Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Biologies 326 (2003) 949–957
COMPTES RENDUS BIOLOGIES

Molecular biology and genetics

Transcriptome study in China

Ze-Guang Han [a], Guo-Ping Zhao [a,b,c,*], Zhu Chen [a,d,*]

[a] *Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Zhangjiang High-Tech Park, Pudong, Shanghai 201203, China*
[b] *Shanghai Engineering Center for Biochips. 684 Song Tao Road, Zhangjiang High-Tech Park, Pudong, Shanghai 201203, China*
[c] *Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China*
[d] *Shanghai Institute of Hematology, Rui Jin Hospital, 197 Rui Jin Road II, Shanghai 200025, China*

Received 16 September 2003; accepted 23 September 2003

Presented by François Gros

Abstract

The Chinese genome project was initiated in 1993 with the goal of contributing 1% to the Human Genome Program. The study of gene expression profiles with cDNA microarrays, and large-scale sequencing and analysis of 130 928 expressed sequence tags (ESTs), allowed isolation and characterization of over 1000 novel full-length human cDNAs derived from human hematopoietic stem/progenitor cells, neuroendocrine tissues, liver, and cardiovascular cells. In addition, EST sequencing for model organisms, including rat, zebrafish, *Schistosoma japonicum* and rice was performed, aiming at identifying genes associated with physiological and/or pathological characteristics. ***To cite this article: Z.-G. Han et al., C. R. Biologies 326 (2003).***

Résumé

Étude du transcriptome en Chine. Le projet Génome chinois a débuté en 1993 avec l'objectif de contribuer à 1 % du Projet Génome Humain. L'étude de profils d'expression utilisant des microréseaux d'ADNc, et le séquençage et l'analyse à grande échelle de 130 928 étiquettes d'ADNc, ont permis d'isoler et de caractériser plus de 1000 nouveaux ADNc complets humains de cellules souches progénétrices hématopoïétiques et de tissus neuroendocriniens, hépatiques et cardio-vasculaires. De plus, des étiquettes d'ADNc ont été obtenues pour des organismes modèles comme le rat, le poisson zèbre, *Schistosoma japonicum* et le riz, dans le but d'identifier les gènes associés à des caractères physiologiques et/ou pathologiques. ***Pour citer cet article : Z.-G. Han et al., C. R. Biologies 326 (2003).***

Keywords: expressed sequence tags; full-length cDNA; Human Genome Project; model organism

Mots-clés : ADNc complet ; étiquette d'ADNc ; organisme modèle ; projet Génome humain

1. General situation

China initiated its human genome project (HGP) in 1993 with a clear long-term vision and feasible

* Corresponding author.
E-mail addresses: gpzhao@sibs.ac.cn (G.-P. Zhao), zchen@stn.sh.cn (Z. Chen).

1631-0691/$ – see front matter
doi:10.1016/j.crvi.2003.09.021

short-term goals. There were two lines of general rationale for the Chinese HGP. First, the highly diversified but relatively isolated large Chinese population (22% of the world's total) is a precious genetic resource for studying the human genetic diversity and disease genes through linkage analysis and association study in large number of patients and families. Second, the development of medicine, agriculture, biotechnology and pharmaceutical industry in China needs information generated not only from abroad, but also from its own HGP.

For medical genomics, it was the broader and deeper awareness about the importance of genetic material towards the research that significantly facilitated the establishment and expanding of a nationwide genetic material collection and conservation network. This included the collection of DNA samples and establishment of immortalized cell lines. Meanwhile, high-throughput technical platforms used for genomics research, including whole genome genotyping, fine mapping of genetic loci, cDNA cloning, mutation screening, and large-scale DNA sequencing, were successfully established in some centers of excellence. For the first phase of the medical genomics research, dozens of projects were carried out emphasizing the identification of genes related to important biological functions or diseases either inherited or multifactorial in nature. From 1994 to this date, many disease-related genes were isolated and identified by Chinese scientists.

For genomic information, Chinese scientists proposed a 'two-one percentage' strategy to contribute to the international HGP, i.e., large-scale sequencing for 30 Mb of human genomic DNA (1% of the human genome) and cloning of 500–1000 full-length human cDNAs for previously undefined genes (1% of the human genes according to the estimation of gene numbers in early 1990s) based on gene expression profile analysis. In addition, functional genomics projects including transcriptomics, proteomics, structural genomics, bioinformatics and research of model organisms were gradually initiated. The long-term goal of these new initiatives is, on one hand to develop the discipline of systems biology/biomedicine focusing on resolving disease mechanisms, and on the other, to facilitate research of structure-function relationship of biomolecules aiming at drug target or candidate discovery.

