DATABASES

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Novel LOVD Databases for Hereditary Breast Cancer and Colorectal Cancer Genes in the Chinese Population

Min Pan,[1†] Peikuan Cong,[1†] Yue Wang,[2†] Changsong Lin,[1] Ying Yuan,[3] Jian Dong,[4] Santasree Banerjee,[1] Tao Zhang,[5] Yanling Chen,[1] Ting Zhang,[1] Mingqing Chen,[4] Peter Hu,[2] Shu Zheng,[3*] Jin Zhang,[6*] and Ming Qi[1,5,7*]

[1]Zhejiang University School of Medicine, Hangzhou, Zhejiang, China; [2]Molecular Genetic Technology Program, University of Texas, M. D. Anderson Cancer Center, Houston, Texas; [3]Zhejiang University School of Medicine Second Affiliated Hospital, Hangzhou, Zhejiang, China; [4]First Affiliated Hospital of Kunming Medical College, Kunming, Yunnan, China; [5]James D. Watson Institute of Genome Science, Zhejiang University, Hangzhou, China; [6]Tianjin Medical University Cancer Hospital and Research Institute, China; [7]Department of Pathology and Laboratory Medicine, University of Rochester, Rochester, New York

Communicated by Mark H. Paalman

Received 1 May 2011; accepted revised manuscript 28 July 2011.
Published online 7 September 2011 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.21588

ABSTRACT: The Human Variome Project (HVP) is an international consortium of clinicians, geneticists, and researchers from over 30 countries, aiming to facilitate the establishment and maintenance of standards, systems, and infrastructure for the worldwide collection and sharing of all genetic variations effecting human disease. The HVP-China Node will build new and supplement existing databases of genetic diseases. As the first effort, we have created a novel variant database of *BRCA1* and *BRCA2*, mismatch repair genes (MMR), and *APC* genes for breast cancer, Lynch syndrome, and familial adenomatous polyposis (FAP), respectively, in the Chinese population using the Leiden Open Variation Database (LOVD) format. We searched PubMed and some Chinese search engines to collect all the variants of these genes in the Chinese population that have already been detected and reported. There are some differences in the gene variants between the Chinese population and that of other ethnicities. The database is available online at http://www.genomed.org/LOVD/. Our database will appear to users who survey other LOVD databases (e.g., by Google search, or by NCBI GeneTests search). Remote submissions are accepted, and the information is updated monthly.

Hum Mutat 32:1335–1340, 2011. © 2011 Wiley Periodicals, Inc.

KEY WORDS: breast cancer; Lynch syndrome; FAP; LOVD; tumor suppressor gene

Introduction

"Variome," a word combining "variant + ome", has been created to describe all the genetic variants. The Human Variome Project (HVP) is an international consortium of clinicians, geneticists, and researchers from over 30 countries, and seeks to facilitate the establishment and maintenance of standards, systems, and infrastructure for the worldwide collection and sharing of all genetic variations effecting human disease (http://www.humanvariomeproject.org/). The ultimate goal of the project is to integrate the systematic collection and sharing of genetic variant information into routine clinical practice. With a population of some 1.4 billion people, a quarter of the world population, there is a huge amount of genetic diversity within China. Our newly established HVP-China Node commit to leverage that diversity to build new and supplement existing databases that catalogue genetic variation within genes implicated in hundreds of genetic diseases [Cyranoski, 2011].

Breast cancer is a common malignant tumor for women in the world [Pasche, 2008]. According to statistical data from the Ministry of Health of China, the incidence of breast cancer increased from 7.7 per 100,000 of Chinese women in 1998 to 11.1 per 100,000 of Chinese women in 2008, with mortality from breast cancer at 5.9 per 100,000 women. Colorectal cancer is one of the five most common carcinomas with a mortality of 8.19 per 100,000 in China [MOH, 2009].

BRCA1 (MIM# 113705) and *BRCA2* (MIM# 600185) are the most common genes associated with hereditary and early-onset breast cancer; these two genes account for 20–25% of all the familial breast cancer cases in the world [Pasche, 2008]. The *BRCA1* gene product is a protein consisting of 1,863 amino acid residues [Miki et al., 1994]. Since the known functions of the *BRCA1* protein include DNA repair, cell cycle control, and protein ubiquitination, it is not surprising that *BRCA1* plays an important role in carcinoma development [Parvin, 2004]. *BRCA2* also classified as a tumor suppressor gene, was mapped to chromosome 13q12-q13 [Wooster et al., 1994]. Previous reports have suggested that the inactivation of the *BRCA2* gene can lead to genetic instability and thus promote tumor initiation [Kinzler and Vogelstein, 1997].

