

Analysis of epistasis for diabetic nephropathy among type 2 diabetic patients

Chang-Hsun Hsieh^{1,†}, Kung-Hao Liang^{2,*,†}, 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 $\beta 3$ 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, Jungshing 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.

Table 1. Comparisons of demographic and clinical information of study subjects

	DN (<i>n</i> = 144)		DM (<i>n</i> = 120)		DN versus DM <i>P</i> -value
	<i>n</i>	Statistics	<i>n</i>	Statistics	
Age ^a (years)	144	62.9 ± 8.8	120	63.1 ± 8.7	0.8764
Diabetes duration ^a (years)	144	15.5 ± 7.7	120	13.7 ± 7.1	0.0491
Gender (% male)	144	50	120	41.67	0.1763
BMI (kg/m ²)	143	25.7 ± 3.7	116	24.8 ± 4.3	0.0605
Weight (kg)	143	66.8 ± 11.8	116	63.3 ± 12.5	
Height (cm)	143	160.9 ± 8.5	116	159.7 ± 10.4	
Hip circumference	133	99.3 ± 8.8	115	97.5 ± 8.8	0.1124
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)					
Mean systolic	139	136.7 ± 17.1	115	126.7 ± 16.8	<0.0001
Mean diastolic	139	78.3 ± 12.6	115	73.4 ± 9.2	0.0004
Antihypertensive medication (%)	144	97.92	120	63.33	<0.0001

Statistics are represented as mean ± standard deviation.

n is the sample size.

^aIn December 2002.

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 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 χ^2 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/destruction of the ECM (*BGN*, *COL1A1*, *COL3A1*, *COL6A1*, *MMP9*, *MMP14*, *SDCI1*); (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 χ^2 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 2. The 28 genes (231 SNPs) screened in this study

Gene symbols	No. of SNP	Gene name
<i>ABCC8</i>	17	ATP-binding cassette, sub-family C, member 8
<i>ACE</i>	6	Angiotensin I converting enzyme
<i>AHSG</i>	6	Alpha-2-HS-glycoprotein
<i>AGT</i>	1	Angiotensinogen
<i>AGTR1</i>	1	Angiotensin II receptor, type 1
<i>BGN</i>	4	Biglycan
<i>COL1A1</i>	4	Collagen, type I, alpha 1
<i>COL3A1</i>	1	Collagen, type III, alpha 1
<i>COL6A1</i>	1	Collagen, type VI, alpha 1
<i>EGFR</i>	8	Epidermal growth factor receptor
<i>DGKG</i>	26	Diacylglycerol kinase, gamma
<i>GPR105</i>	13	G-protein-coupled receptor 105
<i>GPR87</i>	5	G-protein-coupled receptor 87
<i>IGF2R</i>	10	Insulin-like growth factor 2 receptor
<i>IL4R</i>	4	Interleukin 4 receptor
<i>KCNS3</i>	3	Potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3
<i>KNG1</i>	3	Kininogen 1
<i>LPL</i>	7	Lipid protein lipase
<i>LRP3</i>	5	Low density lipoprotein receptor-related protein 3
<i>MED12L</i>	1	Mediator of RNA polymerase II transcription
<i>MKKS</i>	3	McKusick–Kaufman syndrome
<i>MMP14</i>	2	Matrix metalloproteinase 14
<i>MMP9</i>	16	Matrix metalloproteinase 9
<i>PCSK2</i>	50	Proprotein convertase subtilisin/kexin type 2
<i>PDGFRB</i>	9	Platelet-derived growth factor receptor, beta polypeptide
<i>RXRG</i>	9	Retinoid X receptor, gamma
<i>SDC1</i>	3	Syndecan 1
<i>SNAP25</i>	13	Synaptosomal-associated protein
Total=28	231	

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' un-translated region of the gene. Tests of Hardy–Weinberg equilibrium (HWE) on the 13 SNPs were also performed on the basis of the χ^2 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:

$$\begin{aligned} &(\text{SNP0} = \text{'CC/GC'}) * (\text{SNP7} = \text{'GG/AG'}) * (\text{SNP11} = \text{'GG'}) \\ &+ (\text{SNP1} = \text{'TT/CT'}) * (\text{SNP2} = \text{'AA/AT'}) \\ &* (\text{SNP6} = \text{'AA/AG'}) * (\text{SNP12} = \text{'TT'}) \end{aligned}$$

