Journal of Genetics, Vol. 96, No. 6, December 2017, pp. 1041–1046
https://doi.org/10.1007/s12041-017-0854-z

RESEARCH NOTE

Genome-based exome sequencing analysis identifies *GYG1, DIS3L* and *DDRGK1* are associated with myocardial infarction in Koreans

JI-YOUNG LEE[1,2], SANGHOON MOON[1], YUN KYOUNG KIM[1], SANG-HAK LEE[3], BOK-SOO LEE[4], MIN-YOUNG PARK[5], JEONG EUY PARK[4]*, YANGSOO JANG[3]* and BOK-GHEE HAN[1]*

[1] *Center for Genome Science, Korea National Institute of Health, KCDC, Cheongju-si, Chungcheongbuk-do 363-951, Korea*
[2] *Cardiovascular Research Institute and Cardiovascular Genome Center, Yonsei University Health System, Seoul 120-752, Korea*
[3] *Cardiology Division, Department of Internal Medicine, Severance Cardiovascular Hospital, Yonsei University College of Medicine, Seoul 120-752, Korea*
[4] *Division of Cardiology, Samsung Medical Center, Seoul 135-710, Korea*
[5] *DNA link, Seoul 121-850, Korea*
* For correspondence. E-mail: Jeong Euy Parks, jepark39@gmail.com; Yangsoo Jang, jangys1212@yuhs.ac; Bok-Ghee Han, bokghee@nih.go.kr.

Received 4 October 2016; revised 1 January 2017; accepted 17 February 2017; published online 28 November 2017

Abstract. Myocardial infarction (MI) is a complex disease caused by combination of genetic and environmental factors. Although genome-wide association studies (GWAS) identified more than 46 risk loci which are associated with coronary artery disease and MI, most of the genetic variability in MI still remains undefined. Here, we screened the susceptibility loci for MI using exome sequencing and validated candidate variants in replication sets. We identified that three genes (*GYG1*, *DIS3L* and *DDRGK1*) were associated with MI at the discovery and replication stages. Further research will be required to determine the functional association of these genes with MI risk, and these associations have to be confirmed in other ethnic populations.

Keywords. combined multivariate and collapsing method; myocardial infarction; exome sequencing.

Introduction

Myocardial infarction (MI) is a complex disease caused by genetic and environmental factors (Yusuf *et al.* 2004; Kessler *et al.* 2013). In the past decade, genomewide association studies (GWAS) have been applied to identify genetic susceptibility factors of MI and more than 46 risk loci, which are associated with coronary artery disease and MI, have been identified (Anderson *et al.* 2007; Schunkert *et al.* 2010; Peden and Farrall 2011; CARDIoGRAMplusC4D Consortium *et al.* 2013; Kessler *et al.* 2013). However, the variants identified by GWAS explain only a small fraction of the heritable component of MI risk (CARDIoGRAMplusC4D Consortium *et al.* 2013), mobilizing interest in rare, low-frequency variants that might contribute to the risk of this complex disease (Altshuler *et al.* 2008; Manolio *et al.* 2009). We hypothesized that genetic influences were typically larger in young onset of MI. To identify the genetic variants that confer susceptibility to MI, we screened the susceptibility loci for MI using exome sequencing and validated candidate variants in replication sets.

Ji-Young Lee and Sanghoon Moon contributed equally to this work.

Materials and methods

Subjects

In this study, we used a two-stage approach, which first employed exome sequencing in a series of individuals and

Table 1. Baseline clinical attributes of the study subjects.

	Case ($n = 167$)	Control ($n = 100$)	P value
Age, years	42.4 (7.9)	49.3 (6.1)	<0.0001
Female (%)	21.6	60	<0.0001
Hypertension (%)	97.6	0	<0.0001
BMI (kg/m^2)	25.4 (4.2)	21.4 (1.2)	<0.0001
Total cholesterol (mg/dL)	190.4 (56.6)	182.7 (30.5)	<0.0001
HDL-cholesterol (mg/dL)	42.3 (15.0)	48.3 (11.1)	<0.0001
LDL-cholesterol (mg/dL)	116 (43.4)	110.2 (27.1)	<0.0001
Triglycerides (mg/dL)	177.0 (184.8)	120.7 (50.9)	<0.0001

Data are presented as mean ± SD or (%).

