Analysis of phenotype–genotype connection: the story of dissecting disease pathogenesis in genomic era in China, and beyond

Yan Shen^{1,2,*}, Qi Xu¹, Zeguang Han³, Han Liu⁴ and Guang-Biao Zhou^{5,*}

¹State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Peking Union Medical College & Chinese Academy of Medical Sciences, 100005 Beijing, People's Republic of China

²Chinese National Human Genome Center at Beijing, 100176 Beijing, People's Republic of China

³Chinese National Human Genome Center at Shanghai, 201203 Shanghai, People's Republic of China

⁴State Key Laboratory for Medical Genomics, Shanghai Institute of Hematology, 200025 Shanghai, People's Republic of China

⁵Guangzhou Institute of Biomedicine and Health, Chinese Academy of Sciences, 510663 Guangzhou, People's Republic of China

DNA is the ultimate depository of biological complexity. Thus, in order to understand life and gain insights into disease pathogenesis, genetic information embedded in the sequence of DNA base pairs comprising chromosomes should be deciphered. The stories of investigating the association between phenotype and genotype in China and other countries further demonstrate that genomics can serve as a probe for disease biology. We now know that in Mendelian disorders, one gene is not only a dictator of one phenotype but also a dictator of two or more distinct disorders. Dissecting genetic abnormalities of complex diseases, including diabetes, hypertension, mental diseases, coronary heart disease and cancer, may unravel the complicated networks and crosstalks, and help to simplify the complexity of the disease. The transcriptome and proteomic analysis for medicine not only deepen our understanding of disease pathogenesis, but also provide novel diagnostic and therapeutic strategies. Taken together, genomic research offers a new opportunity for determining how diseases occur, by taking advantage of experiments of nature and a growing array of sophisticated research tools to identify the molecular abnormalities underlying disease processes. We should be ready for the advent of genomic medicine, and put the genome into the doctors' bag, so that we can help patients to conquer diseases.

Keywords: phenotype; genotype; genome; China

1. INTRODUCTION

China, a great country with 5000 years of history and 9.6 million square kilometres of land area, has 56 minority nationalities (figure 1) and a total population of 1.3 billion, making it one of the richest human genetic resources in the world. China is also the country where the most diversity and complexity of diseases can be seen and is wealthy in species of plants and animals. For a long time, Chinese scientists have contributed great effort to interpret the genetic basis of the biodiversity and elucidate the pathogenesis of diseases. The launch of the Human Genome Project (HGP) in China in 1993 greatly accelerated this process and tremendous advances have been seen in the past decade. In addition, the HGP has promoted the derivation of science branches in China such as transcriptomics, toxicogenomics, pharmacogenomics, systems biology, Cancer Genome Anatomy Project/Human Cancer Genome Project, proteomics and so on. Owing to these efforts, molecular medicine is not only the future focus but also the present focus of medicine in China.

2. THE HUMAN GENOME PROJECT AND THE INTERNATIONAL HUMAN GENOME HAPMAP PROJECT IN CHINA: A BRIEF INTRODUCTION

The notion that most genetic information is embedded in the sequence of DNA base pairs comprising chromosomes is a central tenet of modern genetics. Renato Dulbecco, a Nobel Prize winning molecular biologist, laid out his ideas of sequencing the genome to understand the genetic origins of cancer in a commentary for *Science* (Dulbecco 1986). After a wide discussion in the public, the HGP was launched in October 1990 (Roberts *et al.* 2001) to map out the sequence of the 3 billion nucleotide pairs, thus to unveil the information of human genetics and decode the structure and function of the genes in the human body.

Chinese scientists have been hunting for disease-related genes and key players of biodiversity in China for a long time. The HGP has attracted much attention from both scientists and decision makers because it is believed that HGP can help to promote these studies. In 1994, the first national-level project on the human genome was launched by the National Natural Science Foundation of China under the initiation of scientists including Min Wu, Zhu Chen and Boqin Qiang (Qiang 2004). Owing to limited resources at that time, the goals of this first project were to focus on the establishment of

* Authors for correspondence (sheny@ms.imicams.ac.cn; zhou_guangbiao@gibh.ac.cn)

One contribution of 14 to a Theme Issue 'Biological science in China'.

Figure 1. Minorities in China (from website: http://www.chinaculture.org/gb/cn_zgsk/2004-06/28/content_55993.htm).

infrastructure and core technology, collection and storage of genetic samples, as well as the training of research professionals. Over the period of 1994–1997, after many discussions among the scientists, a comprehensive workplan was made for the HGP in China to include sequencing of 1% of the human genome (30 Mb) and cloning of 1% of the novel human genes (1000 genes, as the number of human genes at that time was estimated to be 100 000), hunting the disease genes, investigating the functions of genes and exploring the human genome diversity (Chen *et al.* 1998). Subsequently, genome research and related studies were supported by the central government through the National High Technology Research and Development Program of China (863 Program), the National Basic Research Program of China (973 Program) and other research programmes by the Ministry of Science and Technology (MOST), the Chinese Academy of Sciences (CAS), as well as some local governments such as Beijing and Shanghai municipals. The Chinese National Human Genome Centers were then founded by the MOST in Beijing and Shanghai, respectively (Qiang 2004).

In September 1999, at the Fifth International Strategy Meeting on Human Genome Sequencing, thanks to the efforts of Dr Huanming Yang, China was accepted as the sixth member and the only developing country in the international HGP by taking on the sequencing of 1% of the human genome working draft, i.e. the region 3pter-D3S3610 containing 30 million bp. Through the joint efforts of the Beijing Genomics Institute (formerly, the Human Genome Center, Institute of Genetics, CAS) led by Dr Yang, the Chinese National Human Genome Center at Shanghai (CHGCS) led by Dr Zhu Chen and the Chinese National Human Genome Center at Beijing (CHGB) led by Dr Boqin Qiang, Chinese scientists declared their completion of China's HGP assignment ahead of schedule on 26 August 2001 (Lander *et al.* 2001; Qiang 2004). More than 1000 full-length cDNAs representing novel human genes, together with over 100 000 human expressed sequence tags (ESTs), were deposited into the public database. A total number of 754 full-length cDNAs generated by CHGCS were included into an international effort for human gene validation by full-length cDNA clones in 2004 (Qiang 2004).

Inherited genetic variation has a critical but as yet largely uncharacterized role in human disease. Aimed at speeding up the discovery of genes related to common illnesses such as asthma, cancer, diabetes and heart disease, the International HapMap Project was launched in 2002 jointly by the United States, United Kingdom, Japan, China and Canada. By comparing genetic differences among individuals, consortium members believed that they could create a tool to help researchers detect the genetic contributions to many diseases, making the results of genomic research applicable to individuals. The Chinese scientists are responsible for around 10% of the whole effort. In 2005, the consortium reported a public database of common variation in the human genome: more than 1 million single nucleotide polymorphisms (SNPs) for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations, including ten 500 kb regions in which essentially all information about common DNA variation has been extracted. These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. The International HapMap Consortium shows how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution (Altshuler *et al.* 2005).

3. DISSECTING DISEASE PATHOGENESIS IN GENOMIC ERA: THE GENES AS DICTATOR OF PHENOTYPE IN MENDELIAN DISORDERS

Through its contribution to the HGP, China has trained its own researchers and developed good technological platform for medical genetics and genomics research, resulting in many of remarkable achievements in this area during the past 10 years, especially in the identification of disease-related genes and relevant functional studies.

(a) *IHH gene mutations in brachydactyly type A-1: one gene, one disease*

Brachydactyly is an inherited disease and can be classified into five types—A, B, C, D and E—on the basis of malformation of the digits. Brachydactyly type A-1 (BDA1; figure 2a) is characterized by shortening of the middle phalanges which may be fused to the distal ones, and is the first recorded example of a human anomaly with Mendelian autosomal dominant inheritance. Two large families, the affected members of which were radiographed, were studied by Yang *et al.* (2000). Two-point linkage analysis for pedigree 1 and pedigree 2 mapped the locus for BDA1 to chromosome 2q in the two families. Haplotype analysis of pedigree 1 confined the locus for family 1 within an interval of less than 8.1 cM flanked by markers D2S2248 and D2S360, which was mapped to chromosome 2q35–q36 on the cytogenetic map. Haplotype analysis of pedigree 2 confined the locus for family 2 within an interval of less than 28.8 cM flanked by markers

Figure 2. Brachydactyly type A-1 and IHH mutation. (a) Short middle phalanges (i) and relevant radiographs (ii). (b) Structure of IHH and location of mutations. (c) Schematic of IHH protein and sites of amino acid residue transition and cleavage. Reproduced with permission from Gao *et al.* (2001) *Nat. Genet.* © Macmillan Magazines Ltd.

Figure 3. (a) Dentinogenesis imperfecta 1, (b) hearing loss and (c) mutations in DSPP. Figure (a) was reproduced with permission from Zhang *et al.* (2001) *Nat. Genet.*; figure (b) was reproduced with permission from Xiao *et al.* (2001) *Nat. Genet.*; figure (c) was modified with permission from Xiao *et al.* (2001) *Nat. Genet.* © Macmillan Magazines Ltd.

GATA30E06 and D2S427, which was localized to chromosome 2q35–q37. Gao *et al.* (2001) showed that mutations in IHH, which encodes Indian hedgehog, cause BDA-1. They have identified three heterozygous missense mutations in the region encoding the amino-terminal signalling domain (figure 2b,c) in all affected members of three large, unrelated families. The three mutant amino acids, which are conserved across all vertebrates and invertebrates studied so far, are predicted to be adjacent on the surface of IHH.

(b) Cataract: gene mutations result in blindness

Congenital cataracts are an important cause of blindness worldwide, causing 10–30% of all blindness in children, with one-third of cases estimated to have a genetic cause. Lamellar cataract is the most common type of infantile cataract. Bu *et al.* (2002) carried out whole-genome linkage analysis of Chinese individuals

with lamellar cataract and found that the disorder is associated with inheritance of a 5.11 cM locus on chromosome 16. This locus coincides with the one previously described for Marner cataract. Bu *et al.* (2002) screened individuals of three Chinese families for mutations in HSF4 (a gene at this locus that encodes heat shock transcription factor 4) and discovered that in each family, a distinct missense mutation, predicted to affect the DNA-binding domain of the protein, segregates with the disorder. They also discovered an association between a missense mutation and Marner cataract in an extensive Danish family. Their results suggest that HSF4 is critical to lens development and HSF mutation may cause blindness. Meanwhile, in a family of Chinese descent, a dominant congenital nuclear cataract locus was mapped to chromosome 17q11.1–12. The maximum logarithm of odds (LOD) score, 2.49, at recombination fraction $\theta=0$, was

obtained for marker D17S1294. The results of both linkage and haplotype analyses defined a disease gene to an 11.78 cM region harbouring the gene coding for betaA1/A3 crystallin (CRYBA1/A3). Mutation analysis of the CRYBA1/A3 gene identified a 3 bp deletion in exon 4, which co-segregated with the disease risk in this family and was not observed in 100 normal chromosomes. This mutation resulted in the deletion of a highly conserved glycine at codon 91 (DeltaG91) and could be associated with an incorrect folding of betaA1/A3 crystallin. It highlights the physiological importance of crystallin and supports the role of CRYBA1/A3 in human cataract formation (Qi *et al.* 2004). These results demonstrate that mutations in certain genes are directly involved in blindness.

(c) Dentinogenesis imperfecta associates with hearing loss: studies of dentin sialophosphoprotein

Dentinogenesis imperfecta 1 (DGI1; figure 3a) is an autosomal dominant dental disease characterized by abnormal dentin production and mineralization. Zhang *et al.* (2001) identified a nonsense mutation (Gln45stop) in exon 3 of the dentin sialophosphoprotein (DSPP) gene in a Chinese family with dentinogenesis imperfecta Shields type II (DGI-II), in which the affected members showed discoloration and severe attrition of their teeth, with obliterated pulp chambers. Meanwhile, Xiao *et al.* (2001) studied three Chinese families carrying DGI1 and found that the affected individuals of two families also presented with progressive sensorineural high-frequency hearing loss (gene DFNA39). They identified three disease-specific mutations within the DSPP in these three families and detected a G → A transition at the donor-splicing site of intron 3 in one family without DFNA39, a mutation predicted to result in the skipping of exon 3. However, in two other families affected with both DGI1 and DFNA39, they identified two independent nucleotide transversions in exons 2 and 3 of DSPP, respectively, which cause missense mutations of two adjacent amino acid residues in the predicted transmembrane region of the protein. Moreover, transcripts of DSPP previously reported to be expressed specifically in teeth are also detected in the inner ear of mice. These data demonstrated for the first time that distinct mutations in DSPP are responsible for the clinical manifestations of DGI1 with or without DFNA39, and mutation in the DSPP gene could also cause high-frequency hearing loss at the same time. These two groups identified a total of four different mutations, making a strong case that the DSPP gene is the one causing DGI (figure 3b).

