European Journal of Human Genetics (2011) 19, 672–675

www.nature.com/ejhg

ARTICLE

Heterogeneity of genetic associations of *CDKAL1* and *HHEX* with susceptibility of type 2 diabetes mellitus by gender

Hyunju Ryoo[1], Jiyoung Woo[1], Younyoung Kim[1] and Chaeyoung Lee*[,1]

We examined the genetic associations of previously identified sequence variants with type 2 diabetes mellitus (T2DM) and its potentially genetic heterogeneity by gender in a large-scale cohort. A total of 613 T2DM patients and 8221 control subjects from the Korea Association REsource (KARE) cohort were included in the analysis of genetic association of T2DM with 33 nucleotide polymorphic markers identified by previous studies. The association analysis was further conducted with data partitioned by gender. The association analysis resulted in five nucleotide sequence variants associated with the susceptibility of T2DM after Bonferonni correction ($P<0.0015$). One was located near the gene of hematopoietically expressed homeobox (*HHEX*), and the others were all in the gene of cyclin-dependent kinase 5 regulatory subunit-associated protein 1-like 1 (*CDKAL1*). Further analysis revealed that the sequence variant (rs5015480) near *HHEX* and two SNPs (rs7756992 and rs9465871) in *CDKAL1* were associated with the susceptibility of T2DM in females ($P<0.005$), but not in males ($P>0.005$). We suggested heterogeneous genetic associations of the T2DM susceptibility with the *CDKAL1* and *HHEX* genes by gender.
European Journal of Human Genetics (2011) **19,** 672–675; doi:10.1038/ejhg.2011.6; published online 2 February 2011

Keywords: genetic association; genetic polymorphisms; genetic risk factors

INTRODUCTION

Type 2 diabetes mellitus (T2DM) has been a threatening public health risk as it gains increased prevalence. Vigorous efforts have been made to identify genetic factors for the susceptibility to T2DM using genetic association analyses. Especially, several genome-wide association studies were recently conducted.[1–5] Nevertheless, their results revealed inconsistencies in associations of many nucleotide sequence variants.[6] A serious publication bias was suspected because a fewer publications with negative results were likely to be reported from candidate gene association analyses. Another concern was that retrospective studies (eg, case–control study) might have resulted in some genetic effects confounded with other effects. More prospective studies based on genome-wide association analysis should be replicated to determine whether such inconsistencies were caused by genetically different underlying population structures and various environmental exposures or by false findings. This study was aimed to examine genetic associations of previously identified sequence variants with T2DM using a large-scale Korean cohort data.

Incidence rates of T2DM have been known different between males and females, but the diversity varied among studies.[7] In the second vein of the current study, different genetic effects were examined to explain the heterogeneous incidence rate of T2DM by gender.

METHODS

Subjects and data

The Korea Association REsource (KARE) Analysis Consortium was established to understand the human genetic basis by conducting a large-scale genome-wide association study. It has cohort data of 10 038 unrelated Korean individuals collected by the Korean Genome Epidemiology Study (KoGES). The data collection was initiated in 2001, and thereafter follow-up examinations for each participant have been conducted for every 2 years. Genotypic data of the KARE were obtained using the Affymetrix Genome-Wide Human SNP Array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA). For details, see Cho *et al*.[8] An underlying set of unphased genotypes for each individual in the cohort were imputed with the Japanese and Chinese HapMap phase 2 haplotype panel using IMPUTE software program (version 2, http://mathgen.stats.ox.ac.uk/impute). A total of 8842 individuals were from the Ansung (2374 men and 2263 women) and Ansan (1809 men and 2396 women) population-based cohorts in Gyeonggi Province. The data were obtained after screening by genotype calling and quality control.[8] However, eight individuals without phenotype or covariate information were excluded in the current analysis.

Mean age of the remained 8834 subjects was 52.2 ± 8.9 years and their mean BMI was $24.6 \pm 3.1\,kg/m^2$. Six hundred thirteen out of 8834 subjects were self-reported as the patients with T2DM, and they were considered as patients diagnosed with T2DM based on the ADA criteria fasting plasma glucose ≥ 126 mg/dl or 2-h plasma glucose ≥ 200 mg/dl. The other subjects in the cohort were all used as controls. The characteristics of the patients were compared with those of controls by gender in Table 1.

