Application of Back Propagation Artificial Neural Network on Genetic Variants in Adiponectin *ADIPOQ*, *Peroxisome Proliferator-Activated Receptor- γ* , and *Retinoid X Receptor- α* Genes and Type 2 Diabetes Risk in a Chinese Han Population

Hui Shi, B.S.,^{1,*} Ying Lu, M.S.,^{1,*} Juan Du, B.S.,¹ Wencong Du, M.S.,¹ Xinhua Ye, B.S.,²
Xiaofang Yu, B.S.,³ Jianhua Ma, M.S.,⁴ Jinluo Cheng, B.S.,² Yanqin Gao, M.S.,³
Yuanyuan Cao, M.S.,¹ Ling Zhou, M.D.,¹ and Qian Li, M.D.⁴

Abstract

Aims: Our study was designed to explore the applied characteristics of the back propagation artificial neural network (BPANN) on studying the genetic variants in adiponectin *ADIPOQ*, *peroxisome proliferator-activated receptor (PPAR)- γ* , and *retinoid X receptor- α* (*RXR- α*) genes and type 2 diabetes mellitus (T2DM) risks in a Chinese Han population.

Subjects and Methods: We used BPANN as the fitting model based on data gathered from T2DM patients ($n=913$) and normal controls ($n=1,001$). The mean impact value (MIV) for each input variables were calculated, and the sequence of the factors according to their absolute MIVs was sorted.

Results: The results from BPANN were compared with multiple logistic regression analysis, and the generalized multifactor dimensionality reduction (GMDR) method was used to calculate the joint effects of *ADIPOQ*, *PPAR- γ* , and *RXR- α* genes. By BPANN analysis, the sequence according to the importance of the T2DM risk factors was in the order of serum adiponectin level, rs3856806, rs7649121, hypertension, rs3821799, rs17827276, rs12495941, rs4240711, age, rs16861194, waist circumference, rs2241767, rs2920502, rs1063539, alcohol drinking, smoking, hyperlipoproteinemia, gender, rs3132291, T2DM family history, rs4842194, rs822394, rs1801282, rs1045570, rs16861205, rs6537944, body mass index, rs266729, and rs1801282. However, compared with multiple logistic regression analysis, only 11 factors were statistically significant. After overweight and obesity were taken as environment adjustment factors into the analysis, model A2 B4 C5 C6 C8 (rs3856806, rs4240711, rs7649121, rs3821799, rs12495941) was the best model (coefficient of variation consistency = 10/10, $P=0.0107$) in the GMDR method.

Conclusions: These results suggested the interactions of *ADIPOQ*, *PPAR- γ* , and *RXR- α* genes might play a role in susceptibility to T2DM. BPANN could be used to analyze the risk factors of diseases and provide more complicated relationships between inputs and outputs.

Introduction

TYPE 2 DIABETES MELLITUS (T2DM), which accounts for more than 90% of diabetes worldwide, is a common, complex, and chronic disease with rapidly growing global importance. The number of people with T2DM worldwide

will rise to 300 million by 2025, and it is considered to a major medical burden on society.¹ A type of plasma hormone protein, adiponectin, which is newly discovered and is secreted specifically by adipose tissue, can resist inflammation, increase insulin sensitivity, and protect the cardiovascular system,² which has been confirmed with a close relationship to

¹Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China.

²Department of Endocrinology, Affiliated Changzhou 2nd Hospital of Nanjing Medical University, Changzhou, Jiangsu, China.

³Department of Endocrinology, 3rd Affiliated Hospital of Nanjing Medical University, Yizheng, Jiangsu, China.

⁴Department of Endocrinology, Affiliated Nanjing 1st Hospital of Nanjing Medical University, Nanjing, Jiangsu, China.

*The first two authors contributed equally to this study.

high-risk T2DM. Its plasma levels are paradoxically decreased in obesity and may play important roles in the development of T2DM.^{3,4} The adiponectin *ADIPOQ* gene has been identified and is located on chromosome 3q27.^{5–7} In this chromosome region, genome-wide scans have revealed a susceptibility locus for T2DM,⁸ coronary heart disease,⁹ and obesity.¹⁰ The *ADIPOQ* gene was reported to be associated with obesity, T2DM, and insulin sensitivity in many populations.^{11–13} Peroxisome proliferator-activated receptor (PPAR)- γ /retinoid X receptor (RXR) heterodimer, which binds directly to the functional PPAR-responsive element, could increase the human adiponectin promoter activity in cells. Therefore, the genes coding PPAR- γ and RXR- α located in the adiponectin pathway are candidates for T2DM and obesity.^{3,4,14,15} As of now, studies investigating the interaction among polymorphisms in the genes coding for adiponectin *ADIPOQ*, PPAR- γ , and RXR- α with the susceptibility of T2DM have not been reported yet.

Back propagation artificial neural network (BPANN) is an unconventional multidimensional data analysis that is mainly applied in the problem of prediction and discriminant analysis. It can master the essence feature, through learning and training the representative examples. A well-trained BPANN, with a very strong sense of self-organization, adaptive ability, and higher fault tolerance, theoretically is capable of approaching any nonlinear mapping between input (independent variables) and output (dependent variables). The variables of BPANN do not require normality, independence, and other conditions. Thus, it broke through the limitations of case-control studies and traditional logistic regression and provided a new method for etiology research.⁴ We used the BPANN method to analyze our data and compared our findings with the results of logistic regression analysis, to explore the applied characteristics of BPANN on studying the genetic variants among *ADIPOQ*, PPAR- γ , and RXR- α genes with T2DM risk in a Chinese Han population and also to better understand and use the artificial neural network in the approach to epidemiology.

