Journal of Genetics, Vol. 97, No. 5, December 2018, pp. 1213–1223
https://doi.org/10.1007/s12041-018-1031-8

RESEARCH ARTICLE

Impact of variants on type-2 diabetes risk genes identified through genomewide association studies in polycystic ovary syndrome: a case–control study

INTISSAR EZZIDI[1,2]*, NABIL MTIRAOUI[1,2], MOHAMMED ELTIGANI MOHMMED ALI[3], AQEEL AL MASOUDI[4] and FAISEL ABU DUHIER[1]

[1] *Prince Fahd Bin Sultan Research Chair, Using Advance Technology for Diseases Detection and Treatment, Faculty of Applied Medical Sciences, University of Tabuk, Tabuk, Saudi Arabia*
[2] *Laboratory of Human Genome and Multifactorial Diseases, LR12ES07, Faculty of Pharmacy of Monastir, University of Monastir, Monastir, Tunisia*
[3] *King Salman Military Hospital, Tabuk, Saudi Arabia*
[4] *King Abdulaziz Airbase Armed Forces Hospital, Dhahran, Saudi Arabia*
*For correspondence. E-mail: iezzidi@ut.edu.sa.

Received 26 February 2018; revised 24 April 2018; accepted 3 May 2018; published online 30 November 2018

Abstract. Polycystic ovary syndrome (PCOS) is a common endocrine disorder in females, and is associated with altered metabolic processes in particular insulin resistance and diabetes mellitus. PCOS shares with type-2 diabetes (T2D) a number of features, including beta cell dysfunction, impaired glucose tolerance and dyslipidaemia. Recently, genomewide association studies (GWAS) have reported a number of genes with reproducible associations and susceptibilities to T2D. To address this, we examined the association between the T2D GWAS candidate genes (*CDKAL1*, *CDKN2B*, *COL8A1*, *HHEX*, *IGF2BP2*, *KCNJ1*, *KCNQ1* and *SLC30A8*) and PCOS in Saudi women. A case–control study, includes 162 cases and 162 controls was enrolled. Genotyping was carried out by the allelic discrimination method. Our results showed that the variants including rs792837 of *COL8A1*, rs61873498 of *KCNQ1* and rs13266634 of *SLC30A8* genes to be significantly more frequent in PCOS patients than in controls. Our results suggest that *COL8A1*, *KCNQ1* and *SLC30A8*, which are previously identified through GWAS as T2D-associated genes, are associated with PCOS.

Keywords. genome-wide association studies; single-nucleotide polymorphisms; polycystic ovary syndrome.

Introduction

Polycystic ovary syndrome (PCOS) is a common and heterogeneous endocrine disorder in females with well-established metabolic abnormalities. PCOS is characterized by the presence of two or more of the following features: chronic oligo-ovulation or anovulation, androgen excess and polycystic ovaries (Rotterdam 2004). While PCOS can occur at all ages, its prevalence is highest in women in their reproductive age. The reported prevalence of PCOS is ~5–10% in reproductive-age womendepending on the diagnostic criteria (Goodarzi *et al.* 2011; Trikudanathan 2015). PCOS is a growing epidemic among Saudi Arabian women, and women around the world, although is a highly variable condition with a wide array of presentations (Alsibyani *et al.* 2017). PCOS is a complex genetic disorder caused by interplay between several 'susceptibility' genes and environment factors (Diamanti-Kandarakis *et al.* 2006; Hughes *et al.* 2006; Li and Baek 2015). The genetic predisposition is evidenced by the association of specific gene variants with PCOS (Jones and Goodarzi 2016). Over the past several years, the reproductive medicine field revealed great advances in genome-wide association studies (GWAS) of PCOS leading to identification of several promising genes involved in hormone action, type-2 diabetes (T2D) and cell proliferation (Liu *et al.* 2016). The strongest association established for Southeast Asians including Chinese and Japanese, and later replicated in other ethnic group populations was with the gene of encoding voltage-gated potassium channel, KQT-like member 1 (*KCNQ1*) (Chen *et al.* 2011; Li *et al.* 2012). Since then,

genomewide single-nucleotide polymorphism (SNP) genotype data identified in excess of 17 new loci associated with T2D, including *CDKAL1*, *CDKN2B*, *COL8A1*, *HHEX*, *IGF2B2*, *KCNJ11*, *SLC30A8*, and *JAZF1* and *MNTR1B*. Considering its related traits, the PCOS is among the most challenging issues for women at reproductive age and is associated with an alarming rise in the incidence. The Saudi Arabian women are no exception to this trend. More recently, studies replicated several GWAS-T2D variants and pointed out other variants as plausible susceptible gene for PCOS across many ethnic backgrounds. However, genetic studies to date in Arab cohorts remain limited and in Saudi Arabian women still absent and the most of the PCOS studies were confined to the clinical dimensions. To the best of our knowledge, this study is the first in its type in the Kingdom of Saudi Arabia interested in the genetic analyses of PCOS.

In the present study, we examined the contribution of selected GWAS-T2D variants to the increasing PCOS risk among Saudi women within a case–control study. The obtained results allowed fine investigation of the impact of these variants on lipid profiles and sex hormones of PCOS patients.

Materials and methods

Study subjects

A total of 162 patients with PCOS and 162 control women were recruited at the outpatient Endocrinology and Obstetrics and Gynaecology clinics in King Salman Military Hospital (Tabuk, Saudi Arabia). PCOS definition is established as per the 2003 Rotterdam criteria (Rotterdam 2004). All participants are Saudi Arabs and non-Arab Saudi or recently naturalized Saudi Arabs are excluded.

