The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

Study on Multilocus Interactions for Hypertension by an Information Theoretic Method*

Junhua Zhang[1,2] Wentao Huang[3] Zhiyuan Zhao[4]
Biao Li[5] Yuelan Wang[6] Dongfeng Gu[7]
Guoying Li[1,†] Runsheng Chen[8,‡]

[1]Academy of Mathematics and Systems Science, CAS, Beijing, China
[2]Key Laboratory of Random Complex Structures and Data Science,
Academy of Mathematics and Systems Science, CAS, Beijing, China
[3]College of Electronics and Information Engineering,
South-Central University for Nationalities, Wuhan, China
[4]Actuarial Department, China Life Insurance Company Limited, Beijing, China
[5]School of Informatics and Computing, Indiana University, Bloomington, Indiana, USA
[6]Institute of Biotechnology, Beijing, China
[7]Division of Population Genetics and Prevention, Cardiovascular Institute and Fu Wai Hospital,
Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[8]Institute of Biophysics, CAS, Beijing, China

Abstract Hypertension, among diabetes, obesity and others, is one of the common human diseases that is genetically expressed as complex traits to which genetic, environmental, and demographic factors contribute interactively. Identifying the underlying genes and examining their interactions, a crucial step in understanding the molecular pathogenesis of complex diseases, is both a statistical and a computational challenge, stressing the need for novel strategies to move this process forward. In this paper we propose a new method to study the association of multiple gene interactions for complex diseases. Our method is carried out by two steps. First, we sequentially select additionally associated SNP loci combinations by minimizing the p-value of a test based on an information measure, measure of information discrepancy. Therefore, this approach is called MID method. Second, the significance of the selected associated loci combinations is assessed by an χ^2 independence-test. The MID method is model-free and nonparametric, it is easy to compute and implement. The capability of the MID method is confirmed by applying it to investigate the multiple gene interactions on risk of hypertension in northern Han Chinese, where thirty-three SNP loci with three-genotype in eleven candidate genes are examined. Some results are consistent with

*This work was funded by grants H020220030130 Biomedical Project from the Council of Science and Technology, Beijing. W. Huang was supported by the Natural Science Foundation of Central-South University for Nationalities (YZZ09004), and G. Li was supported by the National Natural Science Foundation of China under grant No.10771126.
†Corresponding author: gyli@iss.ac.cn
‡Corresponding author: crs@sun5.ibp.ac.cn

those of Gu et al(2006). Additionally, we get some other new findings. This indicates that our idea is indeed feasible and useful in practice.

Keywords Hypertension; SNP locus; multilocus interactions; measure of information discrepancy

1 Introduction

As one of the most common complex diseases, hypertension is considered to be a complex trait to which genetic, environmental, and demographic factors contribute interactively(Yagil and Yagil 2005; Hunter 2005). The genetic analysis of hypertension has revealed complex and inconsistent results, making it difficult to draw clear conclusions regarding the impact of specific genes on blood pressure regulation in diverse human populations(Agarwal et al 2005). One plausible explanation is that because individual genes play a modest role in the pathogenesis of hypertension, confounding variables, whether individual (sex, ethnic origin, etc.) or environmental, may decrease the chance of identifying a causative relation between the genes and hypertension, depending on the populations studied.

Along with more and more people suffering from the complex diseases like hypertension, there is an urgent need for uncovering and understanding the molecular pathogenesis of common human diseases. As a crucial step to do it, not only identifying the underlying genes but also examining their interactions is payed more and more attentions by many researchers.

Traditional analysis methods such as logistic regression are, however, not sufficient to detect susceptible loci out of a large number of variants and to detect potential interactions among genes and environmental factors, due to the lack of main effects of the variants involving in the complex diseases(Hahn et al 2003; Thornton-Wells et al 2004). Recently several novel methods have been reported aiming at overcoming this obstacle, such as MDR(Ritchie et al 2001; Hahn et al 2003), combinatorial partitioning method (CPM)(Nelson et al 2001), logic regression(Ruczinski et al 2004), the haplotype analysis approach(Yagil and Yagil 2004) and nonparametric tests(Schaid et al 2005). Nevertheless, none of the methods is regarded as the best of all and every method has limitations of detecting the nonlinear interactions or suffering from intensive computation(Hoh and Ott 2003; Thornton-Wells et al 2004). How to effectively detect the interactions between the candidate genes responsible for hypertension is still a great challenge, stressing the need for novel strategies to move this process forward.

