Analysis of epistasis for diabetic nephropathy among type 2 diabetic patients

Chang-Hsun Hsieh1,†, Kung-Hao Liang2,*,†, Yi-Jen Hung1, Li-Chin Huang2, Dee Pei3, Ya-Tang Liao2, Shi-Wen Kuo3, Monica Shian-Jy Bey2, Jui-Lin Chen2 and Ellson Y. Chen2

1Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General Hospital, Taipei, Taiwan, 2Vita Genomics Inc., Taiwan and 3Department of Internal Medicine, Xindian Buddhist Tzu Chi Hospital, Taipei, Taiwan

Received May 9, 2006; Revised and Accepted July 28, 2006

Diabetic nephropathy (DN) is one of the most serious complications of diabetes, accounting for the majority of patients with end-stage renal disease. The molecular pathogenesis of DN involves multiple pathways in a complex, partially resolved manner. The paper presents an exploratory epistatic study for DN. Association analysis were performed on 231 SNP loci in a cohort of 264 type 2 diabetes patients, followed by the epistasis analysis using the multifactor dimensionality reduction and the genetic algorithm with Boolean algebra. A two-locus epistatic effect of EGFR and RXRG was identified, with a cross-validation consistency of 91.7%.

INTRODUCTION

Excessive plasma glucose in diabetic patients induces microvascular complications such as nephropathy and retinopathy. As the major cause of end-stage renal disease (1), resulting in patient morbidity and mortality, diabetic nephropathy (DN) is a growing public health concern. Among diabetic patients, ~5% of Caucasians, 10% of Asian Americans, 20% of African Americans and Mexican Americans, and 50% of native Americans eventually require kidney dialysis (2). The decline of renal function on diabetic patients is correlated with hypertrophy of mesangium and glomerular basement membrane, from excessive deposition of extracellular matrix (ECM) (3,4); and interstitial fibrosis (4). Such structural lesions precede any clinically observable changes in blood or urine (1). Hence, a risk-assessment method of nephropathy, based on the genotypes, would benefit the patient care of diabetes.

Multiple pathways have been implicated in the molecular pathogenesis of DN (1,4–7). Excessive ambient glucose induces reactive oxygen species, which then activate protein kinase C and mitogen-activated protein kinases (3,5,7), resulting in mesangial cell hypertrophy and the accumulation of ECM (4). The renin–angiotensin system (RAS) is also involved, which has been shown to be up-regulated in hyperglycaemia. Intra-glomerular hypertension induces fibrogenic cytokines such as transforming growth factor β (TGF-β), which augments mesangial ECM expansion and renal insufficiency (6,8,9).

Genetic variations has also been implicated in the etiology of DN among type 1 and type 2 diabetic patients (4,7,10). Familial aggregation of DN occurred in the Pima Indians (11). The genetic pathogenesis of DN comprises both hypertension-related and non-hypertension-related mechanisms (4,12). Concerning hypertension-related mechanisms, a polymorphism in the angiotensin-converting enzyme (ACE), an element of RAS, has been associated with DN. Both ACE inhibitors and angiotensin receptor I antagonist attenuate DN (4), hence, they are used as treatments for diabetes, aiming to prevent the onset and progression of complications (12). As for the non-hypertension-related factors, several candidate alleles have been reported. Shcherbak and Schwartz (13) have conducted a single nucleotide polymorphism (SNP) association study on a G-protein β3 subunit gene (GNB3) from Russian type 1 diabetic patients, and concluded that no significant association of this gene to DN could be identified. In positive studies, an SNP in the manganese superoxide dismutase gene was found to be associated with DN among Japanese type 2 diabetic patients (14). A meta-analysis of six independent association studies supported the association of SNP in the glucose transporter type 1 gene (GLUT1) with DN (15).