Biochemical and hormonal analyses

Fasting blood glucose was measured in fluoride oxalate tubes by the hexokinase method (Cobas Integra 800; Roche, Mannheim, Germany). Total insulin was measured by enzyme-linked immunosorbent assay (ELISA), using the DRG EIA kit. Each sample was tested in duplicate, and results were scored as the mean of two determinations (IU/mL), as per the manufacturer's specification. Homeostasis model assessment-insulin resistance (HOMA-IR) index was calculated by using the HOMA calculator (www.dtu.ox.ac.uk/homa/index) as follows: HOMA-IR = [insulin (mU/L) × glucose (mmol/L)]/22.5; HOMA-IR values >2.5 is considered to be elevated. Hyperinsulinaemia level is determined by calculating the area under the curve of insulin response (AUCI). Luteinizing hormone (LH), follicle-stimulating hormone (FSH), thyroid-stimulating hormone, testosterone and oestradiol levels' determination were performed by ELISA with some modifications. Bioavailable testosterone levels are calculated based on the measured testosterone, sex hormone binding globulin and albumin levels, as detailed elsewhere (Vermeulen *et al.* 1999). Serum lipid measurements (total cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) and triglycerides) were performed by the enzymatic colorimetric method (Integra 800; Roche).

Molecular analysis

Genomic DNA preparation: Blood samples were collected by venipuncture in EDTA tubes, and buffy coat layer is extracted into clean nuclease-free microfuge tubes. DNA extraction is performed using a Qiagen kit and DNA is dissolved in nuclease-free water, and stored at 4°C.

SNP genotyping: Genotyping of all patient and control specimens for the selected 11 SNPs was performed using real-time polymerase chain reaction (PCR). Primer selection was performed using SNPbrowser 4.0 software. For each SNP, FAM-labelled and VIC-labelled probes, corresponding to wild-type and mutant alleles were used. Amplification and detection of the genotypes was performed in the same session (i.e., closed system setting); individual genotypes were marked using the manufacturer's default cluster settings using StepOne software.

Statistical analysis: Statistical analysis was performed using SPSS v. 22.0 software (SPSS, Chicago, USA). Data were expressed as mean ± SD for continuous variables which were normally distributed. A Pearson's χ^2 test was used to assess inter-group significance, Student's *t*-test was used to determine differences in means and Mann–Whitney *U*-test was used to determine the distribution of the dependent continuous variable. Allele frequencies were calculated by the gene-counting method, and each polymorphism was tested for the Hardy–Weinberg equilibrium (HWE) by the χ^2 goodness-of-fit test using HPlus 2.5 software (http://qge.fhcrc.org/hplus). Differences in allele and genotype frequencies of gene variants were tested by Pearson's χ^2 test. The association of genotypes with PCOS was determined under log-additive, additive, dominant and recessive models, using SNPstats (http://bioinfo.iconcologia.net/SNPstats).

The calculation of the power for detecting an association between the GWAS-T2D variants and PCOS was performed with the CaTS power calculator (www.sph.umich.edu/csg/abecasis/cats). The parameters used were 162 women with PCOS and 162 control women, genotypic relative risk for heterozygote (1/2) and homozygous minor allele (2/2), and minor-allele frequencies (MAF) for PCOS cases and controls for the nine tested SNPs, and assuming a 19.5% prevalence of PCOS in the general

Table 1. Clinical and biochemical characteristics of study subjects.

Characteristic	Controls[a]	Cases[a]	P^b
Age at inclusion in study[c]	28.44 ± 4.24	27.81 ± 4.98	0.221
BMI (kg/m^2)[c]	23.62 ± 2.29	27.69 ± 4.80	< 0.001
Fasting glucose (mmol/L)[c]	4.88 ± 0.88	7.64 ± 2.30	< 0.001
Free insulin (μU/mL)[c]	7.23 ± 2.63	14.39 ± 6.43	< 0.001
HOMA-IR[c]	1.56 ± 0.62	5.20 ± 3.23	< 0.001
HDL (mmol/L)[c]	1.52 ± 0.27	1.30 ± 0.35	< 0.001
LDL (mmol/L)[c]	1.73 ± 0.58	3.55 ± 1.34	< 0.001
Triglycerides (mmol/L)[c]	1.52 ± 0.51	1.68 ± 0.84	0.034
Cholesterol (mmol/L)[c]	3.81 ± 0.48	5.47 ± 1.40	< 0.001
FSH levels (mIU/mL)[c]	1.71 ± 1.98	5.54 ± 5.80	< 0.001
LH levels (mIU/mL)[c]	0.70 ± 0.97	5.85 ± 6.81	< 0.001
Oestradiol (pmol/L)[c]	338.32 ± 247.49	386.65 ± 338.88	0.159
Progesterone (pmol/L)[c]	14.23 ± 15.49	45.82 ± 68.28	< 0.001
Testosterone (ng/dL)[c]	24.49 ± 20.98	68.02 ± 33.77	< 0.001

[a]A total of 162 PCOS cases and 162 control subjects were included.
[b]Student's *t*-test for continuous variables (variables with normal distribution), Mann–Whitney *U*-test (variables that were not normally distributed).
[c]Mean ± SD.

population according to the Rotterdam criteria (Jalilian *et al.* 2015). With these parameters, we calculated the overall power (80%) as the average power of the tested SNPs.

Logistic regression analysis was performed to determine the adjusted odds ratios (OR) and 95% confidence intervals (95% CI) associated with the risk of PCOS, after controlling for a number of covariates, taking control women as the reference group.

All P values were two-tailed; P values < 0.05 were considered statistically significant.

Results

Study subjects

The clinical characteristics of study subjects are reported in table 1. While age at examination and oestradiol were comparable between cases and controls, significant differences between them were noted in mean body mass index (BMI) ($P<0.001$), fasting glucose ($P<0.001$), free insulin ($P<0.001$) and HOMA-IR ($P<0.001$), and serum lipids (HDL, LDL, total cholesterol and triglycerides) ($P<0.001$) and sex hormone level (FSH, LH, progesterone and testosterone) ($P<0.001$).

