Structural Study of the Microtubule-associated Protein Tau Locus of Alzheimer's Disease in Taiwan

Chun-Wei Chang[1], Wen-Chuin Hsu[1], Alan Pittman[2], Yih-RuWu[1], John Hardy[3], Hon-Chung Fung[1]

Background: Haplotype structure of the microtubule-associated protein tau (*MAPT*) gene is associated with various tauopathies in the Caucasian population. With the knowledge that the association between *MAPT* structure and disease may be distinct in different ethnics, we intend to investigate the haplotype structure of *MAPT* in Taiwanese and test it for association with Alzheimer's disease (AD).

Methods: One hundred and eight AD patients and 108 sex- and-age matched healthy controls were recruited from the dementia outpatient clinic of Chang Gung Medical center. We genotyped the *del-In9* marker that defines the extended H1 and H2 clades. We selected 21 single-nucleotide polymorphisms (SNPs) in the extended *MAPT* region from Japanese SNPs database and dbSNP database. Using the software TagIt, we analyzed the linkage disequilibrium structure of *MAPT* and compared the allele and genotype distribution between patient group and control group.

Results: All the Taiwanese participants were H1 haplotypes. Linkage disequilibrium analysis showed the haplotype blocks in Taiwanese population had a smaller size in comparison to that of the Caucasian population. Single locus association showed significant *p* value in one of the tagging variants (rs242557) in our Taiwanese AD case–control cohorts.

Conclusion: *MAPT* gene has four haplotype blocks in the Taiwanese population, each of around 40 kbp. In both European study and our study, the SNP rs242557 showed association with AD. Given the position of this SNP, the most possible explanation is that genetic variability in tau expression contributes to the risk of developing AD.
(*Biomed J 2014;37:127-132*)

Key words: alzheimer's disease, haplotype, tau, tauopathy

At a Glance Commentary

Scientific background of the subject

Haplotype structure of the microtubule-associated protein tau (*MAPT*) gene is associated with various tauopathies in the Caucasian population. With the knowledge that the association between *MAPT* structure and disease may be distinct in different ethnics, we intend to investigate the haplotype structure of *MAPT* in Taiwanese and test it for association with Alzheimer's disease (AD).

What this study adds to the field

With our findings, the pattern of Taiwanese *MAPT* haplotype block were different from that of European one; while the single locus association of rs242557 remained. Given the position of this SNP, the most possible explanation is that genetic variability in tau expression contributes to the risk of developing AD

The microtubule-associated protein tau (*MAPT*) gene encodes for tau, a phosphorylated protein that assists in the stabilization of the cytoskeleton.[1] The *MAPT* locus is unusual in the Caucasians for having complete linkage disequilibrium (LD) pattern that covers a maximal region of ~ 1.6 million bases (Mb), forming two distinct haplotype clades, H1 and H2.[2,3] There is no recombination between the H1 and H2 haplotype clades over this region. Recently,

---

a large inversion polymorphism was found to be responsible for establishment of these two haplotype clades.[4-6] The H1 haplotype clade is the most common, having an allele frequency of about 70% in European populations and a greater frequency in other populations.[2,7] The H2 haplotype shows almost complete association with Caucasian ancestry, with essentially zero frequencies in east Asian, African, and Native American populations. The common H1 clade shows considerable diversity in haplotype structure,[3,4,6] but the H2 haplotype seemingly derives from a single founder.[5-8] The LD pattern over this region in Asian population has not been established.

It has been reported that the variation of haplotype structure, represented as haplotype tagging single nucleotide polymorphisms (htSNPs), is associated with several major neurodegenerative diseases, including Parkinson's disease, coronary artery disease, and age-related macular degeneration.[9-11] In a previous study on the tauopathies, several H1 haplotype-specific single nucleotide polymorphisms in the *MAPT* gene were found significantly associated with progressive supranuclear palsy and corticobasal degeneration.[12] Though the neurofibrillary tangles, comprising paired helical filaments of phosphorylated tau protein, are a pathognomic feature of Alzheimer's disease (AD),[1] it is not clear whether the *MAPT* locus is genetically involved.