In this paper we propose a new method to study the association of multiple gene interactions for complex diseases. Our method is carried out by two steps. First, we sequentially select additionally associated SNP loci combinations by minimizing the p-value of a test based on an information measure, *m*easure of *i*nformation *d*iscrepancy (MID). Therefore, this approach is called MID method. Second, the significance of the selected associated loci combinations is assessed by an χ^2 independence-test. The MID method is model-free and nonparametric. Moreover, because the significant SNP loci combinations are obtained sequentially, the computation for this method is sparing time and its implementation is easy.

Our method is proposed based on the MID measure. The main principle is that the larger the information discrepancy between the hypertensive cases and the normotensive

controls at a locus, the more associated with hypertension the locus is. In fact, the information theoretic method already has some satisfactory applications in bioinformatics such as the reconstruction of phylogenetic trees(Li et al 2002), multiple sequence alignment(Zhang et al 2005), community detection in complex networks(Zhang et al 2008), and other fields.

The capability of the MID method is confirmed by applying it to investigate the multiple gene interactions on risk of hypertension in northern Han Chinese, where thirty-three SNP loci with three-genotype in eleven candidate genes are examined. Some results are consistent with those of Gu et al(2006). Additionally, we get some other new findings. Moreover, in order to uncover the relationship of hypertension with sex, the data with or without gender stratification are all analyzed.

2 Materials and data

The present study was based on the International Collaborative Study of Cardiovascular Disease in Asia (InterASIA) from which all DNA samples and clinical data were obtained(Gu et al 2002). The local bioethical committee approved the study protocol and informed consent from participation was obtained from all subjects. The sample consists of 478 unrelated essential hypertensive cases and 468 unrelated normotensive controls from northern Han Chinese population. Speaking in more details, the ratio of cases and controls are 248:242 and 226:225 for males and females respectively. All measurements and interviews were taken under standard conditions as previously described(Gu et al 2002). Genotypes of variants chosen were determined by PCR and restriction digestion.

We selected eleven genes which involve in the important blood pressure levels regulating processes. These genes locate on eight different chromosomes on which totally thirty-three SNPs were genotyped (Table 1).

3 Method

In this paper we propose a new method to study the association of multiple gene interactions for complex diseases. We sequentially select additionally associated SNP loci combinations by minimizing the p-value of a test based on an information measure, MID. Therefore, this approach is called MID method. Then the significance of the selected associated loci combinations is assessed by an χ^2 independence-test. The MID method is model-free and nonparametric, it is easy to compute and implement.

In the following we introduce the details of our method. For convenience we simply use locus to denote SNP locus hereafter. First, the statistics based on MID for single locus association and for an additional locus association given a combination of loci are given, respectively. Then the whole procedure of the MID method for the association study of multiple gene interactions for complex diseases is presented.

Let n_1 and n_2 represent the number of the individuals in the case group(called group 1) and the control group(called group 2), respectively. Suppose there are M candidate loci with r_i genotypes at locus i ($1 \leq i \leq M$).

3.1 Statistic for single locus association

For a particular locus with r genotypes, let $(\hat{p}(1,1),\ldots,\hat{p}(1,r))$ and $(\hat{p}(2,1),\ldots,\hat{p}(2,r))$ represent the genotype frequencies in group 1 and group 2, respectively. Then, the MID

Table 1: Candidate genes and SNPs assessed

Gene	SNP	Location
AGTR1	A1166C rs275650 C512T	3q21-q25
GRK4	R65L A142V A486V	4p16.3
ADRB2	T(-47)C Arg16Gly Gln27Glu	5q31-q32
NOS3	T(-786)C Intron4b/a G894T	7q36
ADRA1A	C2564T G2547C C2254G C2238T T1991C D465E C347R	8p21-p11.2
CYP11B2	T(-344)C Lys173Arg IC	8q21-q22
LPL	IVS-214C>T 7754C>A S447X	8p22
TH	rs6356 rs6357 rs2070762	11p15.5
GNB3	A(-350)G C825T C1429T	12p13
ACE	ACE I/D	17q23
WNK4	G1662A	17q21-q22