Subjects and Methods

Subjects

Our study selected a total of 1,914 subjects, consisting of 913 T2DM patients and 1,001 T2DM-free controls, with their informed consent. All subjects were genetically unrelated ethnic Han Chinese. The cases included all eligible patients newly diagnosed with T2DM according to the diagnosis criteria of the World Health Organization for diabetes, who were consecutively recruited between March 2008 and August 2010 at the diabetes outpatient clinic from three affiliated hospitals of Nanjing Medical University (NJMU) (The Affiliated Changzhou 2nd Hospital of NJMU, the 3rd Affiliated Hospital of NJMU, and the Affiliated Nanjing 1st Hospital of NJMU) without the restrictions of age and sex (448 men and 465 women; 57.21 ± 10.99 years old). A diagnosis of T2DM required either fasting plasma glucose ≥ 7.0 mmol/L (126 mg/dL) or 2-h glucose ≥ 11.1 mmol/L (200 mg/dL) after an oral glucose tolerance test.¹⁶ All the patients were tested by glutamic acid decarboxylase autoantibodies and islet-cell antibodies (ICA512) (RSI Company, Oxford, UK) to remove patients with type 1 diabetes. T2DM-free control subjects who had no history of T2DM, frequency-matched to the cases on

age (± 5 years) and gender, were randomly selected from the outpatient clinic within the same geographical area and the period of the cases. Controls were enrolled with routine annual health examinations and did not have diabetes as determined by an oral glucose tolerance test (75 g of glucose), performed according to World Health Organization criteria (472 men and 529 women; 57.48 ± 11.21 years old). Those with medical illnesses such as hyperthyroidism, pituitary diseases, and chronic liver diseases and those taking glucocorticoids or other medications affecting glucose metabolism were excluded from the study. After signing an informed consent, participants were administered standard questionnaires, in order to collect information on demographic data and environmental exposure history, including smoking and drinking. After interview, an approximately 5-mL venous blood sample was collected from each participant.

We used BPANN as the fitting model based on data gathered from T2DM patients and normal controls. The mean impact value (MIV) for each input variable was calculated, and the sequence of the factors according to their absolute MIVs was sorted.

Measurements

The anthropometric measurements included body weight, height, waist circumference (WC), and hip circumference. Body mass index (BMI) was calculated according to the Quetelet equation. Blood pressure was measured on the right arm with the individual in a sitting position after a 10-min rest using a standard sphygmomanometer of appropriate cuff size. The mean value of two consecutive blood pressure readings was taken into account. After an overnight fast, venous blood samples were collected and promptly centrifuged, and the serum was stored at -20°C until the adiponectin assay was performed. All samples were tested in the same assay. Serum adiponectin was measured by enzyme-linked immunosorbent assay (human adiponectin enzyme-linked immunosorbent assay kit; RapidBio Co., Seattle, WA, USA). Fasting plasma glucose was measured in the laboratories in the three affiliated hospitals of NJMU with the glucose oxidase method. Total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglycerides were determined in the three affiliated hospitals by an enzymatic colorimetric method (Au5400; Olympus, Tokyo, Japan). DNA was extracted from peripheral blood by the use of proteinase K and phenol/chloroform. This study was approved by the Research Ethics Committee of NJMU.

Single nucleotide polymorphism selection and genotyping assays

Two potentially functional single nucleotide polymorphisms (SNPs) (rs16861194 and rs266729) of the *ADIPOQ* gene, three of the PPAR- γ gene (rs2920502, rs3856806, and rs1801282), and four of RXR- α (rs1045570, rs3132291, rs4240711, and rs4842194) with minor allele frequency of ≥ 0.05 in the Chinese Han population were identified from the National Center for Biotechnology Information dbSNPs database (www.ncbi.nlm.nih.gov/). Eight tagging SNPs (rs2241767, rs822394, rs1063539, rs12495941, rs16861205, rs182052, rs3821799, and rs7649121) of the *ADIPOQ* gene, two of the PPAR- γ gene (rs17817276 and rs3856806), and two of the RXR- α gene (rs1045570 and rs6537944) were selected

based on the pairwise tagging ($r^2 \geq 0.80$, minor allele frequency ≥ 0.05) using genotype data from the unrelated HapMap CHB individuals.

Genomic DNA was extracted from peripheral blood samples of all subjects. 5'-Nuclease TaqMan[®] assays were used to genotype the polymorphisms in 384-well plates by the ABI PRISM 7900HT Sequence Detection system (Applied Biosystems, Foster City, CA). The primers and probes of TaqMan assays were designed using Primer Express Oligo Design software version 2.0 (ABI PRISM) and are available upon request as TaqMan Pre-Designed SNP Genotyping Assays. The primer and probe sequences used are shown in Table 1. Polymerase chain reactions were performed in a 5- μ L reaction mixture containing 5 ng of DNA, 2.5 μ L of 2* TaqMan Universal PCR Master Mix, and 0.083 μ L of 40* Assay Mix. The polymerase chain reaction conditions were 50°C for 2 min, 95°C for 10 min, 95°C for 15 s, and then 60°C for 1 min; 40 cycles of real-time polymerase chain reaction were performed. Individual genotypes identification was performed by SDS

software version 2.0 (ABI). Each plate contained two samples from a same individual as positive controls and two blank samples as negative controls for the genotyping quality confirmation. In addition, 10% of the samples were randomly selected to perform repeated assays; the results were 100% concordant.