(d) Hearing loss and skin disease: one connexin, two diseases

The complexity of human genetic disease continues to confuse and it sometimes seems remarkable that much progress has been made in identifying disease genes when subsequent work shows that the story is much more involved than was at first imagined. An example comes from the study of GJB3, which locates on chromosome 1p34 and encodes the connexin 31 component of gap junctions (figure 3). Gap junctions connect adjacent cells, allowing small molecules to pass

from one cell to the next, and are believed to play an important role in intercellular communication. Members of the connexin family have highly conserved sequences and four transmembrane domains separating two extracellular loops and one cytoplasmic loop, with cytoplasmic carboxy- and amino-terminal ends. Functions of each domain are reported (figure 3). Six connexin molecules assemble to form one connexon, which docks with its counterpart in the neighbouring cell to form the gap junction channel (Bruzzone *et al.* 1996; Goodenough *et al.* 1996; Steel 1998). Mutations in the gene encoding connexin 26 (GJB2) cause both autosomal recessive and dominant forms of hearing impairment (Kelsell *et al.* 1997; Denoyelle *et al.* 1998). To study the possible involvement of other members of the connexin family in hereditary hearing impairment, Xia *et al.* (1998) cloned the GJB3 using homologous EST searching and nested PCR. Mutation analysis revealed that a missense mutation and a nonsense mutation of GJB3 (figure 3a) were associated with high-frequency hearing loss in two families. Moreover, expression of GJB3 was identified in the rat inner ear tissue by RT-PCR. These findings suggest that mutations in GJB3 may be responsible for bilateral high-frequency hearing impairment.

Erythrokeratoderma variabilis (EKV; figure 4) is an autosomal dominant genodermatosis with considerable intra- and interfamilial variability. It has a disfiguring phenotype characterized by the independent occurrence of two morphological features: transient figurate red patches and localized or generalized hyperkeratosis. Both features can be triggered by external factors such as trauma to the skin. Interestingly, Richard *et al.* (1998) detected heterozygous missense mutations in GJB3 in four EKV families leading to substitution of a conserved glycine by charged residues (G12R and G12D) or change of a cysteine (C86S). These mutations are predicted to interfere with normal connexin 31 structure and function, possibly due to a dominant inhibitory effect. These results implicate connexin 31 in the pathogenesis of EKV and provide evidence that intercellular communication mediated by connexin 31 is crucial for epidermal differentiation and response to external factors.

How can the same gene underlie two such different diseases? Both the mutations associated with hearing loss affect residues in an extracellular loop that is thought to be involved in regulating the specificity of connexon–connexon interactions. One of these, the nonsense mutation, should truncate the protein, eliminating the fourth transmembrane domain and the C-terminal region with its four potential sites for phosphorylation, which are believed to have a role in controlling gating of the whole channel. In contrast, the N-terminal domain harbouring two of the missense mutations that cause the skin disorder is thought to be involved in determining the polarity of the voltage gating (i.e. it determines whether the channel opens when the cytoplasm is at negative or positive potential). The third ‘skin’ mutation affects a residue that sits next to a conserved proline residue in the second transmembrane domain which is critical for voltage-gating activity. Thus, the mutations in GJB3 may affect different aspects of channel function, which might explain the different phenotypes (Steel 1998).

(e) Role of KCNQ1 in cardiac arrhythmia: different mutations, different risks

Genetic factors contribute to the risk of sudden death from cardiac arrhythmias. The long QT syndrome (LQTS) is a cardiac disorder characterized by prolongation of the QT interval on electrocardiograms, syncope and sudden death caused by a specific ventricular tachyarrhythmia known as torsade de pointes, while atrial fibrillation (AF) is a common cardiac arrhythmia whose molecular aetiology is poorly understood. In 1996, positional cloning methods established KVLQT1 (KCNQ1) as the chromosome 11-linked LQT1 gene responsible for the most common inherited cardiac arrhythmia. KVLQT1 is strongly expressed in the heart and encodes a protein with structural features of a voltage-gated potassium channel. KVLQT1 mutations are present in affected members of 16 arrhythmia families, including one intragenic deletion and 10 different missense mutations. These data defined KVLQT1 as a novel cardiac potassium channel gene and show that mutations in this gene cause susceptibility to ventricular tachyarrhythmias and sudden death (Wang *et al.* 1996). Chen *et al.* (2003b) studied a family with hereditary persistent AF and identified the causative mutation (S140G) in the KCNQ1 (KVLQT1) gene on chromosome 11p15.5. The KCNQ1 gene encodes the pore-forming α subunit of the cardiac I(Ks) channel (KCNQ1/KCNE1), the KCNQ1/KCNE2 and the KCNQ1/KCNE3 potassium channels. Functional analysis of the S140G mutant revealed a gain-of-function effect on the KCNQ1/KCNE1 and KCNQ1/KCNE2 currents, which contrasts with the dominant negative or loss-of-function effects of the KCNQ1 mutations previously identified in patients with LQTS. Thus, the S140G mutation is likely to initiate and maintain AF by reducing action potential duration and effective refractory period in atrial myocytes. Taken together, KCNQ1 is critical to maintain regular cardiac rhythm and its mutations may cause LQTS or AF.

Yang *et al.* (2004b) evaluated 28 unrelated Chinese kindreds with AF and sequenced eight genes of potassium channels (KCNQ1, HERG, KCNE1, KCNE2, KCNE3, KCNE4, KCNE5 and KCNJ2). An arginine-to-cysteine mutation at position 27 (R27C) of KCNE2, the β subunit of the KCNQ1–KCNE2 channel responsible for a background potassium current, was found in 2 out of the 28 probands. The mutation was present in all affected members in the two kindreds, but was absent in 462 healthy unrelated Chinese subjects. Similar to KCNQ1 S140G, the mutation had a gain-of-function effect on the KCNQ1–KCNE2 channel; unlike LQTS-associated KCNE2 mutations, it alters neither HERG–KCNE2 current nor the functions of the HCN channel family. Thus, KCNE2 R27C is a gain-of-function mutation associated with the initiation and/or maintenance of AF.

(f) Skin diseases and gene mutations: strong linkage

Disseminated superficial actinic porokeratosis (DSAP) is a chronic cutaneous disorder characterized by multiple superficial keratotic lesions surrounded by a slightly raised keratotic border. It develops in teenagers in

sun-exposed areas of skin and usually follows an autosomal dominant inheritance pattern. To determine the locus of DSAP and identify the candidate gene(s) of the disease, Zhang *et al.* (2005) performed genome-wide scan and linkage analysis in a six-generation Chinese family with DSAP. The coding exons of the candidate genes were sequenced to analyse and detect the nucleotide variations. Linkage analysis showed that the maximum two-point LOD score of 5.56 was obtained with the marker D12S79 at a recombination fraction $\theta=0.00$. Haplotype analysis defined the critical region for DSAP between D12S330 and D12S1612 on 12q24.1–24.2. By sequence analysis, they found a Val591Met mutation in SART3 in all affected individuals of the family. They concluded that SART3 is a candidate gene for DSAP and is possibly involved in the pathogenesis of DSAP. In 2002, to determine whether DSAP is a genetically heterogeneous disorder and to identify the disease gene locus in a three-generation Chinese family with DSAP, genetic linkage analysis was carried out by Xia *et al.* (2002) in this family using 15 microsatellite markers between D12S1671 and D12S369 on chromosome 12q, followed by a genome-wide scan with 382 microsatellite markers from the autosomes. Genetic linkage analysis with chromosome 12q markers suggested that the locus in this family is not linked to chromosome 12q. A genome-wide scan and fine mapping finally localized the locus for DSAP in this family to a 6.4 cM region between markers D15S1023 and D15S1030 at chromosome 15q25.1–26.1. This DSAP locus was named DSAP2. A locus at chromosome 12q23.2–24.1 was also found to be responsible for DSAP; the related gene might be located within a 9.6 cM region between markers D12S1727 and D12S1605, with a maximum two-point LOD score of 20.53 ($\theta=0.00$) at D12S78 (Xia *et al.* 2000; Wu *et al.* 2004).

Epidermolytic palmoplantar keratoderma (EPPK) is an autosomal dominantly inherited disease. Lu *et al.* (2003) studied a family from Shandong, China, having patients suffering from EPPK with a unique symptom—knuckle pads. They noticed that both the hyperkeratosis and knuckle pads in the Chinese family were friction related. Candidate gene analysis was carried out using linkage analysis and direct sequencing. A novel L160F mutation in keratin 9 was found and its effects on the secondary structure of keratin 9 were studied. They predicted that the L160F mutation is also responsible for the knuckle pads in the family. Their study provides a new clue for the study of the function of keratin 9.

Primary erythralgia is a rare autosomal dominant disease characterized by intermittent burning pain with redness and heat in the extremities. A previous study established the linkage of primary erythralgia to a 7.94 cM interval on chromosome 2q, but the causative gene was not identified. Yang *et al.* (2004a) performed linkage analysis in a Chinese family with primary erythralgia and screened the mutations in the two candidate genes, SCN9A and GCA, in the family and a sporadic patient. Linkage analysis yielded a maximum LOD score of 2.11 for both markers D2S2370 and D2S2330. Based on critical recombination events in two patients in the family, they further limited the genetic region to 5.98 cM between D2S2370 and

Figure 4. One connexin, two diseases. (a) Predicted arrangement of the connexin 31 molecule within the cell membrane. Four transmembrane domains (M1–M4) separate the intracellular N-terminal domain (N), the two extracellular loops (E1 and E2), the cytoplasmic loop (CL) and the C-terminal domain. The locations of mutations causing hearing loss are shown by green spots, while mutations associated with the EKV are indicated by red spots. (b) Audiogram of patients with high-frequency hearing loss. (c) Clinical features of EKV. Figure (a) was reproduced with permission from Steel (1998) *Nat. Genet.*; figure (b) was reproduced with permission from Xia *et al.* (1998) *Nat. Genet.*; figure (c) was reproduced with permission from Richard *et al.* (1998) *Nat. Genet.* © Macmillan Magazines Ltd.

D2S2345. They then identified two missense mutations in SCN9A in the family (T2573A) and the sporadic patient (T2543C). These data suggest that mutations in SCN9A cause primary erythralgia. SCN9A, encoding a voltage-gated sodium channel α subunit predominantly expressed in sensory and sympathetic neurons, may play an important role in nociception and vasomotor regulation.

Dyschromatosis symmetrica hereditaria (DSH) is a hereditary skin disease characterized by the presence of hyper- and hypopigmented macules on extremities and face. The gene, or even its chromosomal location, for DSH has not yet been identified. In a study by Xing *et al.* (2003), two Chinese families with DSH were identified and subjected to a genome-wide screen for linkage analysis. Two-point linkage analysis for pedigree A (maximum LOD score ($Z(\max)$) = 7.28 at recombination fraction θ = 0.00) and pedigree B ($Z(\max)$ = 2.41 at θ = 0.00) mapped the locus for DSH in the two families to chromosome 6q. Subsequent multipoint analysis of the two families also provided additional support for the DSH gene being located within the region 6q24.2–q25.2, with $Z(\max)$ = 10.64. Haplotype analysis confined the locus within an interval of 10.2 Mbp, flanked by markers D6S1703 and D6S1708. The two families had no identical haplotype within the defined region, which suggests that the two families were different in origin. Further work on identification of the gene for DSH will open new avenues to exploration of the genetics of pigmentation.

(g) Nervous system and genes

Benign familial neonatal convulsions (BFNC) are a rare autosomal dominant inherited epilepsy syndrome.