ADRA1A, *adrenergic, alpha-1A, receptor*; ADRB2, *adrenergic, beta-2, receptor*; CYP11B2, *cytochrome P450, family 11, subfamily B, polypeptide 2*; GNB3, *guanine nucleotide binding protein (G protein), beta polypeptide 3*; LPL, *lipoprotein lipase*; AGTR1, *angiotensin II receptor, type 1*; ACE, *angiotensin I converting enzyme*; TH, *tyrosine hydroxylase*; NOS3, *nitric oxide synthase 3*; GRK4, *G protein-coupled receptor kinase 4*; WNK4, *WNK lysine deficient protein kinase 4*.

is defined by

$$B = (n_1+n_2)\sum_{k=1}^{2}\sum_{j=1}^{r}\hat{p}(k,j)\ln\frac{\hat{p}(k,j)}{(\hat{p}(1,j)+\hat{p}(2,j))/2}. \quad (1)$$

Originally the MID is introduced by Fang(1994) to measure the degree of disagreement among multiple information sources. It has been proven that B possesses many good properties, such as non-negativity, symmetry, boundedness, uniform continuity, monotonicity, convexity and so on (Fang(1994,2000)). Here the statistic B in (1) measures the difference between the case and control groups in genotype frequency at the locus considered. Hence B can be used to test the following hypothesis:

$$H_0: \quad p(1,j) = p(2,j), \quad j = 1,\ldots,r,$$

where $p(k,j)$ is the probability of an individual who belongs to group k and possesses the j-th genotype at the locus. Under H_0, B is asymptotically distributed as χ^2 with $r-1$ degrees of freedom when $n_1, n_2 \to \infty$(Zhang and Fang(2003)). Let b denote the observed value of the corresponding MID B, we get the p-value

$$p = P(\chi^2(r-1) \geq b). \quad (2)$$

It is obvious, the smaller p, the more associated the locus is.

3.2 Statistic for an additional locus association given a combination of loci

Given a combination of t loci, say $T = \{1, \cdots, t\}$ without loss of generality. We now want to examine whether adding a locus $Y (t+1 \le Y \le M)$ will yield additional information in association. First, we can view T as a single locus which possesses $r_T \triangleq r_1 r_2 \ldots r_t$ genotypes. The above problem can be formulated by the following hypothesis

$$H_{T,Y}^{(0)}: \quad p(k,y|j) = p(k|j)p(y|j) \quad (1 \le j \le r_T,\ 1 \le y \le r_Y,\ k = 1,2),$$

where $p(k,y|j)$ represents the conditional probability of an individual who belongs to group k and possesses the y-th genotype at locus Y with the condition of being the j-th genotype at locus T, and similarly for $p(k|j)$ and $p(y|j)$. Let $p(k,j,y)$ represent the probability of an individual who belongs to group k and possesses the j-th genotype at locus T and the y-th genotype at locus Y. The null hypothesis $H_{T,Y}^{(0)}$ can be tested by the statistic $B(Y;T)$, which is similar to the MID in (1) and defined as

$$B(Y;T) = 4(n_1+n_2) \sum_{k=1}^{2} \sum_{j=1}^{r_T} \sum_{y=1}^{r_Y} \Big[\hat{p}(k,j,y) \ln \frac{\hat{p}(k,j,y)}{(\hat{p}(k,j,y) + \hat{p}(k|j)\hat{p}(y|j)\hat{p}(j))/2}$$

$$+ \hat{p}(k|j)\hat{p}(y|j)\hat{p}(j) \ln \frac{\hat{p}(k|j)\hat{p}(y|j)\hat{p}(j)}{(\hat{p}(k,j,y) + \hat{p}(k|j)\hat{p}(y|j)\hat{p}(j))/2} \Big],$$

where $\hat{p}(k,j,y)$ represents the estimated value of $p(k,j,y)$ and so forth.

It is known that under $H_{T,Y}^{(0)}$, $B(Y;T)$ is asymptotically distributed as χ^2 with $r_T(r_Y-1)$ degrees of freedom when $n_1, n_2 \to \infty$(Zhang and Fang(2003)). Let $b(Y;T)$ denote the observed value of $B(Y;T)$, now we get the p-value

$$p_{T,Y} = P(\chi^2(r_T(r_Y-1)) \ge b(Y;T)). \tag{3}$$

And the smaller the $p_{T,Y}$ is, the more additional information the locus Y yields.