Furthermore, the association between the *MAPT* structure and tauopathies may be varied in different ethnics. In an analysis of *MAPT* in Parkinson's disease of different ethnics, the association between htSNPs and the disease was found in the Finnish and Greek populations, but not in Taiwanese cohort.[13]

Our group has reported the association between one variant of H1 clade at the *MAPT* locus (named H1c) and AD and other tangle diseases in the Caucasian population.[12,14] With this background, we were interested in extending our study to the Asian population in which the H2 haplotype is absent. Thus, we sought to investigate the haplotype structure of *MAPT* in an Asian population (Taiwanese) and test it for association with AD.

METHODS

Selection of Taiwanese AD patients and controls

Patients were recruited from the dementia outpatient clinic of Chang Gung Medical Center. Each subject was informed of the aim of the study, and all gave their consent to participate in the study. Patients with previous clinical history of neurological, psychiatric, somatic, or toxic causes for dementia were excluded. Evaluation included general physical and neurological assessment, the Mini-Mental State Examination (MMSE),[15] and Hachinski ischemia

score.[16] Laboratory studies included complete blood cell count, biochemistry analysis, erythrocyte sedimentation rate, vitamin B12 and folic acid levels, thyroid stimulating hormone level, and syphilis serological testing. Each patient underwent brain computerized tomography (CT) scan. All the patients were examined by at least two neurologists and confirmed to fulfill the DSM-IV criteria for dementia. The diagnosis of AD [$n = 108$, 45.4% females; age (Mean ± SD) 70.7 ± 8.52 years] was made by consensus according to the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) for probable AD.[17] Age- and sex-matched subjects without stroke and cognitive impairment represented the control group [$n = 108$, 45.4% females; age (Mean ± SD) 70.7 ± 8.69 years].

Genotyping and statistical analysis

SNP selection

For the Taiwanese population, there was no genotyping information available on the known single-nucleotide polymorphisms (SNPs) in the extended *MAPT* region that covers 45 kb upstream and downstream of the gene. We selected all the Japanese SNPs with frequency information published on the JSNP database at http://snp.ims.u-tokyo.ac.jp in the region. We then tried to use any SNP(s) from dbSNP to fill the gap, when two adjacent JSNPs were more than 14 kb apart. This led us to the following SNPs: rs962885, rs2301689, rs3744457, rs2280004, rs1467967, rs3785880, rs1001945, rs242557, rs242562, rs2303867, rs3785882, rs3785883, rs3785885, rs2258689, rs2471738, rs916896, rs7521, rs2074432, rs2277613, rs876944, and rs2301732.

Genotyping

The 238-bp *MAPT* H2 deletion in intron 9 (*del-In9*) was tested to confirm that the H2 haplotype does not occur in the Taiwanese.[2] Oligonucleotide primer pairs were designed to specifically amplify the *MAPT* haplotype SNPs of interest (rs227613, rs962885, rs3785880, rs3744457, rs1467967, rs242557, rs3785883, rs2471738, and rs7521) with conventional polymerase chain reaction (PCR) procedure. The reactions contained 25 ng of template DNA, 1 U of Taq DNA polymerase, 10x PCR buffer, 10 mM dNTP mix, 10 pmol of each oligonucleotide primer pair, and 5x Q solution (Qiagen, Crawley, UK). Thermocycling conditions were carried out using the "touchdown" method. Genotyping of the SNPs was carried out either by pyrosequencing (Pyrosequencing Inc., Westborough, MA, USA) or by restriction digests (RFLP). PCR and pyrosequencing primer sequences and thermocycling conditions are available on request.