Two voltage-gated potassium channel genes, KCNQ2 on chromosome 20q13.3 and KCNQ3 on chromosome 8q24, have been identified as the genes responsible for BFNC. By linkage and mutation analyses of KCNQ2 gene, Tang *et al.* (2004) found a novel frameshift mutation of KCNQ2 gene, 1931delG, in a large Chinese family with BFNC. This mutation is located in the C-terminus of KCNQ2, in codon 644, predicting the replacement of the last 201 amino acids with a stretch of 257 amino acids showing a completely different sequence. An unusual clinical feature of this family is that the seizures of every patient did not remit until 12–18 months.

Direct sequencing of exons 3–35 and the exon–intron boundaries of the CACNA1H gene was conducted in 118 childhood absence epilepsy patients of Han ethnicity recruited from North China. Sixty-eight variations have been detected in the CACNA1H gene and among them, 12 were missense mutations and found only in 14 out of the 118 patients in a heterozygous state, but not in any of 230 unrelated controls. The identified missense mutations occurred in the highly conserved residues of the T-type calcium channel gene. These results suggest that CACNA1H might be an important susceptibility gene involved in the pathogenesis of childhood absence epilepsy (Chen *et al.* 2003a).

4. HOW TO SIMPLIFY THE COMPLEXITY OF DISEASES: GENES, GENES, GENES...

The molecular studies of complex disorders in China began in early 1990s. In the beginning of the new century, the study of complex disorders in China showed a new look. An integrated programme consisting of genome-wide linkage analysis and

multipoint linkage-disequilibrium analysis was set up to track down susceptibility genes for complex polygenic diseases. Owing to the gratifying genetic resources in China and novel biotechnologies used for work on the HGP, researchers have achieved several unique successes in studying complex disorders and publications of their work in well-known peer-reviewed journals have rapidly been increasing.

(a) *Diabetes*

Diabetes mellitus or diabetes can be divided into two types, type I and type II. Type I diabetes mellitus is characterized by the marked inability of the pancreas to secrete insulin owing to autoimmune destruction of the beta cells. Type II (non-insulin-dependent) diabetes, which represents 90% of all diabetes, is characterized by peripheral insulin resistance with an insulin-secretory defect that varies in severity. For type II diabetes to develop, both defects must exist: all overweight individuals have insulin resistance, but only those with an inability to increase beta-cell production of insulin develop diabetes. Type II diabetes is a highly heterogeneous multifactorial disease with both genetic and environmental determinants and an uncertain mode of inheritance. At present, there are 143 million people with the disease and more than 15 million diabetic patients in China. In addition, the prevalence of diabetes is still increasing. The belief that type II diabetes has strong genetic determinants is based on several lines of evidence, including the high concordance rate among twins, the marked difference in disease rate between populations and the close correspondence between admixture rate and disease prevalence in hybrid populations. In addition, there are evidences for major gene(s) influencing diabetes or its specific clinical manifestations, such as glucose concentration, 2 h postprandial insulin level and age at onset of diabetes. However, the mode of inheritance of type II diabetes appears to vary across populations, suggesting a complex genetic mechanism underlying the disease. In 2001, a genome-wide scan was performed by [Luo *et al.* \(2001\)](#) to search for the type II diabetes susceptibility genes in a Chinese population. They studied 102 families (478 family members), who were Chinese Hans residing in east and southeast China, including 282 diabetic patients; among them, 142 independently affected sib pairs were available for genotyping. A total of 247 fluorescence-labelled microsatellite markers, with an average resolution of 15 cM, were amplified. GENEHUNTER was used for the non-parametric linkage (NPL) analyses. Two loci on chromosome 9, D9S171 and D9S175, showed suggestive evidence for linkage, with an NPL score of 3.286 and 2.939, respectively, and a p -value of 1.19×10^{-4} and 4.47×10^{-4} , respectively. A locus on the long arm of chromosome 20, D20S196, showed a rise in the NPL score (from 1.517 to 2.922) and a corresponding decrease in the p -value from 0.04 to 6.5×10^{-4} when families with lower body mass index (BMI) were analysed alone. Other loci with weaker evidence for linkage were also observed. Their results suggest that chromosome 9 contains genes involved in the susceptibility to type II diabetes in an eastern and

southeastern Chinese Han population, and chromosome 20 could hide genes linked to type II diabetes in families with a lower BMI. Other regions could also hide susceptibility genes with minor effects.

Through an autosomal genomic scan, [Zhao *et al.* \(2005\)](#) identified four regions on chromosome 1 showing evidence of a link to type II diabetes in the northern Chinese Han population. [Du *et al.* \(2001\)](#) also conducted a genome-wide scan in a northern Chinese population to identify type II diabetes susceptibility genes and found that 1p36.3-1p36.23 region, spanning a long range of 16.9 cM, may contain multiple susceptibility genes. To identify the susceptible gene(s) in this region, 23 SNPs located in 10 candidate genes in the mapped region were chosen from public SNP domains with bioinformatic methods and the single base extension method to genotype the loci for 192 sporadic type II diabetes patients and 172 normal individuals, all with Han ethnic origin, using a case-control study. The haplotypes with significant difference in the gene(s) were further analysed. Among the 23 SNPs, eight were found to be common in the Chinese Han population. Allele frequency of one SNP, rs436045 in the protein kinase C/zetogene (PRKCZ), was statistically different between the case and control groups ($p < 0.05$). Furthermore, haplotypes at five SNP sites of the PRKCZ gene were identified. Thus, the PRKCZ gene may be associated with type II diabetes in the Han population in North China. The haplotypes at five SNP sites in this gene may be responsible for this association. [Liu *et al.* \(2003\)](#) studied co-inheritance of specific HSPG and ApoE genotypes in the development of type II diabetic nephropathy in Chinese samples and found that co-inheritance of the T allele of HSPG/E2 allele of ApoE confers a high risk of type II diabetes mellitus progression to diabetic nephropathy in them.

(b) *Hypertension*

Hypertension is persistent abnormally high blood pressure (BP) which affects 1 billion people worldwide and is implicated in 7.1 million deaths each year from ischaemic heart disease and stroke. It is clear from family and epidemiological studies that hypertension arises from a complex interplay between genetic and environmental lifestyle exposures, including dietary sodium intake, excess alcohol consumption and body weight. In addition, twin studies and segregation analyses have shown that between one-third and one-half of the inter-individual variation of BP is heritable ([Lifton *et al.* 2001](#); [Mein *et al.* 2004](#)). To identify chromosome regions containing hypertension susceptibility genes in the Chinese population, a three-stage study was carried out in their siblings ascertained through outpatient clinics. In the first stage, 283 affected sib pairs from 79 nuclear families were subjected to a genome-wide scan with 240 microsatellite marker loci. The second stage focused on chromosome 2 with additional markers resulting in an average distance of 5 cM and used an independent sample of 637 affected sib pairs from 161 families. In the third stage, a fine-scale mapping study on the suggestive region was performed in an independent set of 777 affected sib pairs from 106 families. Fourteen

markers were used with an average distance less than 2 cM. NPL analyses, parametric linkage analyses and transmission-disequilibrium tests were used to assess the evidence for linkage and association. Three markers (D2S168 at 27.06 cM, D2S151 at 152.04 cM and D2S142 at 161.26 cM) on chromosome 2 with suggestive linkage to hypertension susceptibility genes were identified in the genome-wide scan. In stage II, the suggestive region around D2S151 and D2S142 was replicated, while the linkage around D2S168 was not. In the stage III fine-scale mapping study, multipoint linkage analyses showed LOD scores greater than 2.0 throughout a region between 157.16 and 162.46 cM (all $p < 0.001$), with a maximum peak of 2.24 ($p = 0.00067$) at 160.52 cM. The authors also observed a NPL Z-score peak of 3.27 at 157.55 cM ($p = 0.00086$). These results of a suggestive region on chromosome 2q14–q23 were consistent between each of the three studies. Interestingly, this region overlaps a syntenic region that contains BP quantitative trait loci identified in the rat models of hypertension. These data suggest that the region near D2S142 and D2S151 deserves to be further screened for hypertension susceptibility genes (Zhu *et al.* 2001). To examine the linkage of qualitative and BP quantitative traits in essential hypertension with these genomic regions in a large sample of Chinese hypertensive families, Ge *et al.* (2003) performed a genetic analysis on 148 randomly ascertained families containing 328 affected sib pairs, grouped into two geographically distinct subsets. Five highly informative microsatellite markers (D2S151, D2S142, D5S2090, D5S413 and D5S2013) were genotyped and linkage analyses were performed with different genetic models. They did not observe consistent evidence for excessive allele sharing identity by descent in either of the qualitative or the quantitative test. However, higher LOD scores were found at D5S2013 in North Group subset with Haseman–Elston and maximum-likelihood (ML) variance (no dominance variance (NDV)) algorithms. With the ML (NDV) algorithm, the LOD was 1.410 for diastolic BP at this locus, although it was not statistically significant. These findings provide no evidence to support a significant linkage of 2q14–q23 or 5q32 with essential hypertension or BP quantitative traits in the ethnic Chinese and indicate the aetiologic diversity and complexity of hypertension. Caulfield *et al.* (2003) phenotyped 2010 affected sibling pairs drawn from 1599 severely hypertensive families and completed a 10 cM genome-wide scan. After rigorous quality control, they analysed the genotypic data by NPL, which tests whether genes are shared in excess among the affected sibling pairs. LOD scores, calculated at regular points along each chromosome, were used to assess the support for linkage. Linkage analysis identified a principal locus on chromosome 6q, with a LOD score of 3.21 that attained genome-wide significance ($p = 0.042$). The inclusion of three additional loci with LOD scores higher than 1.57 (2q, 5q and 9q) also shows genome-wide significance ($p = 0.017$) when assessed under a locus-counting analysis. These findings imply that human essential hypertension has an oligogenic element (few genes may be involved in the determination of the trait), possibly superimposed on more minor

genetic effects, and that several genes may be tractable to a positional cloning strategy. So far all human chromosomes except 13 and 20 have been shown to have BP, hypertension or pre-eclampsia loci. Indeed, chromosomes 1, 2, 8, 11, 12, 15, 16, 18 and 19 have regions that have been found in more than one study, yielding some encouraging consistent support for BP loci. Chromosome 2 has three cytogenetic intervals, 2p25.3–p16.3, 2p16.1–p12 and 2q23.3–q24.3, where more than one study has mapped a BP locus. Some regions directly overlap (Mein *et al.* 2004).

Scientists have also contributed great efforts to identify genes involved in hypertension. The lipoprotein lipase (LPL) gene has been reported to have an association with hypertension. Li *et al.* (2004a) identified 22 sequence variants by directly sequencing 10 exons and flanking regions of the LPL gene and investigated the occurrence of three of these variants, in a case–control study including 501 normotensive subjects and 497 essential hypertension subjects. The individual SNP analysis did not provide substantial evidence of an association between polymorphisms in the LPL gene and hypertension status and/or BP levels in this cohort, but the more powerful haplotype analysis suggested an association between the LPL gene and hypertension. Yang *et al.* (2003) used two different weakly parametric linkage tests, sib-pair linkage (SAGE/SIBPAL2) and pedigree-based (SOLAR) approaches, to estimate the quantitative or qualitative contribution of variation at the four candidate genes or their near genomic regions to BP and essential hypertension. Their results also found that the LPL gene may influence individual BP variation in the Chinese population. Through family-based association analyses using the quantitative and sib-transmission/disequilibrium tests (QTDT and Sib-TDT), Xu (Xu *et al.* 2004b) found that angiotensin-converting enzyme (Schunkert *et al.* 1994) insertion/deletion polymorphism might play a role in the development of obesity and hypertension, which are closely linked to cardiovascular risk.

(c) *Studies on schizophrenia, autism and Parkinson's disease: a brief introduction*

Schizophrenia is a common complex disorder characterized by psychosis, cognitive dysfunction and negative symptoms, whose aetiology involves interactions between both genetic and environmental vulnerability factors. Best supported schizophrenia loci include 1q21–q22, 1q42, 5q21–q33, 6p24–p22, 6q21–q25, 8p22–p21, 10p15–p11, 10q25–q26, 13q32–q34 and 22q11–q13 (Sanders *et al.* 2004).