3.3 Procedure of the MID method for association study of multiple gene interactions for complex diseases

1. Choose a significant level α_1, and use the statistic for single locus association to compute the p-value p_i in (2) for each locus i. Then let

   $$C_1 = \{i : p_i \le \alpha_1\}.$$

   These loci in C_1 are the significantly associated single loci with level α_1. In this paper we choose $\alpha_1 = 0.05/33 \approx 0.0015$, for which the Bonferroni correction is considered.
2. Choose a significant level α_2. For each given locus j, let $T_j = \{j\}$, use the statistic for an additional locus association given a combination of loci to compute the p-value $p_{T_j,i}$ in (3) for each $i \notin T_j (j = 1, \cdots, M)$. Denote

   $$C_2 = \{\{i,j\} : p_{T_j,i} \le \alpha_2\},$$

here C_2 is called the set of 2-order additionally associated loci combinations with level α_2. Noticing that α_2 is used to assess the significance of the additional information furnished by adding a locus(which is called the additional level), in this paper we take $\alpha_2 = 0.10/32 \approx 0.0031$, for which Bonferroni correction is also considered.

If $C_2 \neq \emptyset$, then go to step 3; otherwise, the procedure is finished.

3. Suppose C_{t-1} is a set of $t-1$-order additionally associated loci combinations ($t \geq 3$) with level α_2. For each $T \in C_{t-1}$, compute the p-value $p_{T,i}$ in (3) for each $i \notin T$. Denote
$$C_t = \{\{T,i\} : T \in C_{t-1}, i \notin T, p_{T,i} \leq \alpha_2\},$$
here C_t is called the set of t-order additionally associated loci combinations with level α_2.
4. Repeat step 3 until $C_s = \emptyset$ for some $s \geq 3$. Then denote
$$C = C_2 \cup C_3 \cup \cdots \cup C_{s-1}.$$
All the elements $\{i_1, \cdots, i_t\} \in C, 2 \leq t \leq s-1$ are thought as additionally associated loci combinations with the disease considered.
5. For each $T \in C$, let r be its number of genotypes. Based on the $2 \times r$ table of the genotype frequency in case and control groups to perform an χ^2 independence-test with significant level α_3(in this paper, 0.05). If the p-value of this test is less than α_3, then we say that T is significantly associated loci combinations with the disease.

4 Results

Here we use the MID method to investigate the multiple gene interactions on risk of hypertension in northern Han Chinese, where thirty-three SNP loci with three-genotype in eleven candidate genes are examined. The results for single-locus analysis and for multilocus analysis are obtained. In order to uncover the relationship of hypertension with sex, we have analyzed not only the data for the mixed population of male and female but also the monosexual data for each of them.

4.1 Significantly associated single loci

For the significant level $\alpha_1 = 0.05/33 \approx 0.0015$, we compute the p-value p_i in (2) for each locus $i(1 \leq i \leq 33)$, then the significantly associated loci are selected whose p-value are less then α_1(listed in Table 2).

From Table 2 we can see that the significant results are: TH*rs2070762 (for Male), GRK4*A486V (for Female), and TH*rs2070762, GRK4*A486V and ADRB2*Gln27Glu (for Male & Female). All other loci are not significant in three different cases. The results for Male & Female are consistent with those in the one-way study by Gu et al(2006).

4.2 Significantly associated loci combinations

According to our method described in last section, first the additionally associated loci combinations with the additional level $\alpha_2 = 0.10/32 \approx 0.0031$ are sequentially selected by computing the p-value in (3), then significant association with hypertension of these selected combinations is assessed by an χ^2 independence-test with significant level $\alpha_3 = 0.05$. All the analyses are carried out in three cases for males ,females and both of them.

Table 2: Associated single loci and the corresponding p-values

Male			Female		
Gene	SNP	p-value	Gene	SNP	p-value
TH	rs2070762	1.8130e-006	GRK4	A486V	3.3367e-006

Male & Female		
Gene	SNP	p-value
TH	rs2070762	7.8556e-009
GRK4	A486V	9.0905e-007
ADRB2	Gln27Glu	1.0852e-005