Marker selection

We analyzed the genetic association of T2DM with SNP markers identified by previous studies. The SNPs were selected based on peer-reviewed scientific publications using SNPedia (http://www.snpedia.com), a wiki-based database of SNPs associated with human diseases. The selection criteria were the associations previously reported from multiple studies, or from a study with a meta-analysis, with a large consortium (at least 500 patients), or with a multi-laboratory consortium. The associations of the SNPs with susceptibility to T2DM were all confirmed through original scientific articles that identified the associations.

[1]School of Systems Biomedical Science, Soongsil University, Seoul, Korea
*Correspondence: Professor C Lee, School of Systems Biomedical Science, Soongsil University, 511 Sangdo-dong, Dongjak-gu, Seoul 156-743, Korea. Tel +822 820 0455; Fax +822 824 4383; E-mail clee@ssu.ac.kr
Received 25 August 2010; revised 29 December 2010; accepted 5 January 2011; published online 2 February 2011

Table 1 Characteristics of subjects studied in the current study

	Control			*T2DM*		
	Male (n=3848)	*Female (n=4373)*	*Total (n=8221)*	*Male (n=329)*	*Female (n=284)*	*Total (n=613)*
Age	51.4 ± 8.7	52.3 ± 9.0	51.9 ± 8.9	56.0 ± 8.6	57.7 ± 8.4	56.8 ± 8.5
BMI	24.2 ± 2.9	24.8 ± 3.3	24.6 ± 3.1	24.7 ± 2.7	25.9 ± 3.2	25.2 ± 3.0

Table 2 Associations of previously identified SNPs with T2DM susceptibility in Koreans[a]

		SNP			*MAF*				
Chr	*Position*	*ID*	*Gene*	*Allele*[c]	*Case*	*Control*	χ^2	P	P_{corr}[b]
6	**20787687**	**rs7756992**	**CDKAL1**	**A/G**	**0.38**	**0.45**	**25.73**	**2.58E−06**	**8.51E−05**
6	**20825233**	**rs9465871**	**CDKAL1**	**T/C**	**0.39**	**0.45**	**24.47**	**4.85E−06**	**1.60E−04**
10	**94465559**	**rs5015480**	**HHEX**[d]	**C/T**	**0.24**	**0.19**	**19.94**	**4.69E−05**	**1.55E−03**
6	**20765542**	**rs4712523**	**CDKAL1**	**G/A**	**0.53**	**0.48**	**15.99**	**3.37E−04**	**1.11E−02**
6	**20769012**	**rs10946398**	**CDKAL1**	**C/A**	**0.53**	**0.48**	**15.32**	**4.72E−04**	**1.56E−02**
10	94481917	*rs7923837*	—	G/A	0.27	0.23	9.76	7.58E−03	2.50E−01
9	22134094	rs10811661	—	C/T	0.4	0.44	9.53	8.54E−03	2.82E−01
1	66558759	rs4655595	PDE4B	G/A	0.29	0.33	8.33	1.56E−02	5.15E−01
11	17408630	rs5215	KCNJ11	C/T	0.42	0.39	6.42	4.04E−02	1.00E−00
2	160922420	rs6718526	RBMS1	T/C	0.12	0.15	5.61	6.06E−02	1.00E−00
10	114754088	rs7901695	TCF7L2	C/T	0.04	0.03	4.35	1.14E−01	1.00E−00
3	12393125	rs1801282	PPARG	G/C	0.04	0.05	4.12	1.28E−01	1.00E−00
10	114758349	*rs7903146*	TCF7L2	T/C	0.04	0.03	3.88	1.43E−01	1.00E−00
10	114756041	rs4506565	TCF7L2	T/A	0.04	0.03	3.79	1.51E−01	1.00E−00
16	52368186	rs7193144	FTO	C/T	0.13	0.13	3.68	1.59E−01	1.00E−00
16	55495790	*rs2289116*	SLC12A3	A/G	0.09	0.09	3.64	1.62E−01	1.00E−00
16	52373775	rs8050136	FTO	A/C	0.13	0.13	3.41	1.82E−01	1.00E−00
16	52378027	rs9939609	FTO	A/T	0.13	0.13	3.34	1.88E−01	1.00E−00
3	12351744	*rs17036314*	PPARG	C/G	0.27	0.29	3.12	2.11E−01	1.00E−00
10	43388563	rs9326506	—	A/C	0.44	0.46	3.08	2.14E−01	1.00E−00
12	69863367	rs1495377	—	G/C	0.33	0.31	2.7	2.59E−01	1.00E−00
4	122884963	rs7659604	—	T/C	0.33	0.33	2.67	2.63E−01	1.00E−00
11	41915366	*rs9300039*	—	A/C	0.22	0.23	2.63	2.69E−01	1.00E−00
3	186994380	rs4402960	IGF2BP2	T/G	0.31	0.29	2.52	2.84E−01	1.00E−00
3	187011773	rs1470579	IGF2BP2	C/A	0.32	0.31	1.68	4.33E−01	1.00E−00
3	55288439	rs358806	—	A/C	0.23	0.24	1.66	4.36E−01	1.00E−00
8	118253963	*rs13266634*	SLC30A8	T/C	0.4	0.41	0.56	7.56E−01	1.00E−00
6	149763382	rs237025	SUMO4	G/A	0.31	0.31	0.48	7.86E−01	1.00E−00
15	78200438	rs2903265	ZFAND6	G/A	0.47	0.46	0.47	7.91E−01	1.00E−00
12	49643808	rs12304921	—	A/G	0.49	0.5	0.4	8.19E−01	1.00E−00
17	17675464	*rs1889018*	SREBF1	A/G	0.07	0.07	0.33	8.47E−01	1.00E−00
7	36884519	*rs741301*	ELMO1	C/T	0.32	0.33	0.3	8.62E−01	1.00E−00
15	72391886	rs2930291	—	G/A	0.26	0.26	0.01	9.94E−01	1.00E−00