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:

$$\begin{aligned} &((\text{SNP0} = \text{'GG'}) + (\text{SNP7} = \text{'AA'}) + (\text{SNP11} = \text{'AG/AA'})) \\ &* ((\text{SNP1} = \text{'CC'}) + (\text{SNP2} = \text{'TT'}) + (\text{SNP6} = \text{'GG'})) \\ &+ (\text{SNP12} = \text{'TC/CC'}) \end{aligned}$$

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

$$(\text{SNP0} = \text{'CC/GC'}) * (\text{SNP11} = \text{'GG'}).$$

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

Table 3. Result of comparisons on allelic and genotypic frequencies between the DN and DM groups. The *P*-values were derived from χ^2 test. Only those SNPs which have at least one *P*-value smaller than 0.05 (shown in bold face) are presented here. Tests on HWE of these SNPs were also presented

SNP ID	Gene symbols	dbSNP ID	ABI assay ID	Allele	Number		Association (χ^2 <i>P</i> -value)				
					DN	DM	Allele	Genotype	HWE (All)	HWE (DN)	HWE (DM)
SNP0	<i>EGFR</i>	rs12671550	C__2678627_0	G/C	142	119	0.0084	0.0181	0.0077763	0.376743784	0.006697498
SNP1	<i>EGFR</i>	rs2072454	C__2678638_1_	C/T	141	117	0.0185	0.0576	0.0131299	0.289707177	0.025391149
SNP2	<i>MED12L</i>	rs6782313	C__268273_10	A/T	141	120	0.4599	0.0469	0.5119526	0.036166936	0.207240256
SNP3	<i>IGF2R</i>	rs600324	C__1898214_10	C/T	142	118	0.0361	0.0715	0.2084132	0.948250432	0.112617427
SNP4	<i>IGF2R</i>	rs1803989	C__1979425_10	C/T	142	119	0.0415	0.0897	0.6121855	0.338480013	0.957354183
SNP5	<i>IL4R</i>	rs6498012	C__2769593_10	C/G	143	119	0.0148	0.0444	0.363717	0.933816969	0.293973266
SNP6	<i>LPL</i>	rs326	C__1843005_1_	A/G	144	119	0.5403	0.0481	0.0121221	0.982285123	0.000327745
SNP7	<i>PDGFRB</i>	rs6865659	C__11263234_10	A/G	141	118	0.0228	0.0676	0.7574324	0.611732254	0.77248226
SNP8	<i>PDGFRB</i>	rs2304058	C__11260229_1_	C/G	143	117	0.0257	0.0841	0.9176754	0.945138055	0.997455653
SNP9	<i>PDGFRB</i>	rs740750	C__2277878_1_	A/G	142	115	0.0397	0.1181	0.1013249	0.116600065	0.57564379
SNP10	<i>RXRG</i>	rs746332	CC__1007816_20	A/C	143	116	0.0013	0.0141	0.5393246	0.16604897	0.195027125
SNP11	<i>RXRG</i>	rs3818569	CC__3194630_10	G/A	144	117	0.0151	0.0363	0.2555648	0.676766882	0.182034255
SNP12	<i>SNAP25</i>	rs1051312	CC__339356_10	T/C	144	119	0.0043	0.0055	0.3820841	0.719791965	0.10232779