Association studies

Table 2 summarizes the association between the T2D-GWAS SNPs and PCOS in case–control subjects, under log-additive, additive, dominant and recessive genetic models. No significant deviation of allele frequencies from the HWE was found in PCOS and control groups ($P > 0.05$).

After adjustment for the covariates, age, BMI and lipid profiles (total cholesterol, HDL, LDL and triglycerides), the association between rs792837 of *COL8A1* gene and PCOS among case and control subjects was examined and the findings are summarized in table 2. The allele frequencies of the rs792837 were comparable between the PCOS and healthy controls ($P = 0.18$, OR (95% CI) = 1.24 (0.91–1.69)).

Table 2 summarizes the results of association between rs792837 *COL8A1* genotypes and PCOS, under log-additive, additive, dominant and recessive genetic models adjusted for age, BMI and lipid profiles as dependent covariates. We found that the T/C genotype increased PCOS risk under the additive model ($P = 0.008$, OR (95% CI) = 1.96 (1.19–3.23)), same as for the T/C+C/C genotypes under the dominant model ($P = 0.018$, OR (95% CI) = 1.75 (1.10–2.80)).

Table 2 summarizes the association between rs61873498 *KCNQ1* and PCOS risk among case–control subjects. We showed a significant association between *KCNQ1* rs61873498 and PCOS risk. The differences of rs61873498 allele frequencies were statistically significant between the PCOS and healthy controls ($P = 0.004$, OR (95% CI) = 1.62 (1.16–2.25)). We reported that the G/G genotype of rs61873498 under the additive model ($P = 0.002$, OR (95% CI) = 3.39 (1.53–7.48)) as well as under the recessive model ($P = 0.002$, OR (95% CI) = 3.04 (1.42–6.51)) increased clearly the risk of PCOS.

Two common variants of *SLC30A8* were analysed in association with PCOS risk among case–control subjects as shown in table 2. The allele frequencies of the two SNPs

Table 2. Associations between PCOS and the T2D GWAS SNPs.

Gene/SNP	Genotype	Control/PCOS $n = (162/162)$	Additive model[a] OR (95% CI)	$P^{\dagger}$	Dominant model[b] OR (95% CI)	$P^{\ddagger}$	Recessive model[c] OR (95% CI)	P^{*}
CDKAL1	GG (1/1)	93/82	Reference		Reference		Reference	
rs7754840	GC (1/2)	61/67	1.25 (0.79–1.97)	0.340	1.31 (0.85–2.04)	0.220		
	CC (2/2)	8/13	1.84 (0.73–4.67)	0.190			1.68 (0.68–4.17)	0.260
	MAF (C)	0.24/0.29	1.29 (0.91–1.84)	0.150#				
	HWE-*P*	0.83/1.00						
CDKAL1	AA (1/1)	86/82	Reference		Reference		Reference	
rs7756992	AG (1/2)	66/67	1.06 (0.68–1.68)	0.790	1.10 (0.71–1.71)	0.660		
	GG (2/2)	10/13	1.36 (0.57–3.28)	0.490			1.33 (0.56–3.12)	0.520
	MAF (G)	0.27/0.29	1.11 (0.79–1.57)	0.540#				
	HWE-*P*	0.69/1.00						
CDKN2B	TT (1/1)	97/102	Reference		Reference		Reference	
rs1081161	TC (1/2)	54/56	0.99 (0.62–1.57)	0.950	0.88 (0.56–1.37)	0.570		
	CC (2/2)	11/4	0.35 (0.11–1.12)	0.070			0.35 (0.11–1.12)	0.060
	MAF (C)	0.22/0.20	0.80 (0.55–1.17)	0.250#				
	HWE-*P*	0.38/0.33						
COL8A1	TT (1/1)	64/44	Reference		Reference		Reference	
rs792837	TC (1/2)	66/89	1.96 (1.19–3.23)	0.008	1.75 (1.10–2.80)	0.0180		
	CC (2/2)	32/29	1.32 (0.70–2.48)	0.390			0.89 (0.51–1.55)	0.670
	MAF (C)	0.40/0.45	1.24 (0.91–1.69)	0.180#				
	HWE-*P*	0.052/0.21						
HHEX	CC (1/1)	72/73	Reference		Reference		Reference	
rs1111875	CT (1/2)	76/77	1.00 (0.63–1.57)	0.990	0.98 (0.63–1.51)	0.910		
	TT (2/2)	14/12	0.85 (0.37–1.95)	0.690			0.89 (0.51–1.55)	0.680
	MAF (T)	0.32/0.31	0.96 (0.69–1.33)	0.800#				
	HWE-*P*	0.37/0.20						
IGF2BP2	GG (1/1)	81/79	Reference		Reference		Reference	
rs4402960	GT (1/2)	63/62	1.01 (0.63–1.61)	0.970	1.05 (0.68–1.62)	0.820		
	TT (2/2)	18/21	1.20 (0.59–2.41)	0.620			1.19 (0.61–2.33)	0.610
	MAF (T)	0.31/0.32	1.07 (0.77–1.50)	0.670#				
	HWE-*P*	0.27/0.15						

Table 2 (*contd*)