Statistical analysis

All the statistical analyses were performed by SPSS software (version 13.0) (SPSS, Inc., Chicago, IL). The associations between genotypes and T2DM were estimated by computing the standardized partial regression coefficient (β), the odds ratios, and 95% confidence intervals from both univariate and multinomial logistic regression analysis. Among controls, genotype frequencies for each SNP were tested for Hardy-Weinberg equilibrium. Generalized multifactor dimensionality reduction (GMDR) software (version 1.0.1) (www.healthsystem.virginia.edu/internet/Addiction-Genomics/)

TABLE 1. PRIMERS AND PROBE SEQUENCES FOR AMPLIFICATION OF ADIPONECTIN *ADIPOQ*, *PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR- γ* , AND *RETINOID X RECEPTOR- α* GENE SINGLE NUCLEOTIDE POLYMORPHISMS

SNP		Primer	Probe	
rs266729	Sense	ATTCTGTTTGGATGCTTGTG	Probe 1	ATCCTGCCCTTCAA
	Antisense	CTTGGACTTTCTTGGCAGC	Probe 2	TCCTGCGCTTCAA
rs16861194	Sense	TGTCTTGTGAAGTTGGTGCTG	Probe 1	TGAATTAAATTACGACCCC
	Antisense	GACTTCTTGGCAGCTCAT	Probe 2	TGAATTAAACTACGACCC
rs2241767	Sense	TCTTTCATCACAGACCTCCTACA	Probe 1	CAACCTGAAGTGATT
	Antisense	GCACCATCTACACTCATCCTTG	Probe 2	CAACCTGAGGTGATT
rs822394	Sense	TCCTTTACAATCAGAGTCCGTTT	Probe 1	TGCTAATCACACTCTT
	Antisense	CCACTAATAGGTGCGATCAGC	Probe 2	TGCTAATCCCCTCTT
rs7649121	Sense	GAGACATTCTTGGAGTTGAGTATTG	Probe 1	TATACTACCTCTTCACGTG
	Antisense	CTGGAATCCCACTCCTATGC	Probe 2	TATACTACCCACTTCACGTG
rs3821799	Sense	CTGCCTTTGGGGAAGTCTT	Probe 1	TTTCTTGTGGTAACCC
	Antisense	CATCAGGTCCACGGTGAGTAT	Probe 2	CTTCTTGTAGTAACCC
rs16861205	Sense	GAGAAAAGCCTGGCATAATAGTG	Probe 1	TGTTTCTAAGGCATCC
	Antisense	TAACCTTCAGCATCCACAGC	Probe 2	TGTTTCTGAGGCATC
rs12495941	Sense	GCAGTGAGGTACCATTTATTTCC	Probe 1	AGGGCATACCTTAACCTA
	Antisense	CTCCTGATACATATCCCCACAT	Probe 2	AAGGGCATAACTTAACCTA
rs182052	Sense	CAGCCCCAAGAGAGAAAGG	Probe 1	CTGAATTTTACCCAGTTC
	Antisense	GAATTGGACTTCTCTGTGGAC	Probe 2	CTGAATTTTCCCGAGTT
rs1063539	Sense	CTCTGGGGCAGGGTTATTC	Probe 1	ACAGAGACAGTCAACT
	Antisense	TCAAAGCATCACAGGACCATT	Probe 2	ACAGAGAGAGTCAACT
rs17817276	Sense	CTCCCTGACAGCAGCTATCC	Probe 1	AAATAGTAATATATGACAACCT
	Antisense	TTCCAGGATTATCCTAACAGA	Probe 2	AATAGTAATACATGACAACC
rs3856806	Sense	TGTTTGCCAAGCTGCTCC	Probe 1	CTGCACGTGTTC
	Antisense	TTGGCAGTGGCTCAGGAC	Probe 2	CTGCACATGTTCC
rs1801282	Sense	TGCTGTTATGGGTGAAACTCTG	Probe 1	CTATTGACCCAGAAAG
	Antisense	ATAGCCGTATCTGGAAGGAAT	Probe 2	CTATTGACGCAGAAAG
rs2920502	Sense	GCACAGTAGGGCCACG	Probe 1	CCACTCTCTGCC
	Antisense	GGATCCCTCCTCGGAAATG	Probe 2	CCACTGTCTGCC
rs6537944	Sense	CGTGAATGCTGCTCTCTGT	Probe 1	CGTCCGTCAGGCA
	Antisense	AACCTGGATATGGCAGCACT	Probe 2	CGTCCATCAGGCA
rs1045570	Sense	AGCCTTGCTCTGTTGTGTCC	Probe 1	CACCTGCGGCCAC
	Antisense	ACTTCTCCCTTTGCGTGTTC	Probe 2	CACCTGAGGCCAC
rs4842194	Sense	TGGTGGAATGGCAGGAG	Probe 1	TGCCTTCTGCAGCC
	Antisense	CCCTGGGCTTTTTCCTCT	Probe 2	TGCCTTCTGCAGCC
rs3132291	Sense	CTTCAGTGTCTGTGGTGCCTC	Probe 1	AGGGCTCCGGGCA
	Antisense	GCATTGTCTCCTGTGATAAACG	Probe 2	AGGGCTCTGGGCA
rs4240711	Sense	GACTCCCCGTTTCAGACCAG	Probe 1	AGGACAAGTCTCAGC
	Antisense	CTCCAGCAAGGCCAGTGA	Probe 2	AGGACAAGCCTCAGC

SNP, single nucleotide polymorphism.

was applied for detecting gene–gene and gene–environment interactions. Based on the MATLAB version 7.0 software platform (MathWorks, Natick, MA), the artificial neural network prediction model was established, and each input neuron's MIV could be calculated.