*To whom correspondence should be addressed at: Vita Genomics Inc., 7F, No. 6, Sec. 1, Junghsing Road., Wugu Shiang, Taipei County, 248 Taiwan. Tel: +886 289769123 ext. 7702; Fax: +886 289769523; Email: kunghao.liang@vitagenomics.com
†The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

© The Author 2006. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
A mutation of the TGF-β gene was also found to be associated with DN (16). Roles of TGF-β and GLUT1 in DN, inferred from in vitro studies, are summarized in (8). Recently, the association between the IGF-binding protein 1 (IGFBP1) and the impaired renal function has been reported (17).

The analysis of epistasis is very important for revealing the genetic basis of complex traits (18,19). It requires the genotypes of multiple loci from each study subject. However, the aforementioned studies were mostly on the single gene/single SNP basis. A large-scale investigation of nephropathy and diabetes is currently in progress on four ethnic groups: European Americans, African Americans, Mexican Americans and American Indians (20); and the results have not yet been reported. Considering the complex mechanisms involved, the susceptibility of DN may be adequately evaluated on the basis of the epistasis of multiple genes. Hence, we conducted a retrospective study to investigate the epistatic effects among genes. The epistatic effect will be presented in a terse model as the antihypertensive medication is not controlled in our patients, respectively, received antihypertensive medication. As the antihypertensive medication is not controlled in our patients, respectively, received antihypertensive medication. The analysis of epistasis is very important for revealing the genetic basis of complex traits (18,19). It requires the genotypes of multiple loci from each study subject. However, the aforementioned studies were mostly on the single gene/single SNP basis. A large-scale investigation of nephropathy and diabetes is currently in progress on four ethnic groups: European Americans, African Americans, Mexican Americans and American Indians (20); and the results have not yet been reported. Considering the complex mechanisms involved, the susceptibility of DN may be adequately evaluated on the basis of the epistasis of multiple genes. Hence, we conducted a retrospective study to investigate the epistatic effects among genes. The epistatic effect will be presented in a terse model which can be consistently verified in independent datasets. The patients were type 2 diabetes patients of the Han Chinese population in Taiwan.

A two-stage methodology was employed for the epistasis analysis. First, the allele and genotype frequencies of 231 SNP loci were compared between the case (DN) and control (type 2 diabetes mellitus, DM) groups using χ² statistics (21). Those loci with smaller P-values were further analyzed in the second stage, employing the multifactor dimensionality reduction (MDR) method (22), as well as the genetic algorithm with Boolean algebra method (GABA) (23). The purpose of using two methods is to combine their strengths for this analysis.

RESULTS

Study group comparison

The demographic information of study patients is compared and summarized in Table 1. The average years of diabetic history are 15 for the DN group and 13 for the DM group. Comparing age, body mass index (BMI) and hip circumferences between the two groups, all P-values were larger than 0.05, implying no significant difference were detected. Fasting plasma glucose and HbA1c levels were also similar between the two groups (P > 0.05). In contrast, urinary albumin and blood urea nitrogen (BUN) levels were significantly different between the two groups (P < 0.0001). All the P-values reported in this paper were derived from two-sided tests. In the DM and DN groups, 63.33 and 97.92% of patients, respectively, received antihypertensive medication. As the antihypertensive medication is not controlled in our study design, the hypertension-related genetic association tests should only serve as a reference.

Single SNP association analysis

Five groups of candidate genes were screened in this study due to their putative roles in the development of DN: (i) fibrotic and inflammatory genes (ABCC8, AHSG, GPR87, GPR105, IL4R, KNG1, LPL, MED12L); (ii) genes involved in the construction/destuction of the ECM (BGN, COL1A1, COL3A1, COL6A1, MMP9, MMP14, SDCL); (iii) genes involved in cell growth and proliferation (DGKG, EGFR, MKKS, PDGFRB, RXRG); (iv) insulin and diabetes-related genes (IGF2R, PCSK2, SNAP25) and (v) the hypertension-related genes (ACE, AGT, AGTR1, KCNS3, LRP3). The 28 genes are summarized in Table 2. Gene names and symbols were provided according to the HUGO gene nomenclature committee (24). A set of 231 SNPs in these genes were investigated, where the allelic and genotypic frequencies between the case and control groups were compared using standard χ² statistics for contingency tables (21). The genotypic comparison employs a three by two contingency table, comparing three (one heterozygous and two homozygous) genotypes with two traits (DN and DM). Among the 231 SNPs, 13 SNPs obtained P-values smaller