Lynch syndrome is caused by heterozygous variants in mismatch repair (*MMR*) genes. Variants in the *MSH2* (MIM# 609309), *MLH1* (MIM# 120436), *MSH6* (MIM# 600678), *MLH3* (MIM# 604395), *PMS2* (MIM# 600259), and *PMS1* (MIM# 600258) genes have been identified in Lynch syndrome. *MSH2* maps on locus 2p22-p21, comprises 16 exons and encodes a protein of 934 amino acids. *MLH1* is

†These authors contributed equally to this work.

*Correspondence to: Ming Qi, Center for Genetic and Genomic Medicine, Zhejiang University School of Medicine First Affiliated Hospital, 79 Qingchun Road, Hangzhou 310003, Zhejiang, China. E-mail: ming_qi@urmc.rochester.edu or Jin Zhang, Tianjin Medical University Cancer Hospital and Research Institute, China; Tianjin Breast Cancer Prevention, Treatment and Research Center, Breast Cancer Research Key Laboratory of Education Ministry, Tianjin, China. E-mail: Davidz9132002@yahoo.com or Shu Zheng, Zhejiang University School of Medicine Second Affiliated Hospital, Hangzhou 310003, Zhejiang, China. E-mail: zhengshu@zju.edu.cn

Contract grant sponsors: National Natural Science Foundation of China (30872955); 985 Project Grant from the Ministry of Education of China (Dr. Ming Qi); "Qiangjiang Research Talent" (Dr. Ming Qi) from the Science and Technology Department of Zhejiang Province; Zhejiang Provincial Natural Science Foundation (Ying Yuan) (R2090353).

located on 3p21.3 with 19 exons and has 57,357 kb in length, 756 amino acids in the protein. *MSH6* has 10 exons that encode a protein of 1,360 amino acids. *MLH3* encompassed 13 exons encoding for a 1,453 amino acid protein. *PMS2* and *PMS1* comprise 15 and 13 exons encoding proteins of 862 and 932 amino acids, respectively.

Familial adenomatous polyposis (FAP) is caused by variants in the *APC* (MIM# 611731) gene. *APC* gene maps on locus 5q21-q22, comprises 16 exons spanning 8.5 kb and encodes for a 2843 amino acid protein. However, the last exon is 6.5 kb in length and encodes a protein of 2,190 amino acids.

Variant databases are essential during the work of genetic testing for Chinese patients. Although there are some variant databases for breast cancer (Breast Cancer Information Core [BIC] http://research.nhgri.nih.gov/bic/, and Leiden Open Variation Database [LOVD] http://chromium.liacs.nl/LOVD2/cancer/home.php) and colorectal cancer (InSiGHT http://www.insight-group.org/, MRGVD http://www.med.mun.ca/mmrvariants/default.aspx, UMD http://www.umd.be/APC/, and UMCG http://www.mmrmissense.net/), most of the data were collected from Caucasian and Ashkenazi Jewish populations. The data about Chinese people are still deficient. It is our goal to create systematic Chinese specific variant databases for all genes/diseases, in addition to some international databases as part of the HVP [Zhang et al., 2010]. Thus, the breast cancer and colorectal cancer databases were created to provide the guidance for genetic screening of Chinese patients as our first effort. These databases are based on the LOVD system [Fokkema et al., 2005], which is an internet-based database system designed to collect and display DNA variants for specific genes and are accessible via http://www.genomed.org/LOVD/. Our database will appear to users who survey other LOVD databases (e.g., by Google search, or by NCBI GeneTests search).

Data Collection

The bulk of data on gene variants is derived from the published articles and some laboratory unpublished data, including PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed), Wanfang Data (http://www.wanfangdata.com.cn/), and Vip Information (http://www.cqvip.com/) using the key words gene name and Chinese. All relevant indexed articles published in Chinese were collected and reviewed. There are about 2,300 reported patients from different geographic regions of the country have been tested for the *BRCA1/2* genes with several small to mid-size (up to 100 people the largest) studies of control populations by seven different institutions in China. There are about 1,400 reported patients from different geographic regions of the country have been tested for the *MMR/APC* genes with several small to mid-size (up to 137 people the largest) studies of control populations by four different institutions in China.

