npg

ORIGINAL ARTICLE

A Family-based Association Study of *DIO2* and children mental retardation in the Qinba region of China

Kejin Zhang[1,4], Heng Xi[1,2,4], Xiying Wang[1], Yale Guo[3], Shaoping Huang[3], Zijian Zheng[2], Fuchang Zhang[1,2] and Xiaocai Gao[1,2]

Deiodinase enzyme II (DIO2) has an important role in individuals' thyroid hormones' level, the development of central and peripheral nervous systems and characterized by mental retardation (MR). The *DIO2* gene was genotyped by using five haplotype-tagging single-nucleotide polymorphisms (SNPs) in 157 Chinese MR high-density family pedigrees, including 452 nuclear families and >1460 persons. The single marker and haplotype analyses were performed by Family-based Association Tests (FBAT). Three SNPs had *P*-values <0.05 in at least one inherited model survived with the correction. Several haplotypes composed of these SNPs were also associated with MR. The *in silico* analyses identified that one of the SNPs, rs1388378, may be a functional SNP. However, further *in vitro* studies of this SNP should be considered in elucidating its effect on gene expression and the possible role in MR susceptibility.

INTRODUCTION

Thyroid hormone is essential for the development of the brain and nervous system both in the basic processes of neurogenesis and the processes of terminal brain differentiation.[1,2] Croteau *et al.*[3] noted that thyroid hormone has essential roles in some mammalian tissues, such as the developing brain and the anterior pituitary. If a fetus or newborn is not exposed to optimum levels of thyroid hormone, it may have permanent mental retardation (MR) even if it survives. Their clinical symptoms can cover a wide range, from mild blunting of intellect to cretinism, and a large part of the population may have some intellectual impairment.

Iodine deficiency, which can impair thyroid hormone synthesis and/ or thyroid enlargement, is one of the important reasons for MR.[4] Generally, significant improvement typically occurs after appropriate addition of iodine in salt.[5] However, the positive family history for thyroid disease is still one of the features of deficiencies of thyroid hormones.[6] Not everyone living in the same environment was equally susceptible to the deficiency. This feature also indicated that the genetic base may also have an important role. The presence of some genetic variants may result in the heterogeneity status of individuals for thyroid hereditary disorders, which represents MR.

Deiodinase enzyme II (DIO2) has an important role in the conversion of pro-hormone thyroxine (T4) to the active hormone 3,5,3′-L-triiodothyronine (T3). Inadequate T4 hormone to active T3 is one cause of individual's lower thyroid level, which also influence the regulation of synapses, the activity of serotonin, norepinephrine and γ-aminobutyric acid in brain consequently.[7] More reports focused on the relationship between the polymorphisms or mutations of DIO2 gene and the causes of MR and other psychiatric diseases in populations from iodine-deficient regions.[8–10] The positive relation between DIO2 gene and MR was reported in a previous work,[8] but, according to the association study, the unwarranted false positives from the bias of population stratification cannot be ruled out. In addition, the environmental factors, however, were another major impact and could not be eliminated/lowered in these case–control study reports. Therefore, further appropriate work should be done.

The family-based population samples may be another reasonable strategy to make clear this complex problem with linkage method,[11] which can avoid the genetic confounding due to population stratification and environmental factors.[12] The family-based association test (FBAT) is an acronym in genetic analyses and builds on the the transmission disequilibrium test (TDT) method. As opposed to case–

control study designs, the FBAT method avoids false associations caused by admixture of populations and environmental factors, and are convenient for investigators interested in refining linkage findings in family samples.[13]

In this study, the MR patients and their relatives were recruited from the Qinba region of the Shaanxi province, which is one of the historical iodine-deficient areas of China. The FBAT software was used to investigate the relationship between genetic variants of the *DIO2* gene and the MR patients.