Results

Univariate logistic regression analysis for the associations between the risk factors and T2DM

The genotype distribution for all the SNPs did not show any deviation from the Hardy–Weinberg equilibrium (all $P > 0.05$ in controls). Binary logistic regression analysis was used to analyze the association among the 29 variables, including 10 SNPs of the *ADIPOQ* gene, four of the *PPAR- γ* gene, and five of the *RXR- α* gene, serum adiponectin concentration and other possible environmental factors, and T2DM. Table 2 displays the sequences of the influence of the various factors with T2DM risk. The results showed that the top 10 sequences of the factors associated with T2DM risk were T2DM family history, hypertension history, WC, rs17817276, smoking, rs3856806, drinking alcohol, serum adiponectin, rs1801282, and BMI.

Multiple logistic regression analysis for the associations between the risk factors and T2DM

In multiple logistic regression analysis, with stepwise regressive method, 11 factors were statistically significant. We

sequenced the influence of various factors on high-risk T2DM. The order was family history, drinking alcohol, hypertension history, smoking, WC, rs7649121, rs3856806, rs12495941, gender, rs16861194, and serum adiponectin (Table 3). All samples had entered the model; there were no missing values.

BPANN multiple analysis for the associations between the risk factors and T2DM

Using 29 factors as input variables and T2DM diagnosis as output variables, we established the BPANN model with all available samples. The transfer function was logsig function. Learning rate was 0.1. Training error was 0.01. Maximum training steps was set to 1,000 steps. After the termination of training, the MIV was obtained. Then, according to the absolute value of MIV, we sequenced the related factors in the order of serum adiponectin level, rs3856806, rs7649121, hypertension, rs3821799, rs17827276, rs12495941, rs4240711, age, rs16861194, WC, rs2241767, rs2920502, rs1063539, drinking alcohol, smoking, hyperlipoproteinemia, gender, rs3132291, T2DM family history, rs4842194, rs822394, rs1801282, rs1045570, rs16861205, rs6537944, BMI, rs266729, and rs1801282 (Table 4).

Locus–locus interactions of ADIPOQ, PPAR- γ , and RXR- α genes

The SNPs rs17817276, rs3856806, rs4240711, rs16861194, rs7649121, rs3821799, and rs12495941, which were located in

TABLE 2. UNIVARIATE LOGISTIC REGRESSION ANALYSIS RESULTS

Variable	β	OR (95% CI)	P value	Rank number
T2DM family history	1.343	3.829 (2.654, 5.524)	<0.001	1
History of hypertension	0.990	2.690 (2.050, 3.531)	<0.001	2
WC	0.832	2.299 (1.753, 3.015)	<0.001	3
rs17817276	0.830	1.086 (0.809, 1.459)	0.582	4
Smoking	0.514	1.672 (1.226, 2.281)	0.001	5
rs3856806	−0.466	0.627 (0.480, 0.821)	0.001	6
Alcohol drinking	−0.407	0.666 (0.472, 0.938)	0.020	7
Serum adiponectin	−0.269	0.764 (0.714, 0.818)	<0.001	8
rs1801282	−0.264	0.768 (0.501, 1.177)	0.225	9
BMI	0.260	1.297 (1.073, 1.568)	0.007	10
rs4240711	0.254	1.289 (0.984, 1.690)	0.066	11
rs4842194	0.252	1.286 (0.986, 1.678)	0.063	12
rs12495941	0.210	1.234 (0.935, 1.629)	0.138	13
Gender	0.201	1.223 (0.940, 1.591)	0.133	14
rs16861194	0.166	1.181 (0.894, 1.560)	0.241	15
rs3821799	0.160	1.174 (0.864, 1.595)	0.306	16
rs6537944	0.152	1.165 (0.851, 1.593)	0.341	17
rs822394	−0.111	0.895 (0.651, 1.230)	0.495	18
rs2920502	−0.101	0.904 (0.696, 1.176)	0.452	19
rs182052	0.080	1.083 (0.819, 1.434)	0.575	20
rs16861205	0.078	1.081 (0.814, 1.434)	0.591	21
rs2241767	0.075	1.078 (0.830, 1.401)	0.574	22
rs3132291	0.051	1.052 (0.808, 1.370)	0.707	23
rs1045570	0.048	1.049 (0.795, 1.384)	0.737	24
rs266729	−0.025	0.975 (0.750, 1.269)	0.853	25
rs7649121	0.023	1.023 (0.785, 1.334)	0.865	26
rs1063539	0.017	1.018 (0.783, 1.323)	0.897	27
Hyperlipidemia history	−0.005	0.995 (0.762, 1.299)	0.972	28
Age	−0.003	0.997 (0.986, 1.008)	0.592	29

BMI, body mass index; CI, confidence interval; OR, odds ratio; T2DM, type 2 diabetes mellitus; WC, waist circumference.

TABLE 3. MULTIPLE LOGISTIC REGRESSION ANALYSIS RESULTS

Variable	β	OR (95% CI)	P value	Rank number
T2DM family history	1.404	4.073 (2.698, 6.150)	<0.001	1
Alcohol drinking	-1.159	0.314 (0.199, 0.496)	<0.001	2
History of hypertension	1.007	2.738 (2.003, 3.741)	<0.001	3
Smoking	0.738	2.092 (1.374, 3.187)	0.001	4
WC	0.615	1.850 (1.357, 2.523)	<0.001	5
rs7649121	0.397	1.488 (1.067, 2.075)	0.019	6
rs3856806	-0.388	0.679 (0.499, 0.923)	0.014	7
rs12495941	0.353	1.423 (1.004, 2.017)	0.047	8
Gender	0.332	1.394 (0.982, 1.980)	0.063	9
rs16861194	0.302	1.353 (0.955, 1.916)	0.089	10
Serum adiponectin	-0.293	0.746 (0.692, 0.804)	<0.001	11

CI, confidence interval; OR, odds ratio; T2DM, type 2 diabetes mellitus; WC, waist circumference.

the top 10 in BPANN multiple analysis, were termed, respectively, A₁, A₂, B₄, C₂, C₅, C₆, and C₈. GMDR software was used to consider the joint effects of the seven SNPs. The results showed that the A₂ B₄ C₅ C₆ C₈ (rs3856806, rs4240711, rs7649121, rs3821799, and rs12495941) model was the best model (cross-validation consistency 10/10, $P=0.0107$) (Table 5). After taking overweight and obesity as environment adjustment factors into the analysis, the results showed that A₂ B₄ C₅ C₆ C₈ was still the best model (cross-validation consistency 10/10, $P=0.0107$) (Table 6).