Database Structure

The home page of variant database consists of four sections: general information, sequence variant tables, search the database, and links to other resources. Taking *BRCA1* for example (Fig. 1), the general information section includes basic information about the *BRCA1* gene and variant database. Users can obtain *BRCA1* gene coding DNA reference sequences in this section. From the sequence variant tables section, users can search and review the variants data from different ways, including unique sequence variants, complete sequence variant listing, and variants with no known pathogenicity. The entries of pathogenic variants can be found in complete sequence variant listing. The first column is about pathogenic. The meanings of the symbols are as follows: "–" = No known pathogenicity; "–?" = Probably no pathogenicity; "?" = Unknown; "+?" = Probably pathogenic; "+" = Pathogenic. "Pathogenicity (reported)" and "pathogenicity (concluded)" are separated by slash, respectively. Users can also search the database by type of variant or based on the patient's origin, or by a simple or advanced search in the third section. The technology (denaturing high performance liquid chromatography, single strand conformation polymorphism, multiplex-ligation–dependent probe amplification, polymerase chain reaction, sequencing, protein truncation test, array comparative genomic hybridization, etc.) applied to identify each specific variant is recorded and indicated. The last section gives some other resources linking, including the External link, Entrez gene (http://www.ncbi.nlm.nih.gov/gene/), Online Mendelian Inheritance in Man (OMIM; http://www.omim.org), Human Gene Mutation Database (HGMD; http://www.hgmd.org/), and GeneTests (http://www.ncbi.nlm.nih.gov/sites/GeneTests/).

Data Submission

Anyone is allowed to submit new variants to the database by clicking "submit" button, but the submitter should sign in first from the upper left corner. The submitter should complete the variant data and patient data forms in detail as much as possible. The variant data contain the following items: variant allele, exon, DNA change, DNA published, RNA change, protein change, Re-site, frequency, patients, control, ID, type of variant (DNA level), location, remark, pathogenicity (reported), and pathogenicity (concluded). The patients data include patient ID, disease, reference, detection template, techniques used, remarks, times reported, tissue, origin of variant, gender, occurrence, if de novo origin of variant, geographic origin, ethnic origin, and patient population. Then the database curator will check each submitted entry, and the variant will be displayed in the sequence variant tables. The submitter's name will also be shown in the list, together with the references that are submitted. More detail introduction can be found in documentation.

The submitters should pay attention to the following instructions. All variants data submitted should be named according to the nomenclature of the Human Genome Variation Society (HGVS) [den Dunnen and Antonarakis, 2000]. Pathogenicity of variants is determined according to the following criteria: (1) variant that leads to grand change of protein structure (frameshift mutations, nonsense mutations, and some in-frame insertions/deletions); (2) variant that leads to incorrect exon splicing; (3) variant that is segregated in a family; (4) animal model or other research data that show that the variants impair protein function. A variant is considered to be pathogenic if it meets at least two of the four criteria. If a variants is found in proband, pedigree analysis is advised to verify it. If a variant was reported in literature as "no pathogenicity," and it dose not meet any one of the four criteria, then it was classified as a "nonpathogenic" variant. A variant classified as "no known pathogenicity" means it is not able to determine if it is pathogenic according to previous study and our criteria. Some variants were classified as "probably pathogenic," "probably no pathogenicity," or "unknown" because the pathogenicity could not be determined by these four criteria; however, some of these variants were reported as pathogenic, nonpathogenic, or unknown variants in the references.

LOVD - China Gene homepage

General information	
Gene name	breast cancer 1, early onset
Gene symbol	BRCA1
Chromosome Location	17q21
Database location	Zhejiang University Center for Genetic and Genomic Medicine
Curator	Ming Qi, PhD, FACMG, Min Pan and Peikuan Cong
Date of creation	March 25, 2009
Last update	July 07, 2010
Version	BRCA1 100707
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Reference sequence	coding DNA reference sequence for describing sequence variants
Total number of unique DNA variants reported	134
Total number of individuals with variant(s)	211
Total number of variants reported	223
NOTE	This gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, and it also acts as a tumor suppressor. The encoded protein combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as the BRCA1-associated genome surveillance complex (BASC). This gene product associates with RNA polymerase II, and through the C-terminal domain, also interacts with histone deacetylase complexes. This protein thus plays a role in transcription, DNA repair of double-stranded breaks, and recombination. Mutations in this gene are responsible for approximately 40% of inherited breast cancers and more than 80% of inherited breast and ovarian cancers. Alternative splicing plays a role in modulating the subcellular localization and physiological function of this gene. Many alternatively spliced transcript variants, some of which are disease-associated mutations, have been described for this gene, but the full-length natures of only some of these variants has been described. A related pseudogene, which is also located on chromosome 17, has been identified. Also known as IRIS; PSCP; BRCAI; BRCC1; PNCA4; RNF53; BROVCA1;