Table 1
Number of human ESTs and novel full-length cDNA from China*

Tissues/cells	Number of ESTs	Number of full-length cDNA
Hematopoietic and dendritic cells	39 048	550
Liver	48 108	620
Hypothalamus-pituitary-adrenal axis	25 972	300
Neural development	500	40
Cardiovascular	5300	192
Signal transduction		200
Others	12 000	
Total	130 928 (2.78%)#	>1000

* Summary in 31 December 2000;
# Percentage of ESTs contributed by Chinese scientists to dbEST in GenBank based on statistics on 27 September 2002.

With regard to the study of transcriptome, effective and relatively reliable technology platforms were established in some laboratories including cDNA library construction, expressed sequence tag (EST) sequencing and cDNA microarray analysis. A few cDNA libraries derived from human tissues and cells have been constructed for EST sequencing using oligo (dT)-primed and directionally cloned procedure or for full-length cDNA cloning employing the CapFinder PCR-based protocol. Since 1999, cDNA microarray technology based on nylon membrane or glass slides has been developed in some laboratories and biotech companies. Those platforms have been applied successfully in studying the gene expression profiles and identifying biomarkers associated with diseases or physiological characteristics of human tissues and cell types, particularly for those functionally important but poorly represented in the public databases. Recently, rat, zebrafish and *Schistosoma japonicum* as model systems were also studied at the transcriptome level, employing a similar strategy, which may turn out to have significant contributions in the development of comparative genomics.

Isolation of novel human full-length cDNA was one of the major goals of the Chinese HGP. Current statistics indicate that Chinese scientists have contributed 130 928 human ESTs and over 1000 novel full-length cDNAs to the public database, numbers that exceed the originally anticipated number of 500–1000 new human genes from different tissues and cells by EST

sequencing (Table 1). The ESTs and cDNA data derived from hematopoietic tissues/cells (CD34-positive cells), immune cells (dendritic cells), and liver (fetal liver, adult liver and liver cancer) were deposited in DDBJ/GenBank database and open to the general public.

2. Transcriptome analysis for physiological characteristics

2.1. Primary gene expression profile of hematopoiesis

Transcriptome analysis of hematopoiesis was the very first case of success in that direction conducted in China. It reflected to a certain extent the aggregated solid foundation in China in areas of experimental and clinical hematology over the years, albeit the transcriptome strategies were applied only in a limited field associated with normal physiology.

Hematopoietic stem/progenitor cells (HSPCs) possess the potentials of self-renewal, proliferation, and differentiation toward different lineages of blood cells. These cells not only play a primordial role in hematopoietic development but also have important clinical applications. Characterization of the gene expression profile in CD34-positive HSPCs may lead to a better understanding of the regulation of normal and pathological hematopoiesis. Zhu Chen and his colleagues in the Shanghai Institute of Hematology at Ruijin Hospital chose specific CD34-positive cells as a model to initiate the first transcriptome project in China. Since 1996, gene expression profiles of CD34-positive cells derived from human umbilical cord blood, bone marrow and leukemia have been established by the EST sequencing strategy.

Among 9866 ESTs obtained from umbilical cord blood, 4697 (47.6%) showed identity to known genes in the GenBank database, 2603 (26.4%) matched to the ESTs previously deposited in a public domain database, while 1415 (14.3%) were previously undescribed ESTs at that time. Integration of the data of ESTs of known genes generated a catalog including 855 genes that could be divided into different categories according to their functions. Some (8.2%) of the genes in this catalog were considered related to early hematopoiesis [1]. Similarly, 3424 meaningful ESTs obtained from healthy adult bone marrow were integrated into 1630 clusters, representing 622 known genes, 522 dbEST entries and 486 novel sequences [2]. Although the CD34-positive cells of this study were obtained from different tissues, the two gene expression profiles of cord blood and adult bone marrow share about 60% genes and ESTs. However, CD34-positive cells from cord blood have more copies of ESTs corresponding to hemoglobin gamma and HLA class II, less ESTs representing thymosin beta-4 and interleukin-8 than that of adult bone marrow, which might reflect the hematopoietic shift between pre- and post-natal processes. In addition, 1985 EST clusters were produced based on 4321 ESTs of CD34-positive cells derived from patients with acute myeloid leukemia, which included 711 known genes, 657 known ESTs and 617 novel sequences. 58 genes showed statistically significant differences in EST frequency hit by comparing EST datasets between bone marrow and leukemia ($P < 0.05$). The distinct gene expression patterns in leukemia cells as compared to normal control cells may contribute to the development and/or maintenance of the malignant phenotypes of leukemia cells [2].

