MULTILOCUS GENOTYPIC ASSOCIATION WITH VASCULAR DEMENTIA USING MULTIFACTOR DIMENSIONALITY REDUCTION AND ENTROPY DECOMPOSITION

Jiyoung Woo, Minyoung Kong, Younyoung Kim, and Chaeyoung Lee*

Department of Bioinformatics and Life Science, Soongsil University
511 Sangdo-dong, Dongjak-gu, Seoul 156-743, Republic of Korea
**Email: clee@ssu.ac.kr*

Genetic associations with the susceptibility to vascular dementia have been most likely examined for individual candidate genes. This limited to understanding genetic variability of such complex diseases because potential epistasis could not be explained in the analysis with a single gene. In the current study, a simultaneous analysis of candidate genetic markers was conducted by using multifactor dimensionality reduction and entropy decomposition. We compared 121 sequence variants of 71 candidate genes between vascular dementia patients and age/sex matched controls. The multifactor dimensionality reduction analysis revealed that the best 2-locus candidate model including CUB and Sushi multiple domains 1 (CSMD1, rs9694214) and family with sequence similarity 134 member B (FAM134B, rs10041159) was the final model model for predicting genetic susceptibility of VD. An interaction analysis by entropy decomposition revealed that the SNPs included in the final model eliminated the uncertainty of 4.41% and 3.35% in case-control status, and additional 13.35% of entropy decreased with the presence of their interaction. A large portion of genetic variability for the susceptibility of VD was attributed to the synergistic epistasis of the two variants. Further studies on the epistasis are in order to elucidate their underlying mechanisms.

1. INTRODUCTION

Vascular dementia (VD) is a degenerative brain disease induced by cerebrovascular lesions with progressive deterioration in memory and cognitive functions. Although gene-by-gene interaction that is what we call epistasis might explain a large portion of the phenotypic variability for such complex disease, most studies for genetic dissection of VD have focused on estimating effects of individual genes and excluded the potential epistatic effects in their analytical models.[1] Recently, an epistasis between the two genes of angiotensinogen (AGT, Thr235Met) and transforming growth factor-β 1 (TGF-β1, Pro10Leu) was suggested for the susceptibility of VD.[2]

In the current study, an intensive association analysis was examined to discover epistatic effects on susceptibility of VD and to identify the best model for predicting the genetic susceptibility.

2. MATERIALS AND METHODS

The study population was composed of 207 patients with VD and age- and sex-matched 207 controls recruited from Hallym University Hospital. Our previous study described in details on the diagnosis for VD.[3] Patients with mixed dementia presenting both vascular and Alzheimer features were excluded. Controls were subjects without any history of dementia or cerebral ischemic events. Written informed consent was obtained from all subjects, and the study protocol was approved by the Ethical Committee.

The 121 single nucleotide polymorphisms in 71 candidate genes were selected to investigate their epistatic association with VD. Genomic DNA was isolated from peripheral blood cells of the subjects by the kit from Qiagen. The sequence polymorphisms were genotyped using the TaqMan polymerase chain reaction assay. Reactions were carried out following the

* Corresponding author.

manufacturer's protocol, and the products were analyzed using ABI PRISM 7900HT. Genotyping was conducted by laboratory personnel blind to case-control status of the samples.

The epistatic analysis was conducted by multifactor dimensionality reduction (MDR).[4] In MDR, 9 out of 10 random subsets were used as a training set to build a genetic model for predicting risk and protection for VD. The remaining subset was used as a test set to evaluate the models built by using a training set. The training set was analyzed with 3^n possible multigenotype classes for n loci. Each multigenotype class was determined as a high or low risk by ratio of case to control. The best candidate model among n-locus models was defined as the model with the smallest classification error. Testing accuracy (TA) was estimated by using a test set for each selected best n-locus model. All procedures starting with the random partitioning were repeated 10 times, and we called it 10-fold cross-validation. The 100 replicates of MDR were conducted for each 10-fold cross-validation. Selection criteria for a final candidate model were the maximum cross-validation consistency (CVC) and the maximum testing balanced accuracy. The statistical significance of the best candidate model was determined with TA by a permutation test.