[a]The analytical model included age, sex, and BMI as covariates for the combined data and age and BMI for the partitioned by gender. Boldface indicates significant association by Bonferonni test ($P_{corr}<0.05$). Italic ID indicates the locus analyzed with imputed genotypes.
[b]*P*-value corrected by Bonferonni multiple testing.
[c]Major allele/minor allele.
[d]10 Kb apart from *HHEX* gene.

Statistical analysis

Genotypic association of each SNP with susceptibility to T2DM was tested by χ^2 statistics with two degrees of freedom. The analytical model included age, gender, and BMI as covariates. Threshold of false-positive error in the significance test was 0.05 and Bonferonni multiple testing corrections were introduced to correct for occurrence of false positives. The association analysis was further conducted with data partitioned by gender with adjustment for age and BMI. All the association analyses were conducted using PLINK (version 1.06, http://pngu.mgh.harvard.edu/purcell/plink) and SPSS (version 12.0, SPSS Inc., Chicago, IL, USA) software programs.

RESULTS

We obtained 41 previously identified SNPs using SNPedia. Twenty-four out of 41 SNPs were included in this genome-wide association study using the Korean cohorts, and 10 SNPs with imputed genotypes were additionally used in the current association analysis. As rs12255372 in *TCF7L2* gene was monomorphic, it was excluded in the current association analysis. A total of 33 SNPs were analyzed, and none of them were deviated ($P>0.05$) from Hardy–Weinberg equilibrium except for rs12304921. The association analysis revealed

Table 3 Associations of SNPs with T2DM susceptibility by gender[a]

SNP			Male			Female		
ID	*Gene*	*Allele*[b]	χ^2	P	P_{corr}[c]	χ^2	P	P_{corr}[c]
rs7756992	**CDKAL1**	**A/G**	10.44	5.42E−03	5.42E−02	**15.93**	**3.48E−04**	**3.48E−03**
rs9465871	**CDKAL1**	**T/C**	10.21	6.07E−03	6.07E−02	**15.09**	**5.29E−04**	**5.29E−03**
rs5015480	**HHEX**[d]	**C/T**	8.11	1.74E−02	1.74E−01	**12.79**	**1.67E−03**	**1.67E−02**
rs4712523	CDKAL1	G/A	7.29	2.62E−02	2.62E−01	9.64	8.07E−03	8.07E−02
rs10946398	CDKAL1	C/A	6.86	3.24E−02	3.24E−01	9.29	9.62E−03	9.62E−02

[a]The analytical model included age and BMI as covariates. Boldface indicates significant association by Bonferonni test ($P_{corr}<0.05$). Italic ID indicates the locus analyzed with imputed genotypes.
[b]Major allele/minor allele.
[c]*P*-value corrected by Bonferonni multiple testing.
[d]10 Kb apart from *HHEX* gene.

nine SNPs associated with the susceptibility of T2DM ($P<0.05$) and five SNPs after Bonferonni correction ($P<0.0015$, $P_{corr}<0.05$, Table 2). One was an intergenic sequence variant on chromosome 10 and located ~10 kb apart from 3′-end of hematopoietically expressed homeobox (*HHEX*) gene. The other four SNPs were all located within intron 5 of cyclin-dependent kinase 5 (*CDK5*) regulatory subunit-associated protein 1-like 1 (*CDKAL1*) gene.