a large-scale follow-up genotyping of identified candidate variants in larger case–controls. The first phase of screening included 167 MI patients and 100 control subjects. The diagnosis of MI was based on typical chest pain over 430 min, characteristic electrocardiographic patterns of acute myocardial infarction, and elevated creatine kinase-MB and troponin I levels. Patients with familial hypercholesterolaemia, known vasculitides, end-stage renal disease and congenital heart disease were excluded from this study. Young onset of MI is defined as age of first diagnosis less than 45 in male and less than 55 in female (Malmberg *et al.* 1994; Shah *et al.* 2016). Controls were eligible if they were not diagnosed with subclinical disease assessing anthropometric and biochemical examinations including ECG. In detail several studies have been described in the same subjects (Cho *et al.* 2009; CARDIoGRAMplusC4D Consortium *et al.* 2013; Lee *et al.* 2013). The mean (±SD) ages were 42.4 (±7.9) years for the patients and 49.3 (±6.1) years for the control subjects. At the replication stage, we selected 252 MI patients and 1000 control individuals, currently participating in the KoGES to confirm the results of the screening stage. The 252 patients were not included in discovery stage, and the controls were from a large urban cohort. Demographic characteristics, including age, sex, ethnicity, disease history, lifestyle, and medication information were obtained from a detailed questionnaire. Genomic DNA was extracted from peripheral blood leukocyte pellets using a DNA extraction kit (Qiagen, Valencia, USA). The study protocol was approved by the institutional review boards at the Korea National Institute of Health and at each collaborating institute. Informed consent was obtained from all participants.

Target exon capture and exome sequencing

Genomic DNA was extracted from peripheral blood using a QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany). Exome sequencing was performed by the Agilent SureSelect Human All Exon 50Mb Capture kit and 100-bp paired-end sequencing onto Illumina HiSeq 2000 according to the manufacturer's protocol. Aligning to the reference genome was conducted by the Burrow-Wheeler Aligner (ver. 0.6.1). In addition, removal of duplicate PCR read and recalibrating of quality scores were performed by Picard (ver. 1.6.7) (http://broadinstitute.github.io/picard/), and Genome Analysis Toolkit (ver. 2.3.6), respectively (McKenna *et al.* 2010). Variant calls were retained with a minimum of 20× coverage using GATK genotyper (ver. 2.3.6). An average of 6.93 billion bases of high-quality sequence was generated per individual. Though the targeted capture method enables a high rate of enrichment for the targeted regions, nonspecific capture also occurred. A 52% of aligned reads mapped to the targeted regions of the genome.

Statistical analysis

Fisher's exact tests were used for the association test of annotated variants to compare the frequencies of variants identified from the sequence data between patients and controls subjects. Existing association methods for common variants may not work for rare variants (Bansal *et al.* 2010). We therefore deployed the combined multivariate and collapsing (CMC) method (Li and Leal 2008) and C-alpha methods (Neale *et al.* 2011). These methods, provide robust results when dealing with low frequencies and rare variants, which might predispose an individual to a given risk (Pan and Shen 2011). In addition, we applied a permutation to avoid the inflation of type I error rates. All statistics was applied using the PLINK/SEQ (ver. 0.10) (https://atgu.mgh.harvard.edu/plinkseq) and R packages (ver. 3.2.1) (http://cran.r-project.org).

Results

The demographic and baseline clinical characteristics are shown in table 1. Age, body mass index, and HDL-cholesterol differed significantly between MI patients and control individuals. This confirms that the control individuals we selected were very healthy and therefore any confounding risk factors could be ruled out.

Table 2. Gene-based statistical significance analyses at the discovery and replication stages.

						Discovery				Replication			
Gene	Chr.	BP	Ref.	Alt.	SNP[a]	No. of SNP	No. of rare SNP*	*P* value[b]	*P* value[c]	No. of SNP**	No. of rare SNP*	Permutated *P* value[b]	Permutated *P* value[c]
CYP4A22	1	47603172	C	G	47603172**	20	13	8.84×10^{-3}	2.18×10^{-8}	11	6	0.163671	0.170592
		47603188	C	T	rs76011927								
		47603297	C	G	47603297								
		47603312	C	T	47603312								
		47606476	C	T	47606476								
		47609492	A	G	rs191687330**								
		47609531	C	A	47609531								
		47610065	A	C	rs149718343**								
		47610100	C	T	47610100**								
		47610221	G	A	47610221**								
		47610274	C	T	47610274**								
		47610275	G	A	47610275								
		47610293	C	T	rs968322**								
		47610590	A	C	rs2405593**								
		47611765	G	A	rs150794228**								
		47611833	G	A	47611833**								
		47614376	C	A	47614376								
		47614379	C	A	rs185005287								
		47614434	C	T	rs4926600								
		47614485	A	G	rs72684327**								
GYG1	3	148711957	C	T	148711957**	8	6	2.88×10^{-3}	1.15×10^{-6}	5	3	$<5 \times 10^{-6}$	$<5 \times 10^{-6}$
		148712028	T	C	148712028**								
		148714060	G	A	rs143942810								
		148714146	C	T	rs149479866**								
		148714238	G	T	148714238								
		148727127	C	A	148727127								
		148727133	G	A	rs4938**								
		148744731	T	C	148744731**								