Table 3: Significantly associated two-locus combinations for male

loci combinations		test p-value	loci combinations		test p-value
C512T	rs2070762	1.4314e-004	C825T	rs2070762	5.9332e-005
C2564T	rs2070762	5.3445e-005	C1429T	rs2070762	1.7449e-004
C2254G	rs2070762	4.6751e-005	*C1429T*	*A142V*	0.0058
C2238T	rs2070762	9.8334e-005	IVS-214C>T	rs2070762	1.0574e-004
C347R	rs2070762	3.1010e-005	rs6356	rs2070762	5.7114e-005
Arg16Gly	rs2070762	5.8461e-005	T(-786)C	rs2070762	2.7748e-004
T(-344)C	rs2070762	5.2098e-005	*T(-786)C*	*IVS-214C>T*	0.0160
T(-344)C	*A142V*	0.0084	Intron4b/a	rs2070762	5.2980e-005
Lys173Arg	rs2070762	9.1528e-006	A486V	rs2070762	6.1990e-005
Lys173Arg	*A142V*	0.0047	A142V	rs2070762	3.4023e-005
ACE I/D	rs2070762	3.3799e-004			

In the following we list the results for two-locus analysis and for three-locus analysis, respectively. Moreover, only the significant results are reported for sparing space. For the 4-locus analysis, although we obtain some additionally associated loci combinations, for example, 8 combinations for female and 29 combinations for male & female, but none of them is significant by the successive test.

4.2.1 Two-locus combinations

For each locus $j(1 \leq j \leq 33)$, the p-value $p_{\{j\},i}$ in (3) is computed for every $i \neq j$, $\{i, j\}$ is thought as an additionally associated loci combination if $p_{\{j\},i} \leq \alpha_2$. Then the significance of each selected combination with hypertension is assessed by an χ^2 independence-test with significant level $\alpha_3 = 0.05$. The corresponding results are listed in Tables 3-5 for male, female and male & female, respectively.

From Table 3 we can see that most combinations contain the significantly associated single loci TH*rs2070762. But four other combinations, i.e., GRK4*A142V-CYP11B2*Lys173Arg, GRK4*A142V-GNB3*C1429T, GRK4*A142V-CYP11B2*T(-344)C and NOS3 *T(-786)C-LPL*IVS-214C>T, may be worth more attention. We speculate that it it the interaction between the two insignificant single loci that makes the combination significantly associated to hypertension.

Analogous to the situation in Table 3, most combinations contain the corresponding significantly associated single loci in Tables 4 and 5. Thus we think more attention should be paid to the seven combinations in italic type especially ADRA1A*C2254G-LPL*IVS-

Table 4: Significantly associated two-locus combinations for female

loci combinations		test p-value	loci combinations		test p-value
rs275650	A486V	2.3716e-004	ACE I/D	A486V	1.3741e-004
C512T	*rs2070762*	0.0040	rs2070762	A486V	5.5410e-008
C512T	A486V	1.7101e-004	*rs6356*	*rs2070762*	0.0042
C2254G	*Gln27Glu*	0.0018	rs6356	A486V	0.0012
C2254G	A486V	5.1823e-006	C825T	A486V	4.9538e-004
Arg16Gly	A486V	2.0405e-004	C1429T	A486V	7.8324e-004
T(-344)C	A486V	6.9861e-006	*IVS-214C>T*	*C2254G*	0.0117
Lys173Arg	*C2254G*	0.0122	IVS-214C>T	A486V	1.3763e-005
Lys173Arg	*rs2070762*	0.0140	7754C>A	A486V	4.7921e-004
Lys173Arg	A486V	1.2309e-004	*ACE I/D*	*rs2070762*	0.0019
IC	A486V	8.1602e-005	A142V	A486V	2.4565e-004