TABLE 4. INPUT VARIABLES AND SORTING OF MEAN IMPACT VALUES

Variable	MIV	Rank number
Adiponectin	-0.072891	1
rs3856806	-0.020205	2
rs7649121	0.019820	3
History of hypertension	0.017439	4
rs3821799	0.017177	5
rs17817276	0.017077	6
rs12495941	0.017071	7
rs4240711	0.015706	8
Age	-0.015282	9
rs16861194	0.014651	10
WC	0.014116	11
rs2241767	0.013265	12
rs2920502	-0.012454	13
rs1063539	-0.009443	14
Alcohol drinking	0.008053	15
Smoking	-0.007925	16
Hyperlipidemia history	-0.006975	17
Gender	0.006889	18
rs3132291	-0.006306	19
T2DM family history	0.006036	20
rs4842194	-0.003814	21
rs822394	-0.003488	22
rs182052	0.002370	23
rs1045570	0.002037	24
rs16861205	-0.001718	25
rs6537944	-0.001475	26
BMI	0.001164	27
rs266729	0.000328	28
rs1801282	0.000138	29

BMI, body mass index; MIV, mean impact value; T2DM, type 2 diabetes mellitus; WC, waist circumference.

Discussion

T2DM is a typical multifactor disease. Because of T2DM's own restrictions such as confounding factors and multiple collinearities, the interactions among factors could not be fully reflected when we analyzed the associations between genes and T2DM by traditional model analysis. T2DM is caused by both genetic and environmental factors. We had to consider not only SNPs, haplotypes, and gene-gene interaction, but also environmental elements such as behaviors, living habits, family history, and so on. The traditional model often had limitations to analyze the gene-environment interactions, so there was no appropriate method to deal with this kind of problem for the moment. The BPANN with self-study ability did not require the distribution form and independence of variables and also could handle the problem of collinearity better. It not only helped us to find the unknown relationships among variables, but also contributed to analysis of the complex relationships, especially showing its unique advantage in gene-environment interaction analysis. Therefore, we discussed the associations among ADIPOQ, PPAR- γ , and RXR- α gene polymorphism with susceptibility to T2DM in China's southern Han population with the BPANN method and contrasted the results with logistic regression analysis, for a better understanding of using the artificial neural network from the view of epidemiology.

In this study, we preliminarily screened factors, especially genetic factors, using univariate logistic regression, multiple logistic regression, and BPANN analysis. According to the values of MIV and β , we could determine the ranking of importance of each factor and provide the basis for further analysis. We analyzed 29 factors that may be involved with the risk of T2DM by univariate logistic regression. According to the standardized partial regression coefficient (β) of each variable, we sequenced the influence of various factors on high-risk T2DM. The results showed that the positions of environmental factors such as T2DM family history, hypertension history, smoking, and drinking alcohol were the greatest, and most of the genetic SNPs were less important. Because of the relative presence of environmental factors, the effects of gene polymorphism on the disease were concealed and faint. Univariate logistic regression could not consider the combined action of elements and only provided the references for the establishment of the multifactor model.

TABLE 5. GENERALIZED MULTIFACTOR DIMENSIONALITY REDUCTION MODELS FOR ADIPONECTIN *ADIPOQ*, *PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR- γ* , AND *RETINOID X RECEPTOR- α* GENE INTERACTION ON TYPE 2 DIABETES MELLITUS

Model	Training balanced accuracy	Test balanced accuracy	CV consistency	P value
A ₂	0.5297	0.5297	10/10	0.0107
A ₁ A ₂	0.5451	0.5223	7/10	0.0547
A ₂ B ₄ C ₆	0.5644	0.5119	6/10	0.1719
A ₂ B ₄ C ₅ C ₆	0.5930	0.5415	7/10	0.0547
A ₂ B ₄ C ₅ C ₆ C ₈	0.6295	0.5407	10/10	0.0107
A ₁ A ₂ B ₄ C ₅ C ₆ C ₈	0.6734	0.5568	10/10	0.0547
A ₁ A ₂ B ₄ C ₂ C ₅ C ₆ C ₈	0.7056	0.5404	10/10	0.0107

A₁, rs17827276; A₂, rs3856806; B₄, rs4240711; C₂, rs16861194; C₅, rs7649121; C₆, rs3821799; C₈, rs12495941; CV, cross-validations.