Entropy decomposition was further conducted to identify the size and direction (synergy or redundancy) of epistasis among genetic attributes, and an interaction graph was presented. The graph composed of nodes and their pair-wise connections, and the estimates of nodes and connections were calculated by using information gain (IG) with entropy. The graph was drawn with SNPs selected from MDR.

3. RESULTS AND DISCUSSION

The best candidate model was selected by the MDR analysis based on the smallest classification error and summarized in Table 1. The CVC estimate of the best 2-locus candidate model including CUB and Sushi multiple domains 1 (CSMD1, rs9694214) and family with sequence similarity 134 member B (FAM134B, rs10041159) were the largest among the best n-locus candidate models. Furthermore, the testing accuracy estimate of the best 2-locus candidate model was also the largest. Although its training estimate was smaller than that of the best 3- and 4-locus candidate models, their differences were negligible ($P < 0.05$). Both training and testing accuracy estimates of the best 2-locus candidate model were statistically significant ($P < 0.05$). We suggest the best 2-locus candidate model as the final model for predicting genetic susceptibility of VD.

Table 1. Best candidate model estimated by multifactor dimensionality reduction.

Best Model	Training Accuracy	Testing Accuracy	CVC
OXTR	0.6143*	0.5320	6/10
CSMD1 FAM134B	0.7831*	0.7806*	10/10
LRMP CSMD1 FAM134B	0.7862*	0.7726*	8/10
OXTR LRMP CSMD1 FAM134B	0.7869*	0.7663*	8/10

OXTR - oxytocin receptor, rs237895; CSMD1 - CUB and Sushi multiple domains 1, rs9694214; FAM134B - family with sequence similarity 134 member B, rs10041159; LRMP - lymphoid-restricted membrane protein, rs17326622

* $P < 0.05$

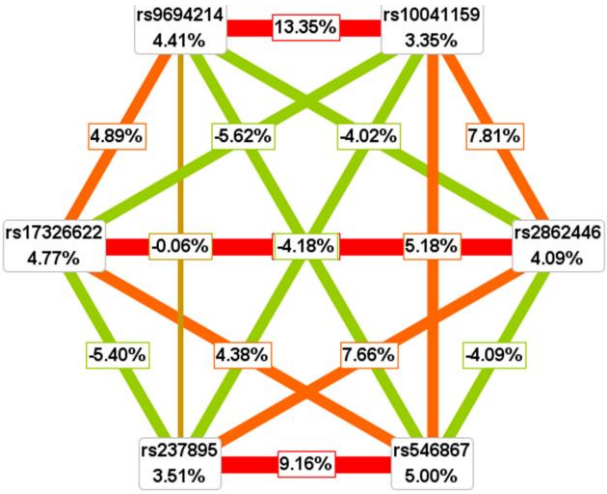

Fig. 1. Entropy decomposition with the 6 loci which were contributing to the susceptibility of vascular dementia. This hierarchical interaction graph shows the percentage of entropy removed in case-control by main individual locus effect (node) and by their pairwise interaction effect (connected line). Positive value on the line indicates synergistic epistasis between the two loci, and negative value indicates redundant epistasis.

An interaction graph drawn by entropy decomposition showed that the SNPs included in the final model (CSMD1 rs9694214 and FAM134B rs10041159) eliminated the uncertainty of 4.41% and 3.35% in case-control status (Fig. 1). Their individual

effects were smaller than those of SNPs such as rs17326622 and rs546867. Nevertheless, additional 13.35% of entropy decreased with the presence of their interaction. A large portion of genetic variability for the susceptibility of VD was attributed to the synergistic epistasis of the two variants.

Acknowledgments

This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A080042).

References

1. Lee C. Vascular dementia. In: Lang F (ed.), *Encyclopedia of Molecular Mechanisms of Disease*. Springer, Berlin, Germany. 2009: 512–514.
2. Kim Y, Park J, Lee C. Multi-locus genotypic association with vascular dementia by multifactor dimensionality reduction and entropy-based estimation. *Psychiatr Genet* 2009; doi: 10.1097/YPG.0b013e32832ceebd.
3. Kim Y, Nam YJ, Lee C. Haplotype analysis of single nucleotide polymorphisms in VEGF gene for vascular dementia. *Am J Med Genet Part B* 2006; **141**: 332-335.
4. Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 2006; **241**: 252-261.