MATERIALS AND METHODS

Subjects

There are 157 MR pedigrees in this study, including 452 nuclear families and 1461 subjects from the Qinba region of Shaanxi province, China. On the basis of the Chinese Classification of Mental Disorders 2nd Revision (CCMD-2-R)[14] and the classification of mental and behavioral disorders from the (WHO) World Health Organization,[15] the clinical psychiatric pediatricians diagnosed, identified and classified the MR patients from >20 000 children in the Qinba region. All MR children were seen as probands, their pedigrees and relatives were tracked, investigated and recruited.

The children's intelligence was tested with the Chinese-Wechsler Young Children Scale of Intelligence (C-WYCSI) for 4–5 years,[16] and the Chinese-Wechsler Intelligence Scale for Children (C-WISC) for 6–16 years.[17] The Adaptive Scale of Infant and Children revised by Zuo et al.[18] was used to screen the social adaptive score or the mental handicap score of participants. The mental health condition of an adult relative more than 16 years of age was estimated by the Chinese Revision of Wechsler Adult Intelligence Scale (WAIS-RC).[19]

The protocol was reviewed and approved by the Ethical Committee of the National Human Genome Center. All subjects were Han Chinese in origin.

Single-nucleotide polymorphism (SNP) selection and genotyping

The *DIO2* gene located in 14q24.3 is about 15 kb length, and includes two exons and one intron. To minimize the genotyping load while maximizing association information, the implementation of haplotype-tagging SNP (tSNP) selection was performed. For this study, tSNPs were identified by using HAPLOVIEW 4.2[20] and the method described by Gabriel et al.[21]

Five SNPs of the *DIO2* gene (rs225015, rs225014, rs225012, rs2267872 and rs1388378) were chosen from Han Chinese in the HCB_Asian population data in the HapMap SNP set (version 22) (http://hapmap.ncbi.nlm.nih.gov/), at a r^2 threshold of 0.80 with a minor allele frequency >0.1. Multiple methods for SNP genotyping were used for five SNPs (Supplementary Table 1). A routine quality control procedure to assess genotyping reproducibility was performed by repeating 10% of samples selected randomly.

Statistical analysis

The PEDCHECK program was used to test the SNP data for Mendelian inconsistencies in the family sample before analysis.[22] A genotyping error rate of 0.3% was observed, and all erroneous genotypes were removed. The Hardy–Weinberg equilibrium testing was performed in the HAPLOVIEW 4.2 software[20] using $P < 0.001$ as the cutoff point. The single marker tests were analyzed in the FBAT[13] program. The Haplotype blocks were generated using the subject genotype data and the full genotype data downloaded from the HapMap in Haploview, defined by the confidence interval method.

First, we tested for excessive transmission of alleles at each marker separately, and then for excessive transmission of multiple-marker haplotypes. Haplotypes were reconstructed using the HBAT command in FBAT for FBAT analyses, which uses the EM algorithm. In these analyses, we used the empirical variance estimator option (-e) as our primary hypothesis. Empiric P-values were generated by permutation tests for individual SNPs and haplotypes, and results reported in the Tables 1 and 2 as uncorrected permuted P-values. The Bonferroni method was used for correction of the P-value for multiple testing. The statistical power of this study with the present sample size was also estimated by Quante 1.2.4.[23]

Table 1 Markers characteristic and single-SNP association result in the family sample

SNPs	Polymorphism (major/minor)	Risk allele	FBAT Fre.[a]	Var(S)[b]	Z	P
Additive model						
rs225015	G/A	G	0.582	25.731	1.2.325	**0.020**[c]
rs225014	C/T	T	0.561	27.750	1.046	0.296
rs225012	A/G	G	0.741	26.682	1.492	0.136
rs2267872	G/A	A	0.314	31.719	2.113	**0.034**[c]
rs1388378	C/A	A	0.293	30.504	2.405	**0.016**[c]
Recessive model						
rs225015	G/A	A	0.418	8.113	−1.897	0.058
rs225014	C/T	T	0.561	13.825	0.524	0.600
rs225012	A/G	G	0.741	20.186	1.479	0.139
rs2267872	G/A	G	0.686	16.890	−1.974	**0.048**[c]
rs1388378	C/A	C	0.707	17.307	−2.951	**0.003**[c]

Abbreviations: FBAT, Family Based Association Tests; SNP, single nucleotide polymorphism.
[a]Fre., is the frequency of risk allele.
[b]Var(S), is a matrix, calculated under the null and used to standardize S.
[c]Bold type denotes $P < 0.05$.