| Table 1. Comparisons of demographic and clinical information of study subjects |
|------------------------------------------------------------------|------------------|------------------|------------------|
|                                | DN (n = 144)     | DM (n = 120)     | DN versus DM     |
|                                | Statistics       | Statistics       | P-value          |
| Age* (years)                  | 144 62.9 ± 8.8   | 120 63.1 ± 8.7   | 0.8764           |
| Diabetes duration* (years)    | 144 15.5 ± 7.7   | 120 13.7 ± 7.1   | 0.0491           |
| Gender (%) male               | 144 50           | 120 41.67        | 0.1763           |
| BMI (kg/m²)                   | 143 25.7 ± 3.7   | 116 24.8 ± 4.3   | 0.0605           |
| Weight (kg)                   | 143 66.8 ± 11.8  | 116 63.3 ± 12.5  | 0.0124           |
| Height (cm)                   | 143 160.9 ± 8.5  | 116 159.7 ± 10.4 | 0.0124           |
| Hip circumference             | 133 99.3 ± 8.8   | 115 97.5 ± 8.8   | 0.1069           |
| Fasting plasma glucose (mmol/l)| 144 9.9 ± 3.7    | 120 9.3 ± 0.2    | 0.1069           |
| HbA1c (%)                     | 110 8.8 ± 2.0    | 96 8.5 ± 1.3     | 0.1649           |
| Urinary ACR (mg/mmol)         | 124 249.2 ± 244.4| 120 2.2 ± 1.4    | <0.0001          |
| BUN (mg/dl)                   | 142 32.5 ± 16.9  | 120 15.0 ± 3.2   | <0.0001          |
| Blood pressure (mmHg)         | 139 136.7 ± 17.1 | 115 126.7 ± 16.8 | <0.0001          |
| Mean systolic                 | 139 78.3 ± 12.6  | 115 73.4 ± 9.2   | 0.0004           |
| Mean diastolic                | 144 97.92        | 120 63.33        | <0.0001          |

Statistics are represented as mean ± standard deviation. 

n is the sample size.

*aIn December 2002.
than 0.05 in either the allelic or genotypic comparison. These SNPs were summarized in Table 3. However, since 231 SNPs were assessed altogether, issues of multiple comparisons may be considered. The P-value is adjusted as 0.05/231 = 0.0002 using the Bonferroni correction method. None of the 13 SNPs is significant enough according to the Bonferroni correction method (Table 3).

An SNP in the SNAP25 gene (rs1051312, SNP12) showed the strongest associations in both the allelic comparisons (P = 0.0043, power = 98.2%) and genotypic comparisons (P = 0.0055), among all the SNPs in this study. This SNP resides on the 3′ untranslated region of the gene. Tests of Hardy–Weinberg equilibrium (HWE) on the 13 SNPs were also performed on the basis of the χ² statistics (25). The tests were conducted on the DN and DM groups, as well as the whole study population, DN + DM (Table 3).

### Analysis of epistasis

MDR and GABA were employed for the analysis of epistatic effects on the dichotomous, qualitative traits, DN versus DM. The epistatic effect is depicted by the models. The MDR is employed to enumerate all possible combinations of SNPs, with various model lengths between one and 10. The optimum model of each length is presented in Table 4. A 12-fold cross-validation test was conducted such that the model constructed on the training data (comprising 11/12 of the entire dataset) is validated (on the remaining 1/12 portion of the dataset). The cross-validation consistency is the primary indication of model performances, based on the assumption that a reliable model should be consistently detected regardless of which portion of the dataset was used (22). The prediction accuracy, an average of accuracies calculated on all the validation datasets, is the secondary performance index. When single-locus SNPs were assessed by MDR, SNP12 was considered optimum (Table 4). This is consistent with the result in the previous section. The two-locus MDR model, comprising SNPs zero and 11, has the highest cross-validation consistency (11/12 = 91.7%) and prediction accuracy (62.7%) among all the models. It also achieved the smallest P-value of the sign test for the cross-validation. The two-locus MDR model is presented in Figure 1.