The SNPs (rs227613, rs962885, rs3785880, rs3744457, rs1467967, and rs2280004) were analyzed by pyrosequencing. Twelve microliters of the biotinylated PCR product was immobilized on streptavidin-sepharose™ HP (Amersham Pharmacia Biotech, Piscataway, NJ, USA). The gel slurry (4 μl) was resuspended in binding buffer [10 mM Tris-HCl, 2 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% Tween-20]. Template and beads were mixed continuously for > 5 min at room temperature. The immobilized DNA template was transferred to a 96-well filter plate attached to a vacuum manifold (Millipore Inc., Bedford, MA, USA), then immersed for 10 s in ethanol followed by denaturing buffer (0.2 M NaOH) and finally wash buffer (10 mM Tris-acetate, pH 7.6). The sequencing primer (15 pmol) was then annealed to the single-stranded template in 12 μl of annealing buffer (20 mM Tris-acetate, 2 mM magnesium acetate, pH 7.6) at 80 C for 2 min before cooling to room temperature. Samples were analyzed using a PSQ 96 system together with SNP software and SNP reagent kits (Pyrosequencing Inc.) following the manufacturer's instructions.

For SNP analysis by RFLP, 15 μl of PCR product – SNPs (rs242557, rs3785883, rs2471738, rs7521, and *del-In9*) – was digested by 1 U of the corresponding restriction e0ndonuclease (New England Biolabs, Hertfordshire, UK) [ApaL I(G), BsaH I(G), BstE II(T), Pst I(A)] in a reaction volume of 20 μl for 4 h. The PCR products were all cleaved by the corresponding enzyme once at the indicated (N) allele. Digests were run on a 4% agarose gel for analysis. Genotype scoring was carried out blindly by two individuals. Any discrepancies between the two were resolved by repeating the assay.

LD and statistical analysis

For each SNP, the allele and genotype distributions in the group of patients were compared with those in the control group. Statistical assessments for the allele and genotype frequencies and Hardy–Weinberg equilibrium (HWE) were made using the genetics software program TagIt (http://popgen.biol.ucl.ac.uk/software.html).[18] The square of the correlation coefficient (r^2) and D' for LD were calculated pairwise between each. Both measures of LD are based on D, the basic pairwise disequilibrium coefficient, the difference between the probabilities of observing them independently in the population: $D = f(A1B1) - f(A1)f(B1)$, where A and B refer to two genetic markers and f is their frequency. D' is obtained from D/Dmax, and a value of 0.0 suggests independent assortment, whereas 1.0 means that all copies of the rarer allele occur exclusively with one of the possible alleles at the other marker. Also, r^2 has a more strict interpretation than that of D', and therefore, $r^2 = 1.0$ only when the marker loci have identical allele frequencies; the allele at one locus can always

be predicted by the allele at the other locus. In contrast, D can reach a value of 1.0 when the allele frequencies vary. An $r^2 > 0.5$ represents a high level of LD and $r^2 < 0.1$ represent essentially no LD. Haplotype predictions were made using an EM algorithm using TagIT. Case–control locus-by-locus association was calculated statistically using a χ^2 distribution and the significance was calculated using a Monte–Carlo approach as implemented by CLUMP software.[19]

RESULTS

LD and haplotype structure of MAPT in Taiwanese

For the haplotype structure investigation of *MAPT* gene, we used the above 21 selected SNPs to genotype 21 normal Taiwanese individuals [52.4% females, age (Mean ± SD) 68.3 ± 2.6 years]. We discarded SNPs that had a minor allele frequency of less than 5%. The average density of the markers is one SNP every 20 kb. None of the polymorphisms deviated from HWE. We genotyped the *del-In9* marker that defines the extended H1 and H2 clades. All the Taiwanese individuals including cases and controls were H1 haplotypes.