Importantly, nearly 60% of the cDNA clones of mRNA under 2 kb in HSPCs cDNA libraries had 5′ ends upstream of the first ATG codon of the open reading frame (ORF). It implied that the cDNA libraries were enriched for full-length cDNAs and might be used to develop an efficient working system for full-length cDNA isolation. Subsequently, 300 cDNAs containing putative entire ORFs for previously undefined genes were obtained from CD34+ hematopoietic stem/progenitor cells, based on EST cataloging, clone sequencing, *in silico* cloning, and rapid amplification of cDNA ends (RACE). The sizes of these cDNAs range from 360 to 3496 bp and the corresponding ORFs may encode peptides of 58–752 amino acid residues. Public database search indicated that 225 cDNAs exhibited sequence similarities to genes identified across a variety of species. Homology analysis led to the recognition of 50 basic structural motifs/domains among these cDNAs. Genomic exon-intron organization could be established in 243 genes by integration of cDNA data with genome sequence information. Chromosomal localizations were obtained using electronic mapping for 192 genes and with radiation hybrid (RH) for 38 genes. cDNA microarray technique was ap-

plied to screen the gene expression patterns in five hematopoietic cell lines (NB4, HL60, U937, K562, and Jurkat) and a number of genes with differential expression were found. The resource work has provided a wide range of information useful not only for expression genomics and annotation of genomic DNA sequence, but also for further research on the function of genes involved in hematopoietic development and differentiation [3].

Fetal liver consists of hepatic cells and hematopoietic stem/progenitor cells. Human fetal liver aged 22 weeks of gestation corresponds to the turning point between immigration and emigration of the hematopoietic system. To gain further molecular insight into its developmental and functional characteristics, Fuchu He and his colleagues of Beijing Institute of Radiation Medicine in Beijing studied the gene expression profile of the 22-week fetal liver by generating ESTs and analyzing the compiled expression profiles of liver at distinct developmental stages. Among the available 13 077 ESTs, 5819 (44.5%) matched to known genes, and 5460 (41.8%) exhibited no significant homology to known genes. Integration of ESTs of known human genes generated an expression profile including 1660 genes that could be divided into different gene categories according to their functions. Liver-specific genes and ESTs associated with hematopoiesis were highly expressed, reflecting the unique physiological characteristics of fetal liver. By comparing the expression profiles, six gene groups that were associated with development of liver, tumorigenesis, physiological functions of Itoh cells against the other types of hepatic cells, and fetal hematopoiesis were identified. Meanwhile, cloning 110 full-length cDNAs of novel genes in that work contributed to our understanding of the unique functional characteristics of the human fetal liver [4].

Zebrafish is a useful model for studying on embryogenesis and development, in particular hematopoiesis. Recently, Zhu Chen and his colleagues started to catalog gene expression by EST sequencing from hematopoietic tissue, i.e. kidney marrow of zebrafish. A total of 20 672 ESTs, including 2150 3′-ESTs, were generated and integrated into 13 555 clusters, of which 11 757 ESTs matched with the public zebrafish EST database while 8915 (5617 clusters) are novel ESTs. Interestingly, 463 EST sets have homology with entries in OMIM database at amino acids level (> 50 amino acid residues, $> 45\%$ Identity) [5].

In conclusion, a systematical survey of hematopoiesis gene expression profile in human fetal liver, umbilical cord blood, adult bone marrow, and even zebrafish for the study of normal cell physiology and leukemia pathology represents a meaningful contribution to the progress of transcriptomics. The EST data and full-length cDNA isolated from these studies laid down a solid foundation for further researches, in both functional genomics and hematology. These works, particularly the transcriptomics of CD34-positive cells, is a pioneering milestone of transcriptome study in China. It is also served as a reference in the following studies on HSPCs worldwide [6–8].