A further analysis with the significant SNPs showed heterogeneous results with data partitioned by gender (Table 3). There were three sequence variants of rs7756992 (*CDKAL1*), rs9465871 (*CDKAL1*), and rs5015480 (*HHEX*) significantly associated with the susceptibility of T2DM in females ($P<0.005$, $P_{corr}<0.05$). On the other hand, no significant variants were observed in males ($P_{corr}>0.05$).

DISCUSSION

The current replication study revealed associations of five previously identified sequence variants with the susceptibility of T2DM. Especially, they were all located within *CDKAL1* gene except for an intergenic variant of rs5015480 near *HHEX* gene. The association of the *CDKAL1* gene concurred with the results from previous studies with American,[1] British,[5] French,[3] Danish,[4] Finnish,[2] Swedish,[1] Icelandic,[4] Korean,[9] Japanese,[10] and Chinese[11–13] populations. The significantly associated variants of the gene in the current study were specifically corresponding to those found in previous studies of Horikawa *et al*[14] and Wu *et al*[11] for rs7756992 and Zeggini *et al*[5] and Wu *et al*[11] for rs9465871 and rs10946398, respectively. The replicated results strengthened the finding that the gene and its sequence variants conferred risk of T2DM. This could be explained by the function of the *CDKAL1* on insulin secretion. The *CDKAL1* has a domain similar to *CDK5* regulatory subunit-associated protein 1 (*CDK5RAP1*), a neuronal protein that specifically inhibits activation of *CDK5*.[15] The reduced expression of *CDKAL1* enhances activity of *CDK5* in β cells and thus decreases insulin secretion.[16] The function of the *CDKAL1* on insulin secretion would be influenced by alternative splicing. A variety of modified functional sites in splicing process were predicted by allelic substitutions of its sequence variants (Supplementary Data), and this *in silico* prediction supported the putative roles of the sequence variants on insulin secretion.

The fact that rs5015480 was located outside known protein-coding sequences essentially tells little about its function. Nevertheless, its potential function of regulating the *HHEX* gene was suspected from previous genome-wide association studies in which several variants including the rs5015480 in and around the gene have been associated with T2DM of Europeans.[2,3,5] The associations were replicated also in the Japanese[17] and the Chinese[12] populations. Furthermore, other variants of the gene were associated with impaired pancreatic β-cell function and thus with decreased insulin secretion.[18,19]

A further analysis showed associations of three (rs7756992, rs9465871, and rs5015480) out of the five significant SNPs with the susceptibility of T2DM in females. No associations were observed in males. The most significant two SNPs were sequence polymorphisms located in the *CDKAL1*. The heterogeneity by gender was a novel finding for T2DM. The larger effect of the *CDKAL1* in females concurred with the study of Steinthorsdottir *et al*,[4] showing a larger association of rs7756992 with insulin secretion in females ($P<0.00001$) than in males ($P<0.001$). The differential effect would lead to a deflated association by adding males in the association analysis. As an example, such deflation was suspected in a nominal effect resulted with a large ratio of males to females (1.5 for both patients and control subjects).[10] The genetic heterogeneity by gender might be attributed to sexual hormones which considerably influenced insulin sensitivity and secretion.[20,21] Especially, as estrogen could affect the susceptibility of T2DM[22] and regulate the activity of *cdk5* in adult rat uterus,[23] it could serve as an important candidate resulting in the gender-specific effect of *CDKAL1* on T2DM susceptibility. Also, estrogen could enhance serum level of thyroid hormone[24] whose production was regulated by *HHEX*, a crucial transcription factor.[25] A potential interactive action of *HHEX* with the sex hormone might produce the heterogeneous genetic effect of *HHEX* on T2DM susceptibility by gender.

The current study provided the first evidence of a heterogeneous association by gender between susceptibility of T2DM and the two genes, *CDKAL1* and *HHEX*. The identified genetic variants conferring risk for the disease were the commonly present SNPs, which can be a great concern on practice of medical application. Functional studies on the heterogeneous association are warranted to elucidate their underlying mechanisms.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (Grant No. 2009-0071063). We thank the National Institute of Health, Korea for providing the genotypic and epidemiological data to the KARE Analysis Consortium.

AUTHOR CONTRIBUTIONS

HR researched data, contributed to discussion, and wrote manuscript. JW researched data. YK researched data, contributed to discussion, and reviewed/edited manuscript. CL contributed to discussion, wrote manuscript, and reviewed/edited manuscript.