Table 2 Haplotype-based association results in the family sample

Haplotypes[a]		Fre.[b]	FBAT Z	P	S[c]	E(S)	Global P
rs225015(A/G)–rs2267872(A/G)–rs1388378(A/C)							
H1	A-G-C	0.399	−2.995	**0.003**[d]	−12.249	16.724	**0.022**[d]
H2	G-G-C	0.275	0.042	0.966	0.191	20.491	
H3	G-A-C	0.230	2.189	**0.028**[d]	11.422	27.222	
rs225015(A/G)–rs2267872(A/G)							
H1	A-G	0.402	−3.257	**0.001**[d]	−14.121	18.799	**0.012**[d]
H2	G-G	0.299	0.499	0.618	2.431	23.731	
H3	G-A	0.280	1.964	**0.050**[d]	10.946	31.075	
rs2267872(A/G)–rs1388378(A/C)							
H1	G-C	0.673	−2.067	**0.039**[d]	−11.783	32.482	0.100
H2	A-A	0.237	2.210	**0.027**[d]	11.640	27.753	
H3	A-C	0.060	−0.626	0.532	−1.219	3.798	

Abbreviation: FBAT, Family Based Association Tests.
[a]The haplotypes frequency <0.05 were excluded in global P-value was calculated;
[b]Fre., is the frequency of risk allele.
[c]S, is the FBAT marker score, and E(S) is the expected values of the marker score;
[d]Bold type denotes $P < 0.05$.

Three aspects of bio-function of these SNPs were evaluated in silicon modeling: (1) The transcription factor-binding site or a promoter site analysis, which verifies whether the SNP can modify the expression level of the *DIO2* gene; (2) the conservative property estimation and the splice site research, which confirmed whether the SNP can influence the product's function, feature, structure and so on and (3) whether a new product will come into being because of these SNPs? How about its function? Structure?

We used the time delay neural networks (TDNN) method[24] and the TFSEARCH program[25] to analyze the 50 bp sequence surrounding each SNP, and identify gain or loss of physical transcription factor-binding sites and promoter sites generated by the nucleotide exchange corresponding to alternative SNP alleles. Two analytical softwares were used to verify whether each

SNP of the *DIO2* gene has putative regulatory activity that may affect its transcription and expression. In both cases, the vertebrate promotor sites' features, transcription factor matrices and optimized matrix threshold were used to reduce the incidence of false positives.

VISTA was used to define the conserved regions in the genomic sequence of *DIO2* surrounding specific SNPs[26] and were visualized as added tracks on the University of California Santa Cruz genome browser (http://genome.ucsc.edu/). A total of 50 bp of DNA sequence including each SNP allele was analyzed for putative splice sites using the predictor from the Berkeley Drosophila genome project (http://www.fruitfly.org/seq_tools/splice.html). Database similarity research, and protein open-reading frame identification were performed using BLAST and ORF-finder via the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov). The Phobius and InterProScan (http://www.ebi.ac.uk/interpro/) programs, integrated database of predictive protein "signatures" and structure feature were used for classification and automatic annotation of proteins.