2.2. Primary gene expression profile of neuroendocrine systems

Hypothalamus and pituitary, together with adrenals, constitute the major neuroendocrine axis (H-P-A axis) responsible for the maintenance of homeostasis and the response to the perturbations in the environment. Potential new hormones or novel functions of the axis as well as the molecular mechanisms of the regulatory networks need to be uncovered by new strategies. In 2000, Zeguang Han and Jialun Chen and their colleagues at the Chinese Human Genome Center at Shanghai (CHGCS) and Shanghai Institute of Endocrinology established a primary gene expression profile for the human H-P-A axis by generating a large number of ESTs, followed by bioinformatics analysis. Totally, 20 626 EST sequences of high quality were obtained from cDNA libraries of the hypothalamus, pituitary, and adrenal glands. These ESTs could be assembled into 9175 clusters (3979, 3074, and 4116 clusters in hypothalamus, pituitary, and adrenal glands, respectively) when overlapping ESTs were integrated. Of these clusters, 2777 (30.3%) correspond to known genes, 4165 (44.8%) to registered ESTs, and 2233 (24.3%) to novel ESTs. The gene expression profiles reflect well the functional characteristics of the three levels in the hypothalamus-pituitary-adrenal axis, because most of the 20 genes with highest expression showed statistical difference in terms of tissue distribution, including a group of tissue-specific functional markers. The order of gene expression levels of the 5 classical hormones in pitu-

itary coincides with their protein levels in the serum, while all enzymes associated with biosynthetic pathway of corticosteroids are encountered in EST data of adrenal gland. Some findings were made with regard to the physiology of the axis. Except for producing the known hormones, hypothalamus and pituitary could secrete many other cytokines that may play important role in interface between neuroendocrine and immune systems. In addition to the gene expression profile, this work contributed 200 full-length cDNAs of novel genes for which comparative genomics analysis revealed interesting features [9].

To this date, these data, which constituted the majority of the ESTs from human H-P-A axis in the public database, is still a unique resource for hormonal genomics approach. The data may facilitate the isolation of novel genes involved in hormonal signaling as well as their characterization, such as chromosomal location, polymorphism, splicing variants, differential expression, and physiological functions. These data may also enhance our insights into the complexity of intercellular communications, and understand the physiology and pathology of the human neuroendocrine system [10].

Further studies with regard to the gene expression profiles in the axis are underway. That includes: EST sequencing of 17 000 cDNA clusters after fingerprinting 54 000 clones from the pituitary library; isolation of more previously undefined full-length cDNAs; identifying regulatory/response proteins by functional genomics and systems biology strategies; medical application of the knowledge and platform of transcriptome analysis, such as analyzing the cell and tissue effect of Traditional Chinese Medicine (TCM) at the level of transcriptome.

3. Transcriptome analysis for pathological characteristics and medicine

Except for addressing the primary gene expression profiles of some biological systems under normal physiological conditions, comparative transcriptomes were established aiming at elucidating the molecular mechanism associated with pathogenesis and therapy.

3.1. Transcriptome analysis of all-trans-retinoic acid (ATRA)-induced differentiation of acute promyelocytic leukemia (APL) cells

Zhenyi Wang and his colleagues in the Shanghai Institute of Hematology were the first to successfully treat patients with acute promyelocytic leukemia (APL) using all-trans-retinoic acid (ATRA) [11]. The revolutionary strategy of cancer differentiation therapy inspired numerous scientists to address the molecular mechanisms of ATRA treatment. In an attempt to identify the events downstream of the targeting of aberrant RA receptor by ligand, Zhu Chen and his co-workers compared the gene expression patterns in the APL cell line NB4 before and after ATRA treatment using cDNA array, suppression-subtractive hybridization, and differential display-polymerase chain reaction (DD-PCR). Among 169 genes modulated by ATRA estimated to represent 1–2% of all genes expressed in APL cells, 100 genes, including 8 novel ones, were up-regulated while 69 ones were down-regulated by ATRA. The ATRA-induced gene expression profiles were in high accord with the differentiation and proliferation status of the NB4 cells. The time courses of their modulation were interesting. The 100 up-regulated genes occurred most frequently 12–48 hours after ATRA treatment, while 59 of 69 down-regulated genes had their expression suppressed within 8 h. The transcriptional regulation of 8 induced and 24 repressed genes was not blocked by cycloheximide, which suggests that these genes may be direct targets of the ATRA signaling pathway. A balanced functional network seemed to emerge, forming the foundation of decreased cellular proliferation, maintenance of cell viability, increased protein modulation, and promotion of granulocytic maturation. Several cytosolic signaling pathways, including JAKs/STAT and MAPK, may also be implicated in the symphony of differentiation [12].