Using the same dataset with 13 SNPs, the GABA software detects an optimum model comprising seven loci in six genes, based on the sensitivity and specificity of predictions. A diabetic patient susceptible to nephropathy is identified if the following statement is true:

$$(SNP0 = ‘CC’/‘GC’)*(SNP7 = ‘GG’/‘AG’)*(SNP11 = ‘GG’) + (SNP1 = ‘TT’/‘CT’)*(SNP2 = ‘AA’/‘AT’)*(SNP6 = ‘AA’/‘AG’)*(SNP12 = ‘TT’).$$

where the multiplicative operator ‘*’ corresponds to the logical ‘AND’ and the additive operator ‘+’ the logical ‘OR’. In contrast, the insusceptible patients are identified if the following complement statement is true:

$$(SNP0 = ‘GG’)*(SNP7 = ‘AA’)*(SNP11 = ‘AG’/‘AA’)/(SNP1 = ‘CC’)*(SNP2 = ‘TT’)*(SNP6 = ‘GG’)/(SNP12 = ‘TC’/‘CC’).$$

When the seven SNPs are combined for the prediction of patients in the study group, the result is as shown in Table 5. The sample size is 246, because the other 18 subjects have missing genotypes occurring in the seven loci. The performance indexes of the prediction are as follows: the sensitivity is 81.5%, the specificity is 61.3%, the positive predictive value (PPV) is 71.9% and the negative predictive value (NPV) is 73.1%. The concordance rate between the clinical status and the prediction is 72.4%. Considering other non-genetic factor involved in the development of nephropathy, the prediction performance is reasonably acceptable.

The combination of SNP0 (on EGFR) and SNP11 (on RXRG) in the two-locus MDR model also appeared as a portion of the seven-locus model of GABA. The repeated appearance of SNP0 and SNP11 motivated our further investigation on their epistatic effects. Simplifying the seven-locus GABA model for susceptible patients, we obtain a two-locus model in the Boolean statement as

$$(SNP0 = ‘CC’/‘GC’)*(SNP11 = ‘GG’).$$

The prediction performance of the above model is shown in Table 6. The sensitivity is 62.7%, the specificity is 68.1%,