Table 4. The models detected by MDR and their performance

Locus no.	List of locus in the optimum model	Cross-validation consistency	Prediction accuracy	Sign test (<i>P</i> -value)
1	SNP12	7/12	0.5250	8 (0.1938)
2	SNPs 0, 11	11/12	0.6270	11 (0.0032)
3	SNPs 0, 2, 11	9/12	0.5966	10 (0.0193)
4	SNPs 0, 2, 5, 11	5/12	0.5322	8 (0.1938)
5	SNPs 0, 2, 5, 9, 11	2/12	0.4497	6 (0.6128)
6	SNPs 1, 2, 4, 6, 8, 9	4/12	0.4568	5 (0.8062)
7	SNPs 1, 2, 4, 5, 6, 7, 9	5/12	0.4531	3 (0.9807)
8	SNPs 0, 1, 2, 3, 5, 8, 9, 10	2/12	0.4387	5 (0.8062)
9	SNPs, 0, 1, 2, 3, 4, 5, 6, 9, 11	2/12	0.2308	1 (0.9998)
10	SNPs, 0, 1, 2, 3, 4, 5, 6, 7, 9, 11	10/12	0.3750	4 (0.9270)

the PPV is 70.6%, and the NPV is 59.8%. The concordance rate between the clinical status and the prediction is 65.1%. Although the two-locus model does not reach a prediction performance as high as the seven-locus model, the two-locus model may demonstrate a consistent epistatic effect in independent datasets, based on its parsimonious format, as well as the cross-validation evidences of MDR.

Analysis of linkage disequilibrium on *egfr* and *rxrg*

Linkage disequilibrium (LD) among the SNPs of the *EGFR* and *RXRG* 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. The association study based on haplotypes is presented in Table 7. It is shown that the haplotype 'TA' of Block2 is associated to DN, with a *P*-value of 0.0330. In comparison, no blocks were detected from the eight SNPs of *EGFR*, because the adjacent SNPs were in linkage equilibrium. SNP0 and SNP1 of Table 3 correspond to numbers 3 and 4 of Figure 3, respectively. The distance between SNP0 and SNP1 is 40 kb, a large distance that reduces the possibility for them to reside in the same LD block.

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*.

Among the genes pertaining to fibrosis and inflammation, *IL4R*, *LPL* and *MED12L* were identified. The Interleukin 4 receptor (*IL4R*) is a pro-inflammatory cytokine produced by many different cell types, including glomerular mesangial cells. It is related to fibroblast proliferation, collagen production, chemotaxis, fibrosis as well as inflammation, which could contribute to interstitial fibrosis and renal failure. No association was found in the genes involved in the construction/destruction of ECM.

Among the genes related to cell growth and proliferation, *EGFR*, *PDGFRB* and *RXRG* were identified. *PDGF* is known to be a significant mitogen for mesangial cells (27). The epistasis effect of *EGFR* and *RXRG* was also identified and will be discussed in the next section. Among the genes related to insulin functions, *IGF2R* and *SNAP25* were identified. The insulin-like growth factor receptor gene (*IGF2R*) was reported to bind to the *TGF- β* complex (28).

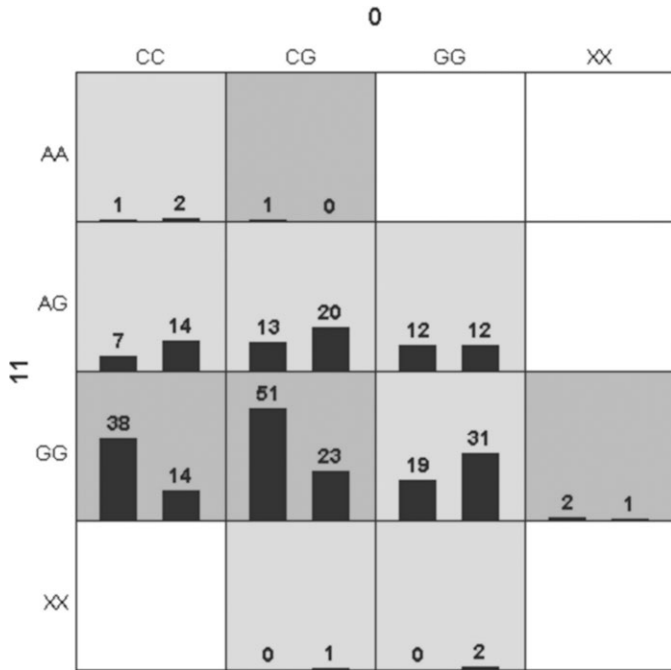


Figure 1. The optimum two-locus model detected by MDR. This model comprises SNP0 and SNP11. 'XX' denotes the missing genotypes.