RESULTS

Five SNPs in the *DIO2* gene were genotyped. These five SNPs had a reasonable heterozygosity (observed HET >0.35), and the rare allele frequency was also >0.29 in the present samples. The deviation from Hardy–Weinberg equilibrium tests was calculated by HAPLOVIEW 4.2, and no significant results were found. More than 169 trio-families can support the useful genetic information, which also guarantee the statistical power of this study. Genetic statistical power estimated that this sample had $>90\%$ power for the FBAT test (MAF >0.25, OR value >2.0), estimated by QUANTE 1.2.4.[23]

In the family samples, three SNPs (rs225015, rs2267872 and rs1388378) were significantly associated at the 0.05 significance level, for at least one of three inherited models, as displayed in Table 1, and survived the multiple test correction ($P<0.01$). The genotype data of HapMap were downloaded for the HapMap Genome Browser (Phase 1 and 2—full data set), which indicated a strong linkage disequilibrium between them ($D' >0.94$). The five SNPs formed one haplotype block (Supplementary Figure 1). The halpotype structure, which was estimated with the genotype information of five SNPs in the present samples by the HAPLOVIEW 4.2 package, displayed a very similar result.

On the basis of these results, the haplotype association tests were conducted. The haplotype combinations constructed by the rs225015, rs2267872 and rs1388378 were analyzed. The haplotype analysis was performed, and indicated that the several individual haplotypes constructed by three SNPs had a significant association. Haplotypes constructed by rs225015–rs2267872–rs1388378 and rs225015–rs2267872 all showed a substantial global significant association (global $P=0.022$ and 0.012, respectively; Table 2).

TFSEARCH program analysis showed that rs225015 and rs1388378 all perturbed transcription factor-binding sites. More than three binding sites were influenced by rs1388378, the most significantly associated SNP in the family sample (Supplementary Table 2). Although functionality for these binding sites is currently putative, it is possible that these SNPs could affect the regulation of *DIO2* gene expression. There was no evidence of cryptic splice site creation for any of the associated SNPs.

No higher similarities were found for these SNPs against the database of expressed sequence target by BLASTN sequence similarity search, whereas the incorporation of selected VISTA tracks onto the UCSC genome browser showed this genomic region to be conserved in mammals, for example, mouse and dog (Supplementary Figure 2). Open reading frame finder estimated that the sequence may encode a 48-amino-acid multi-peptide, which functions as a signal peptide with $>25\%$ possibility, whereas the SNP5's alternative can influence its

secondary structure (a 6–23 amino acid transmembrane structure) and the putative function (Supplementary Figure 3). Therefore, we cannot rule out the possibility that rs1388378 would be a biologically meaningful SNP.

DISCUSSION

The Qinba mountain region is one of the higher MR prevalence areas in China. The previous epidemic investigations suggested that there were 218 000 MR patients including 66 000 children in the Qinba mountain region according to official statistics based on the WHO (World Health Organization) standards. The incidence ratio is about 2.78% and rises to 8.23% when the number of children in the low range of normal IQ is included.[27] MR represents an important socioeconomic and medical issue, given that MR patients need continuous support from families and health-care providers. The epidemiological survey and pedigree's investigation revealed a higher heritability and distinct familial aggregation for MR in this region,[28,29] indicating that the genetic factors and regional nutritional deficiencies may be the important causes. All these demographic, high heritability and familial aggregation features impel us to focus on the relationship between the individual's genetic variants and the MR susceptibility in this region.

Iodine is an essential element that enables the thyroid gland to produce thyroid hormones. Fetal iodine deficiency is the commonest cause of preventable MR,[30] because of inadequate amounts of thyroid hormones, which regulate the processes of terminal brain differentiation such as dendritic and axonal growth, synaptogenesis, neuronal migration and myelination. Variants come from the elements involved in the synthetic process maybe because of the individual's thyroid hormone deficiency, and the presence of MR. Moreover, we also wondered whether the genetic variants of candidate gene products take part in involving or regulating the thyroid hormones' metabolism process, for instance, the iodine transport, thyroid dysgenesis, dyshormonogenesis and iodine circulation. The previous work also confirmed this hypothesis; the positive associations were found between *DIO2* and *POU1F1* genes with an independent case–control MR sample.