We evaluated pairwise LD across *MAPT* for all 21 SNPs in the 21 normal Taiwanese individuals both by D prime (D') and the square of the correlation coefficient (r^2), calculated from the expectation-maximization inferred haplotypes. The entire *MAPT* gene is featured by LD as is particularly evident by the measured r^2. This pattern of LD across the extended *MAPT* region has shown that the haplotype blocks in Taiwanese population have a smaller size, rather a single clade that we have previously defined in the Caucasian population.[8]

Selection, performance, assessment, and association analysis of MAPT haplotype tagging SNPs

We used an association-based criterion (criterion 11 in Tagit, haplotype r^2) in order to select the haplotype tagging SNPs (htSNPs). Ten htSNPs out of the 21 selected SNPs were required to represent all the genetic variations in our Taiwanese cohort. In addition, the bi-allelic *del-In9* marker was used to unambiguously confirm that all in the Taiwanese population in our study is of the H1 clade. The performance value for the 10 htSNPs was interpreted at an average haplotype r^2 value of 0.95 (95%). That is, these 10 htSNPs captured 95% of all genetic variations in the *MAPT* region [Figure 1].

We genotyped the *MAPT* htSNPs in our Taiwanese AD case–control cohorts. In none of the groups were there any significant deviations from HWE at any of the htSNPs. Only one of the tagging variants (rs242557) had a significant p value in our series [Table 1].

Table 1: Single locus association analysis of MAPT with Taiwanese AD cohort

dbSNP	Chromosomal location	Major allele	Major allele frequency		*p* value (FDR-adjusted *p* value) AD	
			Controls	AD	Allelic	Genotypic
rs962885	41291420	T	80.8	75.0	0.156 (0.312)	0.351 (0.702)
rs3744457	41328711	C	62.0	53.8	0.095 (0.308)	0.238 (0.595)
rs2280004	41340959	T	84.7	84.4	0.951 (0.951)	0.992 (0.992)
rs1467967	41342006	G	63.7	60.3	0.470 (0.588)	0.505 (0.741)
rs3785880	41349204	T	70.4	63.2	0.123 (0.308)	0.040 (0.200)
rs242557	41375573	A	64.6	50.6	*0.002 (0.020)*	0.006 (0.060)
rs3785883	41410269	G	86.0	83.0	0.398 (0.588)	0.638 (0.798)
rs2471738	41431900	C	75.2	77.3	0.614 (0.682)	0.826 (0.918)
rs7521	41461242	G	88.9	91.2	0.422 (0.588)	0.519 (0.741)
rs2277613	41472982	G	62.9	55.3	0.122 (0.308)	0.088 (0.293)

The relative locations, major allele frequencies, *p* values for all tagging loci tested in Taiwanese cases and control groups. Significant values are shown in bold italics, Abbreviations: AD: Alzheimer's disease; FDR: False discovery rate
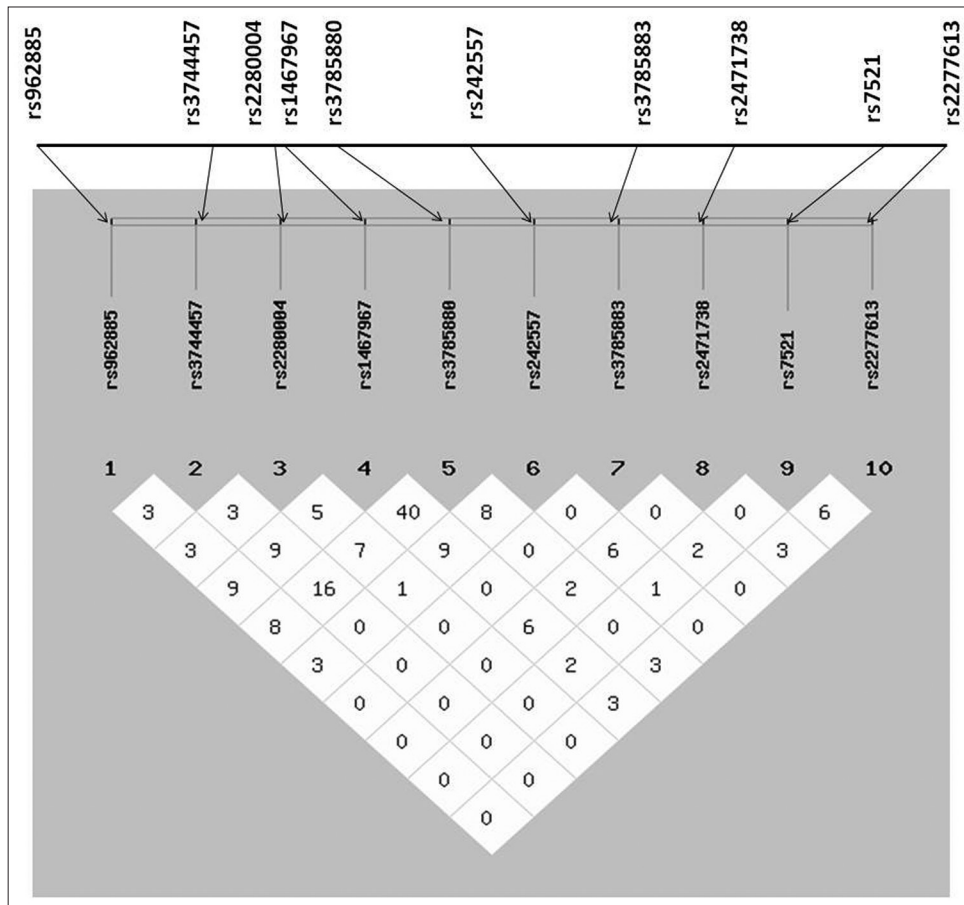