Table 5. The prediction performance of the 7-locus model detected by GABA

		Clinical status	
		Positive	Negative
Prediction	Positive	110	43
	Negative	25	68

Table 6. The prediction performance of the two-locus model described by a Boolean statement

		Clinical status	
		Positive	Negative
Prediction	Positive	89	37
	Negative	53	79

The synaptosomal-associated protein, *SNAP25*, is a pre-synaptic plasma membrane protein that functions in the synaptic 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.

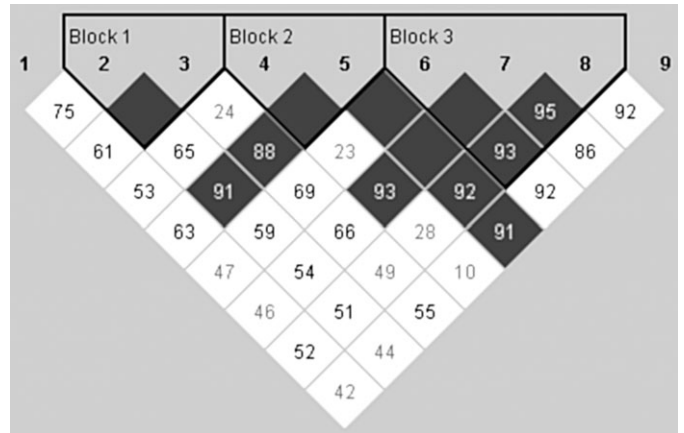


Figure 2. The LD, shown as D prime numbers, as well as the LD blocks among the nine SNPs of *RXRG*.

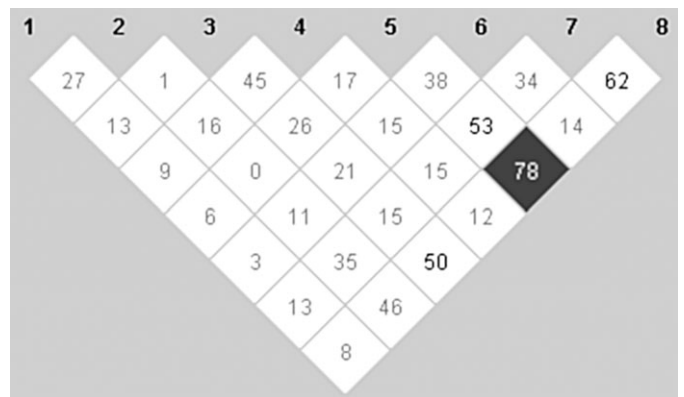


Figure 3. The LD, shown as D prime numbers, among the eight SNPs of *EGFR*.

When investigated individually, the 13 SNPs were of interest, because the null hypotheses of no association were rejected. However, when the 231 SNPs were assessed together, the *P*-values of these 13 SNPs cannot pass the criteria set by the Bonferroni correction method. Thus, their role of disease association need to be further confirmed, particularly on their epistasis. The null hypothesis of no association could not be rejected for the remaining SNPs, thus, they were excluded from the epistasis analysis. The removal of the remaining SNPs can avoid obscuring the true epistatic effects. It can also reduce the search space of epistasis models.

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).

Table 7. Associations of haplotypes of the *RXRG* gene with DN were shown as χ^2 *P*-values. The *P*-value smaller than 0.05 is shown in bold face

Block	Haplotype	<i>P</i> -value	Haplotype frequency
Block 1	GA	0.0643	0.248106061
	GC	0.2854	0.409090909
	AC	0.7182	0.327651515
Block 2	TA	0.0330	0.174242424
	TG	0.9265	0.185606061
	CG	0.2011	0.623106061
	CA	0.2733	0.001893939
Block 3	CGG	0.8551	0.007575758
	CGT	0.4236	0.21780303
	TGG	0.5312	0.284090909
	TGT	0.4450	0.132575758
	TAG	0.2344	0.35515873
	TAT	0.9552	0.003968254