Table 5: Significantly associated two-locus combinations for male & female

loci combinations		test p-value	loci combinations		test p-value
rs275650	rs2070762	7.7444e-007	ACE I/D	Gln27Glu	0.0015
rs275650	A486V	9.5120e-005	ACE I/D	rs2070762	8.2228e-007
C512T	rs2070762	3.4645e-008	ACE I/D	A486V	1.0316e-004
C512T	A486V	4.3606e-005	rs2070762	A486V	1.7379e-012
C2564T	rs2070762	9.2342e-008	rs6357	A486V	1.1671e-004
C2564T	A486V	4.2902e-005	rs6356	rs2070762	1.1316e-006
C2254G	Gln27Glu	2.8746e-004	rs6356	A486V	3.3631e-004
C2254G	rs2070762	1.9199e-006	C825T	rs2070762	4.6518e-007
C2254G	A486V	8.3125e-005	C825T	A486V	1.2972e-004
C2238T	rs2070762	1.9380e-006	C1429T	rs2070762	3.4087e-006
T1991C	rs2070762	6.8697e-006	C1429T	A486V	7.8757e-005
T1991C	A486V	1.2032e-004	IVS-214C>T	Gln27Glu	0.0011
C347R	rs2070762	1.2409e-007	IVS-214C>T	rs2070762	7.5058e-006
C347R	A486V	3.8456e-005	IVS-214C>T	A486V	3.0972e-006
T(-47)C	A486V	1.5720e-005	7754C>A	Gln27Glu	4.9937e-004
Gln27Glu	A486V	4.9826e-008	7754C>A	A486V	1.4232e-004
Arg16Gly	Gln27Glu	2.5028e-005	T(-786)C	rs2070762	1.4276e-006
Arg16Gly	rs2070762	2.4583e-007	T(-786)C	A486V	6.2276e-005
Arg16Gly	A486V	1.0703e-005	Intron4b/a	rs2070762	1.5313e-007
T(-344)C	*Lys173Arg*	0.0145	Intron4b/a	A486V	5.1153e-005
T(-344)C	rs2070762	1.4826e-006	IC	A486V	3.4393e-006
T(-344)C	A486V	8.6719e-006	IC	rs2070762	2.5000e-007
Lys173Arg	rs2070762	2.1465e-007	*A142V*	*C347R*	0.0126
Lys173Arg	A486V	4.7370e-006	A142V	rs2070762	1.4913e-006
A142V	A486V	2.3836e-005			

Table 6: Significantly associated three-locus combinations for male

loci combinations			test p-value
C825T	rs2070762	Lys173Arg	1.2598e-004

Table 7: Significantly associated three-locus combinations for female

loci combinations			test p-value
C2254G	A486V	rs2070762	9.7568e-006
C2254G	A486V	IVS-214C>T	2.7469e-006
Arg16Gly	A486V	rs2070762	8.3997e-006
Lys173Arg	A486V	C2254G	1.4253e-004
C825T	A486V	IC	2.4027e-004
C825T	A486V	rs2070762	1.9221e-004
IVS-214C>T	A486V	rs2070762	1.7895e-006

214C>T and ADRA1A*C2254G-CYP11B2*Lys173Arg in Table 4 as well as GRK4*A142V-ADRA1A*C347R and CYP11B2*T(-344)C-CYP11B2* Lys173Arg in Table 5.

4.2.2 Three-locus combinations

Similar to two-locus analysis above, first we pick out the three-order additionally associated loci combinations, then the significant combinations identified from them are listed in Tables 6-8.

5 Discussion

Gene-gene interactions are increasingly found to play critical roles in the etiology of complex diseases(Hunter 2005; Moore 2003), and traditional methods such as multivariable regression analysis have some problem in localizing the underlying genetic variants of complex diseases(Ritchie et al 2001). Thus, we develop a method named MID for identifying the effect of gene-gene interactions in complex diseases such as essential hypertension (EH). In comparison with analysis using the CART and MARS methods on

Table 8: Significantly associated three-locus combinations for male & female

loci combinations			test p-value
C2254G	rs2070762	Lys173Arg	1.6705e-006
C2254G	rs2070762	A486V	1.6377e-008
Arg16Gly	rs2070762	A486V	1.0654e-010
T(-344)C	rs2070762	A486V	8.7280e-010
T(-344)C	A486V	IVS-214C>T	3.4599e-005
Lys173Arg	rs2070762	A486V	8.4167e-010
ACE I/D	rs2070762	A486V	4.1533e-009
C825T	rs2070762	A486V	1.3687e-008
IVS-214C>T	rs2070762	C512T	2.0910e-005
IVS-214C>T	rs2070762	A486V	3.9073e-009
A142V	A486V	Lys173Arg	1.5737e-005

the same real data(Gu et al 2006), there is some difference in the actual outcome of the analyses. However, substantially more information is obtained with the MID method.