Figure 1: Linkage disequilibrium (LD) across the *MAPT* in Taiwanese population. Pairwise LD plot for depicting the r^2 value. The numbers in the square represent the pairwise $r^2 \times 100$ value. The bar at top of the plot stands only for the relative chromosome position of the 10 *MAPT* SNPs. The figure was created by SHEsis online version.[21]

DISCUSSION

The *MAPT* gene has four haplotype blocks in the Taiwanese population, each of around 40 kbp. This haplotype structure is entirely normal for the human genome and differs from the haplotype structure in European populations, which is distorted by the occurrence of the non-recombining and inverted H2 haplotype.[3] The 21 normal Taiwanese genomes had been analyzed with genotyping of 21 SNPs in the *MAPT* region. Twenty-one

individuals had been recruited in regard to the predetermined minor allele frequency of greater than 5%. The minimal number of the samples needed is 20; in order to avoid the mistaken genotyping, 1 sample in excess had been recruited. As in our case, a significant LD picture had been shown with this set of samples. Assessment of association with AD reveals that rs242557 shows an association with AD. This SNP is 5′ of exon 2 and is the same SNP that showed a strong association with disease in the European population. We have found that both the European and Asian populations share the same risk allele with the A-allele. Nevertheless, the A-allele is the major allele in Asian population (65.9%) while it is the minor allele in European population (32%) [Table 2].[14] We proposed that the risk domain of *MAPT* region associated with AD should be in the region with the complete LD of rs242557. A real association indicates that the variability that leads to predisposition is within a short distance (~30 kb) of this SNP in intron 1. The false discovery rate-corrected *P* value showed no significance in the genotypic analysis at rs242557. Consequently, the bias from multiple comparisons cannot be excluded. Therefore, this association needs further validation.

Furthermore, in the case of our previous report, the association was with the same haplotype, which also showed a robust association with progressive supranuclear palsy.[12] This indicates that this risk domain might be involved in a common pathway of development of various neurodegenerative diseases.

The SNP rs242557 lies within a ~80 bp of sequence conservation. Sequencing of this region in all these Taiwanese cases failed to reveal additional variability (authors' unpublished data). This suggests that either rs242557 itself or variability that lies outside this 180 bp conserved region but within this haplotype block is likely to contribute to the risk of development of neurodegenerative disease. Given these data, and the position of this SNP, the most plausible explanation of this work and previous studies is that genetic variability in tau expression contributes to the risk of developing AD.[20] Such an interpretation would be favored by either the confirmation of this association in another Asian population or, indirectly, through the analysis of progressive supranuclear palsy or corticobasal degeneration since, in the Caucasians, these share the same haplotypic association, but have a higher relative risk. The genetic investigation of this region with fine-mapping approach and the functional approach may lead to a partial elucidation of the pathogenesis of those neurodegenerative diseases.