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-proliferate 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 reference (34). This is certainly an important direction for further investigations. However, this project was initiated in 2002, and the study design was based on the concept of diplotype, which was defined as the collection of single-locus genotypes whose phase is unknown (35). The SNPs were selected without the consideration of haplotype structures (and Tag-SNPs) of the human genome. Consequently, the SNPs of many genes (e.g. *EGFR*) cannot be used to construct haplotypes. This has limited our in-depth analysis on haplotypes

using the current raw data. It is our future work to genotype *EGFR* and many other genes on the Tag-SNPs, taking advantage of the resources from the International HapMap project (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 52.8 mg/mmol (300 μ g/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 μ l, 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 $144/120 = 1.2$, a value 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 the varying numbers of SNPs jointed together by different Boolean operators. Models were continuously generated by the mutations and cross-over operations on the previous models with above-average performance. GABA carries out a 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 Haploview 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

- Caramori, L. and Mauer, M. (2003) Diabetes and nephropathy. *Curr. Opin. Nephrol. Hypertens.*, **12**, 273–282.

- Reasner, C.A. (1998) Diabetic nephropathy—can it be prevented? *Disease Prevention News*, **58**, 1–4.
- Goruppi, S., Bonventre, J.V. and Kyriakis, J.M. (2002) Signaling pathways and late-onset gene induction associated with renal mesangial cell hypertrophy. *EMBO J.*, **21**, 5427–5436.
- Mason, R.M. and Wahab, N.A. (2003) Extracellular matrix metabolism in diabetic nephropathy. *J. Am. Soc. Nephrol.*, **14**, 1358–1573.
- Lee, H.B., Yu, M.R., Yang, Y., Jiang, Z. and Ha, H. (2003) Reactive oxygen species-regulated signaling pathways in diabetic nephropathy. *J. Am. Soc. Nephrol.*, **14**, S241–S245.
- Schnaper, H.W., Hayashida, T., Hubchak, S.C. and Poncelet, A.C. (2003) TGF-beta signal transduction and mesangial cell fibrogenesis. *Am. J. Physiol. Renal. Physiol.*, **284**, 243–252.
- Wolf, G. and Ritz, E. (2003) Diabetic nephropathy in type 2 diabetes prevention and patient management. *J. Am. Soc. Nephrol.*, **14**, 1396–1405.
- Mogyorosi, A. and Ziyadeh, F.N. (1999) *GLUT1* and *TGF- β* : the link between hyperglycaemia and diabetic nephropathy. *Nephrol. Dial. Transplant.*, **4**, 2827–2829.
- Wang, W., Koka, V. and Lan, H.Y. (2005) Transforming growth factor and Smad signalling in kidney diseases. *Nephrology*, **10**, 48–56.
- Canani, L.H., Gerchman, F. and Gross, J.L. (1999) Familial clustering of diabetic nephropathy in Brazilian type 2 diabetic patients. *Diabetes*, **48**, 909–913.
- Imperatore, G., Knowler, W., Pettitt, D., Kobes, S., Bennett, P. and Hanson, R. (2000) Segregation analysis of diabetic nephropathy in Pima Indians. *Diabetes*, **49**, 1049–1056.
- Nesbitt, K.N. (2004) An overview of diabetic nephropathy. *J. Pharmacy Practice.*, **17**, 75–79.
- Shcherbak, N.S. and Schwartz, E.I. (2001) The C825T polymorphism in the G-protein β subunit gene and diabetic complications in IDDM patients. *J. Hum. Genet.*, **36**, 188–191.
- Nomiyama, T., Tanaka, Y., Piao, L., Nagasaka, K., Sakai, K., Ogihara, T., Nakajima, K., Watada, H. and Kawamori, R. (2003) The polymorphism of manganese superoxide dismutase is associated with diabetic nephropathy in Japanese type 2 diabetic patients. *J. Hum. Genet.*, **48**, 138–141.
- Zintzaras, E. and Stefanidis, I. (2005) Association between the *GLUT1* gene polymorphism and the risk of diabetic nephropathy: a meta-analysis. *J. Hum. Genet.*, **50**, 84–91.
- Pociot, F., Hansen, P.M., Karlsen, A.E., Langdahl, B.L., Johannesen, J. and Nerup, J. (1998) TGF-beta1 gene mutations in insulin-dependent diabetes mellitus and diabetic nephropathy. *J. Am. Soc. Nephrol.*, **9**, 2302–2307.
- Stephens, R.H., McElduff, P., Heald, A.H., New, J.P., Worthington, J., Ollier, W.E. and Gibson, J.M. (2005) Polymorphisms in IGF-binding protein I are associated with impaired renal function in type 2 diabetes. *Diabetes*, **54**, 3547–3553.
- Bell, J.T., Wallace, C., Dobson, R., Wiltshire, S., Mein, C., Pembroke, J., Brown, M., Clayton, D., Samani, N., Dominiczak, A. *et al.* (2006) Two-dimensional genome-scan identifies novel epistatic loci for essential hypertension. *Hum. Mol. Genet.*, **15**, 1365–1374.
- Carlborg, O. and Haley, C.S. (2004) Epistasis: too often neglected in complex trait studies. *Nature*, **5**, 618–625.
- Knowler, W.C., Coresh, J., Elston, R.C., Freedman, B.I., Iyengar, S.K., Kimmel, P.L., Olson, J.M., Plaecke, R., Sedor, J.R. and Seldin, M.F. Family Investigation of Nephropathy Diabetes Research Group. (2005) The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. *J. Diabetes Complications.*, **19**, 1–9.
- Schlesselman, J. (1982) *Case-control studies.*, Oxford University Press, NY.
- Ritchie, M., Hahn, L., Roodi, N., Bailey, R., Dupont, W., Parl, F. and Moore, J. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.*, **69**, 138–147.
- Liang, K.H., Hwang, Y., Shao, W.C. and Chen, E.Y. (2006) An Algorithm for model construction and its applications to pharmacogenomic studies. *J. Hum. Genet.* DOI 10.1007/s10038-006-0016-2.
- Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M. and Wain, H. (2001) The HUGO Gene Nomenclature Committee (HGNC). *Hum. Genet.*, **109**, 678–680.
- Schaid, D.J. and Jacobsen, S.J. (1999) Biased tests of association: comparison of allele frequencies when departing from Hardy-Weinberg proportions. *Am. J. Epidemiology*, **149**, 706–711.

26. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
27. Kawano, H., Kim, S., Ohta, K., Nakao, T., Miyazaki, H., Nakatani, T. and Iwao, H. (2003) Different contribution of three mitogen-activated protein kinases to PDGF-BB-induced mesangial cell proliferation and gene expression. *J. Am. Soc. Nephrol.*, **14**, 584–592.
28. Oklu, R. and Hesketh, R. (2000) The latent transforming factor β binding protein family. *Biochem. J.*, **352**, 601–610.
29. Nevins, A.K. and Thurmond, D.C. (2005) A direct interaction between Cdc42 and vesicle-associated membrane protein 2 regulates SNARE-dependent insulin exocytosis. *J. Biol. Chem.*, **280**, 1944–1952.
30. Huang, X., Wheeler, M.B., Kang, Y.H., Sheu, L., Lukacs, G.L., Trimble, W.S. and Gaisano, H.Y. (1998) Truncated SNAP-25 (1-197), like botulinum neurotoxin A, can inhibit insulin secretion from HIT-T15 insulinoma cells. *Mol. Endocrinol.*, **12**, 1060–1070.
31. Cordell, H.J. (2004) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.*, **11**, 2463–2468.
32. Tsai, C.T., Lai, L.P., Chiang, F.T., Fallin, D., Hwang, J.J., Ritchie, M.D., Moore, J.H., Hsu, K.L., Tseng, C.D., Liao, C.S. *et al.* (2004) Renin-angiotensin system gene polymorphisms and atrial fibrillation. *Circulation*, **109**, 1640–1646.
33. Oda, K., Matsuoka, Y., Funahashi, A. and Kitano, H. (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.*, 1–17. msb410014.
34. The International HapMap Consortium (2003) The international HapMap project. *Nature*, **426**, 789–796.
35. Schaid, D.J. (2005) Power and sample size for testing associations of haplotypes with complex traits. *Ann. Hum. Genet.*, **70**, 116–130.
36. Excoffier, L. and Slatkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.