In this study, we tested the candidate gene *DIO2* for association with MR in 157 high-density pedigrees in the Qinba region, which is one of the iodine-deficient areas of China. As the previous work indicated, our family-based association tests yielded considerable evidence for association in several markers, which were in significant linkage equilibrium (LD) with each other. This was supported by the fact that three SNPs (rs225015, rs2267872 and rs1388378) had a solid association between the genetic variants of *DIO2* gene and MR. One of the SNPs (rs1388378) survived the Bonferroni correction, regardless of whether in single SNP analysis or haplotype analysis (Tables 1 and 2). The haplotypes constructed by rs225015, rs2267872 and rs1388378 also displayed a significant global association ($P=0.022$ and 0.012, respectively). This result also confirmed our previous research work[8] in an independent family-based sample.

Guo *et al.*[8] reported the positive association between rs225012 and MR, although only three SNPs were investigated with a case–control sample. Because of lower genotyping ratio and heterogeneity, we did not obtain the effective data of rs225012. However, the best positive SNP rs1388378 was also located in the same intron (Supplementary Figure 2). The true LD block indicates the genomic region across recombination hot spots, which also indicated that the tSNPs may not be located within the target gene region sometimes. However, given the fact that more tSNPs were tested, the other functional or strong association SNPs may be found out. All of these indicated that maybe

we overlooked some more strong associations for hidden SNPs and MR, because of the limited SNP density in this study. Therefore, further work should be done.

Although five tSNPs in this study were spread around the *DIO2* gene, including two exons and one intron (Table 1), no specific function of these tSNPs were reported, unfortunately. The functional studies of the associated variants observed in this study may be informative, given the weight of previous studies. It is therefore of interest that the most significant SNP (rs1388378) in the family sample was observed to occur in five transcription factor-binding sites in our *in silico* analyses. The conserved sequence surrounding rs1388378 may encode a 48-amino-acid multi-peptide, a signal peptide. The assumed function and structure of this peptide were perturbed by the alternative of rs1388378. However, all of these were only in silico analysis; *in vitro* studies of this SNP may further elucidate its effect on gene expression, and possibly its role in MR susceptibility.