Liu *et al.* (2004b) reported a result obtained in the study of 547 schizophrenia cases and 536 controls in the Chinese population. Six SNPs were genotyped at and around the enzyme D-amino acid oxidase (DAAO) locus (Chumakov *et al.* 2002; Schumacher *et al.* 2004), covering a 10 kb region entirely encompassing the complementary DNA sequences of DAAO. They found statistically significant differences in allele distributions on one marker: SNP rs3741775 ($p = 0.0000001$). In the haplotype analysis based on the information of linkage-disequilibrium block across this gene locus, they demonstrated a highly significant

association between schizophrenia and a DAAO haplotype, which therefore provides an independent statistical support for association of the DAAO gene with schizophrenia and indicates that the DAAO gene may play a significant role in the aetiology of schizophrenia in the Chinese Han population. Case-control studies also revealed the relationship between the metabotropic glutamate receptor 3 gene (Chen *et al.* 2005b), frizzled 3 (FZD3; Zhang *et al.* 2004) and ZDHHC8 on 22q11 (Chen *et al.* 2004) and schizophrenia in the Chinese population.

Since the conventional approaches to study of polygenic disease have shown some limitations in several aspects, Xu *et al.* (2004a) had recently introduced a novel idea into the genetic analysis of paranoid schizophrenia. Instead of examining one candidate at a time, an entire dopamine (DA) metabolism pathway is tested. In this work, 85 SNPs present in 23 genes for the DA metabolism pathway were genotyped in patients with paranoid schizophrenia and controls. Their study focused on investigating the combined effect of schizophrenia susceptibility genes. The strategy consists of systematical and thorough study of genes, strict clinical diagnosis and subgroups of patients, and new mathematical methods for genotype-phenotype association study. Their work showed an association of the combination patterns of network-related gene SNPs with schizophrenia. To capture the combined effects of susceptibility genes from the data warehouse, they developed two novel multilocus-based analysis methods, named 'potential effective SNP combination pattern' and 'potential effective dynamic effects'. Based on this strategy, they found that three genotype combinations were associated with a susceptibility to schizophrenia. These results were also validated in an independent cohort sample consisting of 95 paranoid schizophrenic family trios.

Apoptosis is thought to play a role in neuronal pathology in schizophrenia. Recently, the GSN gene was reported to have anti-apoptotic properties. In a genome-wide expression analysis on schizophrenia, GSN was also found to be significantly downregulated in schizophrenia. All the hints suggest that GSN is a novel candidate gene in the occurrence of schizophrenia. Xi *et al.* (2004) genotyped three SNPs around the GSN locus in 493 sets of the Chinese Han trio sample using allele-specific PCR. A weak association or a marginally positive result was detected.

Autism is a complex brain disorder that often inhibits a person's ability to communicate, respond to surroundings and form relationships with others. Several genome-wide screens indicated that chromosome 7q was linked to autistic disorder. FOXP2, located on 7q31, encodes a putative transcription factor containing a polyglutamine tract and a forkhead DNA-binding domain. It is one of the members of the forkhead family who are known to be key regulators of embryogenesis. A point mutation at a highly conserved residue within the forkhead domain co-segregated with affected status in the KE family, which was a unique three-generation pedigree with a severe speech and language disorder, and FOXP2 was directly disrupted by a translocation in an individual who had similar deficits as those of the KE

family. Several studies have investigated the role of FOXP2 polymorphisms in autism and none of them found positive association. Gong *et al.* (2004) performed a family-based association study of three SNPs of FOXP2 in 181 Chinese Han trios using the analyses of transmission/disequilibrium test (TDT) and haplotype. They found a significant association between autistic disorder and one SNP, as well as with specific haplotypes formed by this SNP with two other SNPs they investigated. Their findings suggest that the FOXP2 gene may be involved in the pathogenesis of autism in the Chinese population.

Parkinson's disease is chronic, progressive neuro-degenerative movement disorder characterized by an insidious onset with slowing of emotional and voluntary movement, muscular rigidity, postural abnormality and tremor. NR4A2, encoding a member of nuclear receptor superfamily, is essential for the differentiation of the nigral dopaminergic neurons. To determine whether NR4A2 is a susceptibility gene for Parkinson's disease, Le *et al.* (2003) carried out genetic analyses in 201 individuals affected with Parkinson's disease and 221 age-matched unaffected controls. They identified two mutations in NR4A2 associated with Parkinson's disease (-291Tdel and -245T→G), which map to the first exon of NR4A2 and affect one allele in 10 out of 107 individuals with familial Parkinson's disease, but not in any individuals with sporadic Parkinson's disease ($n=94$) or in unaffected controls ($n=221$). The age at onset of disease and clinical features of these 10 individuals were not different from those of individuals with typical Parkinson's disease. The mutations resulted in a marked decrease in NR4A2 messenger RNA (mRNA) levels in transfected cell lines and in lymphocytes of affected individuals. Additionally, mutations in NR4A2 affect transcription of the gene encoding tyrosine hydroxylase. These data suggest that mutations in NR4A2 can cause dopaminergic dysfunction, associated with Parkinson's disease.

(d) *Coronary heart disease*

Coronary heart disease (CHD) (or coronary artery disease) is a narrowing of the small blood vessels that supply blood and oxygen to the heart, and is one of the leading causes of death in China (He *et al.* 2005). Serum C-reactive protein (CRP) has been strongly implicated in the pathogenesis of CHD. Chen *et al.* (2005a) investigated the association between the gene coding for CRP and CHD in the ethnic Han population of China. Two polymorphisms of -717A/G and +2147A/G of the CRP gene were identified by direct sequencing of genomic DNA derived from 48 randomly selected patients and were further investigated for associations with CHD in 619 male patients and 615 age-matched male normal controls. The frequency of A allele carriers of -717A/G polymorphism was significantly higher in patients than in controls by univariate analysis. After controlling for other risk factors, the association between this polymorphism and CHD remained significant by multivariate logistic regression analysis. Individuals carrying the -717A allele had an approximately 6.8-fold higher risk of developing CHD. Haplotype analysis confirmed the results of individual

Figure 5. Age-standardized mortality for the five leading causes of death (a) among study participants who were rural or urban residents and (b) from malignant neoplasms.

polymorphism analyses. However, the resolution of effect size is poor, which may be due to the deficiency in sample size of this study. Neither polymorphism was observed to have an influence on serum CRP level. Since the frequency difference between CHD cases and controls for the $-717A$ allele carriers is only 2.28% and homozygosity for $-717G$ occurs in only 1.78% of subjects, $-717 A/G$ polymorphism is not a major determinant of population genetic risk of CHD in the Chinese population. The association of this polymorphism with CHD supports the belief that carriers of $-717A$ allele of the CRP gene are genetically predisposed to CHD in the Chinese Han population, and it remains possible that this polymorphism is in disequilibrium with one as yet unidentified functional polymorphism in the vicinity (Chen *et al.* 2005a).

5. HOW TO UNRAVEL THE MYSTERY OF CANCER: DISSECTING THE CANCER GENOME

Cancer represents the leading cause of death in China. After three decades of rapid advances, cancer research has generated a rich and complex body of knowledge, revealing cancer to be a disease involving dynamic changes in the genome. The foundation has been set in the discovery of mutations that produce oncogenes with dominant gain of function and tumour suppressor genes with recessive loss of function; both classes of cancer genes have been identified through their alteration in human and animal cancer cells and by their elicitation of cancer phenotypes in experimental models (Hanahan & Weinberg 2000; He *et al.* 2005; figure 5).

(a) Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is a cancer that arises from hepatocytes, the major cell type of the liver, and is a leading cause of death worldwide, especially in Asia and Africa. To understand the genetic mechanisms

underlying the progression of HCC metastasis, differences in genomic alterations between 10 pairs of primary HCC tumours and their matched metastatic lesions were analysed by comparative genomic hybridization. Several chromosomal alterations including loss of 8p, 4q, 17p and 19p, gain of 5p and high-level amplification of 1q12–q22 were detected in two or more cases. The most significant finding is the loss of 8p which was detected in eight metastatic tumours, but only in three corresponding primary tumours ($p=0.03$). This result suggests that the deletion of chromosome 8p might contribute to the development of HCC metastasis. Another interesting result is the detection of a minimum high-level amplification region at 1q12–q22 in HCC. This result provides a candidate amplification region in HCC for further study to identify amplified oncogenes related to the development or progression of HCC. Finally, this study provides a practicable model to detect specific genetic alterations related to the tumour metastasis by comparing the primary tumour and its corresponding metastatic lesion using the comparative genomic hybridization technique (Qin *et al.* 1999).

Previous studies have shown that there is a high frequency of loss of heterozygosity (LOH) on chromosome 17p13.3 in HCC (Fujimori *et al.* 1991). The minimum region of LOH on chromosome 17p13.3 in HCC has been defined within the region between D17S643 and D17S1574. Moreover, D17S926 in the minimum region of LOH has the highest frequency of LOH and its sequencing analysis has been accomplished. In this region, 6 out of 13 novel genes have been characterized.

To systematically explore the genes that contribute to cancer development and progression, a large-scale assay was performed by transfecting 29 910 individual cDNA clones derived from human placenta, foetus and

normal liver tissues into human hepatoma cells and 22 926 cDNA clones into mouse NIH 3T3 cells. Based on the results of colony formation in hepatoma cells and foci formation in NIH 3T3 cells, 3806 cDNA species (8237 clones) were found to possess the ability of either stimulating or inhibiting cell growth. Among them, 2836 (6958 clones) were known genes, 372 (384 clones) were previously unrecognized genes and 598 (895 clones) were unigenes of uncharacterized structure and function. A comprehensive analysis of the genes and the potential mechanisms for their involvement in the regulation of cell growth is provided. The genes were classified into four categories: (i) genes related to the basic cellular mechanism for growth and survival; (ii) genes related to the cellular microenvironment; (iii) genes related to host-cell systemic regulation; and (iv) genes of miscellaneous function. The extensive growth-regulatory activity of genes with such highly diversified functions suggests that cancer may be related to multiple levels of cellular and systemic controls. This study provides a direct genome-wide functional screening method. It offers a better understanding of the basic machinery of oncogenesis, including previously undescribed systemic regulatory mechanisms, and also provides a tool for gene discovery with potential clinical applications (Wan *et al.* 2004).

(b) *Oesophageal carcinoma*

Oesophageal cancer is one of the prevalent cancers in northern China. In order to explore the mode of inheritance of oesophageal cancer in a moderately high-incidence area of northern China, Zhang *et al.* (2000b) conducted a pedigree survey on 225 patients affected by oesophageal cancer in Yangquan, Shanxi Province. Segregation analysis was performed using the REGTL program of S.A.G.E. Their results showed that Mendelian autosomal recessive inheritance of a major gene that influences susceptibility to oesophageal cancer provided the best fit to the data. In the best-fitting recessive model, the frequency of the disease allele was 0.2039. There was a significant sex effect on susceptibility to the disease. The maximum cumulative probability of oesophageal cancer among males with the AA genotype was 100%, but among females it was 63.5%. The mean age at onset for both men and women was 62 years. The age-dependent penetrances for males with the AA genotype at 60 and 80 years of age were 41.6 and 95.2%, respectively, whereas for females they were 26.4 and 60.5%, respectively. Incorporating environmental risk factors such as cigarette smoking, pipe smoking, alcohol drinking, eating hot food and eating pickled vegetables into the models did not provide significant improvement of the fit of the models to these data. These results suggest a major locus underlying susceptibility to oesophageal cancer with sex-specific penetrance. Afterwards, the genetic polymorphisms of some genes, including methylenetetrahydrofolate reductase, hOGG1, CYP2A6, matrix metalloproteinases (MMPs) 2 and 3, STK15, oesophageal cancer-related gene (ECRG2) and some base excision repair (BER) genes, were considered to be associated with the risk of susceptibility, occurrence and metastasis of oesophageal

squamous cell carcinoma (ESCC; Song *et al.* 2001; Tan *et al.* 2001; Xing *et al.* 2001; Hao *et al.* 2004; Miao *et al.* 2004; Yu *et al.* 2004; Yue *et al.* 2004).