<table>
<thead>
<tr>
<th>Gene symbols</th>
<th>No. of SNP</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCG8</td>
<td>17</td>
<td>ATP-binding cassette, sub-family C, member 8</td>
</tr>
<tr>
<td>ACE</td>
<td>6</td>
<td>Angiotensin I converting enzyme</td>
</tr>
<tr>
<td>AHSG</td>
<td>6</td>
<td>Alpha-2-HS-glycoprotein</td>
</tr>
<tr>
<td>AGT</td>
<td>1</td>
<td>Angiotensinogen</td>
</tr>
<tr>
<td>AGTR1</td>
<td>1</td>
<td>Angiotensin II receptor, type 1</td>
</tr>
<tr>
<td>BGN</td>
<td>4</td>
<td>Biglycan</td>
</tr>
<tr>
<td>COL1A1</td>
<td>4</td>
<td>Collagen, type I, alpha 1</td>
</tr>
<tr>
<td>COL3A1</td>
<td>1</td>
<td>Collagen, type III, alpha 1</td>
</tr>
<tr>
<td>COL6A1</td>
<td>1</td>
<td>Collagen, type VI, alpha 1</td>
</tr>
<tr>
<td>EGF</td>
<td>8</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>DGK</td>
<td>26</td>
<td>Diacylglycerol kinase, gamma</td>
</tr>
<tr>
<td>GPR105</td>
<td>13</td>
<td>G-protein-coupled receptor 105</td>
</tr>
<tr>
<td>GPR87</td>
<td>5</td>
<td>G-protein-coupled receptor 87</td>
</tr>
<tr>
<td>IGF2R</td>
<td>10</td>
<td>Insulin-like growth factor 2 receptor</td>
</tr>
<tr>
<td>IL4R</td>
<td>4</td>
<td>Interleukin 4 receptor</td>
</tr>
<tr>
<td>KCNS3</td>
<td>3</td>
<td>Potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3</td>
</tr>
<tr>
<td>KNG1</td>
<td>3</td>
<td>Kinogen 1</td>
</tr>
<tr>
<td>LPL</td>
<td>7</td>
<td>Lipid protein lipase</td>
</tr>
<tr>
<td>LRP3</td>
<td>5</td>
<td>Low density lipoprotein receptor-related protein 3</td>
</tr>
<tr>
<td>MED12L</td>
<td>1</td>
<td>Mediator of RNA polymerase II transcription</td>
</tr>
<tr>
<td>MKKS</td>
<td>3</td>
<td>McKusick–Kaufman syndrome</td>
</tr>
<tr>
<td>MMP14</td>
<td>2</td>
<td>Matrix metalloproteinase 14</td>
</tr>
<tr>
<td>MMP9</td>
<td>16</td>
<td>Matrix metalloproteinase 9</td>
</tr>
<tr>
<td>PCSK2</td>
<td>50</td>
<td>Proprotein convertase subtilisin/kexin type 2</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>9</td>
<td>Platelet-derived growth factor receptor, beta polypeptide</td>
</tr>
<tr>
<td>RXRG</td>
<td>9</td>
<td>Retinoid X receptor, gamma</td>
</tr>
<tr>
<td>SDC1</td>
<td>3</td>
<td>Syndecan 1</td>
</tr>
<tr>
<td>SCN15</td>
<td>13</td>
<td>Synaptosomal-associated protein</td>
</tr>
<tr>
<td>Total=28</td>
<td>231</td>
<td>Total=28</td>
</tr>
</tbody>
</table>
which have at least one MED12L EGFR SNP0 for them to reside in the same LD block.

SNP1 is 40 kb, a large distance that reduces the possibility of figure 3, respectively. The distance between SNP0 and SNP1 of Table 3 correspond to numbers 3 and 4 because the adjacent SNPs were in linkage equilibrium. Analysis of linkage disequilibrium on genes was further investigated. LD was shown as D prime numbers in figures 2 and 3, which is calculated by the Haploview software using the default algorithm on confidence intervals (26). Three blocks were identified on RXRG. SNP10 and SNP11 of Table 3, corresponding to numbers 3 and 5 of figure 2, respectively, reside in different LD block. Table 7. It is shown that the haplotype 'TA' of Block2 is no blocks were detected from the eight SNPs of Table 4. The association study based on haplotypes is presented in Figure 2, respectively. Among the genes pertaining to fibrosis and inflammation, IL4R, LPL, PDGFRB, RXRG and SNAP25. Among the genes related to cell growth and proliferation, EGFR, MED12L, IGF2R, IL4R, LPL, PDGFRB, RXRG and SNAP25 were identified.

**DISCUSSION**

**Single SNP association analysis**

This research employs a two-stage methodology, where the first stage is the single SNP association study. We screened 231 SNPs, and 13 of them have a P-value less than 0.05 in either the allelic or genotypic tests. The 13 SNPs belong to eight genes EGFR, MED12L, IGF2R, IL4R, LPL, PDGFRB, RXRG and SNAP25.