1 Anderson, G. W. Thyroid hormones and the brain. *Front Neuroendocrinol* **22,** 1–17 (2001).
2 Oppenheimer, J. H. & Schwartz, H. L. Molecular basis of thyroid hormone dependent brain development. *Endocr Rev* **18,** 462–475 (1997).
3 Croteau, W., Davey, J. C., Galton, V. A. & St Germain, D. L. Cloning of the mammalian type II iodothyronine deiodinase. A selenoprotein differentially expressed and regulated in human and rat brain and other tissues. *J Clin Invest* **98,** 405–417 (1996).
4 Zimmermann, M. B., Jooste, P. L. & Pandav, C. S. Iodine-deficiency disorders. *Lancet* **372,** 1251–1262 (2008).
5 International Council for the Control of Iodine Deficiency Disorders (ICCIDD). IODINE AND THE BRAIN. (2004). URL: www.iccidd.org/.
6 Karger, S., Schotz, S., Stumvoll, M., Berger, F. & Fuhrer, D. Impact of pregnancy on prevalence of goitre and nodular thyroid disease in women living in a region of borderline sufficient iodine supply. *Horm Metab Res* **42,** 137–142 (2010).
7 Dratman, M. B. & Gordon, J. T. Thyroid hormones as neurotransmitters. *Thyroid* **6,** 639–647 (1996).
8 Guo, T. W., Zhang, F. C., Yang, M. S., Gao, X. C., Bian, L., Duan, S. W. *et al.* Positive association of the DIO2 (deiodinase type 2) gene with mental retardation in the iodine-deficient areas of China. *J Med Genet* **41,** 585–590 (2004).
9 Davis, J. D., Podolanczuk, A., Donahue, J. E., Stopa, E., Hennessey, J. V., Luo, L. G. *et al.* Thyroid hormone levels in the prefrontal cortex of post-mortem brains of Alzheimer's disease patients. *Curr Aging Sci* **1,** 175–181 (2008).
10 He, B., Li, J., Wang, G., Ju, W., Lu, Y., Shi, Y. *et al.* Association of genetic polymorphisms in the type II deiodinase gene with bipolar disorder in a subset of Chinese population. *Prog Neuropsychopharmacol Biol Psychiatry* **33,** 986–990 (2009).
11 Lin, H. W. & Chen, Y. H. Association analysis under population stratification: a two-stage procedure utilizing population- and family-based analyses. *Hum Hered* **69,** 11 (2010).
12 Tsai, H. J., Kho, J. Y., Shaikh, N., Choudhry, S., Naqvi, M., Navarro, D. *et al.* Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet* **118,** 626–639 (2006).
13 Laird, N. M., Horvath, S. & Xu, X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* **19,** S36–42 (2000).
14 CPA. *The Chinese Classification of Mental Disorders, Second Edition, Revised (CCMD-2-R)* (Chinese Psychiatric Association, Quanzhou, 1995).
15 WHO. *The ICD-10 Classification of Mental and Behacioral Disorders: the Clinical Descriptions and Diagnostic Guiddline* (World Health Organzation, Geneva, 1992).
16 Gong, Y. X. & Dai, X. Y. *Chinese-Wechsler Young Children Scale of Intelligence (C-WYCSI)* (Map Press, Hunan, 1992))).
17 Gong, Y. X. & Cai, T. S. *Wechsler Intelligence Scale for Children Chinese Revision (C-WISC)* (Map Press, Hunan, 1993).
18 Zuo, Q. H., Zhang, Z. X. & Lei, Z. W. *Adaptive Scale of Infant and Children* (Medical University of China, Beijing, 1988).
19 Gong, Y. X. & Dai, X. Y. *The Manual of Chinese Revision of the Wechsler Adult Intelligence Scale (WAIS-RC)* (the Map Press, Chang Sha, 1992).
20 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).
21 Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296,** 2225–2229 (2002).
22 O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63,** 259–266 (1998).
23 Gauderman, W. J. Sample size requirements for matched case-control studies of gene-environment interaction. *Statistics in Medicine* **24,** 35–50 (2002).
24 Hampshire, J. B. & Waibel, A. H. A novel objective function for improved phoneme recognition using time-delay neural networks. *IEEE Trans Neural Netw* **1,** 216–228 (1990).
25 Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V. *et al.* Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* **26,** 362–367 (1998).
26 Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E. *et al.* Strategies and tools for whole-genome alignments. *Genome Res* **13,** 73–80 (2003).
27 Zhang, F. C., Li, R. L., Gao, X. C., Zheng, Z. J., Huang, S. P., Song, H. Y. *et al.* An investigation of mental retarded children aged 0–14 in Qinba Mountain areas. *J Fourth Mil Med Univ* **25,** 511–514 (2004).
28 Zhang, K. J., Zhang, F. C., Zheng, Z. J., Gong, P. Y., He, G., Nan, Y. P. *et al.* An analysis of inheriting type of non-causing mental retarded children in Zhashui experimental station. *J NW Uni (Nat Sci Ed)* **35,** 597–600 (2005).
29 Zhang, S. M., Zhang, F. C., Gong, P. Y., Zhang, K. J., Yhang, X. Z., He, G. *et al.* The discussion of inheriting type of mental retardation in Ankang experimental 'ion. *J Northwest Un (Nat Sci Ed)* **36,** 97–100 (2006).
30 WHO. *National Strategies for Overcoming Microntrient Malnutrition, 89th* (WHO: Geneva, Switzerland, 1992).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)