Sequence variant tables	
Unique sequence variants	Listing of all unique sequence variants in the BRCA1 database, without patient data
Complete sequence variant listing	Listing of all sequence variants in the BRCA1 database
Variants with no known pathogenicity	Listing of all BRCA1 variants reported to have no noticeable phenotypic effect (note: excluding variants of unknown effect)

Search the database	
By type of variant	View sequence variant table after selecting one type of variant
Simple search	Query the database by selecting the most important variables (exon number, type of variant, disease phenotype)
Advanced search	Query the database by selecting a combination of variables
Based on patient origin	View all variants based on your patient origin search terms

Links to other resources	
Homepage	http://www.genomed.org/LOVD/BC/home.php?select_db=BRCA1
External link #1	Breast-ovary
External link #2	Breast Cancer Information Core
Entrez Gene	672
OMIM - Gene	113705
OMIM - Disease #1	Genetic Susceptibility to Breast Cancer (BC)
OMIM - Disease #2	Familial Breast-Ovarian Cancer-1
HGMD	BRCA1
GeneTests.org	BRCA1

Figure 1. Home page of the *BRCA1* variant database.

Database Content

Breast Cancer

The *BRCA1* gene variant database currently has 223 entries and 134 unique variants. Among the 134 unique variants, 73 are pathogenic according to the criteria described above and 57% (42/73) of these pathogenic variants have not been previously reported in the BIC database. About 67% (49/73) of the pathogenic variants of the *BRCA1* gene result in truncated proteins, all of which are 29 frameshift and 20 nonsense mutations (Table 1). For *BRCA2*, there are currently 62 unique variants identified, 40 of which are pathogenic. Of all 40 entries of pathogenic variants, 92.5% (37/40) are frameshift and nonsense mutations, which lead to the formation of truncated proteins (Table 1).

Colorectal Cancer

The *APC* gene consists of 229 variants among those 158 are unique variants, of which 124 are pathogenic variants. The 158 unique variants include 14 missense, 76 frameshift, 48 nonsense, 14 synonymous, and six unclassified variants. Among the 86 unique variants (currently listed in *MLH1* gene database) 51 are disease-causing variants, nine are nonpathogenic variants and for 26 variants the pathogenic significance is unclear yet. *MSH2* displays the pathogenic variants 73.4% (58/79). Only 10 and seven variants are tested in *MLH3* and *MSH6* gene, respectively. No variant has yet been founded in *PMS1* and *PMS2* gene in Chinese patients. Substitution and deletion variants hold most of the variants in all the colorectal cancer genes (Table 1 and Fig. 2).

Discussion

This studyreports the first effort of the HVP-China creating novel variant database of *BRCA1* and *BRCA2*, *MMR*, and *APC* genes for breast cancer, Lynch syndrome, and FAP, respectively, in the Chinese population.

China participated in the HVP in 2008 and the Chinese Consortium officially started at the Beijing International Forum of Genomic Medicine, November 1, 2008 (HVP-CHINA: http://www.genomed.org/LOVD/). The assembly of HVP-China participants includes collaborators in three groups: (1) analytic platforms of genomics; (2) national clinical networks; and (3) database and bioinformatics. For analytic platforms of genomics, we aim to build four to five national central laboratories. For clinical networks, we plan to build teams of different levels: a few national core members with experienced physicians, dozens of provincial

Table 1. Overview of Published Mutations

Gene	*BRCA1*	*BRCA2*	*APC*	*MLH1*	*MSH2*	*MLH3*	*MSH6*
MIM#	113705	600185	611731	120436	609309	604395	600678
Pathogenicity							
Unknown	13	2	16	26	11	3	4
Pathogenic	73	40	124	51	58	0	3
Nonpathogenic	48	20	18	9	10	7	0
Total	134	62	158	86	79	10	7
Sequence variant							
Missense	42	7	14	34	25	7	7
Nonsense	20	7	48	11	6	0	0
Synonymous	12	7	14	3	4	3	0
Frameshift	29	30	76	25	28	0	0
Splice site	12	2	0	5	5	0	0
Unclassified	19	9	6	8	11	0	0
Total	134	62	158	86	79	10	7