(c) *Nasopharyngeal carcinoma*

Nasopharyngeal carcinoma (NPC) is a tumour arising from the epithelial cells that cover the surface and line the nasopharynx, and is also one of the earliest cancers linked to a prior viral infection, Epstein-Barr virus (EBV). NPC has a remarkably distinctive ethnic and geographical distribution. In the year 2000, a total of 64 798 new cases were registered worldwide and more than 80% of those were reported from China, southeast Asia, especially among people of Cantonese ancestry, and some Asian countries (Lo *et al.* 2004). Early linkage analysis on Chinese sib-pair studies of NPC suggested the association of susceptibility HLA haplotypes with NPC development. The investigators hypothesized that some of the HLA antigens have reduced efficiency in activating host immune response to EBV infection, which plays a critical role in the pathogenesis of NPC. Most studies conducted among Chinese showed that individuals with HLA-A2 are at increased risk. A recent high-resolution genotyping study has detected a consistent association between NPC and the prevalent Chinese A2 subtype (HLA-A*0207), but not the prevalent Caucasian subtype (HLA-A*0201) (Hildesheim *et al.* 2002). Supported by affected sib-pair haplotype-sharing analysis and association study on HLA regions, a NPC susceptible gene/locus closely linked to the MHC region but distinct from the HLA genes was proposed (Lu *et al.* 1990).

In areas with high incidence, NPC clusters in families, which suggests that both geography and genetics may influence disease risk. Feng *et al.* (2002) report the results of a genome-wide search carried out in families at high risk of NPC from Guangdong Province, China. Parametric analyses provide evidence of linkage to the D4S405 marker on chromosome 4 with a logarithm of odds for linkage (LOD) score of 3.06 and a heterogeneity-adjusted LOD (HLOD) score of 3.21. Fine mapping with additional markers flanking D4S405 resulted in a LOD score of 3.54 and HLOD score of 3.67 for the region 4p15.1–q12. Multipoint NPL analysis gives LOD scores of 3.54 and 4.2 at D4S405 and D4S3002, respectively, which is positioned 4.5 cM away from D4S405. When EBV antibody titre was included as a covariate, the LOD scores reached 4.70 and 5.36 for D4S405 and D4S3002, respectively. These findings provide evidence of a major susceptibility locus for NPC on chromosome 4 in a subset of families. This group further screened all the genes in this region, with a focus on exons, promoters and the exon-intron boundary to identify NPC-associated mutations or functional variants. They found a novel gene (LOC344967) with a SNP –32G/A in the promoter region. This gene encodes a member of the acyl-CoA thioesterase family that plays an important role in fatty acid metabolism and is involved in the progression of various types of tumours. The –32A variant was found co-segregated with the disease phenotype in the NPC pedigrees that they previously used for the linkage

study. Moreover, this –32A variant creates an activator protein (AP-1)-binding site in the transcriptional regulatory region of LOC344967, which significantly enhances the binding of AP-1 to the promoter region and the transcription activity of the promoter *in vivo*. Moreover, the expression of LOC344967 was significantly upregulated at both mRNA and protein levels in NPC cells sharing the –32G/A genotype compared with those cells with the –32G/G genotype. Collectively, these results provide evidence that the –32A variant is a functional sequence change and may be related to NPC susceptibility in the families studied (Jiang *et al.* 2006). Xiong *et al.* (2004) collected samples from 18 families at high risk of NPC from the Hunan province in southern China, genotyped with a panel of polymorphic markers on short arms of chromosomes 3, 9, and 4p15.1–q12. A locus on 3p21 was identified to link to NPC with a maximum logarithm of odds for linkage score of 4.18. Fine mapping located the locus to a 13.6 cM region on 3p21.31–21.2, where a tumour suppressor gene cluster resided. These results identified a novel locus for NPC and provided a map location for susceptibility gene candidates, but not on the MHC region. Polymorphisms of genes for carcinogen metabolism (CYP2E1), detoxification (GSTM1) and DNA repair (XRCC1 and hOGG1) were also reported to be associated with increased risk of NPC (Hildesheim *et al.* 1997; Nazar-Stewart *et al.* 1999; Cho *et al.* 2003).

The genetic, environmental and viral causative factors, either acting alone or in combination, would lead to multiple genetic and epigenetic alterations. By the comprehensive genome-wide studies, multiple genetic defects have been identified in this EBV-associated cancer. Consistently high frequencies of genetic losses are observed on chromosomes 3p, 9, 11q, 13q, 14q and 16q, while recurrent chromosomal gains were identified on chromosome 12 (Hui *et al.* 1999; Lo *et al.* 2000). In particular, the inactivation of tumour suppressor genes on 3p, 9p and 14q appears to be a critical event, since deletion of these chromosomal regions was detected in almost all microdissected NPC samples (Lo *et al.* 2000). Genes located on chromosomes 9p21 (p14 and p16) and 3p21.3 (RASSF1A) were found to be defective due to the deletion or promoter hypermethylation. The tumour suppressor properties of p16 and RASSF1A have also been demonstrated in NPC cells (Chow *et al.* 2004). Frequent promoter hypermethylation of cancer genes is an important feature of NPC. Apart from the common tumour suppressors, such as RASSF1A and p16, genome-wide aberrant methylation disrupts multiple cellular functions through the inactivation of the retinoid signalling pathway (e.g. RARB2), endothelin-1 pathway (e.g. EDNRB), cell adhesion (e.g. E-cadherin, TSLC-1) and other novel pathways (e.g. HIN-1). Such widespread hypermethylation in the NPC genome may imply a 'methylator' phenotype of this EBV-associated cancer. Induction of epigenetic alterations of cellular genes was proposed as one of the mechanisms for enhancing the transformation of nasopharyngeal epithelial cells by EBV infection (Lu *et al.* 1990).

(d) *Leukaemia*

Leukaemia, a group of haematological malignancies characterized by abnormal proliferation, decreased apoptosis and blocked differentiation of haemopoietic stem/progenitor cells (HSPCs), is a disease involving dynamic change in the genome. Chromosomal translocation and point mutation are the major mechanisms in leukaemia which lead to production of oncogenes with dominant gain of function and tumour suppressor genes with recessive loss of function. Mutations in transcription factors, which impair haemopoietic differentiation and subsequent apoptosis, and mutations in protein tyrosine kinases, which confer a proliferative and/or survival advantage to HSPCs, represent two classes of the most frequently detected genetic anomalies (Kelly & Gilliland 2002).

Leukaemia is also another common cancer in China. In 1993, Chen *et al.* (1993a,b) reported an unusual karyotype 46,XY,t(11;17)(q23;21) without apparent rearrangement of chromosome 15 in a patient with acute promyelocytic leukaemia (APL). Similar to t(15;17) APL, all-trans retinoic acid treatment in this patient produced an early leucocytosis which was followed by a myeloid maturation, but the patient died too early to achieve remission. A rearrangement between the RAR α on 17q22 gene and a newly discovered zinc finger gene named PLZF (promyelocytic leukaemia zinc finger) on 11q23.1 was demonstrated. In a subsequent work, a 201 kb genomic DNA region containing the entire PLZF gene was sequenced (Zhang *et al.* 1999). PLZF contains six exons and five introns, and the exon organization corresponds well with protein domains. There are at least four alternative splicings (AS-I, -II, -III and -IV) within exon 1. AS-I could be detected in most tissues tested, whereas AS-II, -III and -IV were present in the stomach, testis and heart, respectively. Although donor and acceptor splicing signals at exon–intron boundaries for AS-I and exons 1–6 were classical (gt-ag), AS-II, -III and -IV had atypical splicing sites. These alternative splicings, nevertheless, maintained the ORF and may encode isoforms with the absence of important functional domains. In mRNA species without AS-I, there is a relatively long 5' UTR of 6.0 kb. A TATA box and several transcription factor-binding sites were found in the putative promoter region upstream of the transcription start site. PLZF is a well-conserved gene from *Caenorhabditis elegans* to human. PLZF paralogous sequences are found in the human genome (Zhang *et al.* 1999). PLZF encodes a potential transcription factor containing nine zinc finger motifs related to the *Drosophila* gap gene Kruppel and is expressed as at least two isoforms which differ in the sequences encoding the N-terminal region of the protein. Partial exon/intron structure of the PLZF gene flanking the break point on chromosome 11 was also established and the break point within the RAR α gene was mapped approximately 2 kb downstream of the exon encoding the 5' untranslated region and the unique A2 domain of the RAR α 2 isoform. These results demonstrated for the first time the association of a variant chromosomal translocation involving the RAR α gene with APL, further implicating the RAR α in leukaemogenesis and also suggesting an important role for PLZF, as well as

Figure 6. Schematic of feature and molecular pathogenesis of AML M2 with t(8;21). (a) Morphology of AML M2 cells. (b) Mutations in C-KIT oncoprotein reported by SIH (Wang *et al.* 2005). Letters in black represent mutations reported previously, whereas those in red represent mutations first identified in their works. (c) Model for disease pathogenesis of t(8;21) AML.

retinoic acid and its receptors in myeloid maturation (Chen *et al.* 1993a,b). In addition, this group was not only among the first to clone PML-RAR α and identify the variant isoforms of PML-RAR α (Chen *et al.* 1992; Tong *et al.* 1992; Geng *et al.* 1993), but also provides the molecular basis for the formation of PML-RAR α variants (Gu *et al.* 2002). They are also involved in the identification and functional studies of several other leukaemic fusion genes, including NUP98-HOXA9 (Nakamura *et al.* 1996), NUP98-HOXC11 (Gu *et al.* 2003), MLL-EEN (So *et al.* 1997; Liu *et al.* 2004a) and other genes.

To further address the molecular basis of leukaemia in a systematic way, the Leukemia Genome Anatomy Project (LGAP) has been launched in Shanghai Institute of Hematology (SIH) led by Zhu Chen (Zelent *et al.* 2005). This project systematically surveys the abnormalities in transcription factors and protein tyrosine kinases using functional genomics technologies and had achieved some progress in several disease models. For example, to explore the genetic

abnormalities that cooperate with AML1-ETO (AE) fusion gene to cause acute myeloid leukaemia (AML) with t(8;21), Wang *et al.* (2005) screened a number of candidate genes and identified 11 types of mutations in C-KIT gene (mC-KIT), including six previously undescribed ones among 26 out of 54 (48.1%) cases with t(8;21). To address a possible chronological order between AE and mC-KIT, they showed that, among patients with AE and mC-KIT, most leukaemic cells at disease presentation harboured both genetic alterations, whereas in three such cases investigated during complete remission, only AE, but not mC-KIT, could be detected by allele-specific PCR. Therefore, mC-KIT should be a subsequent event on the basis of t(8;21). Furthermore, induced expression of AE in U937-A/E cells significantly upregulated mRNA and protein levels of C-KIT. This may lead to an alternative way of C-KIT activation and may explain the significantly higher C-KIT expression in 81.3% of patients with t(8;21) than in patients with other leukaemias. These data strongly suggest that t(8;21)

AML follows a stepwise model in leukaemogenesis, i.e. AE represents the first, fundamental genetic hit to initiate the disease, whereas activation of the C-KIT pathway may be a second but also a crucial hit for the development of a full-blown leukaemia (figure 6).

Besides these major types of cancer in China, several genes related to other types of cancer are also investigated. For example, tamoxifen, a selective oestrogen receptor modulator, has been used in the treatment of all stages of hormone-responsive breast cancer. However, tamoxifen shows partial oestrogenic activity in the uterus and its use has been associated with an increased incidence of endometrial cancer. The molecular explanation for these observations was not known. Wu *et al.* (2005) show that tamoxifen and oestrogen have distinct but overlapping target gene profiles. Among the overlapping target genes, they identify a paired-box gene, PAX2, which is crucially involved in cell proliferation and carcinogenesis in the endometrium. Their experiments show that PAX2 is activated by oestrogen and tamoxifen in endometrial carcinomas but not in normal endometrium, and that this activation is associated with cancer-linked hypomethylation of the PAX2 promoter. Xiangyin Kong's group showed that mutations in the CYLD gene were the genetic basis for three different Chinese families with multiple familial trichoepithelioma (Zheng *et al.* 2004). Chinese scientists have revealed that the genetic polymorphisms of some genes, such as those encoding MMP-2 (Miao *et al.* 2003), methylenetetrahydrofolate reductase (Shen *et al.* 2005) and some cytokines (IL-8, IL-10 and TNF α ; Lu *et al.* 2005), could play important roles in developing gastric cancer in the Chinese population.