This large-scale transcriptome analysis for APL cells upon treatment with ATRA provided the first global illustration on the molecular characteristics of ATRA-induced differentiation. A model of gene regulatory network modulation during this process could be obtained and precious datasets were developed for further studies in order to ultimately clarify the mechanism of both APL leukemogenesis and the differentiation therapy. This kind of knowledge will certainly

accelerate the identification of drug targets for novel drug discovery.

3.2. Transcriptome analysis of human hepatocellular carcinoma vs. corresponding non-cancerous liver

Human hepatocellular carcinoma (HCC) is one of the most common cancers worldwide, particularly in China. Hepatitis viruses (HBV and HCV) infection is the major factor contributing to hepatocarcinogenesis in China. Zeguang Han and his colleagues conducted a comprehensive characterization of gene expression profiles for hepatitis B virus-positive HCC. A large set of 5′-read EST clusters (11 065 in total) from HCC and non-cancerous liver samples were established, and then a homemade membrane based cDNA microarray system containing 12 393 genes/ESTs was applied for further analysis. The Atlas^{TM} human cancer 1.2 array cDNA microarray from Clontech Inc., which contains 1176 known genes related to oncogenesis, was also used for profiling gene expression. Integrated data from the above approaches identified 2253 genes/ESTs as candidates with differential expression. A number of genes related to oncogenesis and hepatic function/differentiation were selected for further semi-quantitative reverse transcriptase-PCR analysis in 29 paired HCC/non-cancerous liver samples. Many genes involved in cell cycle regulation such as cyclins, cyclin-dependent kinases, and cell cycle negative regulators were deregulated in most patients with HCC. Aberrant expression of the Wnt-beta-catenin pathway and enzymes for DNA replication also could contribute to the pathogenesis of HCC. The alteration of transcription levels was noted for a large number of genes implicated in metabolism, whereas a profile change of others might represent a status of dedifferentiation of the malignant hepatocytes, both considered as potential markers of diagnostic value. Notably, the altered transcriptome profiles in HCC could be correlated to a number of chromosome regions with amplification or loss of heterozygosity, providing one of the underlying causes of the transcription anomaly of HCC [13].

Gengxi Hu and his colleagues of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, also studied the expression profiles in HCC using a cDNA array representing 14 000 cDNA clusters. They found that 72 genes (including 30 novel genes) were down-regulated and 84 genes (including 48 novel genes) were up-regulated in a majority of the cancer samples. It was conspicuous that 21 of 38 HCC down-regulated genes studied previously were regulated by a group of liver-enriched transcription factors (LETFs). Reexamination of the cDNA array data further revealed that most of the genes known to be regulated by LETFs were down regulated in at least a portion of the HCC samples. Among the LETFs, the expression level of CCAAT/enhancer-binding protein (C/EBP) alpha was down-regulated in cancer, whereas hepatocyte nuclear factor 1 (HNF-1), HNF-3beta, HNF-4alpha, and HNF-4gamma were up-regulated. The expression profiling thus suggested multiple regulatory pathways involved in HCC, especially that related to LETFs [14].

Although a comprehensive and systematic understanding of hepatocarcinogenesis and metastasis is yet to be achieved, the transcriptomics approach for HCC conducted by Chinese scientists is one of the earliest trials in that direction. This work provided not only significant amount of relevant publicly available ESTs but also intriguing unique concepts to explore in hepatocarcinogenesis. These works clearly facilitated the process of isolation and identification of novel genes associated with HCC, and thus have laid down the foundation for the discovery of HCC related biomarkers and drug targets [15–18].

3.3. Transcriptome analysis of the rat dorsal root ganglion in peripheral axotomy model of neuropathic pain

Chronic neuropathic pain, as common symptom associated with many diseases, disturbs continually the normal life of patients. Phenotypic modification of dorsal root ganglion (DRG) neurons represents an important mechanism underlying neuropathic pain. However, the nerve injury-induced molecular changes are not fully identified. To determine the molecular alterations in a broader way, Dr. Zhang's and Dr. Han's groups in Shanghai carried out EST sequencing and cDNA array analysis on the genes mainly represented in the cDNA libraries of lumbar DRGs of normal rats and of rats 14 days after peripheral axotomy. Of the 7523 examined genes and ESTs, the expression of 122 genes and 51 ESTs is strongly changed. These genes encompass a large number of members of

distinct families, including neuropeptides, receptors, ion channels, signal transduction molecules, synaptic vesicle proteins, and others. Of particular interest is the up-regulation of gamma-aminobutyric acid (A) receptor alpha5 subunit, peripheral benzodiazepine receptor, nicotinic acetylcholine receptor alpha7 subunit, P2Y1 purinoceptor, Na(+) channel beta2 subunit, and L-type Ca(2+) channel alpha2delta-1 subunit. Those findings therefore reveal dynamic and complex changes in molecular diversity among DRG neurons after axotomy [19].