Unknown (?): variants with unknown pathogenicity; Pathogenic (+): pathogenic variants, including "probably pathogenic (+?)" variants; Nonpathogenic (–): variants with no known pathogenicity, including variants with "probably no pathogenicity (–?)".

representative members, and regular members. We will have one central database but multiple bioinformatics teams in different institutions. Additionally, the Chinese Student Volunteer Club for HVP was officially formed at Zhejiang University, Hangzhou China on November 3, 2008.

In the past decades, many studies about breast cancer and colorectal cancer have been carried out and many variants and polymorphisms have been identified in the relevant genes. To facilitate academic research and clinical genetic testing for women at high risk of developing breast cancer and ovarian cancer, it was necessary to create a database for gene variants and other relevant information about these two cancers. There are some existing variant databases on the internet such as the BIC database, Mismatch Repair Genes Variant Database, InSiGHT database, and HGMD. But most of those variants are variants and polymorphisms in European, Caucasian, and Ashkenazi Jewish populations, with limited data about Chinese populations. Our database is important to complement these existing databases. Currently, many variants databases in LOVD format have been created in different order of columns and versions. All these variants databases should be integrated and discussed in the upcoming international meeting.

Founder mutations of *BRCA1* and *BRCA2* genes have been previously identified in some specific ethnic groups. In women with Ashkenazi Jewish ancestry, the founder mutations 185delAG and 5382insC in *BRCA1* gene and 6174delT in *BRCA2* gene account for 60% of families with breast and ovarian cancer, and these variants were also found in up to 30% of women with early-onset breast cancer [Abeliovich et al., 1997; Bar-Sade et al., 1998; Muto et al., 1996; Tonin et al., 1996]. Founder mutations have also been observed in many populations including Icelandic, French Canadian, and Norwegian populations [Ferla et al., 2007].

Because the current cost of genetic testing is still very high in China, some patients cannot afford the whole-gene testing. And founder mutations represent a majority of the mutation occurrence

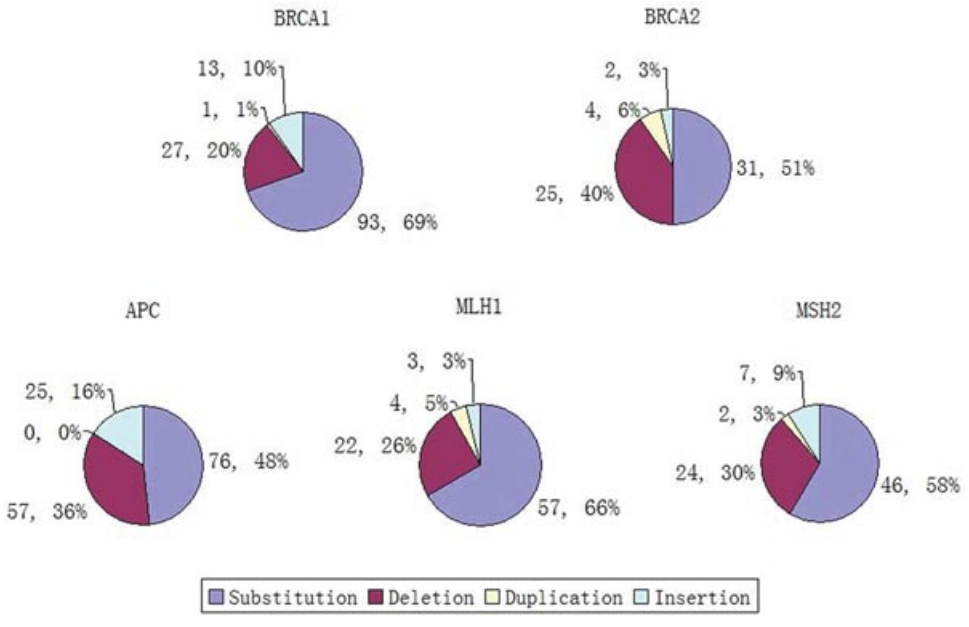

Figure 2. Frequency of pathogenic variant types in the variant database in the Chinese population.