With respect to single locus analysis, our data showed that after Bonferroni correction, three polymorphisms including TH*rs2070762, ADRB2*Gln27Glu and GRK4*A486V were independently associated with EH in northern Han Chinese. The two genes, TH and ADRB2, are components of the sympathetic system which exerts an important pathogenetic role in the development of EH through regulation of regional blood flow and maintenance of blood pressure (BP)(Wood et al 2009). GRK4, a component maintaining the balance of sodium and electrolyte, takes part in the desensitization of the D1 dopamine receptor, and makes D1 dopamine receptor uncouple from the G protein/effector enzyme complex, which may cause the failure of dopamine acting on renal tubules of the kidney to inhibit sodium reabsorption(Felder et al 2002; Watanabe et al 2002). Therefore, the TH and ADRB2 and GRK4 genes are attractive candidate genes for hypertension. Our study found 3 individual significant loci out of 33 polymorphisms, indicating common variants contribute to complex diseases with different risk effects. To some extent, there are some consistencies across ours and other studies(Binder et al 2006; Sethi et al 2005; Speirs et al 2004; Zhu et al 2006). As observation made by Hirschhorn et al(2002), after a comprehensive review of association studies, it may not yield fruitful results and will be difficult to find and prove one gene to be involved in the etiology of complex diseases because of gene-gene interactions and other reasons. Thus, multilocus analysis for multiple genes should be conducted.

Out of a large amount of two-locus and three-locus interactions, our multilocus analysis identified a few combinations displaying significant joint effects. Although many of the joint effects included individual significant locus, it becomes gradually weakened with more loci involved. All the significant interactions are not simply additive effects of individual locus. In particular, two combinations, CYP11B2*T(-344)C-CYP11B2*Lys173Arg and GRK4*A142V-ADRA1A*C347R, both of which are independent of the other 3 individual predictors, were found to be significantly associated with EH in unstratified population. The T(-344)C SNP in the promoter of the CYP11B2 gene may alter transcription factor binding, other variants located in the introns and 3'-untranslated region may affect binding of regulatory microRNA species(Wood et al 2009), and the Lys173Arg variants may influence the normal function of the protein. Together, synergistic effect of the two SNPs and other variants may result in the abnormal expression of CYP11B2 through transcriptional, post-transcriptional and protein level. Similarly, the A142V SNP in the GRK4 gene and the C347R SNP in the ADRA1A gene are not single significant locus, but show joint significance beyond simple additive effect through affecting the respective protein function. After gender stratification, more significant combination effects were observed with strong gender difference. The significant combination effects are not consistent according to gender. Interestingly, there are more significant associations in male than in female independent of single significant locus. Gender specificity related to systolic blood pressure (SBP) and diastolic blood pressure (DBP) has been reported by Sethi et al(2005) and by us. Although these findings are interesting, it is not easy to explain them in biologically satisfying manner nowadays. Our findings indicate that different types of gene-gene interactions with or without individually significant predictors involved may play a critical role in the regulation of BP, and suggest that traditional methods of searching for susceptible or resistant loci one at a time may overlook the contribution of many

important genetic loci. Our findings are consistent with the growing understanding that complex interactions are 'the norm', rather than amounting to a small perturbation to classical Mendelian genetics, however, interactions may be the predominant effect(McKinney et al 2006). The significant interactions we identified involved multiple genes in multiple physiological pathways, including renin-angiotensin-aldosterone system, sympathetic nervous system, sodium and electrolyte balance, lipoprotein metabolism, and intracellular messengers. These interactions demonstrated predominant genetic effects over single loci.

Several studies on multilocus analysis of the relationship between genetic polymorphisms and EH have been published. Naber et al(2000) found that there exists a significant interaction of the GNB3 C825T allele and the ACE I/D allele in myocardial infarction. Hypertension is an independent risk factor and the products of the two genes may be connected in a pathway(Kedzierska et al 2006). However, such effect of interaction was not found in the study of Wang et al(2004) in the Kazakh isolate of northerneast China, or in a study of Caucasians with Polish origin(Kedzierska et al 2006). In accordance with the study of Gu et al(2006), no significant effect was observed for neither two- nor three-way interaction involving these two genes in our analysis. Thus, although the association between gene-gene interactions and hypertension vary to some extent among the different studies, possible due to the ethnic differences and the methods used to analyze data, all studies emphasized the importance of gene-gene interactions in the etiology of EH.

In conclusion, we develop the MID method aiming at identifying the effect of gene-gene interactions in complex diseases. When the method applied to examine the relationship between EH and 33 polymorphisms of 11 candidate genes in the northern Chinese Han population, we found that gene-gene interactions with different weight played an important role in the etiology of EH, and that multiple genes in the multiple physiological pathways were involved in the regulation of BP. In particular, two combinations including no single significant locus were found to be significantly associated with EH. Revealing the mechanisms underlying complex diseases like EH poses a considerable challenge to researchers, but also provides great promise for precise diagnosis prognosis and drug design, which would pave the way towards personalized healthcare.

