | School of Life Sciences                                                    | PhD, Master's |
| Zibo       | Shandong University of Technology                        | School of Life Sciences                                                    | Master's   |

doi:10.1371/journal.pcbi.1000020.t003
References

1. Yu ZG, Anh V, Lau KS (2001) Measure representation and multifractal analysis of complete genomes. Phys Rev E Stat Nonlin Soft Matter Phys 64: 011903.

2. Qi J, Wang B, Hao BI (2004) Whole proteome prokaryote phylogeny without sequence alignment: a K-string comparison approach. J Mol Evol 59: 49–63.

3. Zhang CT, Chou KC (1992) An optimization approach to predicting protein structural class from amino acid composition. Protein Sci 1: 201–209.

4. Zhang R, Zhang CT (1994) Z-curves, an intuitive tool for visualizing and analyzing the DNA sequences. J Biomol Struct Dyn 11: 767–782.

5. Guo FB, Ou HY, Zhang CT (2005) ZCURVE: a new system for recognizing protein-coding genes in bacterial and archival genomes. Nucleic Acids Res 31: 1760–1769.

6. Zhang GT, Wang J (2000) Recognition of protein coding genes in the yeast genome at better than 95% accuracy based on the Z curve. Nucleic Acids Res 28: 2004–2014.

7. Caudin, N. G. R. (2003) An isochore map of the human genome based on the Z curve method. Genome 317: 127–135.

8. Zhang R, Zhang CT (2005) Identification of replication origins in archival genomes based on the Z-curve. ISME J 1: 353–366.

9. Munzy DM, Scherer SE, Kaul R, Wang J, YU, et al. (2006) The DNA sequence, annotation and analysis of human chromosome 3. Nature 440: 1199–1206.

10. He H, Cai L, Skogerbo G, Deng W, Liu T, et al. (2006) Profiling Caenorhabditis elegans non-coding RNA expression with a combined method in 5'-UTRs and 3'-UTRs. Nucleic Acids Res 34: 2976–2983.

11. Deng W, Zhu X, Skogerbo G, Zhao Y, Fu Z, et al. (2006) Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, CT. BMC Bioinformatics 7: ZCURVE: a new system for recognizing protein-coding genes in bacterial and archival genomes. Nucleic Acids Res 31: 1760–1769.

12. He H, Wang J, Liu T, Liu XS, Li T, et al. (2007) Identification of primary transcripts and expressions of mammal micromers from evolutionary stability. Biosystems 65: 23–33.

13. Hua S, Sun Z (2001) A novel predictor machine approach for protein subcellular localization prediction. Bioinformatics 17: 721–728.

14. Guo T, Hua S, Ji X, Sun Z (2008) DBSubLoc: database of protein subcellular localization. Nucleic Acids Res 36: D122–D124.

15. Hua S, Sun Z (2001) A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. J Mol Biol 308: 397–407.

16. Ji X, Li-Ling J, Sun Z (2003) Mining gene expression data using a novel approach based on hidden Markov models. FEBS Lett 542: 125–131.

17. Zheng J, Wu J, Sun Z (2003) An approach to identify over-represented cis-elements in related sequences. Nucleic Acids Res 31: 1995–2005.

18. Landers ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.

19. Zhang B (2004) Human genome research in China. J Mol Med 82: 214–222.

20. Yu J, Wong GK, Liu S, Wang J, Yang H (2007) A comprehensive crop genome research project: the Superhybrid Rice Project in China. Philos Trans R Soc Lond B Biol Sci 362: 1023–1034.

21. Xiao Q, Zhou Z, Lu C, Cheng D, Dai F, et al. (2004) A draft sequence of the genome of the domesticated silkworm (Bombyx mori). Science 306: 1937–1940.

22. Bao Q, Tian Y, Li W, Xu Z, Xuan Z, et al. (2002) A complete sequence of the T. tengcongensis genome. Genome Res 12: 689–700.

23. Zhao W, Wang J, He X, Huang X, Jiao Y, et al. (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomes. Nucleic Acids Res 32: D377–382.

24. Wang J, Xia Q, He X, Dai M, Ruan J, et al. (2005) SiDB: a knowledgebase for silkworm biology and genomics. Nucleic Acids Res 33: D399–402.

25. Wang J, He X, Ruan J, Dai M, Chen J, et al. (2005) ChickVD: a sequence variation database for the chicken genome. Nucleic Acids Res 33: D438–441.

26. Lu H, Liu J-S, Xu X, Ji J, Fang L, et al. (2006) Test data sets and evaluation of gene prediction programs on the rice genome. Journal of Computer Science and Technology 20: 446–453.