4. Further development of transcriptome study in China

Transcriptome analysis has been proven to be essential for functional genomics research and it is expected that together with the efforts of proteomics studies, this discipline is laying down the foundation for the future development of systems biology. On the other hand, continuous technology improvement in its efficiency, accuracy and reliability is essential to make it successful. In addition, innovations in bioinformatics algorithm as well as consistent efforts in developing sophisticated databases are also vital to its development. Nevertheless, one can clearly foresee that transcriptome research will be applied in broader fields with better technology and more sophisticated experimental design in the near future. Here, we will briefly comment on the trends in China.

4.1. Transcriptome for human physiology and disease

Human physiological and pathological characteristics, particularly for diseases with high incidence in China, will be approached initially by profiling gene expression patterns. Based on our previous experiences, systematic sample collection, sophisticated experiment design and innovative technology application are the main factors affecting the outcome of these costly studies. In order to understand the mechanism of human physiology and disease, appropriate model systems are essential. The transcriptome studies should be coordinated with the development of model animals, in particular, mouse and primates. Another major goal for the study of this field is isolation of more full-length cDNAs from different tissues and cells such as hypothalamus, pituitary, adrenal gland, liver, esophagus cancer, hematopoietic tissues and immune cells. Meanwhile, systematic survey of RNA isoforms will gradually become a more important goal in human transcriptome studies.

4.2. Transcriptome of model organisms and comparative genomics

Model organisms are especially useful systems for comparative genomics studies. Besides bacteria, China has been emphasizing studies of the rice genome, a model organism for monocotyl plants. Transcriptome research for rice and other plants is developing quickly in China. It is an important strategy for isolation of genes combining with the genomic information. It will also facilitate the identification of genes associated with diseases and specific characters of plants. Microbes associated with human and plant diseases will also be studied in priority, emphasizing their interaction with their host and environment.

For animals, the metazoan parasite *Schistosoma japonicum*, as the major blood fluke in China, remains a public health problem even today. Gaining knowledge about the *S. japonicum* genome will be important for understanding the biology of the worms with respect to their evolution and pathogenesis. Schistosomes are dioecious digenetic trematode under the phylum platyhelminthes. Unlike *Caenorhabditis elegans*, schistosomes have sexual dimorphism and larger males gripping smaller females in a gynecophoric canal. Recently, The *S. japonicum* genome project was initiated and large-scale EST data, representing about 43 000 ESTs and 13 000 clusters, was generated from male, female, mixed sexes adult worms, and their eggs by EST sequencing a first step.

Recently Chinese scientists were actively involved in the international consortium for the genomics research of *Lepidoptera*, moths and butterflies. The centerpiece of this strategy is the production of a draft genome sequence of *Bombyx mori*, the domesticated silk moth, by the end of 2004. The silk moth is regarded as the central model for lepidopteran genomics and genetics, and is an important source of livelihood for subsistence farmers engaged in silk production in at least 55 countries. Along with the efforts in genomic sequencing, gene expression profiles are being stud-

ied by Chinese, Japanese and French scientists with respect to its development, particularly the development of the posterior silk gland employing techniques of EST sequencing and SAGE. Proteomics work has also started in Japan and China. It is expected that this study will be greatly improved with the help of the whole genome draft sequence.

4.3. Improvement of platform technology in transcriptome analysis

Although the transcriptome study was initiated by large-scale EST sequencing analysis, it will be hard to achieve the essential high-throughput in order to generate meaningful data sets without the technology of cDNA microarray and biochips. However, the biochip technology is obviously still under development and it will be crucial to make this technology accurate and reliable enough for high quality research. It is also crucial to make it economically affordable for most laboratories working on functional genomics. In that direction, SAGE technology, along with the improvement of its protocol as well as the easier availability of large-scale sequencing platforms, is becoming more and more popular. It will become a nice compromise to complement the weaknesses of the biochip technology.