Gene/SNP	Genotype	Control/PCOS $n = (162/162)$	Additive model[a] OR (95% CI)	$P^{\dagger}$	Dominant model[b] OR (95% CI)	$P^{\ddagger}$	Recessive model[c] OR (95% CI)	P^{*}
IGF2BP2	AA (1/1)	70/56	Reference		Reference		Reference	
rs1470579	AC (1/2)	67/85	1.59 (0.99–2.55)	0.060	1.44 (0.92–2.26)	0.110		
	CC (2/2)	25/21	1.05 (0.53–2.07)	0.890			0.82 (0.44–1.53)	0.520
	MAF (C)	0.36/0.39	1.14 (0.83–1.57)	0.420[#]				
	HWE-P	0.18/0.25						
KCNJ11	CC (1/1)	66/59	Reference		Reference		Reference	
rs5219	CT (1/2)	82/82	1.12 (0.70–1.78)	0.640	1.20 (0.77–1.88)	0.420		
	TT (2/2)	14/21	1.68 (0.78–3.60)	0.180			1.57 (0.77–3.22)	0.210
	MAF (T)	0.34/0.38	1.21 (0.87–1.66)	0.250[#]				
	HWE-*P*	0.37/0.20						
KCNQ1	CC (1/1)	84/67	Reference		Reference		Reference	
rs61873498	CG (1/2)	68/68	1.25 (0.79–2.00)	0.340	1.53 (0.98–2.37)	0.060		
	GG (2/2)	10/27	3.39 (1.53–7.48)	0.002			3.04 (1.42–6.51)	0.002
	MAF (G)	0.27/0.38	1.62 (1.16–2.25)	0.004[#]				
	HWE-*P*	0.55/0.18						
SLC30A8	CC (1/1)	58/52	Reference		Reference		Reference	
rs13266634	CT (1/2)	81/71	0.98 (0.60–1.60)	0.930	1.18 (0.74–1.87)	0.480		
	TT (2/2)	23/39	1.89 (1.00–3.58)	0.048			1.92 (1.08–3.39)	0.020
	MAF (T)	0.39/0.46	1.32 (0.97–1.80)	0.080[#]				
	HWE-*P*	0.62/0.15						
SLC30A8	GG (1/1)	105/109	Reference		Reference		Reference	
rs3802177	GA (1/2)	55/51	1.06 (0.68–1.68)	0.630	0.90 (0.57–1.42)	0.640		
	AA (2/2)	2/2	1.36 (0.57–3.28)	0.970			1.00 (0.14–7.19)	1.00
	MAF (A)	0.18/0.17	1.11 (0.79–1.57)	0.680[#]				
	HWE-*P*	0.11/0.17						

n, Number of total subjects; MAF, minor-allele frequency; HWE-*P*, Hardy–Weinberg equilibrium *P*–value. †,‡,*Genotype specific *P*-values and OR are adjusted for age, BMI and lipid profiles in each additive, dominant or recessive genetic model, respectively. [#]Allele-specific *P*-values and OR of the log-additive genetic model are adjusted for age, BMI and lipid profiles. [a]Genetic additive model: 1/1 versus 1/2, 2/2 genotypes; [b]genetic dominant model: 1/1 versus 1/2+2/2 genotypes and [c]genetic recessive model: 1/1+1/2 versus 2/2 genotypes.

Table 3. Minimum effect size detected with a statistical power of 80% in the study sample.

Gene name	SNP rs ID	Log-additive model OR (PCOS)	Additive model OR (PCOS)	Dominant model OR (PCOS)	Recessive model OR (PCOS)
CDKAL1	rs7754840	1.45	1.52	1.78	2.55
CDKAL1	rs7756992	1.45	1.52	1.78	2.55
CDKN2B	rs1081161	1.51	1.57	1.73	3.55
COL8A1	rs792837	1.42	1.52	2.18	1.94
HHEX	rs1111875	1.44	1.52	1.80	2.43
IGF2BP2	rs4402960	1.44	1.51	1.82	2.38
IGF2BP2	rs1470579	1.43	1.51	1.97	2.09
KCNJ11	rs5219	1.43	1.51	1.94	2.12
KCNQ1	rs61873498	1.43	1.51	1.94	2.12
SLC30A8	rs13266634	1.42	1.53	2.23	1.92
SLC30A8	rs3802177	1.54	1.60	1.73	4.19

were significantly comparable between the PCOS patients and healthy controls ($P = 0.08$, OR (95% CI) = 1.32 (0.97–1.80)) and ($P = 0.68$, OR (95% CI)=0.92 (0.61–1.38)), respectively. The T/T genotype of the rs13266634 showed, under the additive model ($P = 0.048$, OR (95% CI) = 1.89 (1.00–3.58)) and under the recessive model ($P = 0.02$, OR (95% CI) = 1.92 (1.08–3.39)), a significant difference between the PCOS patients and control subjects as given in table 2. However, the frequency distributions of the remaining variants were comparable between cases and controls.

We estimated the minimum effect sizes detectable with a statistical power of 80% under different genetic models, and according to the allelic frequencies of each SNP tested in our study, as indicated in table 3. Except for *COL8A1* rs792837, *KCNQ* rs61873498 and *SLC30A8* rs13266634 SNPs that showed a significant effect on PCOS risk with an additive OR value above the threshold estimated in our power calculation, the other eight SNPs are known to have a lower allelic contribution that could not be easily detectable in this middle-sized cohort (table 3).

Correlation studies

Association analyses of the 11 SNPs against quantitative traits were conducted in the PCOS group (table 4). Unfortunately, none of the SNPs illustrated significant association with the quantitative traits after multiple test corrections.

Discussion

To date a number of genes have been reported to be associated with PCOS. Most of these genes are investigated due to relevant pathogenesis of PCOS based on their functions. However, the conclusion of the pathogenesis of PCOS is still controversial.

The present study examined at the differential contributions of T2D susceptibility loci identified through GWAS on PCOS risk and its accompanying features in women of northern region of Saudi Arabia. Our data showed that three genetic variants: rs792837 (*COL8A1*), rs61873498 (*KCNQ1*) and rs13266634 (*SLC30A8*) of 11 SNPs in eight genes pointed out from GWAS-T2D loci were associated with PCOS. However, the distribution frequencies of the remaining variants were comparable between cases and controls.