27. Zhu H, Hu GQ, Yang YF, Wang J, She ZS (2007) MED: a new non-supervised gene prediction algorithm for bacterial and archival genomes. BMC Bioinformatics 8: 97.

28. Cai Z, Mao X, Li W, Liu X (2006) Genome comparison using GeneGO with statistical testing. BMC Bioinformatics 7: 374.

29. Chinese S-MEC (2004) Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 305: 1666–1669.

30. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, et al. (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc Natl Acad Sci U S A 102: 2430–2435.

31. Lu H, Zhao Y, Zhang J, Wang Y, Li W, et al. (2004) Date of origin of SARS coronavirus strains. BMC Infect Dis 4: 3.

32. Yang W, Hao Y, Zhang Y, Peng W, Qin J, et al. (2004) Proteinomic analysis on structural proteins.
of Severe Acute Respiratory Syndrome coronavirus. Proteomics 4: 492–504.
64. Yu XJ, Luo C, Liu JC, Hao P, He YY, et al. (2003) Putative iAPN receptor binding sites in SARS CoV spike protein. Acta Pharmacol Sin 24: 481–488.
65. Qin L, Xiong B, Luo C, Guo ZM, Hao P, et al. (2003) Identification of probable genomic packaging signals from SARS-CoV genome by bioinformatics analysis. Acta Pharmacol Sin 24: 493–496.
66. Li S, Wang R, Zhang Y, Zhang X, Layon AJ, et al. (2007) Combination of evolutionary and functional analysis to identify disease-gene candidates. Bioinformatics 23: 1444–1450.
67. Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165: 666–671.
68. Yu J, Wang J, Liu N, Li S, Li H, et al. (2005) The Genomes of Osarva sativa: a history of duplications. PLoS Biol 3 e38.
69. Paterson AH, Van de Peer Y, Vandepoele K (2003) Ancient duplication of cereal genomes. New Phytol 163: 638–661.
70. Xiong Y, Liu T, Tian C, Sun S, Li J, et al. (2005) Transcription factors in rice: a genome-wide conserved DNA analysis between monocots and eudicots. Plant Mol Biol 59: 191–203.
71. Yang J, Gu H, Yang Z (2004) Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories (Ipsomea). J Mol Evol 58: 54–63.
72. Wang W, Zheng H, Fan C, Li J, Shi J, et al. (2006) High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell 18: 1791–1802.
73. Song T, Zhong Y (2000) Testing hybridization hypotheses based on incongruent gene trees. Syst Biol 49: 413–425.
74. Zhang C-T, Zhang R (1991) Analysis of distribution of bases in the coding sequences by a digrammatic technique. Nucleic Acids Research 19: 6313–6317.
75. Sun J, Chen M, Xu J, Luo J (2003) Relationships among stop codon usage bias, its context, isochores, and gene expression level in eukaryotic organisms. J Mol Evol 44: 156–167.
76. Li Y, Wang X (2006) Mprob 2.0: computer-aided probe design for oligonucleotide microarray. Appl Bioinformatics 5: 181–186.
77. Wang X, He H, Li C, Chen R, Deng XW, et al. (2006) NCAA: a NimhGene microarray data processing pipeline. Bioinformatics 22: 2955–2957.
78. Yang X, Deng M, Fang WK, Qian M (2005) Detection of differentially expressed genes by relative entropy. J Theor Biol 234: 395–402.
79. Li X, Rao S, Wang Y, Gong B (2004) Gene mining: a novel and powerful ensemble decision approach for identifying disease genes using microarray expression profiling. Nucleic Acids Res 32: 2683–2694.
80. Zhu QH, Guo AV, Gao G, Zhong YF, Xu M, et al. (2007) miRAS: a novel database of plant microRNA sequence attributes to predict possible disease association of single amino acid polymorphism in Arabidopsis. Acta Pharmacol Sin 28: 179–185.
81. Guo A, He K, Liu D, Bai S, Gu X, et al. (2005) pFind: a novel database-searching software for protein identification via tandem mass spectrometry. Bioinformatics 21: 3049–3050.
82. Zhao T, Li G, Mi S, Li S, Hannon GJ, et al. (2007) A complex system of small RNAs in the unicellular green algae Chlamydomonas reinhardtii. Genes Dev 21: 1190–1203.
83. Tian F, Zhang H, Zhang X, Song C, Xia Y, et al. (2007) miRAN: a data processing system for microRNA expression profiling study. BMC Bioinformatics 8: 285.
84. Kong L, Zhang Y, Ye ZQ, Liu QX, Zhao SQ, et al. (2007) CPCR: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35: W345–349.
85. Ying X, Luo H, Luo J, Li W (2004) RDiFolder: a web server for prediction of RNA secondary structure. Nucleic Acids Res 32: W130–153.
86. Zhang Y, Liu XS, Liu QR, Wei L (2006) Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res 34: 3463–3475.
87. Li YY, Qin L, Guo ZM, Liu L, Xu H, et al. (2006) In silico discovery of human natural antisense transcripts. BMC Bioinformatics 7: 10.
88. Wang H, Chau GI, Wu JX (2006) Prediction of trans-antisense transcripts in Arabidopsis thaliana. Genome Biol 7: R92.
89. Jiang F (2003) Prediction of protein secondary structure with a reliability score estimated by local sequence clustering. Protein Eng 16: 651–657.
90. Guo J, Chen H, Sun Z, Lin Y (2004) A novel method for protein secondary structure prediction using dual-layer SVM and profiles. Proteins 54: 738–743.
91. Liu X, Zhang LM, Zheng WM (2004) Prediction of protein secondary structure based on residue pairs. J Bioinform Comput Biol 2: 345–352.
92. Liu X, Zheng WM (2006) An amino acid substitution matrix for protein conformation identification. J Bioinform Comput Biol 4: 769–782.
93. Jina L, Fang W, Tanga H (2002) Prediction of protein structural classes by a new measure of information discrepancy. Journal of Experimental Biology and Chemistry 27: 373–380.
94. Liu X, Zhao YP, Zheng WM (2007) CLEM-APS: Multiple alignment of protein structures based on similarity of contact map fractionation and multiple protein separation and identification technology. Mol Cells 25: 1705–1707.
95. Zhang X, Guo Y, Song Y, Sun W, Yu C, et al. (2006) Proteomic analysis of individual variation in normal livers of human beings using differential gel electrophoresis. Proteomics 6: 5260–5268.
96. Chen M, Ying W, Song Y, Liu X, Yang B, et al. (2007) Analysis of human liver proteins using replicate shotgun strategy. Proteomics 7: 2479–2488.
97. Li D, Fu Y, Sun R, Ling CX, Wei Y, et al. (2005) pPhad: a novel database and software for automated peptide and protein identification via tandem mass spectrometry. Bioinformatics 21: 3019–3026.
98. Zhang J, Gao W, Cai J, He S, Zeng R, et al. (2005) Predicting molecular formulas of fragment ions with isotope patterns in tandem mass spectra. IEEE/ACM Trans Comput Biol Bioinformatics 2: 217–229.
99. Xiao X, Wu S, Wang Z, Zhu Y, He F (2006) Protein probabilities in shotgun proteomics: evaluating different estimation methods using a
139. Xia K, Xue H, Zhu Y, He F (2007) A new strategy to filter out false positive identifications of peptides in SEQUEST database search results. Proteomics 7: 4036–4044.