6. TRANSCRIPTOMICS AND PROTEOMICS RESEARCH IN CHINA

The genome is only a source of information. In order to function, it must be expressed. The essential steps of gene expression include the transcription of genes to produce RNA and the translation of RNA into proteins. The regulation of gene expression is the key process for development and for adaptation to changes in environmental conditions and thus for survival. Transcriptomics and proteomics describe the process of transcription and translation in a genome-wide range, thus providing a systematical insight into the genetic information flow.

A comprehensive knowledge of the transcriptome and proteome has become increasingly important for understanding complex diseases. Recently, Chinese scientists used the improved platform technology to address the pathogenesis of these diseases, especially different kinds of cancers. The findings from these transcriptomic and proteomic analyses provide new insights for developing improved interventions for disease (Han *et al.* 2003).

(a) Transcriptome analysis for medicine

Transcriptome analysis of haemopoiesis was the very first case of success in that direction conducted in China. To a better understanding of the regulation of normal and pathological haemopoiesis, Zhu Chen and

his colleagues in the SIH chose specific CD34-positive HSPCs as a model to initiate the first transcriptome project in China. Through the EST sequencing strategy, gene expression profiles of CD34-positive cells, derived from human umbilical cord blood, bone marrow and leukaemia, have been established (Mao *et al.* 1998; Gu *et al.* 2000). Based on this EST project, Chen's group also cloned 300 previously undefined genes expressed in CD34+ HSPCs (Zhang *et al.* 2000a). This is a pioneering milestone in transcriptomic study in China. To gain further molecular insight into the development of the haemopoietic system, Fuchu He and his colleagues of Beijing Institute of Radiation Medicine studied the gene expression profile of the 22-week foetal liver by generating ESTs and analysing the compiled expression profiles of liver at distinct developmental stages (Yu *et al.* 2001).

The hypothalamus and pituitary, together with the adrenals, constitute the major neuroendocrine axis (H-P-A axis) which is responsible for the maintenance of homeostasis and the response to the perturbations in the environment. Hu *et al.* (2000) established a primary gene expression profile for the human H-P-A axis by generating a large number of ESTs, followed by bioinformatics analysis. This work also contributed 200 full-length cDNAs of novel genes and enhanced the insights into the complexity of intercellular communications, and the physiology and pathology of the human neuroendocrine system.

To address the pathogenesis of HCC, Xu *et al.* (2001b) first reported a comprehensive characterization of gene expression profiles of hepatitis B virus-positive HCC through a large dataset of EST clusters from HCC and non-cancerous liver samples, which were then applied to a cDNA microarray. Integrated data indicated that 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. Interestingly, the gene expression pattern of HCC might represent a status of dedifferentiation of the malignant hepatocytes. Notably, the altered transcriptome profiles in HCC could be correlated to a number of chromosome regions with amplification or LOH, providing one of the underlying causes of the transcription anomaly of HCC. Gengxi Hu also used a cDNA array representing 14 000 cDNA clusters derived from EST sequencing to study the expression profiles in paired HCC and non-tumorous liver tissues (Xu *et al.* 2001a). It was conspicuous that 21 out of 38 down-regulated genes in HCC were reportedly regulated by a group of liver-enriched transcription factors, whereas 12 out of 36 upregulated genes studied previously were involved in protein translation. Furthermore, to characterize the molecular signature of metastatic HCC, Ye *et al.* (2003) analysed the expression profiles of HCC samples without or with intra-hepatic metastases through the supervised machine-learning algorithm. Their results indicated that the gene expression signature of primary HCCs with accompanying metastasis was very similar to that of their corresponding metastases, implying that genes favouring metastasis progression were initiated in the primary tumours. Among them, osteopontin was identified as a lead gene in the signature, where an osteopontin-specific antibody

effectively blocked HCC cell invasion *in vitro* and inhibited pulmonary metastasis of HCC cells in nude mice, suggesting that osteopontin could be considered as both a diagnostic marker and a potential therapeutic target for metastatic HCC.

Based on the transcriptomics studies, several novel genes related to the HCC were further investigated. Xu *et al.* (2003) identified a liver-specific gene that encodes a novel zona pellucida (ZP) domain-containing protein named liver-specific ZP domain-containing protein (LZP). Interestingly, human LZP is expressed specifically in liver in the 23 tissues examined, and its mouse counterpart was detected at a very early stage during embryo development. Importantly, LZP is down-regulated in HCC and HCC cell lines; meanwhile, the decreased level of hLZP mRNA could, at least in some HCC samples, be related to the methylation status of the putative LZP promoter. Huang *et al.* (2003) identified a specific splicing variant (SVH-B) of SVH, a novel human armadillo repeat protein that is upregulated in HCCs. SVH-B may accelerate growth rate and tumorigenicity of human liver cell line QSG-7701 in nude mice, whereas its antisense oligodeoxynucleotides can inhibit the growth of HCC cell line BEL-7404 by inducing apoptosis, implying that SVH-B may play an important role in hepatocarcinogenesis. Yan *et al.* (2004) found that signal regulatory protein alpha 1 (SIRP α 1), a member of the SIRP family, was decreased in HCC. Exogenous expression of wild-type SIRP α 1 resulted in the suppression of tumour cell growth of the HCC cell line Huh7 both *in vitro* and *in vivo*, possibly through reducing ERK1 activation and inhibition of the NF- κ B pathway.

To identify genes that are differentially expressed in human ESCC, Luo *et al.* (2004) analysed gene expression profiles in ESCC using cDNA microarray. The resulting data revealed that genes involved in squamous cell differentiation were coordinately down-regulated, including genes for annexin I, small proline-rich proteins, calcium-binding S100 proteins, etc. Interestingly, most of the downregulated genes encoded Ca²⁺-binding or -modulating proteins that constitute the cell envelope. Moreover, genes associated with invasion or proliferation were upregulated, including genes for fibronectin, cathepsin B and KRT17. The data provide new insights into the role of squamous cell differentiation-associated genes in ESCC initiation and progression. In addition, some deregulated genes in ESCC, such as NMES1 and PTTG, were further shown to be related to ESCC (Zhou *et al.* 2002, 2005).

(b) Proteomic analysis for medicine

Besides the transcriptomics strategies applied to medical research, proteomic approaches were recently employed to investigate the qualitative and quantitative disease-related proteomes. Li *et al.* (2004b) analysed the solubilized proteins from microdissected HCC and non-HCC hepatocytes by laser capture microdissection technique, through two-dimensional liquid chromatography tandem mass spectrometry (2D-LC-MS/MS) alone or coupled with cleavable isotope-coded affinity tag labelling technology. A total of 644 proteins were qualitatively identified, and 261 proteins were unambiguously quantitated. To better

understand the mechanism underlying the HCC metastasis and to search for potential markers for HCC prognosis, Ding *et al.* (2004) conducted differential proteomics analysis on two well-established HCC cell strains with high and low metastatic potentials, MHCC97-H and MHCC97-L. Interestingly, cytokeratin 19 (CK19) was identified to be overexpressed in MHCC97-H as compared with MHCC97-L, which was further confirmed in human HCC specimens, suggesting that serum CK19 level might reflect the pathological progression in some HCCs and may be a useful marker for predicting tumour metastasis.

He *et al.* (2004) used a proteomic approach to globally analyse the protein profiles of paired surgical specimens of primary gastric adenocarcinoma and non-tumour mucosa to identify specific disease-associated proteins as potential clinical biomarkers. Some of these deregulated proteins, including cytoskeleton proteins, heat shock proteins and glycolytic enzymes, were observed in tumour tissues. More interestingly, a human stomach-specific protein, 18 kDa antrum mucosa protein, was found to be dramatically underexpressed in cancer tissues, implicating a possible special pathological role for this protein in gastric carcinogenesis.

7. CONCLUSION AND PERSPECTIVES

We apologize to authors whose distinguished works have not been cited in this review. We would like to say that so far our understanding of the complexity of medicine and the biodiversity of China is still at a relatively early stage and genes underlying complex diseases remain largely unknown. However, tremendous advances in the construction of genomic research platforms and in the dissection of disease pathogenesis have been seen. With great efforts of scientists, we believe that we can further elucidate the genotype of the phenotype. With the functional genomic research, we can further unveil the roles that genes play in disease pathogenesis and design molecular mechanism-based therapies which may ultimately improve clinical outcome of patients.

REFERENCES

- Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J. & Donnelly, P. 2005 A haplotype map of the human genome. *Nature* **437**, 1299–1320. (doi:10.1038/nature04226)
- Bruzzzone, R., White, T. W. & Paul, D. L. 1996 Connections with connexins: the molecular basis of direct intercellular signaling. *Eur. J. Biochem.* **238**, 1–27. (doi:10.1111/j.1432-1033.1996.0001q.x)
- Bu, L. *et al.* 2002 Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nat. Genet.* **31**, 276–278. (doi:10.1038/ng921)
- Caulfield, M. *et al.* 2003 MRC British Genetics of Hypertension Study. Genome-wide mapping of human loci for essential hypertension. *Lancet* **361**, 2118–2123. (doi:10.1016/S0140-6736(03)13722-1)
- Chen, S. J., Chen, Z., Chen, A., Tong, J. H., Dong, S., Wang, Z. Y., Waxman, S. & Zelent, A. 1992 Occurrence of distinct PML-RAR- α fusion gene isoforms in patients