REFERENCES

1. Spillantini MG, Goedert M. Tau protein pathology in neurodegenerative diseases. Trends Neurosci 1998;21:428-33.

2. Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, *et al*. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet 1999;8:711-5.

3. Pittman AM, Myers AJ, Duckworth J, Bryden L, Hanson M, Abou-Sleiman P, *et al*. The structure of the tau haplotype in controls and in progressive supranuclear palsy. Hum Mol Genet 2004;13:1267-74.

4. Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, *et al*. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum Mol Genet 2005;14:1753-62.

5. Hardy J, Pittman A, Myers A, Gwinn-Hardy K, Fung HC, de Silva R, *et al*. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. Biochem Soc Trans 2005;33:582-5.

6. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, *et al*. A common inversion under selection in Europeans. Nat Genet 2005;37:129-37.

7. Evans W, Fung HC, Steele J, Eerola J, Tienari P, Pittman A, *et al*. The tau H2 haplotype is almost exclusively Caucasian in origin. Neurosci Lett 2004;369:183-5.

8. Fung HC, Evans J, Evans W, Duckworth J, Pittman A, de Silva R, *et al*. The architecture of the tau haplotype block in different ethnicities. Neurosci Lett 2005;377:81-4.

9. UK Parkinson's Disease Consortium; Wellcome Trust Case Control Consortium 2, Spencer CC, Plagnol V, Strange A, Gardner M, Paisan-Ruiz C, Band G, *et al*. Dissection of the genetics of Parkinson's disease identifies an additional association 5′ of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet 2011;20:345-53.

10. Beckie TM, Beckstead JW, Groer MW. The association between variants on chromosome 9p21 and inflammatory biomarkers in ethnically diverse women with coronary heart disease: A pilot study. Biol Res Nurs 2011;13:306-19.

11. Hadley D, Orlin A, Brown G, Brucker AJ, Ho AC, Regillo CD, *et al*. Analysis of six genetic risk factors highly associated with AMD in the region surrounding ARMS2 and HTRA1 on chromosome 10, region q26. Invest Ophthalmol Vis Sci 2010;51:2191-6.

12. Pittman AM, Myers AJ, Abou-Sleiman P, Fung HC, Kaleem M, Marlowe L, *et al*. Linkage disequilibrium fine-mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. J Med Genet 2005;42:837-46.

13. Fung HC, Xiromerisiou G, Gibbs JR, Wu YR, Eerola J, Gourbali V, *et al*. Association of tau haplotype-tagging polymorphisms with

Table 2: Comparison between allele frequencies of rs242557 in different ethnicities

Ethnicity	Residence	Allelic type	Allele frequency
Caucasian[14]	USA and UK	G	0.68
Caucasian	Western Europe	G	0.59
African	Nigeria (Africa)	G	0.66
Japanese	Tokyo (Asia)	G	0.56
Taiwanese	Taipei (Asia)	A	0.66

Data adapted from the international HapMap consortium. The international HapMap Project. Nature 426, 789-96

Parkinson's disease in diverse ethnic Parkinson's disease cohorts. Neurodegener Dis 2006;3:327-33.

14. Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees A, Fung HC, *et al*. The H1c Haplotype at the MAPT Locus is associated with Alzheimer's Disease. Hum Mol Genet 2005;14:2399-404.

15. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189-98.

16. Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A. Pathological verification of ischemic score in differentiation of dementias. Ann Neurol 1980;7:486-8.

17. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939-44.

18. Weale ME, Depondt C, Macdonald SJ, Smith A, Lai PS, Shorvon SD, *et al*. Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: Implications for linkage-disequilibrium gene mapping. Am J Hum Genet 2003;73:551-65.

19. Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann Hum Genet 1995;59:323-36.

20. Singleton A, Myers A, Hardy J. The law of mass action applied to neurodegenerative disease: A hypothesis concerning the etiology and pathogenesis of complex diseases. Hum Mol Genet 2004;13:R123-6.

21. Shi YY, He L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. Cell Res 2005;15:97-8.