Table 2 (*contd*)

Gene	Chr.	BP	Ref.	Alt.	SNP[a]	Discovery				Replication			
						No. of SNP	No. of rare SNP*	*P* value[b]	*P* value[c]	No. of SNP**	No. of rare SNP*	Permutated *P* value[b]	Permutated *P* value[c]
DIS3L	15	66604032	C	T	rs28616181	11	5	3.16×10^{-4}	4.52×10^{-5}	7	1	$<5 \times 10^{-6}$	0.001483
		66610984	G	A	rs141774650**								
		66612965	T	C	rs17851970**								
		66618143	T	C	66618143								
		66618308	G	A	rs141173452								
		66618342	A	G	rs3803412								
		66618345	A	C	66618345**								
		66618663	G	A	rs117749912**								
		66621347	C	T	66621347**								
		66625161	A	G	rs11071885**								
		66625470	G	A	rs3759785**								
DDRGK1	20	3175957	C	T	3175957	5	4	5.09×10^{-5}	1.43×10^{-5}	2	1	0.46987	0.004
		3175991	C	T	3175991**								
		3176009	T	C	rs2295552**								
		3180666	C	A	rs35327491								
		3180717	C	T	3180717								

[a]Novel SNPs were notified by base position (GRCh37.p10).

[b]*P* value were calculated by CMC.

[c]*P* value were calculated by C-ALPHA; *minor allele frequency<1%; **successfully replicated SNP.

Chr., chromosome; BP, base position; Ref., reference SNP; Alt., alternative SNP; *CYP4A22*, cytochrome P450, family 4, subfamily A, polypeptide 22; *GYG1*, glycogenin 1; *DIS3L*, DIS3-like exosome 3′–5′ exoribonuclease; *DDRGK1*, DDRGK domain containing 1.

We found 3420 SNPs with P values of <0.05 by analysing differences in allele distribution between the patients and the control subjects. In the gene-based analysis, we selected $P < 1 \times 10^{-5}$ (CMC or C-alpha methods) as the cut-off for highly suggestive loci. After ruling out the outliers (number of rare variants (less than 4) or with unknown function), we selected 44 variants to replicate the association with MI. Using combined analysis of the common and rare variants, we confirmed that three gene (*GYG1*, *DIS3L* and *DDRGK1*) were associated with MI at the discovery stage ($P < 0.05$). The results of the statistical analysis of the variants are shown in table 2. In the analysis at the replication stage, variants in *CYP4A22* did not show a significant association with MI risk ($P > 0.05$).

Discussion

This study is the first to explore the presence and effect of rare and common variation in early onset MI. We evaluated specific genes chosen on the basis of CMC methods and confirmed that three genes (*GYG1*, *DIS3L* and *DDRGK1*) were associated with MI at the discovery and replication stages.

GYG1, the strongest gene, encodes a member of the glycogenin family. Glycogenin is a glycosyltransferase that catalyzes the formation of a short glucose polymer from uridine diphosphate glucose in an autoglucosylation reaction. Sequencing of *GYG1*, the glycogenin-1 gene, revealed a nonsense mutation in one allele and a missense mutation; (Thr83Met) in the other. The missense mutation resulted in the inactivation of the autoglucosylation of glycogenin-1, which is necessary for priming glycogen synthesis in muscle (Moslemi *et al.* 2010). Further, mutations in genes involved in glycogen metabolism cause glycogen storage cardiomyopathies (Arad *et al.* 2005).

Notably, *DDRGK1*, encoding *DDRGK* domain containing 1, is a protein-coding gene associated with thrombocytopenia and hepatitis, which explains MI-related platelet pathogenesis (Ochi *et al.* 2010; Tanaka *et al.* 2011). Previous exome sequencing identified rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction at an early age (Do *et al.* 2015). In addition, recent consortium study showed that carriers of novel mutations in *ANGPTL4* were associated with protection from myocardial infarction (Stitziel *et al.* 2016). However, most studies in exome sequencing were conducted among populations of European ancestry suggesting lack of replication in other populations. The size of our study sample limited our ability to detect previously identified variants among European ancestries. Hence, efforts to identify additional variation underlying MI will require much larger study samples. While we include relatively small number of subjects compared to other studies, this is the first evidence of three novel genes (*GYG1*, *DIS3L* and *DDRGK1*) detected using gene-based exome sequencing in an Asian population. Due to a lack of additional functional studies, the mechanisms by these genes influence MI pathogenesis remain unknown.