- **rs6782313 C____268273_10 A/T**
  - Allele: C
  - Genotype: 10/12
  - P-value: 0.4599
  - Confidence interval: (0.9270)
- **rs12671550 C____2678627_0 G/C**
  - Allele: C
  - Genotype: 12/12
  - P-value: 0.5759
  - Confidence interval: (0.0361)
- **rs2072454 C____2678638_1_ C/T**
  - Allele: C
  - Genotype: 10/12
  - P-value: 0.4699
  - Confidence interval: (0.9573)
The synaptosomal-associated protein, SNAP25, is a pre-synaptic plasma membrane protein that functions in the synap- tic vesicle membrane docking and fusion pathway. It is also involved in the regulation of insulin secretion (29,30).

No significant association was found on the hypertension-related genes, probably because the genetic effects were obscured by the antihypertensive treatments of the study groups. Lack of significant association on these genes did not necessarily mean that they have no influence on nephropathy. More sophisticated study design, which stratifies patients according to the longitudinal measurements of blood pressures and antihypertensive medications, is required to reveal the hypertension-related associations.

The synaptosomal-associated protein, SNAP25, is a pre-synaptic plasma membrane protein that functions in the synap- tic vesicle membrane docking and fusion pathway. It is also involved in the regulation of insulin secretion (29,30).

No significant association was found on the hypertension-related genes, probably because the genetic effects were obscured by the antihypertensive treatments of the study groups. Lack of significant association on these genes did not necessarily mean that they have no influence on nephropathy. More sophisticated study design, which stratifies patients according to the longitudinal measurements of blood pressures and antihypertensive medications, is required to reveal the hypertension-related associations.

Analysis of epistasis

Analysis of epistasis has been advocated for the genetic dissection of complex diseases (18,19). A detailed account on various definitions of epistasis is given in (31). Bell et al. (18) presents a two-locus model for hypertension, based on whole-genome sib-pair studies. For examples of higher-order interactions, MDR has successfully identified the interactive roles of four SNPs on the sporadic breast cancer (22), as well as the three-locus epistatic effect on atrial fibrillation (32).
We combined the strength of MDR and GABA for finding the epistatic effect on DN. MDR presents the epistasis effect in a tabular format where each cell of the table represents a particular combination of genotypes (e.g. Fig. 1). This format of presentation is not adequate for higher-order interactions (e.g. involving more than five loci). Unlike MDR, GABA presents models in Boolean expressions, where the interactions of loci are stated in an algebraic equation (23). GABA is capable of detecting higher-order interactions and showing them in terse statements. That is why we employ GABA in addition to MDR for facilitating further biological interpretations.

The pathogenesis of DN involves complicated interactions between multiple genetic and environmental factors. The stringent criteria on the study groups have permitted us to obtain evidences on the novel coupling effect of EGFR and RXRG on DN. SNP0 (rs12671550) resides in the first intron region of EGFR. The EGFR signaling pathway is one of the most important pathways that regulate growth, survival, proliferation and differentiation in mammalian cells (33). The expression of EGFR is also closely related to tumorigenesis (33). RXRG, a member of the retinoid X receptor (RXR), is a nuclear receptor which is involved in mediating the anti-proliferative effects of retinoic acid. Retinoids, including vitamin A and its synthetic and non-synthetic derivatives, modulate fundamental cellular processes, including cell growth, differentiation and apoptosis. EGFR and retinoid-dependent signaling pathways have both been shown to play roles in carcinogenesis.

The epistatic effect of EGFR and RXRG provides a starting point for further fine-mapping and functional studies on these two regions, which is currently in progress. Haplotypes are valuable information for association, particularly, after the dataset of International HapMap project is released for references (34). This will enable the investigation of the epistatic effects based on multiple haplotypes. Nevertheless, our initial findings on the epistatic effect of EGFR and RXRG may lead toward detailed views on the pathophysiological mechanism of DN, and potential points for clinical interventions.

