Analysis of the association between DNA methylation sites and cardiovascular disease-related single nucleotide polymorphisms in the Japanese population

Masahiro Nakatochi[*], Sahoko Ichihara[†], Ken Yamamoto[‡], Tatsuaki Matsubara[§], and Mitsuhiro Yokota[**]

***Abstract*—The development of cardiovascular disease (CVD) depends on environmental and genetic factors. Although many single nucleotide polymorphisms (SNPs) associated with CVD susceptibility have been identified to date, the mechanisms through which these polymorphisms contribute to disease development remain unclear. To investigate the epigenetic basis of myocardial infarction (MI), a type of CVD, we analyzed the associations between DNA methylation (DNAm) status at methylation sites and CVD-associated SNPs in elderly Japanese individuals. A total of 192 patients with MI and 192 controls were recruited from hospital attendees and the general population, respectively. Genome-wide DNAm profiles in whole blood were obtained by analysis with an Infinium HumanMethylation450 BeadChip. The CVD-associated SNPs were genotyped using an SNP array and were imputed using 1000 genomes phase 3 as a reference. Association analysis identified 431 genome-wide or suggestive significant combinations of DNAm sites and SNPs. A total of 225 gene symbols were located near the DNAm sites. These genes were significantly enriched in the 6p21 region. Our results suggested that SNPs located at chromosome 6p21 may influence the development of CVD by affecting DNAm in blood cells.**

Keyword: *Single nucleotide polymorphisms, DNA methylation, cardiovascular disease, myocardial infarction*

1 Introduction

Cardiovascular disease (CVD) is the leading cause of mortality, morbidity, and hospitalization worldwide. The prevalence of CVD is increasing more rapidly in Asia than in Western countries. Improving our understanding of the pathogenesis of CVD may help mitigate further increases in the incidence of CVD.

Genetic factors have been found to contribute to the development of CVD. Genome-wide association studies (GWAS) for CVD have revealed many associated susceptibility genes and single nucleotide polymorphisms (SNPs). Although many SNPs associated with CVD susceptibility have been identified to date, the mechanisms through which these polymorphisms contribute to disease development remain unclear. Recent progress in epigenetic epidemiology has facilitated investigations of the relationships among genomic coding, modifiable exposure, and manifestations of disease phenotypes. DNA methylation (DNAm), a major type of epigenetic modification, may be an important mechanism underlying these relationships.

In this study, we measured the genome-wide DNAm status for DNA samples prepared from whole blood of patients with myocardial infarction (MI), a type of CVD, attending hospitals in Japan and of elderly Japanese participants of the Kita-Nagoya Genomic Epidemiology (KING) study. These samples were also genotyped using an SNP array. We thereby assessed the relationships between DNAm sites identified in our study and SNPs previously found to be associated with CVD in GWAS.

[*] Statistical Analysis Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8560, Japan, E-mail: mnakatochi@med.nagoya-u.ac.jp, Tel: +81-52-744-1957
[†] Graduate School of Regional Innovation Studies, Mie University, Tsu 514-8507, Japan
[‡] Department of Medical Biochemistry, School of Medicine, Kurume University, Kurume 830-0011, Japan
[§] Department of Internal Medicine, School of Dentistry, Aichi Gakuin University, Nagoya 464-8561, Japan
[**] Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya 464-8561, Japan

2 Methods

2.1 Study participants

This was an exploratory study examining the association of DNAm status at various sites with CVD-associated SNPs. A total of 192 male cases with MI and 192 male controls without MI were enrolled. All participants were at least 55 years old, and the two groups were matched by age (within 5 years). Genome-wide DNAm profiles for DNA isolated from whole blood were obtained by analysis with an Infinium HumanMethylation450 BeadChip. The details of DNAm profile analysis were described previously (Nakatochi *et al*., 2015 and Lehn *et al.*, 2015).

2.2 Genotyping and imputing SNPs

All samples were also genotyped using an Illumina HumanOmniExpress-12 BeadChip. Quality control of samples and SNPs was performed. After the quality control process, we imputed genotypes using SHAPEIT2 and minimac3 with data from the 1000 genomes project phase 3 as a reference panel. SNPs with poor imputation quality (Rsq < 0.3) and low minor allele frequency (MAF < 0.01) were excluded. Of SNPs that passed the postimputation process, 146 CVD-associated SNPs were registered in the GWAS catalog database (https://www.ebi.ac.uk/gwas/). Furthermore, we used PLINK software to exclude SNPs with a pairwise genotypic r^2 of greater than 0.2 on the data of East Asian samples from the 1000 genomes project 3. A total of 105 CVD-associated SNPs remained for further analysis.

2.3 Statistical analysis

To explore DNAm sites associated with 105 CVD-associated SNPs, association analysis was performed between each DNAm site and each SNP in the case and control groups. A general linear model was applied for each combination of DNAm site and SNP; the dependent variable was DNAm status at each site, and independent variables included the genotype of each SNP and covariates. The covariates included age, smoking status (not a current smoker = 0, current smoker = 1), the first 30 principal component scores calculated from Infinium 450K assay control probes, the first five principal component scores calculated from the residuals after adjustment for technical and biological factors, and the cell type composition of samples. We coded genotypes as 0, 1, or 2 on the basis of the number of minor alleles.