PCOS shares a number of features with T2D (Bhattacharya 2008) including insulin resistance, beta cell dysfunction, impaired glucose tolerance and dyslipidaemia (Norman *et al.* 2007; Li and Baek 2015).

Novel T2D susceptibility loci within or close to the genes encoding: *CDKAL1*, *CDKN2B*, *COL8A1*, *HHEX*, *IGF2BP2*, *KCNJ11*, *KCNQ1* and *SLC30A8* have identified in GWAS and confirmed in large meta-analyses (Frayling 2007; Zeggini *et al.* 2008; Voight *et al.* 2010; Kim *et al.* 2012) to be associated with the insulin resistant trait of PCOS, not clearly with overall phenotype. Among the identified loci, replication studies for PCOS have been reported in *SLC30A8* and *KCNJ11* with conflicting results (Li *et al.* 2009; Ewens *et al.* 2011; Kim *et al.* 2012). These studies were also replicated in several populations with T2D (Lee *et al.* 2009) and gestational diabetes (Cho *et al.* 2009). Recently, Reddy *et al.* (2016) reported negative association of these T2D genes in PCOS aetiology among women from southern India.

Kim *et al.* (2012) investigated whether rs13266634 (*SLC30A8*) was associated with PCOS. Inconsistent with the results of our study, which showed significant association of this SNP with the development of PCOS and the increased levels of prolactin, these authors found that none of the studied allelic variants were associated with PCOS (Kim *et al.* 2012).

Moreover, several studies have identified and confirmed rs13266634 *SLC30A8* as a risk locus for T2D in European ancestry populations (Saxena *et al.* 2007; Sladek *et al.* 2007; Xiang *et al.* 2008). In our study, we found that rs13266634 has no evident association with PCOS-related clinical features. Our negative result may not be attributed

Table 4. Linear regression analysis between T2D-*GWAS* SNPs with PCOS features.

Trait	Gene	SNP	Effect allele	*P* value	*r*
Fasting glucose	*CDKAL1*	rs7754840	C	0.503	− 0.147
	CDKAL1	rs7756992	G	0.782	− 0.059
	CDKN2B	rs1081161	C	0.725	− 0.077
	COL8A1	rs792837	C	0.406	0.180
	HHEX	rs1111875	T	0.586	− 0.114
	IGF2BP2	rs4402960	T	0.855	0.065
	IGF2BP2	rs1470579	C	0.770	0.061
	KCNJ11	rs5219	T	0.846	0.042
	KCNQ1	rs61873498	G	0.124	− 0.325
	SLC30A8	rs13266634	T	0.234	− 0.257
	SLC30A8	rs3802177	A	0.974	0.007
Free insulin	*CDKAL1*	rs7754840	C	0.971	− 0.012
	CDKAL1	rs7756992	G	0.946	− 0.021
	CDKN2B	rs1081161	C	0.947	− 0.021
	COL8A1	rs792837	C	0.158	0.450
	HHEX	rs1111875	T	0.197	− 0.396
	IGF2BP2	rs4402960	T	0.947	0.034
	IGF2BP2	rs1470579	C	0.551	0.184
	KCNJ11	rs5219	T	0.793	0.083
	KCNQ1	rs61873498	G	0.674	− 0.130
	SLC30A8	rs13266634	T	0.316	− 0.318
	SLC30A8	rs3802177	A	0.310	0.322
HOMA-IR	*CDKAL1*	rs7754840	C	0.727	0.155
	CDKAL1	rs7756992	G	0.789	0.116
	CDKN2B	rs1081161	C	0.796	0.115
	COL8A1	rs792837	C	0.228	− 0.529
	HHEX	rs1111875	T	0.184	0.563
	IGF2BP2	rs4402960	T	0.824	0.093
	IGF2BP2	rs1470579	C	0.661	− 0.187
	KCNJ11	rs5219	T	0.722	− 0.155
	KCNQ1	rs61873498	G	0.569	0.242
	SLC30A8	rs13266634	T	0.269	0.484
	SLC30A8	rs3802177	A	0.564	− 0.252
Cholesterol	*CDKAL1*	rs7754840	C	0.506	− 0.053
	CDKAL1	rs7756992	G	0.330	− 0.077
	CDKN2B	rs1081161	C	0.765	− 0.024
	COL8A1	rs792837	C	0.468	− 0.057
	HHEX	rs1111875	T	0.249	− 0.091
	IGF2BP2	rs4402960	T	0.452	0.060
	IGF2BP2	rs1470579	C	0.222	− 0.096
	KCNJ11	rs5219	T	0.137	− 0.117
	KCNQ1	rs61873498	G	0.969	− 0.003
	SLC30A8	rs13266634	T	0.988	− 0.001
	SLC30A8	rs3802177	A	0.917	0.008
Triglycerides	*CDKAL1*	rs7754840	C	0.421	0.069
	CDKAL1	rs7756992	G	0.480	0.059
	CDKN2B	rs1081161	C	0.329	− 0.084
	COL8A1	rs792837	C	0.886	− 0.012
	HHEX	rs1111875	T	0.131	− 0.123
	IGF2BP2	rs4402960	T	0.966	0.069
	IGF2BP2	rs1470579	C	0.263	− 0.092
	KCNJ11	rs5219	T	0.290	− 0.185
	KCNQ1	rs61873498	G	0.083	− 0.143
	SLC30A8	rs13266634	T	0.835	0.018
	SLC30A8	rs3802177	A	0.850	0.016
HDL	*CDKAL1*	rs7754840	C	0.793	0.023
	CDKAL1	rs7756992	G	0.135	0.127
	CDKN2B	rs1081161	C	0.367	0.078
	COL8A1	rs792837	C	0.776	− 0.024
	HHEX	rs1111875	T	0.064	− 0.165