Other technologies under development are emphasizing the detection of protein-protein, protein-nucleic acid interactions. For other purposes, the interaction between biomolecules with small molecules (ligands) as well as large-scale detection of molecular interaction in tissues and cells will soon be detected by certain kinds of 'biochip' technology, e.g., tissue chips or cell chips. In that direction, the long-waited protein chips and antibody chips are the most important.

Bioinformatics has always been considered as the key component as well as the tool for any 'ome' research, as for the transcriptome studies. Innovations in bioinformatics algorithms should be emphasized while consistent efforts in sophisticated interrogative database development must be encouraged.

4.4. Inter-disciplinary research and innovation

An important direction is to combine biology with mathematics, physics, chemistry, informatics, computing technology and engineering. Many biomedical and systems biology centers and groups have been established in large cities of China, particularly in Shanghai and Beijing. These centers have the advantage of aggregating resources of many universities, institutions and even companies in order to establish inter-disciplinary research platforms and environment. Scientists working in these centers will be fostered to integrate information and technology ranging from genomics, transcriptomics, proteomics, to metabolomics, to get novel insight into the nature of complex biological systems and diseases.

4.5. International collaboration

Besides joining the international HGP, Chinese scientists will contribute continually their data to the public databases accessible by their colleagues worldwide. Human full-length cDNA will be deposited in GenBank and exchanged by collaboration with the international mammalian gene collection (MGC) project. The data of transcriptome analysis will follow the MIAME guidelines of the MGED Society and will be shared with other scientists via the SYSTEMOSCOPE International Consortium.

Acknowledgements

This work was supported by the Chinese High Tech R&D Program (863), Chinese National Key Program on Basic Research (973), National Natural Science Foundation of China, National Foundation for Excellence Doctoral Project, and Shanghai Commission for Science and Technology.

References

[1] M. Mao, G. Fu, J.S. Wu, Q.H. Zhang, J. Zhou, L.X. Kan, Q.H. Huang, K.L. He, B.W. Gu, Z.G. Han, Y. Shen, J. Gu, Y.P. Yu, S.H. Xu, Y.X. Wang, S.J. Chen, Z. Chen, Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning, Proc. Natl Acad. Sci. USA 95 (1998) 8175–8180.

[2] J. Gu, Q.H. Zhang, Q.H. Huang, S.X. Ren, X.Y. Wu, M. Ye, C.H. Huang, G. Fu, J. Zhou, C. Niu, Z.G. Han, S.J. Chen, Z. Chen, Gene expression in CD34(+) cells from normal bone marrow and leukemic origins, Hematol. J. 1 (2000) 206–217.

[3] Q.H. Zhang, M. Ye, X.Y. Wu, S.X. Ren, M. Zhao, C.J. Zhao, G. Fu, Y. Shen, H.Y. Fan, G. Lu, M. Zhong, X.R. Xu, Z.G. Han, J.W. Zhang, J. Tao, Q.H. Huang, J. Zhou, G.X. Hu, J. Gu, S.J. Chen, Z. Chen, Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells, Genome Res. 10 (2000) 1546–1560.

[4] Y. Yu, C. Zhang, G. Zhou, S. Wu, X. Qu, H. Wei, G. Xing, C. Dong, Y. Zhai, J. Wan, S. Ouyang, L. Li, S. Zhang, K. Zhou, Y. Zhang, C. Wu, F. He, Gene expression profiling in human fetal liver and identification of tissue- and developmental-stage-specific genes through compiled expression profiles and efficient cloning of full-length cDNAs, Genome Res. 11 (2001) 1392–1403.

[5] H.D. Song, T.X. Liu, X.Y. Wu, Y. Zhou, X.J. Sun, G.W. Zhang, S.J. Chen, Z. Chen, T.A. Look, L.I. Zon, Gene expression profiling in the zebrafish kidney marrow tissue. 4th HUGO Pacific Meeting & 5th ASIA-Pacific Conference on Human Genetics, October 27–30, 2002, Ambassador City Jomtien, Pattaya, Chonburi, Thailand, DY-15.

[6] G. Zhou, J. Chen, S. Lee, T. Clark, J.D. Rowley, S.M. Wang, The pattern of gene expression in human CD34(+) stem/progenitor cells, Proc. Natl Acad. Sci. USA 98 (2001) 13966–13971.

[7] Z. Lian, L. Wang, S. Yamaga, W. Bonds, Y. Beazer-Barclay, Y. Kluger, M. Gerstein, P.E. Newburger, N. Berliner, S.M. Weissman, Genomic and proteomic analysis of the myeloid differentiation program, Blood 98 (2001) 513–524.