Table 2. Recurrent Pathogenic Mutations of *BRCA1/2* in the Chinese population

Mutation	Exon	Times reported in Chinese	Times reported in other ethnicities (from BIC)
BRCA1			
c.5470_5477del8	24	7	2, Asian
c.981_982delAT	11	6	6, Caucasian
c.5332+1G>C	21	4	0
c.3359_3363del5	11	4	0
c.5521delA	24	3	1, Unknown
c.3330_3331insA	11	3	1, Unknown
c.470_471delCT	8	3	4, Caucasian, Latin American/Caribbean
BRCA2			
c.5574_5577delAATT	11	4	0
c.3109C>T	11	3	5, Pakistani, Caucasian, Asian
c.5722_5723delCT	11	3	42, Caucasian, European, Asian, India, Latin American/Caribbean, Ashkenazi

BIC: Breast Cancer Information Core database; Mutations that were reported less than three times in this database were not listed in the table.

in specific populations. While comprehensive full coding region test for patients are available in China, it is our wish to identify any founder mutations for genetic testing for cost-effect purpose. However, no founder mutations of *BRCA1* or *BRCA2* genes have yet been identified in the Chinese population from this study. Two previous studies have proposed that 1081delG (c.962delG) in *BRCA1* gene and c.3109C>T in BRCA2 gene are founder mutations in the Chinese population [Khoo et al., 2002; Kwong et al., 2009], but these results have not been supported by any other data. Nevertheless, some recurrent pathogenic variants were observed in our database (Tables 1 and 2). But these variants have not been proved as founder mutation. The most frequent pathogenic variants of *BRCA1* are c.981_982delAT, c.3359_3363del5, c.5332+1G>C, and c.5470_5477del8, all of which are frameshift variants that result in truncated proteins, except c.5332+1G>C. Variant c.5332+1G>C is a single nucleotide substitution at exon–intron boundary in intron 21, which may affect exon splicing [Chen et al., 2009; Li et al., 2008]. In *BRCA2*, the most frequent pathogenic variants are c.5574_5577delAATT c.3109C>T and c.5722_5723delCT. The variant c.5574_5577delAATT has been found in at least three families, two of which are kindreds with more than three breast cancers; there was at least one bilateral case in each kindred [Hu et al., 2008; Li et al., 2008; Song et al., 2006ab]. Of note, we compared our database to the BIC database, more than 50% of the recurrent variants in our database have never been reported in populations other than Chinese (Tables 1 and 2). These results indicate that some specific variants of *BRCA1/2* genes may contribute more than others to the development of breast cancer in a Chinese population.

"Hot spots" are the regions of the gene sequences having multiple high incidence of the mutation occurrence. There is apparently no such "hot variant spot" region of *BRCA1/2* genes in the Chinese population. All variants are spread evenly throughout the whole gene (exons and splice sites). Although more than 50% of the pathogenic variants reside within exon 11 in both *BRCA1* and *BRCA2*, a hotspot region is not indicated because exon 11 covers more than 50% of the coding sequence in *BRCA1* and *BRCA2* genes [Smith et al., 1996; Wooster et al., 1995].

APC is the gene associated with FAP, which is a tumor suppressor gene. The protein encoded by *APC* gene plays an important role in tumor suppression [Hoshino et al., 1991]. Most of variants of *APC* gene produce abnormal proteins that cannot suppress the cellular overgrowth that becomes cancerous. Several variants are reported many times and in different populations in our database and other databases. These "hot" variants are c.3927_3931delAAAGA (p.Glu1309AspfsX4), c.4393_4394delAG (p.Ser1465TrpfsX3), c.3183_3187delACAAA (p.Lys1061Lysfsx2), c.4348C>T (p.Arg1450X), c.994C>T (p.Arg332X), c.847C>T (p.Arg283X), c.694C>T (p.Arg232X), and c.423-1G>A. Three variants c.646C>T (p.Arg216X), c.2721delG (p.Gly907GlyfsX9), and c.4057G>T (p.Glu1408X) are only identified in Chinese patients.