Compared with univariate logistic regression, multiple logistic regressions determined the independent factor by the adjustment of other factors and also reflected part of the combined action among various factors. We screened out 11 variables that had positive or negative statistically significant associations ($P < 0.01$) with the risk of T2DM in the order of T2DM family history, drinking alcohol, hypertension, smoking, WC, rs7649121, rs3856806, rs12495941, gender, rs16861194, and serum adiponectin. SNPs rs7649121, rs12495941, and rs16861194 belonged to the *ADIPOQ* gene. Reports have shown that the *ADIPOQ* gene had susceptible associations with diabetes. Sanghera et al.¹⁷ found that two-site haplotype analysis in the *ADIPOQ* locus using only two marginally associated SNPs (rs182052 and rs7649121) revealed a significant protective association of the GA haplotype with T2DM. Multiple logistic regression analysis also revealed significant association of an *ADIPOQ* variant (rs12495941) with total body weight, waist, and hip. Fumeron et al.¹⁸ found the -11391 GA genotype (rs16861194) is a risk for hyperglycemia. SNPs rs3856806 belonged to the *PPAR- γ* gene. Up to now, the studies investigating the association between rs3856806C/T of the *PPAR- γ* gene and the risk of T2DM had been reported rarely, and the conclusions were not identical. In these studies, some researchers found a statistically significant association between rs3856806C/T polymorphism with obese and overweight patients with T2DM.^{19–22} However, Evans et al.²³ did not find a statistically significant association between the rs3856806C > T polymorphism with T2DM in a German population study. We failed to find those results in our study. This may be due to the differences in ethnicity and T2DM susceptibility variants. Although the ef-

fect of serum adiponectin concentration appeared in the model, because of genetic and environmental multiple influences and the complex interaction between variables, the model contribution of adiponectin was weakened.

We performed this analysis using BPANN, which is a computer-based algorithm that can be trained to recognize and categorize complex patterns.^{24–27} It was able to process data containing complex (nonlinear) relationships and interactions that were often too difficult or complex to interpret by conventional linear methods.²⁸ Meanwhile, BPANN did not require variables to satisfy normality and independence conditions and could process collinearity questions between variables. Thus it had a unique advantage in the gene-environment interaction analysis. According to the absolute value of MIV, we sequenced relevant factors. Compared with univariate and multiple logistic regression analysis, the ranking arrangement of serum adiponectin concentration was obviously raised. The position of serum adiponectin concentration was changed to the first from the last one in the sequence for multiple logistic regression. So we considered that adiponectin might have a great influence in the pathogenesis of T2DM. Studies have proven that adiponectin can regulate insulin sensitivity. Treatment with adiponectin can reverse insulin sensitivity in the fat atrophy rat.²⁹ Studies of rhesus monkeys in vivo showed that the adiponectin concentration began to drop early in the beginning of obesity and continuing to decrease with the development of T2DM. Low levels of adiponectin decreased the glucose absorption rate of insulin, and the reduction of adiponectin paralleled the progress of insulin resistance.³⁰

The direct influence of adiponectin to T2DM was embodied in the BPANN model. The *ADIPOQ* gene locus rs3821799, the

TABLE 6. GENERALIZED MULTIFACTOR DIMENSIONALITY REDUCTION MODELS FOR ADIPONECTIN *ADIPOQ*, *PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR- γ* , AND *RETINOID X RECEPTOR- α* GENE-ENVIRONMENT INTERACTION ON TYPE 2 DIABETES MELLITUS

Model	Training balanced accuracy	Test balanced accuracy	CV consistency	P value
A ₂	0.5299	0.5298	10/10	0.0547
A ₁ A ₂	0.5472	0.5256	7/10	0.1719
A ₂ B ₄ C ₆	0.5664	0.5147	5/10	0.0547
A ₂ B ₄ C ₅ C ₆	0.5959	0.5340	8/10	0.0107
A ₂ B ₄ C ₅ C ₆ C ₈	0.6301	0.5303	10/10	0.0107
A ₁ A ₂ B ₄ C ₅ C ₆ C ₈	0.6769	0.5489	10/10	0.0547
A ₁ A ₂ B ₄ C ₂ C ₅ C ₆ C ₈	0.7087	0.5145	10/10	0.0547

Adjusted for overweight and obesity.

A₁, rs17827276; A₂, rs3856806; B₄, rs4240711; C₂, rs16861194; C₅, rs7649121; C₆, rs3821799; C₈, rs12495941; CV, cross-validations.

PPAR- γ gene locus rs17817276, and the RXR- α gene locus rs4240711 had no associations with T2DM in univariate and multiple logistic regression analysis, but ranking arrangements of the three gene polymorphism were in the beginning in BPANN analysis. This result suggested that the three gene interactions may be associated with the risk of T2DM. In order to test this hypothesis, we used the GMDR method to analyze the interaction of seven polymorphism loci (rs7649121, rs3856806, rs12495941, rs16861194, rs3821799, rs17817276, and rs4240711) that screened out from multiple logistic regression and BPANN multifactor analysis. GMDR had been applied to SNP association studies for detecting gene–gene interactions associated with several complex diseases and was applicable to both continuous and dichotomous phenotypes. In addition, GMDR permits adjustment for discrete and quantitative covariates in various population-based studies with unbalanced case–control samples.³¹ Model A₂ B₄ C₅ C₆ C₈ (rs3856806, rs4240711, rs7649121, rs3821799, and rs12495941) was the best model because of the highest test balanced accuracy (0.5407) (cross-validations consistency = 10/10). It showed that interactions existed among the loci of the five SNPs and might point out the interactions among the three genes. PPAR- γ is a member of the nuclear receptor family of ligand-activated transcription factors that heterodimerize with the RXR to regulate gene expression.¹⁵ PPAR- γ /RXR heterodimer directly combined to the PPAR-responsive element and increased the promoter activity. The present study has identified a functional PPAR-responsive element in the adiponectin promoter and has demonstrated that it plays a significant role in the transcriptional activation of the adiponectin gene in adipocytes.¹⁵ The related research just focused on the individual or several sites of the PPAR- γ and ADIPOQ genes. So far, it has not been found on the RXR- α gene and in the genetic interactions of the three genes in the relationship with T2DM.