[8] J.W. Baird, K.M. Ryan, I. Hayes, L. Hampson, C.M. Heyworth, A. Clark, M. Wootton, J.D. Ansell, U. Menzel, N. Hole, G.J. Graham, Differentiating embryonal stem cells are a rich source of haemopoietic gene products and suggest erythroid preconditioning of primitive haemopoietic stem cells, J. Biol. Chem. 276 (2001) 9189–9198.

[9] R.M. Hu, Z.G. Han, H.D. Song, Y.D. Peng, Q.H. Huang, S.X. Ren, Y.J. Gu, C.H. Huang, Y.B. Li, C.L. Jiang, G. Fu, Q.H. Zhang, B.W. Gu, M. Dai, Y.F. Mao, G.F. Gao, R. Rong, M. Ye, J. Zhou, S.H. Xu, J. Gu, J.X. Shi, W.R. Jin, C.K. Zhang, T.M. Wu, G.Y. Huang, Z. Chen, M.D. Chen, J.L. Chen, Gene expression profiling in the human hypothalamus-pituitary-adrenal axis and full-length cDNA cloning, Proc. Natl Acad. Sci. USA 97 (2000) 9543–9548.

[10] C.P. Leo, S.Y. Hsu, A.J. Hsueh, Hormonal genomics, Endocr. Rev. 23 (2002) 369–381.

[11] M.E. Huang, Y.C. Ye, S.R. Chen, J.R. Chai, J.X. Lu, L. Zhoa, L.J. Gu, Z.Y. Wang, Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia, Blood 72 (1988) 567–572.

[12] T.X. Liu, J.W. Zhang, J. Tao, R.B. Zhang, Q.H. Zhang, C.J. Zhao, J.H. Tong, M. Lanotte, S. Waxman, S.J. Chen, M. Mao, G.X. Hu, L. Zhu, Z. Chen, Gene expression networks underlying retinoic acid-induced differentiation of acute promyelocytic leukemia cells, Blood 96 (2000) 1496–1504.

[13] X.R. Xu, J. Huang, Z.G. Xu, B.Z. Qian, Z.D. Zhu, Q. Yan, T. Cai, X. Zhang, H.S. Xiao, J. Qu, F. Liu, Q.H. Huang, Z.H. Cheng, N.G. Li, J.J. Du, W. Hu, K.T. Shen, G. Lu, G. Fu, M. Zhong, S.H. Xu, W.Y. Gu, W. Huang, X.T. Zhao, G.X. Hu, J.R. Gu, Z. Chen, Z.G. Han, Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver, Proc. Natl Acad. Sci. USA 98 (2001) 15089–15094.

[14] L. Xu, L. Hui, S. Wang, J. Gong, Y. Jin, Y. Wang, Y. Ji, X. Wu, Z. Han, G. Hu, Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma, Cancer Res. 61 (2001) 3176–3181.

[15] H. Okabe, S. Satoh, T. Kato, O. Kitahara, R. Yanagawa, Y. Yamaoka, T. Tsunoda, Y. Furukawa, Y. Nakamura, Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, Cancer Res. 61 (2001) 2129–2137.

[16] Y. Shirota, S. Kaneko, M. Honda, H.F. Kawai, K. Kobayashi, Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays, Hepatology 33 (2001) 832–840.

[17] N. Iizuka, M. Oka, H. Yamada-Okabe, N. Mori, T. Tamesa, T. Okada, N. Takemoto, A. Tangoku, K. Hamada, H. Nakayama, T. Miyamoto, S. Uchimura, Y. Hamamoto, Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method, Cancer Res. 62 (2002) 3939–3944.

[18] X. Chen, S.T. Cheung, S. So, S.T. Fan, C. Barry, J. Higgins, K.M. Lai, J. Ji, S. Dudoit, I.O. Ng, M. Van De Rijn, D. Botstein, P.O. Brown, Gene expression patterns in human liver cancers, Mol. Biol. Cell 13 (2002) 1929–1939.

[19] H.S. Xiao, Q.H. Huang, F.X. Zhang, L. Bao, Y.J. Lu, C. Guo, L. Yang, W.J. Huang, G. Fu, S.H. Xu, X.P. Cheng, Q. Yan, Z.D. Zhu, X. Zhang, Z. Chen, Z.G. Han, X. Zhang, Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain, Proc. Natl Acad. Sci. USA 99 (2002) 8360–8365.