1 Saxena R, Voight BF, Lyssenko V *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
2 Scott LJ, Mohlke KL, Bonnycastle LL *et al*: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
3 Sladek R, Rocheleau G, Rung J *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
4 Steinthorsdottir V, Thorleifsson G, Reynisdottir I *et al*: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007; **39**: 770–775.
5 Zeggini E, Weedon MN, Lindgren CM *et al*: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
6 Dehwah MA, Wang M, Huang QY: CDKAL1 and type 2 diabetes: a global meta-analysis. *Genet Mol Res* 2010; **9**: 1109–1120.
7 Franconi F, Seghieri G, Canu S, Straface E, Campesi I, Malorni W: Are the available experimental models of type 2 diabetes appropriate for a gender perspective? *Pharmacol Res* 2008; **57**: 6–18.
8 Cho Y, Go M, Kim Y *et al*: A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 2009; **41**: 527–534.
9 Lee YH, Kang ES, Kim SH *et al*: Association between polymorphisms in SLC30A8, HHEX, CDKN2A/B, IGF2BP2, FTO, WFS1, CDKAL1, KCNQ1 and type 2 diabetes in the Korean population. *J Hum Genet* 2008; **53**: 991–998.
10 Omori S, Tanaka Y, Takahashi A *et al*: Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. *Diabetes* 2008; **57**: 791–795.
11 Wu Y, Li H, Loos RJ *et al*: Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. *Diabetes* 2008; **57**: 2834–2843.
12 Han X, Luo Y, Ren Q *et al*: Implication of genetic variants near SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, FTO, TCF2, KCNQ1, and WFS1 in type 2 diabetes in a Chinese population. *BMC Med Genet* 2010; **11**: 81.
13 Wen J, Rönn T, Olsson A *et al*: Investigation of type 2 diabetes risk alleles support *CDKN2A/B, CDKAL1*, and *TCF7L2* as susceptibility genes in a Han Chinese cohort. *PLoS One* 2010; **5**: e9153.
14 Horikawa Y, Miyake K, Yasuda K *et al*: Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J Clin Endocrinol Metab* 2008; **93**: 3136–3141.
15 Lew J, Huang QQ, Qi Z *et al*: A brain-specific activator of cyclin-dependent kinase 5. *Nature* 1994; **371**: 423–426.
16 Ubeda M, Rukstalis JM, Habener JF: Inhibition of cyclin-dependent kinase 5 activity protects pancreatic beta cells from glucotoxicity. *J Biol Chem* 2006; **281**: 28858–28864.
17 Takeuchi F, Serizawa M, Yamamoto K *et al*: Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009; **58**: 1690–1699.
18 Pascoe L, Tura A, Patel SK *et al*: Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* 2007; **56**: 3101–3104.
19 Staiger H, Stancáková A, Zilinskaite J *et al*: A candidate type 2 diabetes polymorphism near the HHEX locus affects acute glucose-stimulated insulin release in European populations: results from the EUGENE2 study. *Diabetes* 2008; **57**: 514–517.
20 Buffington CK, Givens JR, Kitabchi AE: Opposing actions of dehydroepiandrosterone and testosterone on insulin sensitivity. *In vivo* and *in vitro* studies of hyperandrogenic females. *Diabetes* 1991; **40**: 693–700.
21 Jones ME, Thorburn AW, Britt KL *et al*: Aromatase-deficient (ArKO) mice have a phenotype of increased adiposity. *Proc Natl Acad Sci USA* 2000; **97**: 12735–12740.
22 Margolis KL, Bonds DE, Rodabough RJ *et al*: Effect of oestrogen plus progestin on the incidence of diabetes in postmenopausal women: results from the Women's Health Initiative Hormone Trial. *Diabetologia* 2004; **47**: 1175–1187.
23 Altucci L, Addeo R, Cicatiello L *et al*: Estrogen induces early and timed activation of cyclin-dependent kinases 4, 5, and 6 and increases cyclin messenger ribonucleic acid expression in rat uterus. *Endocrinology* 1997; **138**: 978–984.
24 Arafah BM: Increased need for thyroxine in women with hypothyroidism during estrogen therapy. *N Engl J Med* 2001; **344**: 1743–1749.
25 Pellizzari L, D'Elia A, Rustighi A, Manfioletti G, Tell G, Damante G: Expression and function of the homeodomain-containing protein Hex in thyroid cells. *Nucleic Acids Res* 2000; **28**: 2503–2511.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)