Variants in the *MSH2* and *MLH1* genes are detected in up to approximately 60% and 30% of families with Lynch syndrome, respectively [Fishel et al., 1993; Papadopoulos et al., 1994]. *MSH6* gene variants account for approximately 7–10% of such families [Miyaki et al., 1997]. The share of *PMS2* is less than 5% [Thompson et al., 2004]. *MLH3* and *PMS1* genes variants were only reported in very limited families and the clinical significance of these genes in Lynch syndrome has not been determined [Lu et al., 1998; Nicolaides et al., 1994; Peltomaki, 2003]. No variant hot area is found in the MMR genes, but c.1168C>T (p.Leu390Phe), c.1255C>A (p.Gln419Lys), and c.1886A>G (p.Gln629Arg) in *MSH2* gene and c.199G>A (p.Gly67Arg), c.1151T>A (p.Val384Asp), and c.2041G>A (p.Ala681Thr) in *MLH1* gene are reported relatively more times. The variants c.610G>T (p.Gly204X) in *MSH2* gene and c.265G>T (p.Glu89X) in *MLH1* gene are only identified in Chinese populations. Only 10 and seven variants are founded in *MLH3* and *MSH6* genes, respectively, and no variant has been founded in *PMS1* and *PMS2* genes in Chinese patients.

China is a large country with a population of about one-fourth of the whole world. Ministry of Health of the People's Republic of China issued an approved List of Clinical Genetic Testing for Medical Institutions (180) in 2007. The list granted the approval for some genetic testing, including mutation screening for *BRCA1*, *BRCA2*, and other genes in familiar breast cancer, mutation screening for *MLH1*, *MSH2*, *PMS1*, and *PMS2* genes for Lynch syndrome, mutation screening for *APC* and *DCC* genes in colon cancer. More tests are expected to be approved. We have built 32 genes database based on our LOVD system and will finish other 60 genes related to malignant arrhythmia and some genetic tumor disorders. In addition, China spurs quest for HVP and will invest significant funding support to this project over 10 years to take on important role in the medical genetics field. At present, most of the genetic testing is conducted on the research base with no national clinical standards and certification system yet. China will set up a new institute in Beijing to adopt the international standards and coordinate activities across the country and to give training in genetic counseling and testing. Currently, the Ministry of Health, PRC, has organized several genetic counseling workshops. The International HVP is the global initiative to collect and curate all human genetic variation affecting human health. This is a huge and long-term project. It will require continuous efforts from generations of scientists and physicians of genomic medicine. This students club aims to attract

the future scientists and physicians to participate in this important program.

In conclusion, the novel LOVD Databases for Hereditary Breast Cancer and Colorectal Cancer will provide a great convenience to researchers and clinicians who study, test, and diagnose breast cancer, colorectal cancer, and other cancers caused by variants on the involved genes. The databases can be of especially use for facilitating future genetic tests in the Chinese population and others worldwide.

Acknowledgments

We would like to thank all the members of the Chinese Student Volunteer Club of the International HVP for their work during the set up of this database.

References

Abeliovich D, Kaduri L, Lerer I, Weinberg N, Amir G, Sagi M, Zlotogora J, Heching N, Peretz T. 1997. The founder mutations 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashkenazi women. Am J Hum Genet 60:505–514.

Bar-Sade RB, Kruglikova A, Modan B, Gak E, Hirsh-Yechezkel G, Theodor L, Novikov I, Gershoni-Baruch R, Risel S, Papa MZ, Ben-Baruch G, Friedman E. 1998. The 185delAG BRCA1 mutation originated before the dispersion of Jews in the diaspora and is not limited to Ashkenazim. Hum Mol Genet 7:801–805.

Chen W, Pan K, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Lu Y, You W, Xie Y. 2009. BRCA1 germline mutations and tumor characteristics in Chinese women with familial or early-onset breast cancer. Breast Cancer Res Treat 117:55–60.

Cyranoski D. 2011. China spurs quest for human variome. Nature 469:455.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

Ferla R, Calo V, Cascio S, Rinaldi G, Badalamenti G, Carreca I, Surmacz E, Colucci G, Bazan V, Russo A. 2007. Founder mutations in BRCA1 and BRCA2 genes. Ann Oncol 18 Suppl 6:vi93–vi98.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75:1027–1038.

Fokkema IF, den Dunnen JT, Taschner PE. 2005. LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. Hum Mutat 26:63–68.

Hoshino Y, Horikawa I, Oshimura M, Yuasa Y. 1991. Normal human chromosome 5, on which a familial adenomatous polyposis gene is located, has tumor suppressive activity. Biochem Biophys Res Commun 174:298–304.