Compared with traditional analysis methods, BPANN provided much more abundant information about the relationships of variables and a model that was more practical to use. In practical application, BPANN analysis and multiple logistic regression analysis would be used to select common influence factors, especially in the early screening of genetic factors, according to the ranking list of MIV and standardized partial regression coefficient (β). For instance, in the GMDR analysis of our study, we added all the 19 SNPs into the analysis without screening. We failed to get an optimal model. Accordingly, we might naturally arrive at the conclusion that there was no interaction among the three genes. However, when we put in seven SNPs, which were selected by multiple logistic regression and BPANN analysis, into the model analysis, we could get an optimal model with statistical significance (cross-validation consistency = 10/10, $P = 0.0107$). These results suggested that an interaction existed among the ADIPOQ, PPAR- γ , and RXR- α genes. This study suggested that the actual application significance of BPANN mainly lies in the output MIV of BPANN analysis, which, as a relative stable reference, can be used as a reference for the early screening of variables, especially the early screening of SNPs, which provide a basis for interaction of genes for further analysis. Even so, BPANN analysis only gave the weights and MIV values of each variable. It could not reject or choose some variables clearly as logistic regression analysis or give a exact P value for judging the effect of independent variable to de-

pendent variable, which indicated whether the association had statistical significance or not. It is necessary to seek a solution in our further study.

Acknowledgments

We would like to thank the National Natural Science Foundation of China for grant 30771858 and the Jiangsu Provincial Natural Science Foundation for grant BK2007229.

Author Disclosure Statement

No competing financial interests exist.

References

1. Prokopenko I, McCarthy MI, Lindgren CM: Type 2 diabetes: new genes, new understanding. *Trends Genet* 2008;24:613–621.
2. Gil-Campos M, Canete R, Gil A: Adiponectin, the missing link in insulin resistance and obesity. *Clin Nutr* 2004;23:963–974.
3. Neumeier M, Weigert J, Schäffler A, Weiss T, Kirchner S, Laberer S, Schölmerich J, Buechler C: Regulation of adiponectin receptor 1 in human hepatocytes by agonists of nuclear receptors. *Biochem Biophys Res Commun* 2005;334:924–929.
4. Kintscher U, Law RE: PPAR γ -mediated insulin sensitization: the importance of fat versus muscle. *Am J Physiol Endocrinol Metab* 2005;288:287–291.
5. Nakano Y, Tobe T, Choi-Miura NH, Mazda T, Tomita M: Isolation and characterization of GBP28, a novel gelatin-binding protein purified from human plasma. *J Biochem* 1996;120:803–812.
6. Saito K, Tobe T, Minoshima S, Asakawa S, Sumiya J, Yoda M, Nakano Y, Shimizu N, Tomita M: Organization of the gene for gelatin-binding protein (GBP28). *Gene* 1999;229:67–73.
7. Schäffler A, Langmann T, Palitzsch KD, Schölmerich J, Schmitz G: Identification and characterization of the human adipocyte apM-1 promoter. *Biochim Biophys Acta* 1998;1399:187–197.
8. Vionnet N, Hani E-H, Dupont S, Gallina S, Francke S, Dotte S, De Matos F, Durand E, Leprêtre F, Lecoœur C, Gallina P, Zekiri L, Dina C, Froguel P: Genome wide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21–q24. *Am J Hum Genet* 2000;67:1470–1480.
9. Francke S, Manraj M, Lacquemant C, Lecoœur C, Leprêtre F, Passa P, Hebe A, Corset L, Yan SL, Lahmidi S, Jankee S, Gunness TK, Ramjuttun US, Balgobin V, Dina C, Froguel P: A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 2001;10:2751–2765.
10. Kissebah AH, Sonnenberg GE, Myklebust J, Goldstein M, Broman K, James RG, Marks JA, Krakower GR, Jacob HJ, Weber J, Martin L, Blangero J, Comuzzie AG: Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A* 2000;97:14478–14483.
11. Populaire C, Mori Y, Dina C, Vasseur F, Vaxillaire M, Kadowaki T, Froguel P: Does the –11377 promoter variant of