- with acute promyelocytic leukemia detected by reverse transcriptase/polymerase chain reaction. *Oncogene* **7**, 1223–1232.
- Chen, S. J., Zelent, A., Tong, J. H., Yu, H. Q., Wang, Z. Y., Derre, J., Berger, R., Waxman, S. & Chen, Z. 1993a Rearrangements of the retinoic acid receptor alpha and promyelocytic leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a patient with acute promyelocytic leukemia. *J. Clin. Invest.* **91**, 2260–2267.
- Chen, Z., Brand, N. J., Chen, A., Chen, S. J., Tong, J. H., Wang, Z. Y., Waxman, S. & Zelent, A. 1993b Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *EMBO J.* **12**, 1161–1167.
- Chen, Z., Li, W., Yu, M., Xiong, H. & Chen, S. J. 1998 Opportunities and challenges in Human Genome Project. *Chem. Life* **18**, 5–13.
- Chen, Y. *et al.* 2003a Association between genetic variation of CACNA1H and childhood absence epilepsy. *Ann. Neurol.* **54**, 239–243. (doi:10.1002/ana.10607)
- Chen, Y. H. *et al.* 2003b KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* **299**, 251–254. (doi:10.1126/science.1077771)
- Chen, W. Y., Shi, Y. Y., Zheng, Y. L., Zhao, X. Z., Zhang, G. J., Chen, S. Q., Yang, P. D. & He, L. 2004 Case-control study and transmission disequilibrium test provide consistent evidence for association between schizophrenia and genetic variation in the 22q11 gene ZDHHC8. *Hum. Mol. Genet.* **13**, 2991–2995. (doi:10.1093/hmg/ddh322)
- Chen, J., Zhao, J., Huang, J., Su, S., Qiang, B. & Gu, D. 2005a –717A>G polymorphism of human C-reactive protein gene associated with coronary heart disease in ethnic Han Chinese: the Beijing atherosclerosis study. *J. Mol. Med.* **83**, 72–78. (doi:10.1007/s00109-004-0585-5)
- Chen, Q., He, G., Chen, Q., Wu, S., Xu, Y., Feng, G., Li, Y., Wang, L. & He, L. 2005b A case-control study of the relationship between the metabotropic glutamate receptor 3 gene and schizophrenia in the Chinese population. *Schizophr. Res.* **73**, 21–26. (doi:10.1016/j.schres.2004.07.002)
- Cho, E. Y. *et al.* 2003 Nasopharyngeal carcinoma and genetic polymorphisms of NA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol. Biomarkers Prev.* **12**, 1100–1104.
- Chow, L. S., Lo, K. W., Kwong, J., To, K. F., Tsang, K. S., Lam, C. W., Dammann, R. & Huang, D. P. 2004 RASSF1A is a target tumor suppressor from 3p21.3 in nasopharyngeal carcinoma. *Int. J. Cancer* **109**, 839–847. (doi:10.1002/ijc.20079)
- Chumakov, I. *et al.* 2002 Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc. Natl Acad. Sci. USA* **99**, 13 675–13 680. (doi:10.1073/pnas.182412499)
- Denoyelle, F., Lina-Granade, G., Plauchu, H., Bruzzone, R., Chaib, H., Levi-Acobas, F., Weil, D. & Petit, C. 1998 Connexin 26 gene linked to a dominant deafness. *Nature* **393**, 319–320. (doi:10.1038/30639)
- Ding, S. J. *et al.* 2004 From proteomic analysis to clinical significance overexpression of cytokeratin 19 correlates with hepatocellular carcinoma metastasis. *Mol. Cell Proteomics* **3**, 73–81. (doi:10.1074/mcp.M300094-MCP200)
- Du, W. *et al.* 2001 Confirmation of susceptibility gene loci on chromosome 1 in northern China Han families with type 2 diabetes. *Chin. Med. J.* **114**, 876–878.
- Dulbecco, R. 1986 A turning point in cancer research: sequencing the human genome. *Science* **231**, 1055–1056. (doi:10.1126/science.3945817)
- Feng, B. J. *et al.* 2002 Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat. Genet.* **31**, 395–399.
- Fujimori, M. *et al.* 1991 Allelotype study of primary hepatocellular carcinoma. *Cancer Res.* **51**, 89–93.
- Gao, B. *et al.* 2001 Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat. Genet.* **28**, 386–388. (doi:10.1038/ng577)
- Ge, D. *et al.* 2003 Linkage analysis of 2q14–q23 and 5q32 with blood pressure quantitative traits in Chinese sib pairs. *J. Hypertens* **21**, 305–310. (doi:10.1097/00004872-200302000-00021)
- Geng, J. P., Tong, J. H., Dong, S., Wang, Z. Y., Chen, S. J., Chen, Z., Zelent, A., Berger, R. & Larsen, C. J. 1993 Localization of the chromosome 15 breakpoints and expression of multiple PML-RAR alpha transcripts in acute promyelocytic leukemia: a study of 28 Chinese patients. *Leukemia* **7**, 20–26.
- Gong, X. *et al.* 2004 Association between the FOXP2 gene and autistic disorder in Chinese population. *Am. J. Med. Genet B Neuropsychiatr. Genet* **127**, 113–116. (doi:10.1002/ajmg.b.20162)
- Goodenough, D. A., Goliger, J. A. & Paul, D. L. 1996 Connexins, connexons and intercellular communication. *Annu. Rev. Biochem.* **65**, 475–502. (doi:10.1146/annurev.bi.65.070196.002355)
- Gu, J. *et al.* 2000 Gene expression in CD34+ cells from normal bone marrow and leukemic origins. *Hematol. J.* **1**, 206–217. (doi:10.1038/sj.thj.6200020)
- Gu, B. W. *et al.* 2002 Variant-type PML-RAR(α) fusion transcript in acute promyelocytic leukemia: use of a cryptic coding sequence from intron 2 of the RAR(α) gene and identification of a new clinical subtype resistant to retinoic acid therapy. *Proc. Natl Acad. Sci. USA* **99**, 7640–7645. (doi:10.1073/pnas.112194799)
- Gu, B. W. *et al.* 2003 Major form of NUP98/HOXC11 fusion in adult AML with t(11;12)(p15;q13) translocation exhibits aberrant trans-regulatory activity. *Leukemia* **17**, 1858–1864. (doi:10.1038/sj.leu.2403036)
- Han, Z. G., Zhao, G. P. & Chen, Z. 2003 Transcriptome study in China. *C. R. Biol.* **326**, 949–957. (doi:10.1016/j.crv.2003.09.021)
- Hanahan, D. & Weinberg, R. A. 2000 The hallmarks of cancer. *Cell* **100**, 57–70. (doi:10.1016/S0092-8674(00)81683-9)
- Hao, B. *et al.* 2004 Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. *Cancer Res.* **64**, 4378–4384. (doi:10.1158/0008-5472.CAN-04-0372)
- He, Q. Y., Cheung, Y. H., Leung, S. Y., Yuen, S. T., Chu, K. M. & Chiu, J. F. 2004 Diverse proteomic alterations in gastric adenocarcinoma. *Proteomics* **4**, 3276–3287. (doi:10.1002/pmic.200300916)
- He, J. *et al.* 2005 Major causes of death among men and women in China. *J. Med.* **353**, 1124–1134.
- Hildesheim, A. *et al.* 1997 CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J. Natl Cancer Inst.* **89**, 1207–1212. (doi:10.1093/jnci/89.16.1207)
- Hildesheim, A. *et al.* 2002 Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J. Natl Cancer Inst.* **94**, 1780–1789.
- Hu, R. M. *et al.* 2000 Gene expression profiling in the human hypothalamus–pituitary–adrenal axis and full-length cDNA cloning. *Proc. Natl Acad. Sci. USA* **97**, 9543–9548. (doi:10.1073/pnas.160270997)
- Huang, R., Xing, Z., Luan, Z., Wu, T., Wu, X. & Hu, G. 2003 A specific splicing variant of SVH, a novel human armadillo repeat protein, is up-regulated in hepatocellular carcinomas. *Cancer Res.* **63**, 3775–3782.
- Hui, A. B. *et al.* 1999 Detection of recurrent chromosomal gains and losses in primary nasopharyngeal carcinoma

- by comparative genomic hybridisation. *Int. J. Cancer* **82**, 498–503. (doi:10.1002/(SICI)1097-0215(19990812)82:4<498::AID-IJC5>3.0.CO;2-S)
- Jiang, R. C. *et al.* 2006 A functional variant in the transcriptional regulatory region of gene LOC344967 cosegregates with disease phenotype in familial nasopharyngeal carcinoma. *Cancer Res.* **66**, 693–700. (doi:10.1158/0008-5472.CAN-05-2166)
- Kelly, L. M. & Gilliland, D. G. 2002 Genetic's of myeloid leukemias. *Annu. Rev. Genomics Hum. Genet.* **3**, 179–198. (doi:10.1146/annurev.genom.3.032802.115046)
- Kelsell, D. P., Dunlop, J., Stevens, H. P., Lench, N. J., Liang, J. N., Parry, G., Mueller, R. F. & Leigh, I. M. 1997 Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* **387**, 80–83. (doi:10.1038/387080a0)
- Lander, E. S. *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)
- Le, W. D., Xu, P., Jankovic, J., Jiang, H., Appel, S. H., Smith, R. G. & Vassilatis, D. K. 2003 Mutations in NR4A2 associated with familial Parkinson disease. *Nat. Genet.* **33**, 85–89. (doi:10.1038/ng1066)
- Li, B., Ge, D., Wang, Y., Zhao, W., Zhou, X., Gu, D. & Chen, R. 2004a Lipoprotein lipase gene polymorphisms and blood pressure levels in the northern Chinese Han population. *Hypertens. Res.* **27**, 373–378. (doi:10.1291/hyres.27.373)
- Li, C. *et al.* 2004b Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **3**, 399–409. (doi:10.1074/mcp.M300133-MCP200)
- Lifton, R. P., Gharavi, A. G. & Geller, D. S. 2001 Molecular mechanisms of human hypertension. *Cell* **104**, 545–556. (doi:10.1016/S0092-8674(01)00241-0)
- Liu, L., Xiang, K., Zheng, T., Zhang, R., Li, M. & Li, J. 2003 Co-inheritance of specific genotypes of HSPG and ApoE gene increases risk of type 2 diabetic nephropathy. *Mol. Cell. Biochem.* **254**, 353–358. (doi:10.1023/A:1027364121738)
- Liu, H., Chen, B., Xiong, H., Huang, Q. H., Zhang, Q. H., Wang, Z. G., Li, B. L., Chen, Z. & Chen, S. J. 2004a Functional contribution of EEN to leukemogenic transformation by MLL-EEN fusion protein. *Oncogene* **23**, 3385–3394. (doi:10.1038/sj.onc.1207402)
- Liu, X. *et al.* 2004b Association of DAAO with schizophrenia in the Chinese population. *Neurosci. Lett.* **369**, 228–233. (doi:10.1016/j.neulet.2004.07.078)
- Lo, K. W., Teo, P. M., Hui, A. B., To, K. F., Tsang, Y. S., Chan, S. Y., Mak, K. F., Lee, J. C. & Huang, D. P. 2000 High resolution allelotype of microdissected primary nasopharyngeal carcinoma. *Cancer Res.* **60**, 3348–3353.
- Lo, K. W., To, K. F. & Huang, D. P. 2004 Focus on nasopharyngeal carcinoma. *Cancer Cell* **5**, 423–428. (doi:10.1016/S1535-6108(04)00119-9)
- Lu, S. J. *et al.* 1990 Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* **346**, 470–471. (doi:10.1038/346470a0)
- Lu, Y. *et al.* 2003 A novel mutation of keratin 9 in epidermolytic palmoplantar keratoderma combined with knuckle pads. *Am. J. Med. Genet. A* **120**, 345–349. (doi:10.1002/ajmg.a.20090)
- Lu, W., Pan, K., Zhang, L., Lin, D., Miao, X. & You, W. 2005 Genetic polymorphisms of interleukin (IL)-1B, IL-1RN, IL-8, IL-10 and tumor necrosis factor α and risk of gastric cancer in a Chinese population. *Carcinogenesis* **26**, 631–636. (doi:10.1093/carcin/bgh349)
- Luo, T. H., Zhao, Y., Li, G., Yuan, W. T., Zhao, J. J., Chen, J. L., Huang, W. & Luo, M. 2001 A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* **44**, 501–506. (doi:10.1007/s001250051649)
- Luo, A. *et al.* 2004 Discovery of Ca⁽²⁺⁾-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. *Oncogene* **23**, 1291–1299. (doi:10.1038/sj.onc.1207218)
- Mao, M. *et al.* 1998 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**, 8175–8180. (doi:10.1073/pnas.95.14.8175)
- Mein, C. A., Caulfield, M. J., Dobson, R. J. & Munroe, P. B. 2004 Genetics of essential hypertension. *Hum. Mol. Genet.* **13**, R169–R175. (doi:10.1093/hmg/ddh078)
- Miao, X., Yu, C., Tan, W., Xiong, P., Liang, G., Lu, W. & Lin, D. 2003 A functional polymorphism in the matrix metalloproteinase-2 gene promoter (–1306C/T) is associated with risk of development but not metastasis of gastric cardia adenocarcinoma. *Cancer Res.* **63**, 3987–3990.
- Miao, X., Sun, T., Wang, Y., Zhang, X., Tan, W. & Lin, D. 2004 Functional STK15 Phe31Ile polymorphism is associated with the occurrence and advanced disease status of esophageal squamous cell carcinoma. *Cancer Res.* **64**, 2680–2683. (doi:10.1158/0008-5472.CAN-04-0651)
- Nakamura, T. *et al.* 1996 Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat. Genet.* **12**, 154–158. (doi:10.1038/ng0296-154)
- Nazar-Stewart, V., Vaughan, T. L., Burt, R. D., Chen, C., Berwick, M. & Swanson, G. M. 1999 Glutathione S-transferase M1 and susceptibility to nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **8**, 547–551.
- Qi, Y. *et al.* 2004 A deletion mutation in the beta-A1/A3 crystallin gene (CRYBA1/A3) is associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Hum. Genet.* **114**, 192–197. (doi:10.1007/s00439-003-1049-7)
- Qiang, B. 2004 Human genome research in China. *J. Mol. Med.* **82**, 214–222. (doi:10.1007/s00109-003-0515-y)
- Qin, L. X., Tang, Z. Y., Sham, J. S., Ma, Z. C., Ye, S. L., Zhou, X. D., Wu, Z. Q., Trent, J. M. & Guan, X. Y. 1999 The association of chromosome 8p deletion and tumor metastasis in human hepatocellular carcinoma. *Cancer Res.* **59**, 5662–5665.
- Richard, G., Smith, L. E., Bailey, R. A., Itin, P., Hohl, D., Epstein Jr, E. H., DiGiovanna, J. J., Compton, J. G. & Bale, S. J. 1998 Mutations in the human connexin gene GJB3 cause erythrokeratoderma variabilis. *Nat. Genet.* **20**, 366–369. (doi:10.1038/3840)
- Roberts, L., Davenport, R. J., Pennisi, E. & Marshall, E. 2001 A history of the Human Genome Project. *Science* **291**, 1195. (doi:10.1126/science.291.5507.1195)
- Sanders, A. R., Duan, J. & Gejman, P. V. 2004 Complexities in psychiatric genetics. *Int. Rev. Psychiatry* **16**, 284–293. (doi:10.1080/09540260400014393)
- Schumacher, J. *et al.* 2004 Examination of G72 and D-amino acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder. *Mol. Psychiatry* **9**, 203–207. (doi:10.1038/sj.mp.4001421)
- Schunkert, H., Hense, H. W., Holmer, S. R., Stender, M., Perz, S., Keil, U., Lorell, B. H. & Riegger, G. 1994 Association between a deletion polymorphism of the angiotensin-converting enzyme and left ventricular hypertrophy. *N. Engl. J. Med.* **330**, 1634–1638. (doi:10.1056/NEJM199406093302302)