Table 4 (*contd*)

Trait	Gene	SNP	Effect allele	*P* value	*r*
	IGF2BP2	rs4402960	T	0.086	0.166
	IGF2BP2	rs1470579	C	0.847	0.016
	KCNJ11	rs5219	T	0.533	− 0.053
	KCNQ1	rs61873498	G	0.445	0.064
	SLC30A8	rs13266634	T	0.776	− 0.024
	SLC30A8	rs3802177	A	0.491	− 0.059
LDL	*CDKAL1*	rs7754840	C	0.729	− 0.032
	CDKAL1	rs7756992	G	0.221	− 0.109
	CDKN2B	rs1081161	C	0.574	0.051
	COL8A1	rs792837	C	0.117	− 0.141
	HHEX	rs1111875	T	0.564	− 0.050
	IGF2BP2	rs4402960	T	0.364	0.046
	IGF2BP2	rs1470579	C	0.350	− 0.082
	KCNJ11	rs5219	T	0.167	− 0.125
	KCNQ1	rs61873498	G	0.095	0.147
	SLC30A8	rs13266634	T	0.611	0.046
	SLC30A8	rs3802177	A	0.854	0.016
FSH	*CDKAL1*	rs7754840	C	0.336	0.085
	CDKAL1	rs7756992	G	0.101	0.141
	CDKN2B	rs1081161	C	0.805	− 0.021
	COL8A1	rs792837	C	0.355	0.080
	HHEX	rs1111875	T	0.361	0.076
	IGF2BP2	rs4402960	T	0.987	0.010
	IGF2BP2	rs1470579	C	0.933	0.007
	KCNJ11	rs5219	T	0.757	0.027
	KCNQ1	rs61873498	G	0.844	0.017
	SLC30A8	rs13266634	T	0.597	0.046
	SLC30A8	rs3802177	A	0.263	− 0.097
LH	*CDKAL1*	rs7754840	C	0.727	0.034
	CDKAL1	rs7756992	G	0.730	0.032
	CDKN2B	rs1081161	C	0.589	0.052
	COL8A1	rs792837	C	0.236	− 0.113
	HHEX	rs1111875	T	0.389	− 0.080
	IGF2BP2	rs4402960	T	0.633	0.010
	IGF2BP2	rs1470579	C	0.220	0.114
	KCNJ11	rs5219	T	0.744	− 0.031
	KCNQ1	rs61873498	G	0.125	0.144
	SLC30A8	rs13266634	T	0.507	− 0.064
	SLC30A8	rs3802177	A	0.896	− 0.013
Progesterone	*CDKAL1*	rs7754840	C	0.358	− 0.092
	CDKAL1	rs7756992	G	0.528	− 0.061
	CDKN2B	rs1081161	C	0.599	− 0.052
	COL8A1	rs792837	C	0.483	0.069
	HHEX	rs1111875	T	0.512	0.062
	IGF2BP2	rs4402960	T	0.146	0.001
	IGF2BP2	rs1470579	C	0.504	− 0.064
	KCNJ11	rs5219	T	0.404	0.082
	KCNQ1	rs61873498	G	0.230	0.115
	SLC30A8	rs13266634	T	0.061	− 0.185
	SLC30A8	rs3802177	A	0.209	− 0.124
Oestradiol	*CDKAL1*	rs7754840	C	0.656	− 0.038
	CDKAL1	rs7756992	G	0.306	− 0.086
	CDKN2B	rs1081161	C	0.890	0.011
	COL8A1	rs792837	C	0.226	− 0.103
	HHEX	rs1111875	T	0.052	0.159
	IGF2BP2	rs4402960	T	0.303	0.000
	IGF2BP2	rs1470579	C	0.116	0.130
	KCNJ11	rs5219	T	0.644	0.039
	KCNQ1	rs61873498	G	0.189	0.109
	SLC30A8	rs13266634	T	0.979	− 0.002
	SLC30A8	rs3802177	A	0.181	− 0.113

Table 4 (*contd*)

Trait	Gene	SNP	Effect allele	*P* value	*r*
Prolactin	*CDKAL1*	rs7754840	C	0.744	0.029
	CDKAL1	rs7756992	G	0.340	0.082
	CDKN2B	rs1081161	C	0.851	− 0.016
	COL8A1	rs792837	C	0.538	− 0.053
	HHEX	rs1111875	T	0.372	− 0.075
	IGF2BP2	rs4402960	T	0.210	0.004
	IGF2BP2	rs1470579	C	0.500	0.242
	KCNJ11	rs5219	T	0.768	0.025
	KCNQ1	rs61873498	G	0.319	0.084
	SLC30A8	rs13266634	T	0.340	0.083
	SLC30A8	rs3802177	A	0.144	0.127
Testosterone	*CDKAL1*	rs7754840	C	0.967	− 0.004
	CDKAL1	rs7756992	G	0.940	− 0.006
	CDKN2B	rs1081161	C	0.810	0.021
	COL8A1	rs792837	C	0.469	− 0.064
	HHEX	rs1111875	T	0.206	0.108
	IGF2BP2	rs4402960	T	0.426	0.002
	IGF2BP2	rs1470579	C	0.100	0.142
	KCNJ11	rs5219	T	0.996	0.000
	KCNQ1	rs61873498	G	0.144	− 0.126
	SLC30A8	rs13266634	T	0.232	− 0.106
	SLC30A8	rs3802177	A	0.617	− 0.044

r, Spearman correlation (2-tailed).

to the ethnic stratifications because the association of this variant with T2D has also been observed in the Chinese population (Scott *et al.* 2007).