For the road ahead, further research will be required to determine the functional association of these genes with MI risk, and these associations will need to be confirmed in other ethnic populations.

Acknowledgements

We thank all participants and investigators of the Korea Genome Epidemiology Study (KoGES). This work was supported by grants from the Korea Centers for Disease Control and Prevention (4845-301), an intramural grant from the Korea National Institute of Health (2012-N73003-00).

References

Altshuler D., Daly M. J. and Lander E. S. 2008 Genetic mapping in human disease. *Science* **322**, 881–888.

Anderson J. L., Carlquist J. F., Horne B. D. and Hopkins P. N. 2007 Progress in unraveling the genetics of coronary artery disease and myocardial infarction. *Curr. Atheroscler. Rep.* **9**, 179–186.

Arad M., Maron B. J., Gorham J. M., Johnson Jr W. H., Saul J. P., Perez-Atayde A. R. *et al.* 2005 Glycogen storage diseases presenting as hypertrophic cardiomyopathy. *N. Engl. J. Med.* **352**, 362–372.

Bansal V., Libiger O., Torkamani A. and Schork N. J. 2010 Statistical analysis strategies for association studies involving rare variants. *Nat. Rev. Genet.* **11**, 773–785.

CARDIoGRAMplusC4D Consortium, Deloukas P., Kanoni S., Willenborg C., Farrall M., Assimes T. L. *et al.* 2013 Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33.

Cho Y. S., Go M. J., Kim Y. J., Heo J. Y., Oh J. H., Ban H. J. *et al.* 2009 A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534.

Do R., Stitziel N. O., Won H. H., Jorgensen A. B., Duga S., Angelica Merlini P. *et al.* 2015 Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106.

Kessler T., Erdmann J. and Schunkert H. 2013 Genetics of coronary artery disease and myocardial infarction–2013. *Curr. Cardiol. Rep.* **15**, 368.

Lee J. Y., Lee B. S., Shin D. J., Woo Park K., Shin Y. A., Joong Kim K. *et al.* 2013 A genome-wide association study of a coronary artery disease risk variant. *J. Hum. Genet.* **58**, 120–126.

Li B. and Leal S. M. 2008 Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321.

Malmberg K., Bavenholm P. and Hamsten A. 1994 Clinical and biochemical factors associated with prognosis after myocardial infarction at a young age. *JACC* **24**, 592–599.

Manolio T. A., Collins F. S., Cox N. J., Goldstein D. B., Hindorff L. A., Hunter D. J. *et al.* 2009 Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A. *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

Moslemi A. R., Lindberg C., Nilsson J., Tajsharghi H., Andersson B. and Oldfors A. 2010 Glycogenin-1 deficiency and inactivated priming of glycogen synthesis. *N. Engl. J. Med.* **362**, 1203–1210.

Neale B. M., Rivas M. A., Voight B. F., Altshuler D., Devlin B., Orho-Melander M. *et al.* 2011 Testing for an unusual distribution of rare variants. *PLoS Genet.* **7**, e1001322.

Ochi H., Maekawa T., Abe H., Hayashida Y., Nakano R., Kubo M. *et al.* 2010 ITPA polymorphism affects ribavirin-induced anemia and outcomes of therapy–a genome-wide study of Japanese HCV virus patients. *Gastroenterology* **139**, 1190–1197.

Pan W. and Shen X. 2011 Adaptive tests for association analysis of rare variants. *Genet. Epidemiol.* **35**, 381–388.

Peden J. F. and Farrall M. 2011 Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* **20**, R198–R205.

Schunkert H., Erdmann J. and Samani N. J. 2010 Genetics of myocardial infarction: a progress report. *Eur. Heart J.* **31**, 918–925.

Shah N., Kelly A. M., Cox N., Wong C. and Soon K. 2016 Myocardial infarction in the "Young": risk factors, presentation, management and prognosis. *Heart Lung Circ.* **25**, 955–960.

Stitziel N. O., Stirrups K. E., Masca N. G. D., Erdmann J., Ferrario P. G., König I. R. *et al.* 2016 Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N. Engl. J. Med.* **374**, 1134–1144.

Tanaka Y., Kurosaki M., Nishida N., Sugiyama M., Matsuura K., Sakamoto N. *et al.* 2011 Genome-wide association study identified ITPA/DDRGK1 variants reflecting thrombocytopenia in pegylated interferon and ribavirin therapy for chronic hepatitis C. *Hum. Mol. Genet.* **20**, 3507–3516.

Yusuf S., Hawken S., Ounpuu S., Dans T., Avezum A., Lanas F. *et al.* 2004 Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 937–952.

Corresponding editor: KUNAL RAY