- Shen, H. *et al.* 2005 Methylenetetrahydrofolate reductase polymorphisms/haplotypes and risk of gastric cancer: a case-control analysis in China. *Oncol. Rep.* **13**, 355–360.
- So, C. W., Caldas, C., Liu, M. M., Chen, S. J., Huang, Q. H., Gu, L. J., Sham, M. H., Wiedemann, L. M. & Chan, L. C. 1997 *EEN* encodes for a member of a new family of proteins containing an Src homology 3 domain and is the third gene located on chromosome 19p13 that fuses to *MLL* in human leukaemia. *Proc. Natl Acad. Sci. USA* **94**, 2563–2568. (doi:10.1073/pnas.94.6.2563)
- Song, C., Xing, D., Tan, W., Wei, Q. & Lin, D. 2001 Methylenetetrahydrofolate reductase polymorphisms increase risk of esophageal squamous cell carcinoma in a Chinese population. *Cancer Res.* **61**, 3272–3275.
- Steel, K. P. 1998 One connexin, two diseases. *Nat. Genet.* **20**, 319–320. (doi:10.1038/3781)
- Tan, W., Chen, G. F., Xing, D. Y., Song, C. Y., Kadlubar, F. F. & Lin, D. X. 2001 Frequency of CYP2A6 gene deletion and its relation to risk of lung and esophageal cancer in the Chinese population. *Int. J. Cancer* **95**, 96–101. (doi:10.1002/1097-0215(20010320)95:2<96::AID-IJC1017>3.0.CO;2-2)
- Tang, B., Li, H., Xia, K., Jiang, H., Pan, Q., Shen, L., Long, Z., Zhao, G. & Cai, F. 2004 A novel mutation in *KCNQ2* gene causes benign familial neonatal convulsions in a Chinese family. *J. Neurol. Sci.* **221**, 31–34. (doi:10.1016/j.jns.2004.03.001)
- Tong, J. H. *et al.* 1992 Molecular rearrangements of the *MYL* gene in acute promyelocytic leukemia (APL, M3) define a breakpoint cluster region as well as some molecular variants. *Oncogene* **7**, 311–316.
- Wan, D. *et al.* 2004 Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc. Natl Acad. Sci. USA* **101**, 15 724–15 729. (doi:10.1073/pnas.0404089101)
- Wang, Q. *et al.* 1996 Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause cardiac arrhythmias. *Nat. Genet.* **12**, 17–23. (doi:10.1038/ng0196-17)
- Wang, Y. Y. *et al.* 2005 *AML1*-ETO and *c-KIT* mutation/overexpression in t(8;21) leukemia: implication in stepwise leukemogenesis and response to Gleevec. *Proc. Natl Acad. Sci. USA* **102**, 1104–1109. (doi:10.1073/pnas.0408831102)
- Wu, L. Q. *et al.* 2004 Confirmation and refinement of a genetic locus for disseminated superficial actinic porokeratosis (DSAP1) at 12q23.2-24.1. *Br. J. Dermatol.* **150**, 999–1004. (doi:10.1111/j.1365-2133.2004.05912.x)
- Wu, H. *et al.* 2005 Hypomethylation-linked activation of *PAX2* mediates tamoxifen-stimulated endometrial carcinogenesis. *Nature* **438**, 981–987. (doi:10.1038/nature04225)
- Xi, Z. R., Qin, W., Yang, Y. F., He, G., Gao, S. H., Ren, M. S., Peng, Y. W., Zhang, Z. & He, L. 2004 Transmission disequilibrium analysis of the *GSN* gene in a cohort of family trios with schizophrenia. *Neurosci. Lett.* **372**, 200–203. (doi:10.1016/j.neulet.2004.09.041)
- Xia, J. H. *et al.* 1998 Mutations in the gene encoding gap junction protein beta-3 associated with autosomal dominant hearing impairment. *Nat. Genet.* **20**, 370–373. (doi:10.1038/3845)
- Xia, J. H. 2000 Identification of a locus for disseminated superficial actinic porokeratosis (DSAP) at Chromosome 12q23.2-24.1. *J. Invest. Dermatol.* **114**, 1071–1074. (doi:10.1046/j.1523-1747.2000.00978.x)
- Xia, K. *et al.* 2002 A novel locus (DSAP2) for disseminated superficial actinic porokeratosis maps to chromosome 15q25.1–26.1. *Br. J. Dermatol.* **147**, 650–654. (doi:10.1046/j.1365-2133.2002.05058.x)
- Xiao, S. *et al.* 2001 Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in *DSPP*. *Nat. Genet.* **27**, 201–204. (doi:10.1038/84848)
- Xing, D. Y., Tan, W., Song, N. & Lin, D. X. 2001 Ser326Cys polymorphism in hOGG1 gene and risk of esophageal cancer in Chinese population. *Int. J. Cancer* **95**, 140–143. (doi:10.1002/1097-0215(20010520)95:3<140::AID-IJC1024>3.0.CO;2-2)
- Xing, Q. H. *et al.* 2003 A gene locus responsible for dyschromatosis symmetrica hereditaria (DSH) maps to chromosome 6q24.2-q25.2. *Am. J. Hum. Genet.* **73**, 377–382. (doi:10.1086/377007)
- Xiong, W. *et al.* 2004 A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res.* **64**, 1972–1974. (doi:10.1158/0008-5472.CAN-03-3253)
- Xu, L. *et al.* 2001a Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma. *Cancer Res.* **61**, 3176–3181.
- Xu, X. R. *et al.* 2001b 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**, 15 089–15 094. (doi:10.1073/pnas.241522398)
- Xu, Z. G. *et al.* 2003 A novel liver-specific zona pellucida domain containing protein that is expressed rarely in hepatocellular carcinoma. *Hepatology* **38**, 735–744. (doi:10.1053/jhep.2003.50340)
- Xu, Q. *et al.* 2004a Association study of an SNP combination pattern in the dopaminergic pathway in paranoid schizophrenia: a novel strategy for complex disorders. *Mol. Psychiatry* **9**, 510–521. (doi:10.1038/sj.mp.4001472)
- Xu, Q. *et al.* 2004b Interaction and relationship between angiotensin converting enzyme gene and environmental factors predisposing to essential hypertension in Mongolian population of China. *Biomed. Environ. Sci.* **17**, 177–186.
- Yan, H. X. *et al.* 2004 Negative regulation of hepatocellular carcinoma cell growth by signal regulatory protein alpha1. *Hepatology* **40**, 618–628. (doi:10.1002/hep.20360)
- Yang, X., She, C., Guo, J., Yu, A. C., Lu, Y., Shi, X., Feng, G. & He, L. 2000 A locus for brachydactyly type A-1 maps to chromosome 2q35-q36. *Am. J. Hum. Genet.* **66**, 892–903. (doi:10.1086/302806)
- Yang, W. J. *et al.* 2003 Evidence for linkage and association of the markers near the *LPL* gene with hypertension in Chinese families. *J. Med. Genet.* **40**, e57. (doi:10.1136/jmg.40.5.e57)
- Yang, Y. 2004a Mutations in *SCN9A*, encoding a sodium channel alpha subunit, in patients with primary erythromalgia. *J. Med. Genet.* **41**, 171–174. (doi:10.1136/jmg.2003.012153)
- Yang, Y. *et al.* 2004b Identification of a *KCNE2* gain-of-function mutation in patients with familial atrial fibrillation. *Am. J. Hum. Genet.* **75**, 899–905. (doi:10.1086/425342)
- Ye, Q. H. *et al.* 2003 Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat. Med.* **9**, 416–423. (doi:10.1038/nm843)
- Yu, Y. *et al.* 2001 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**, 1392–1403. (doi:10.1101/gr.175501)
- Yu, C., Zhou, Y., Miao, X., Xiong, P., Tan, W. & Lin, D. 2004 Functional haplotypes in the promoter of matrix metalloproteinase-2 predict risk of the occurrence and metastasis of esophageal cancer. *Cancer Res.* **64**, 7622–7628. (doi:10.1158/0008-5472.CAN-04-1521)
- Yue, C. M., Bi, M. X., Tan, W., Deng, D. J., Zhang, X. Y., Guo, L. P., Lin, D. X. & Lu, S. H. 2004 Short esophageal cancer-related tandem repeat polymorphism in a novel

- gene (ECRG2) implicates susceptibility to esophageal cancer in Chinese population. *Int. J. Cancer* **108**, 232–236. (doi:10.1002/ijc.11560)
- Zelent, A., Petrie, K., Chen, Z., Lotan, R., Lubbert, M., Tallman, M. S., Ohno, R., Degos, L. & Waxman, S. 2005 Molecular target-based treatment of human cancer: summary of the 10th International Conference on Differentiation Therapy. *Cancer Res.* **65**, 1117–1123. (doi:10.1158/0008-5472.CAN-04-3603)
- Zhang, T. *et al.* 1999 Genomic sequence, structural organization, molecular evolution, and aberrant rearrangement of promyelocytic leukemia zinc finger gene. *Proc. Natl Acad. Sci. USA* **96**, 11 422–11 427. (doi:10.1073/pnas.96.20.11422)
- Zhang, Q. H. *et al.* 2000a 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**, 1546–1560. (doi:10.1101/gr.140200)
- Zhang, W., Bailey-Wilson, J. E., Li, W., Wang, X., Zhang, C., Mao, X., Liu, Z., Zhou, C. & Wu, M. 2000b Segregation analysis of esophageal cancer in a moderately high-incidence area of northern China. *Am. J. Hum. Genet.* **67**, 110–119. (doi:10.1086/302970)
- Zhang, X. *et al.* 2001 DSPP mutation in dentinogenesis imperfecta Shields type II. *Nat. Genet.* **27**, 151–152. (doi:10.1038/84765)
- Zhang, Y. *et al.* 2004 Positive association of the human frizzled 3 (FZD3) gene haplotype with schizophrenia in Chinese Han population. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **129**, 16–19.
- Zhang, Z. H. *et al.* 2005 A mutation in SART3 gene in a Chinese pedigree with disseminated superficial actinic porokeratosis. *Br. J. Dermatol.* **152**, 658–663. (doi:10.1111/j.1365-2133.2005.06443.x)
- Zhao, J. Y. 2005 An autosomal genomic scan for loci linked to type 2 diabetes in northern Han Chinese. *J. Mol. Med.* **83**, 209–215. (doi:10.1007/s00109-004-0587-3)
- Zheng, G. *et al.* 2004 CYLD mutation causes multiple familial trichoepithelioma in three Chinese families. *Hum. Mutat.* **23**, 400. (doi:10.1002/humu.9231)
- Zhou, J. *et al.* 2002 A novel gene, NMES1, downregulated in human esophageal squamous cell carcinoma. *Int. J. Cancer* **101**, 311–316. (doi:10.1002/ijc.10600)
- Zhou, C. *et al.* 2005 Overexpression of human pituitary tumor transforming gene (hPTTG), is regulated by beta-catenin/TCF pathway in human esophageal squamous cell carcinoma. *Int. J. Cancer* **113**, 891–898. (doi:10.1002/ijc.20642)
- Zhu, D. L. *et al.* 2001 Linkage of hypertension to chromosome 2q14-q23 in Chinese families. *J. Hypertens* **19**, 55–61. (doi:10.1097/00004872-200101000-00008)