We also investigated whether any of the SNPs are associated with beta-cell function as measured by HOMA-IR. In our finding the rs10811661, SNP near *CDKN2B* was not associated either with PCOS or with HOMA-IR which is in disagreement with Ewens and collaborators report. They suggested a role for *CDKN2B* locus in the metabolic abnormalities in PCOS, although it evidently does not contribute to the reproductive phenotype (Ewens *et al.* 2011).

In this present study, we tested whether novel risk loci for T2D, rs7756992 and rs7752842 in gene *CDKAL1* may also have contributed to PCOS susceptibility. We found that neither of these SNPs are associated with PCOS nor PCOS-related endocrine and metabolic clinical features in Saudi Arabian women in agreement with results reported in Han Chinese women on the same (Liu *et al.* 2010). The variant within this gene is correlated with insulin secretion although its function remains to be elucidated. *CDKAL1* may facilitate insulin production under glucotoxic conditions through interaction with *CDK5* (Steinthorsdottir *et al.* 2007).

Our findings showed that *KCNJ11* variant was not associated with PCOS. Given the central role of insulin to PCOS pathogenesis, genes such as *KCNJ11*, whose products influence either the secretion or action of insulin, represent important potential candidates. Previous comprehensive tagging studies have demonstrated that the common variant *E23K* (rs5219) is heavily implicated in susceptibility to multifactorial T2D, which is in epidemiological overlap between the PCOS and T2D (thereby obviating the need for retagging of *KCNJ11* in PCOS) (Barber *et al.* 2007; Ewens *et al.* 2011; Saxena and Welt 2013). However, a nominal association at rs5219 in *KCNJ11* was demonstrated in a smaller association study in which PCOS subjects were not defined based on Rotterdam criteria (Ewens *et al.* 2011).

The *HHEX* gene represents the strongest biological candidate given postulated effects on both insulin signalling and islet function. We tested the relationship between *HHEX* rs1111875 and the clinical findings and PCOS. Our results indicated that there was no association of *HHEX* with PCOS occurrence or other relevant clinical manifestations. Our observation supported the notion that *HHEX* rs1111875 cannot be regarded as a fundamental candidate gene for PCOS susceptibility which is in full agreement with Barber *et al.* (2007) study. However, an association of the combined GA+AA genotype of rs1111875 with a reduced of IR in non-PCOS subjects was reported. It has been postulated that the protective effect of *HHEX* allele variants on the risk of IR is masked or prohibited by other alterations in beta-cell function (Tehrani *et al.* 2015).

Since the first report on GWAS in 2005, more than 2000 studies have been added to the catalogue of published GWAS. To date, five GWAS have been performed on women with PCOS, which were conducted in Han Chinese, Korean and European ancestry populations. From these studies, 15 risk loci (*P*-values $<10^{-8}$) have been identified and they have provided new clues for understanding

the genetic components and cellular pathways in PCOS (Zhao *et al.* 2016). Studies using GWAS for PCOS are still in their infancy, and subsequent studies shall be performed to test the novel loci that are found from GWAS. At the same time, GWAS shall be conducted in Arabic populations, which based on their specific genetic background might reveal novel loci for PCOS.

In conclusion, our results pointed out to the evidence that variants rs792837 *COL8A1*, rs61873498 *KCNQ1* and rs13266634 *SLC30A8* genes contributed significantly to the occurrence of PCOS among Saudi Arabian women. However, the frequency distributions of the remaining variants were comparable between cases and controls which might not be prominent genes involved in PCOS pathogenicity.

Acknowledgements

We gratefully acknowledge all of the patients and volunteers for their participation. We also thank the nurses and physicians for their contribution to the collection of all samples in King Salman Military Hospital of Tabuk, Saudi Arabia. This work was supported in part by a grant from National Applied Research Program of King Abdul-Aziz City for Science and Technology, KSA (KACST: AC-34-156), and the Deanship of Scientific Research, University of Tabuk, KSA (DSR: S1436-0288). All funding sources were independent and had no influence on the study design, the data extraction, analyses, interpretation of the data, writing of this article, or in the decision to submit the article for publication.

References

Alsibyani M., Derham A., Jazzar M., Alnefayi R. A. and Alnefaie A. S. 2017 Clinical presentation of polycystic ovary syndrome among Saudi Arabian women – Jeddah – Saudi Arabia. *Int. J. Adv. Res.* **5**, 1872–1876.

Barber T. M., Bennett A. J., Gloyn A. L., Groves C. J., Sovio U., Ruokonen A. *et al.* 2007 Relationship between E23 K (an established type II diabetes-susceptibility variant within KCNJ11), polycystic ovary syndrome and androgen levels. *Eur. J. Hum. Genet.* **15**, 679–684.

Bhattacharya S. M. 2008 Metabolic syndrome in females with polycystic ovary syndrome and International Diabetes Federation criteria. *J. Obstet. Gynaecol. Res.* **34**, 62–66.

Chen Z. J., Zhao H., He L., Shi Y., Qin Y., Shi Y. *et al.* 2011 Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat. Genet.* **43**, 55–59.

Cho Y. M., Kim T. H., Lim S., Choi S. H., Shin H. D., Lee H. K. *et al.* 2009 Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. *Diabetologia* **52**, 253–261.

Diamanti-Kandarakis E., Piperi C., Spina J., Argyrakopoulou G., Papanastasiou L., Bergiele A. *et al.* 2006 Polycystic ovary syndrome: the influence of environmental and genetic factors. *Hormones (Athens)* **5**, 17–34.