Hu Z, Li WF, Liu XY, Zhang B, Cao MZ, Wang YS, Zhao L, Song CG, Lu JS, Wu J, Di GH, Shen KW, Han QX, Shen ZZ, Huang W, Shao ZM. 2008. BRCA1/2 gene mutation in Chinese familial breast cancer patients: a multi-center report of 115 cases. Zhonghua Yi Xue Za Zhi 88:2383–2386.

Khoo US, Chan KY, Cheung AN, Xue WC, Shen DH, Fung KY, Ngan HY, Choy KW, Pang CP, Poon CS, Poon AY, Ozcelik H. 2002. Recurrent BRCA1 and BRCA2 germline mutations in ovarian cancer: a founder mutation of BRCA1 identified in the Chinese population. Hum Mutat 19:307–308.

Kinzler KW, Vogelstein B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 386:761, 763.

Kwong A, Wong LP, Wong HN, Law FB, Ng EK, Tang YH, Chan WK, Ho LS, Kwan KH, Poon M, Wong TT, Leung FC, Chan SW, Ying MW, Ma ES, Ford JM. 2009. A BRCA2 founder mutation and seven novel deleterious BRCA mutations in southern Chinese women with breast and ovarian cancer. Breast Cancer Res Treat 117:683–686.

Li WF, Hu Z, Rao NY, Song CG, Zhang B, Cao MZ, Su FX, Wang YS, He PQ, Di GH, Shen KW, Wu J, Lu JS, Luo JM, Liu XY, Zhou J, Wang L, Zhao L, Liu YB, Yuan WT, Yang L, Shen ZZ, Huang W, Shao ZM. 2008. The prevalence of BRCA1 and BRCA2 germline mutations in high-risk breast cancer patients of Chinese Han nationality: two recurrent mutations were identified. Breast Cancer Res Treat 110:99–109.

Lu SL, Kawabata M, Imamura T, Akiyama Y, Nomizu T, Miyazono K, Yuasa Y. 1998. HNPCC associated with germline mutation in the TGF-beta type II receptor gene. Nat Genet 19:17–18.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66–71.

Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T. 1997. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet 17:271–272.

MOH. 2009. Available at: http://www.moh.gov.cn/publicfiles//business/htmlfiles/zwgkzt/pwstj/index.htm.

Muto MG, Cramer DW, Tangir J, Berkowitz R, Mok S. 1996. Frequency of the BRCA1 185delAG mutation among Jewish women with ovarian cancer and matched population controls. Cancer Res 56:1250–1252.

Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM. 1994. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 371:75–80.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD. 1994. Mutation of a mutL homolog in hereditary colon cancer. Science 263:1625–1629.

Parvin JD. 2004. Overview of history and progress in BRCA1 research: the first BRCA1 decade. Cancer Biol Ther 3:505–508.

Pasche B. 2008. Recent advances in breast cancer genetics. Cancer Treat Res 141:1–10.

Peltomaki P. 2003. Role of DNA mismatch repair defects in the pathogenesis of human cancer. J Clin Oncol 21:1174–1179.

Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, King MC. 1996. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. Genome Res 6:1029–1049.

Song CG, Hu Z, Wu J, Luo JM, Shen ZZ, Huang W, Shao ZM. 2006a. The prevalence of BRCA1 and BRCA2 mutations in eastern Chinese women with breast cancer. J Cancer Res Clin Oncol 132:617–626.

Song CG, Hu Z, Yuan WT, Di GH, Shen ZZ, Huang W, Shao ZM. 2006b. BRCA1 and BRCA2 gene mutations of familial breast cancer from Shanghai in China. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 23:27–31.

Thompson E, Meldrum CJ, Crooks R, McPhillips M, Thomas L, Spigelman AD, Scott RJ. 2004. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. Clin Genet 65:215–225.

Tonin P, Weber B, Offit K, Couch F, Rebbeck TR, Neuhausen S, Godwin AK, Daly M, Wagner-Costalos J, Berman D, Grana G, Fox E, Kane MF, Kolodner RD, Krainer M, Haber DA, Struewing JP, Warner E, Rosen B, Lerman C, Peshkin B, Norton L, Serova O, Foulkes WD, Garber JE. 1996. Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer families. Nat Med 2:1179–1183.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1995. Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789–792.

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D. 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265:2088–2090.

Zhang T, Moss A, Cong P, Pan M, Chang B, Zheng L, Fang Q, Zareba W, Robinson J, Lin C, Li Z, Wei J, Zeng Q, Qi M. 2010. LQTS gene LOVD database. Hum Mutat 31:E1801–E1810.