### MATERIALS AND METHODS

#### Study population

Case/control groups of type 2 diabetic patients, all ethnically Han Chinese, were recruited from the Tri-Service General Hospital in Taipei, Taiwan in 2002. The control group comprised 120 type 2 diabetic patients without nephropathies (DM); the case group comprised 144 diabetic patients with nephropathies (DN). All the recruited patients fulfilled the following criteria: (i) the age was between 30–75 years old; (ii) had been diagnosed with diabetes for more than 5 years; (iii) the fasting plasma glucose was greater than 6.93 mmol/l (126 mg/dl); (iv) the HbA1C was greater than 6%. The rationale of criterion (ii) is that DN is a chronic process that a long-term observation of patients is required to correctly stratify patients. The study subjects were then further classified as DN or DM according to three surrogate endpoints: urinary albumin to creatinine ratio (ACR), BUN and serum creatinine. ACR were measured more than twice at different time points.

Patients fulfilling either one of the following three criteria were classified as the DN group: (i) average ACR was greater than 5.28 mg/mmol (300 μg/mg); (ii) serum creatinine was greater than 1.7 mg/dl; (iii) BUN was greater than 20 mg/dl.

The rest of the patients were classified as the DM group. In this definition, microalbuminuric and macroalbuminuric patients were classified into the DN group.

Approval was obtained from the Internal Review Board of the Tri-Service General Hospital before conducting the study, and the approved informed consent form was signed by each subject.

#### Genes and snps selection

Genes were selected so as to address important aspects of DN: (i) interstitial fibrosis, (ii) ECM disposition and (iii) cell growth and proliferation. The three aspects were all associated with the progression of nephropathy. In addition, insulin and diabetes-related genes were also investigated. Finally, hypertension-related genes were also investigated. All the SNPs chosen had minor allele frequency larger than 5%. They were also confirmed not residing in the repetitive area of the genome, avoiding erroneous associations.

#### DNA extraction and genotyping

DNA was isolated from blood samples using QIAamp DNA blood kit following the manufacture’s instructions (Qiagen). The qualities of isolated genomic DNAs were checked using the agarose gel electrophoresis and the quantities determined using spectrophotometry.
Genotyping was performed using commercial TaqMan® Genotyping assays of Applied Biosystems Inc. (ABI). TaqMan® PCR was performed according to the manufacturer’s standard protocol as follows: 5 ng of genomic DNA was mixed with the 2X TaqMan Universal PCR Master Mix and 20X TaqMan Assay Mix to a final volume of 5 μl, which was then dispensed to a 384-well plate. Each sample underwent 40 amplification cycles on the GeneAmp® PCR System 9700 instrument (ABI). Fluorescent signals of the two probes, corresponding to two different alleles, were analyzed using PRISM® 7900HT Sequence Detection System (ABI). Genotypes were determined automatically by Sequence Detection Software (ABI). SNP IDs in Table 3 were referred to the dbSNP database of the National Center for Biotechnology Information, as well as the ABI assay IDs.

Methods for epistasis analysis
MDR and GABA were chosen for the analysis of epistasis because they are suitable for this study on dichotomous, qualitative traits (DN versus DM). MDR is a non-parametric approach (22). The case/control ratio of this study is ¼, which is reasonably close to 1, fulfilling the assumption of MDR. The open-source Java version MDR software v1.0.0rc1 was used. This software was downloaded from SourceForge.net.

The GABA algorithm was proposed and described in (23). It is briefly summarized here for the ease of readers. It is a combination of a genetic algorithm (GA) and the Boolean algebra (BA). The BA is used to present the model. The GA can systematically evaluate a variety of models, shown as algebra (BA). The BA is used to present the model. The GA carries out the heuristic optimization process with respect to sensitivity and specificity, until the optimum model appears.

Methods for haplotype analysis
The LD blocks were identified by the Haplovie software downloaded from SourceForge.net (26). After the LD blocks were determined, the haplotypes of each block were then derived from the diplotype data (35), using the expectation-maximization method (36).

ACKNOWLEDGEMENTS
The authors would like to thank Chun-Yi Chen and Jun-Jeng Fen for their assistances on data extraction and analysis. They also thank Dr Lawrence Shih-Hsin Wu and Dr David Schlessinger for their critical reading of this manuscript.

Conflict of Interest statement. None declared.

REFERENCES