Next, the results from case and control groups were then combined into a fixed effects meta-analysis with inverse variance weighting.

Finally, we choose DNAm sites suggestive significantly associated with CVD-related SNPs ($p < 1 \times 10^{-6}$), based on the results of the meta-analysis. We inspected gene symbols located near the significantly associated DNAm sites. To identify significantly enriched pathways, the Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david-d.ncifcrf.gov/) was used.

For the association analysis between DNAm sites and SNPs, the genome wide significance level and suggestive significance level were set to 1×10^{-8} and 1×10^{-6}, respectively. For the DAVID analysis, a false discovery rate (FDR) of less than 0.05 was considered significant. All statistical analyses were performed with the R project (version 3.3.0, www.r-project.org).

3 Results

The baseline characteristics of the study participants are shown in Table 1. The frequency of current smokers was significantly higher in the case group than in the control group.

We performed genome-wide DNAm profiling for whole-blood DNA from 192 cases and 192 controls. After initial processing, 188 cases and 190 controls, including 348,595 DNAm sites and 105 CVD-associated SNPs, remained for subsequent analysis. We initially performed an association analysis between DNAm status at each site and each CVD-associated SNP. As a result, 431 combinations of DNAm sites and SNPs showed suggestive significance. Of these combinations, 187 combinations showed genome-wide significance. The combination of DNAm site and SNP located near the *ABO* gene showed the most significant association ($\beta = 1.330 \pm 0.032$, $p < 1 \times 10^{-99}$ in meta-analysis).

We identified 225 gene symbols located near the DNAm sites, which included the 431 combinations. The top 10 gene ontology (GO) terms based on DAVID analysis are shown in Table 2. All of these GO terms, including GO:0071556, GO:0046978, and GO:0032395, were related to the 6p21 region.

4 Discussion and summary

In this study, we performed an association analysis of DNAm sites with CVD-related SNPs and found 431 significant combinations of DNAm sites and SNPs. The most significant combination was the DNAm site and SNP located near the *ABO* gene. This relationship with the *ABO* gene was previously reported by Grunberg et al (2013). The *ABO* gene encodes proteins related to the first discovered blood group system, ABO; thus, the result observed in our blood samples was consistent with their reports. Furthermore, genes for which the DNAm sites had significant associations with CVD-related SNPs were significantly enriched in GO terms related to the 6p21 region. We have previously reported that the *HLA-DQB1* SNP located on chromosome 6p21 is associated with MI (Takeuchi *et al*., 2012). The 6p21 region is known as the most polymorphic region in the human genome. Our results suggested that these SNPs located on chromosome 6p21 may influence the development of CVD by affecting DNAm in blood cells.

Table 1 Characteristics of the study participants

Characteristic	Controls (n = 192)	Cases (n = 192)	*p*
Male, n (%)	192 (100%)	192 (100%)	1.000
Age (years)	65.8 ± 6.0	65.9 ± 6.4	0.915
Current smoker, n (%)	43 (22.4%)	83 (43.2%)	2.01×10^{-5}

Continuous data are means ± SDs. Differences in characteristics between cases and controls were evaluated by Student's *t* tests or Fisher's exact tests.

Table 2 Top five GO terms based on DAVID analysis

Rank	GO ID	GO term	Number of genes	FDR
1	GO:0071556	Integral component of lumenal side of the endoplasmic reticulum membrane	7	7.06×10^{-5}
2	GO:0032395	MHC class II receptor activity	5	6.59×10^{-4}
3	GO:0046978	TAP1 binding	4	1.13×10^{-3}
4	GO:0042605	Peptide antigen binding	6	1.28×10^{-3}
5	GO:0042613	MHC class II protein complex	5	1.94×10^{-3}

Acknowledgments

This study was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, including those for the Funding Program for Next Generation World-Leading Researchers (NEXT Program) to S.I. (LS056), for Scientific Research (B) to M.Y. (24390169 and 16H05250) and to T.M. (25293144), for Scientific Research (C) to K.Y. (15K08290), and for Young Scientists (B) to M.N. (15 K19242).

References

[1] Nakatochi M, Ichihara S, Yamamoto K, et al. (2015) Epigenome-wide association study suggests that SNPs in the promoter region of RETN influence plasma resistin level via effects on DNA methylation at neighbouring sites. *Diabetologia* 58: 2781-2790

[2] Lehne B, Drong AW, Loh M, et al. (2015) A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* 16: 37
[3] Grundberg E, Meduri E, Sandling JK, et al. (2013) Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet* 93: 876-890
[4] Takeuchi F, Yokota M, Yamamoto K, et al. (2012) Genome-wide association study of coronary artery disease in the Japanese. *Eur J Hum Genet* 20: 333-340