140. Guo J, Lin Y, Liu X (2006) GNISBL: a new integrative system to predict the subcellular location for Gram-negative bacteria protein. Proteomics 6: 5099–5105.

141. Shen J, Zhang J, Luo X, Zhu W, Yu K, et al. (2007) Predicting protein-protein interactions based only on sequences information. Proc Natl Acad Sci U S A 104: 4337–4341.

142. Wu X, Zha L, Guo J, Zhang D, Lin K (2006) Prediction of yeast protein-protein interaction: insights from the Gene Ontology and annotations. Nucleic Acids Res 34: 2137–2150.

143. Wu X, Zha L, Guo J, Fu C, Zhou H, et al. (2006) SPIDER: Saccharomyces protein-protein interaction database. BMC Bioinformatics 7 Suppl 5: S16.

144. Bu D, Zhao Y, Cai L, Xue H, Zhu X, et al. (2003) Topological structure analysis of the protein-protein interaction network in budding yeast. Nucleic Acids Res 31: 2483–2490.

145. Mao X, Cai T, Olyarchuk JG, Wei L (2005) Dynamic analysis of optimality in myogenesis of Drosophila melanogaster: large-scale organization and robustness. Proteins 6: 456–461.

146. Liu W, Li D, Zhang J, Zhu Y, He F (2006) NichFlex: a novel network feature to evaluate the importance of proteins in signal transduction networks. BMC Bioinformatics 7: 515.

147. Zhang Z, Liu C, Skogerbo G, Zhu X, Lu H, et al. (2006) Dynamic changes in subgraph preference profiles of crucial transcription factors. PLoS Comput Biol 2: e47.

148. Xue H, Xian B, Dong D, Xia K, Zhu S, et al. (2007) A modular network model of aging. Mol Syst Biol 3: 147.

149. Xia K, Xue H, Dong D, Zhu S, Wang J, et al. (2006) Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. PLoS Comput Biol 2: e145.

150. Luo KY, Liao S, Tao GY, Li YY, Zeng S, et al. (2006) Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. Mol Syst Biol 2: 20060031.