- APM1 gene contribute to the genetic risk for type 2 diabetes mellitus in Japanese families? *Diabetologia* 2003;46:443–445.
12. Hu FB, Doria A, Li T, Meigs JB, Liu S, Memisoglu A, Hunter D, Manson JE: Genetic variation at the adiponectin locus and risk of type 2 diabetes in women. *Diabetes* 2004;53:209–213.
 13. Gu HF, Abulaiti A, Ostenson CG, Humphreys K, Wahlestedt C, Brookes AJ, Efendic S: Single nucleotide polymorphisms in the proximal promoter region of the adiponectin (APM1) gene are associated with type 2 diabetes in Swedish Caucasians. *Diabetes* 2004;53:31–35.
 14. Wan Y-JY, An D, Cai Y, Repa JJ, Chen TH-P, Flores M, Postic C, Magnuson MA, Chen J, Chien KR, French S, Mangelsdorf DJ, Sucov HM: Hepatocyte-specific mutation establishes retinoid X receptor α as a heterodimeric integrator of multiple physiological processes in the liver. *Mol Cell Biol* 2000;20:4436–4444.
 15. Iwaki M, Matsuda M, Maeda N, Funahashi T, Matsuzawa Y, Makishima M, Shimomura I: Induction of adiponectin, a fat-derived antidiabetic and antiatherogenic factor, by nuclear receptors. *Diabetes* 2003;52:1655–1663.
 16. Alwan A, King H: Definition and diagnostic criteria for diabetes mellitus and other categories of glucose intolerance. In: *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications*. Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: Department of Noncommunicable Disease Surveillance, World Health Organization, 1999:1–59.
 17. Sanghera DK, Demirci FY, Been L, Ortega L, Ralhan S, Wander GS, Mehra NK, Singh J, Aston CE, Mulvihill JJ, Kamboh IM: PPARG and ADIPOQ gene polymorphisms increase type 2 diabetes mellitus risk in Asian Indian Sikhs: Pro12Ala still remains as the strongest predictor. *Metabolism* 2010;59:492–501.
 18. Fumeron F, Aubert R, Siddiq A, Betoulle D, Péan F, Hadjadj S, Tichet J, Wilpart E, Chesnier MC, Balkau B, Froguel P, Marre M; Epidemiologic Data on the Insulin Resistance Syndrome (DESIR) Study Group: Adiponectin gene polymorphisms and adiponectin levels are independently associated with the development of hyperglycemia during a 3-year period: the Epidemiologic Data on the Insulin Resistance Syndrome Prospective Study. *Diabetes* 2004;53:1150–1157.
 19. Dong C-P, He L, Li J-N, Ye F, He M, Wang Y: Association of Pro12Ala and C1431T polymorphism of the PPAR- γ 2 gene and their haplotypes with obesity and type 2 diabetes. *China J Med Genet* 2008;25:447–450.
 20. Jaziri R, Lobbens S, Aubert R, Péan F, Lahmidi S, Vaxillaire M, Porchay I, Bellili N, Tichet J, Balkau B, Froguel P, Marre M, Fumeron F; DESIR Study Group: The PPARG Pro12Ala polymorphism is associated with a decreased risk of developing hyperglycemia over 6 years and combines with the effect of the APM1 G-11391A single nucleotide polymorphism: the Data From an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* 2006;55:1157–1162.
 21. Tai ES, Corella D, Deurenberg-Yap M, Adiconis X, Chew SK, Tan CE, Ordovas JM: Differential effects of the C1431T and Pro12Ala PPAR γ gene variants on plasma lipids and diabetes risk in an Asian population. *J Lipid Res* 2004;45:674–685.
 22. Doney AS, Fischer B, Cecil J, Boylan K, McGuigan FE, Ralston SH, Morris AD, Palmer CN: Association of Pro12Ala and C1431T variants of PPARG and their haplotypes with susceptibility to Type 2 diabetes. *Diabetologia* 2004;47:555–558.
 23. Evans D, de Heer J, Hagemann C, Wendt D, Wolf A, Beisiegel U, Mann WA: Association between the P12A and C1431T polymorphisms in the peroxisome proliferator activated receptor γ (PPAR γ) and type 2 diabetes. *Exp Clin Endocrinol Diabetes* 2001;109:151–154.
 24. Ritchie MD, White BC, Parker JS, Hahn LW, Moore JH: Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC Bioinform* 2003;7:1–14.
 25. Yan X, Selaru FM, Jing Y, Zou TT, Shustova V, Mori Y, Sato F, Liu TC, Olaru A, Wang S, Kimos MC, Perry K, Desai K, Greenwald BD, Krasna MJ, Shibata D, Abraham JM, Meltzer SJ: Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. *Cancer Res* 2002;62:3493–3497.
 26. Hanai T, Hibino S, Nagata E, Matsubara M, Fukagawa K, Shirataki T, Honda H, Kobayashi T: Assessment of senile dementia of Alzheimer type using artificial neural networks. *Jpn J Med Electron Biol Eng* 1999;37:178–183.
 27. Tomida S, Hanai T, Koma N, Suzuki Y, Kobayashi T, Honda H: Artificial neural network predictive model for allergic disease using single nucleotide polymorphisms data. *J Biosci Bioeng* 2002;93:470–478.
 28. Lancashire LJ, Lemetre C, Ball GR: An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies. *Brief Bioinform* 2009;10:315–329.
 29. Yamauchi T, Kamon J, Waki H, Terauchi Y, Kubota N, Hara K, Mori Y, Ide T, Murakami K, Tsuboyama-Kasaoka N, Ezaki O, Akanuma Y, Gavrilova O, Vinson C, Reitman ML, Kagechika H, Shudo K, Yoda M, Nakano Y, Tobe K, Nagai R, Kimura S, Tomita M, Froguel P, Kadowaki T: The fat-derived hormone adiponectin reverses insulin resistance associated with both lipodystrophy and obesity. *Nat Med* 2001;7:941–946.
 30. Hotta K, Funahashi T, Bodkin NL, Ortmeier HK, Arita Y, Hansen BC, Matsuzawa Y: Circulating concentrations of the adipocyte protein adiponectin are decreased in parallel with reduced insulin sensitivity during the progression to type 2 diabetes in rhesus monkeys. *Diabetes* 2001;50:1126–1133.
 31. Wu LS-H, Hsieh C-H, Pei D, Hung Y-J, Kuo S-W, Lin E: Association and interaction analyses of genetic variants in ADIPOQ, ENPP1, GHSR, PPAR γ and TCF7L2 genes for diabetic nephropathy in a Taiwanese population with type 2 diabetes. *Nephrol Dial Transplant* 2009;24:3360–3366.

Address correspondence to:

Ling Zhou, M.D.

Department of Epidemiology and Biostatistics

School of Public Health

Nanjing Medical University

140 Hanzhong Road

Nanjing, Jiangsu, 210029, China

E-mail: lzhou@njmu.edu.cn

or

Qian Li, M.D.

Department of Endocrinology

Affiliated Nanjing 1st Hospital of Nanjing Medical University

68 Changle Road

Nanjing, Jiangsu, 210009, China

E-mail: shygu@njmu.edu.cn