Ewens K. G., Jones M. R., Ankener W., Stewart D. R., Urbanek M., Dunaif A. *et al.* 2011 Type 2 diabetes susceptibility single-nucleotide polymorphisms are not associated with polycystic ovary syndrome. *Fertil. Steril.* **95**, 2538–2541, e1–6.

Frayling T. M. 2007 Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat. Rev. Genet.* **8**, 657–662.

Goodarzi M. O., Dumesic D. A., Chazenbalk G. and Azziz R. 2011 Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat. Rev. Endocrinol.* **7**, 219–231.

Hughes C., Elgasim M., Layfield R. and Atiomo W. 2006 Genomic and post-genomic approaches to polycystic ovary syndrome – progress so far: mini review. *Hum. Reprod.* **21**, 2766–2775.

Jalilian A., Kiani F., Sayehmiri F., Sayehmiri K., Khodaee Z. and Akbari M. 2015 Prevalence of polycystic ovary syndrome and its associated complications in Iranian women: a meta-analysis. *Iran J. Reprod. Med.* **13**, 591–604.

Jones M. R. and Goodarzi M. O. 2016 Genetic determinants of polycystic ovary syndrome: progress and future directions. *Fertil. Steril.* **106**, 25–32.

Kim J. J., Choi Y. M., Cho Y. M., Hong M. A., Chae S. J., Hwang K. R. *et al.* 2012 Polycystic ovary syndrome is not associated with polymorphisms of the TCF7L2, CDKAL1, HHEX, KCNJ11, FTO and SLC30A8 genes. *Clin. Endocrinol. (Oxf)* **77**, 439–445.

Lee Y. H., Martin J. M., Maple R. L., Tharp W. G. and Pratley R. E. 2009 Plasma amyloid-beta peptide levels correlate with adipocyte amyloid precursor protein gene expression in obese individuals. *Neuroendocrinology* **90**, 383–390.

Li L. and Baek K. H. 2015 Molecular genetics of polycystic ovary syndrome: an update. *Curr. Mol. Med.* **15**, 331–342.

Li L., Zhao Y., Wang L., Yan J. and Chen Z. J. 2009 Genetic variations of solute carrier family 30 (zinc transporter) member 8 (SLC30A8) are not associated with polycystic ovary syndrome. *Fertil. Steril.* **91**, 1598–1601.

Li T., Zhao H., Zhao X., Zhang B., Cui L., Shi Y. *et al.* 2012 Identification of YAP1 as a novel susceptibility gene for polycystic ovary syndrome. *J. Med. Genet.* **49**, 254–257.

Liu X., Li L., Chen Z. J., Lu Z., Shi Y. and Zhao Y. 2010 Genetic variants of cyclin-dependent kinase 5 regulatory subunit associated protein 1-like 1 and transcription factor 7-like 2 are not associated with polycystic ovary syndrome in Chinese women. *Gynecol. Endocrinol.* **26**, 129–134.

Liu H., Zhao H. and Chen Z. J. 2016 Genome-wide association studies for polycystic ovary syndrome. *Semin. Reprod. Med.* **34**, 224–229.

Norman R. J., Dewailly D., Legro R. S. and Hickey T. E. 2007 Polycystic ovary syndrome. *Lancet* **370**, 685–697.

Reddy B. M., Kommoju U. J., Dasgupta S. and Rayabarapu P. 2016 Association of type 2 diabetes mellitus genes in polycystic ovary syndrome aetiology among women from southern India. *Indian J. Med. Res.* **144**, 400–408.

Rotterdam E. A.-S. P. C. W. G. 2004 Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil. Steril.* **81**, 19–25.

Saxena R. and Welt C. K. 2013 Polycystic ovary syndrome is not associated with genetic variants that mark risk of type 2 diabetes. *Acta Diabetol.* **50**, 451–457.

Saxena R., Voight B. F., Lyssenko V., Burtt N. P., de Bakker P. I., Chen H. *et al.* 2007 Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336.

Scott L. J., Mohlke K. L., Bonnycastle L. L., Willer C. J., Li Y., Duren W. L. *et al.* 2007 A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345.

Sladek R., Rocheleau G., Rung J., Dina C., Shen L., Serre D. *et al.* 2007 A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885.

Steinthorsdottir V., Thorleifsson G., Reynisdottir I., Benediktsson R., Jonsdottir T., Walters G. B. *et al.* 2007 A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775.

Tehrani F. R., Zarkesh M., Tohidi M., Azizi F. and Zadeh-Vakili A. 2015 Is the association between insulin resistance and diabetogenic haematopoietically expressed homeobox (HHEX) polymorphism (rs1111875) affected by polycystic ovary syndrome status? *Reprod. Fertil. Dev.* **29**, 670–678.

Trikudanathan S. 2015 Polycystic ovarian syndrome. *Med. Clin. North Am.* **99**, 221–235.

Vermeulen A., Verdonck L. and Kaufman J. M. 1999 A critical evaluation of simple methods for the estimation of free testosterone in serum. *J. Clin. Endocrinol. Metab.* **84**, 3666–3672.

Voight B. F., Scott L. J., Steinthorsdottir V., Morris A. P., Dina C., Welch R. P. *et al.* 2010 Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589.

Xiang J., Li X. Y., Xu M., Hong J., Huang Y., Tan J. R. *et al.* 2008 Zinc transporter-8 gene (SLC30A8) is associated with type 2 diabetes in Chinese. *J. Clin. Endocrinol. Metab.* **93**, 4107–4112.

Zeggini E., Scott L. J., Saxena R., Voight B. F., Marchini J. L., Hu T. *et al.* 2008 Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645.

Zhao H., Lv Y., Li L. and Chen Z. J. 2016 Genetic studies on polycystic ovary syndrome. *Best Pract. Res. Clin. Obstet. Gynaecol.* **37**, 56–65.

Corresponding editor: H. A. Ranganath