References

[1] Agarwal A, Williams GH and Fisher NDL (2005), Trends Endocrinol. Metab. 16, 127-133.

[2] Binder A, Garcia E, Wallace C, Gbenga K, Ben-Shlomo Y, Yarnell J, Brown P, Caulfield M, Skrabal F, Kotanko P et al (2006), J Hypertens. 24, 471-477.

[3] Fang W (1994), Math. Social Sci. 28, 85-111.

[4] Fang W (2000), Infor. Sci. 125, 207-232.

[5] Felder RA, Sanada H, Xu J, Yu PY, Wang Z, Watanabe H, Asico LD, Wang W, Zheng S, Yamaguchi I et al (2002), Proc. Natl. Acad. Sci. USA 99, 3872-3877.

[6] Gu D, Reynolds K, Wu X, Chen J, Duan X, Muntner P, Huang G, Reynolds RF, Su S, Whelton PK and He J (2002), Hypertension 40, 920-927.

[7] Gu D, Su S, Ge D, Chen S, Huang J, Li B, Chen R and Qiang B (2006), Hypertension 47, 1147-1154.

[8] Hahn LW, Ritchie MD and Moore JH (2003), Bioinformatics 19, 376-382,

[9] Hirschhorn JN, Lohmueller K, Byrne E and Hirschhorn K (2002), Genet. Med. 4, 45-61.

[10] Hoh J and Ott J (2003), Nat. Rev. Genet. 4, 701-709.

[11] Hunter DJ (2005), Nat. Rev. Genet. 6, 287-298.

[12] Kedzierska K, Ciechanowski K, Safranow K, Bober J, Golembiewska E, Kwiatkowska E, Kabat-Koperska J et al (2006), Arch. Med. Res. 37, 150-157.

[13] Li W, Fang W, Ling W, Wang J, Xuan Z and Chen R (2002), J. Biol. Phys. 28, 439-447.

[14] McKinney BA, Reif DM, Ritchie MD and Moore JH (2006), Appl. Bioinformatics 5, 77-88.

[15] Moore JH (2003), Hum. Hered. 56, 73-82.

[16] Naber CK, Husing J, Wolfhard U, Erbel R and Siffert W (2000), Hypertension 36, 986-989.

[17] Nelson M, Kardia SLR, Ferrell RE and Sing CF (2001), Genome Res. 11 458-470.

[18] Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF and Moore JH (2001), Am. J. Hum. Genet. 69, 138-147.

[19] Ruczinski I, Kooperberg C and LeBlanc M (2004), J. Multivari. Anal. 90, 178-195.

[20] Schaid DJ, McDonnell SK, Hebbring SJ, Cunningham JM and Thibodeau SN (2005), Am. J. Hum. Genet. 76, 780-793.

[21] Sethi AA, Tybjaerg-Hansen A, Jensen GB and Nordestgaard BG (2005), Pharmacogenet. Genomics 15, 633-645.

[22] Speirs HJ, Katyk K, Kumar NN, Benjafield AV, Wang WY and Morris BJ (2004), J. Hypertens. 22, 931-936.

[23] Thornton-Wells TA, Moore JH and Haines JL (2004), Trends Genet. 20, 640-647.

[24] Wang X, Wang S, Lin R, Jiang X, Cheng Z, Turdi J, Ding J, Wu G, Lu X and Wen H (2004), J. Hum. Hypertens. 18, 663-668.

[25] Watanabe H, Xu J, Bengra C, Jose PA and Felder RA (2002), Kidney Int. 62, 790-798.

[26] Wood S, Forbes G, MacKenzie S, Stewart P, Connell J and Davies E (2009), Endocrine Abstracts 19, 321.

[27] Yagil Y and Yagil C (2004), J. Hypertens., 22, 1255-1258.

[28] Yagil Y and Yagil C (2005), Curr. Opin. Nephrol. Hypertens. 14, 141-147.

[29] Zhang J and Fang W (2003), Communi. Stat. — Theo. Meth. 32, 435-457.

[30] Zhang J, Zhang S and Zhang XS (2008), Physica A 387, 1675-1682.

[31] Zhang M, Fang W, Zhang J and Chi Z (2005), Comput. Biol. Chem. 29, 175-181.

[32] Zhu H, Lu Y, Wang X, Treiber FA, Harshfield GA, Snieder H and Dong Y (2006), Am. J. Hypertens. 19, 61-66.