

SCIENTIFIC REPORTS

OPEN

Genetic risk variants for metabolic traits in Arab populations

Prashantha Hebbar[†], Naser Elkum[†], Fadi Alkayal, Sumi Elsa John, Thangavel Alphonse Thanaraj* & Osama Alsmadi*

Received: 27 April 2016

Accepted: 14 December 2016

Published: 20 January 2017

Despite a high prevalence of metabolic trait related diseases in Arabian Peninsula, there is a lack of convincingly identified genetic determinants for metabolic traits in this population. Arab populations are underrepresented in global genome-wide association studies. We genotyped 1965 unrelated Arab individuals from Kuwait using Cardio-MetaboChip, and tested SNP associations with 13 metabolic traits. Models based on recessive mode of inheritance identified Chr15:40531386-rs12440118/*ZNF106/W*->R as a risk variant associated with glycated-hemoglobin at close to 'genome-wide significant' p-value and five other risk variants 'nominally' associated (p-value $\leq 5.45E-07$) with fasting plasma glucose (rs7144734/[*OTX2-AS1,RPL3P3*]) and triglyceride (rs17501809/*PLGRKT*; rs11143005/*LOC105376072*; rs900543/[*THSD4,NR2E3*]; and Chr12:101494770/*IGF1*). Furthermore, we identified 33 associations (30 SNPs with 12 traits) with 'suggestive' evidence of association (p-value $< 1.0E-05$); 20 of these operate under recessive mode of inheritance. Two of these 'suggestive' associations (rs1800775-*CETP/HDL*; and rs9326246-*BUD13/TGL*) showed evidence at genome-wide significance in previous studies on Euro-centric populations. Involvement of many of the identified loci in mediating metabolic traits was supported by literature evidences. The identified loci participate in critical metabolic pathways (such as Ceramide signaling, and Mitogen-Activated Protein Kinase/ Extracellular Signal Regulated Kinase signaling). Data from Genotype-Tissue Expression database affirmed that 7 of the identified variants differentially regulate the up/downstream genes that mediate metabolic traits.

Prevalence of obesity and Type 2 diabetes (T2DM) is enormously rising in the Arabian Peninsula¹. Rapid transitions in diet practices and lifestyle factors have resulted in high incidence of lifestyle disorders in the Middle East, making the region an epicenter of the escalating epidemics of diabetes. Our recent study using nationwide health data estimated the prevalence of T2DM to be 25% among Kuwaiti natives². Till date, not many genome-wide association (GWA) studies have been conducted on populations from the Arabian Peninsula. Few that have been reported include: (i) a T2DM case-control study on Saudi Arabian population that validated only some (*WFS1, JAZF1, SLC30A8, CDKN2A/B, TCF7L2, KCNQ1, HMG20A, HNF4A* and *DUSP9*) of the T2DM genes implicated in European population³; however, the study was not adjusted for population stratification. (ii) a T2DM case-control study on Lebanese population that replicated only two (*CDKAL1* and *TCF7L2*) of the genes implicated in European population⁴; and (iii) a study from the United Arab Emirates⁵ on an extended pedigree (consisting of 178 Arab individuals) that implicated association of *KCTD8, COX7B2* and *GABRA4* genes with T2DM for the first time; the implicated markers have not so far been replicated in any other population.

Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations⁶. A significant portion of the ancestry of indigenous Arabs can be traced back to ancient lineages of the Arabian Peninsula^{6,7}. These observations have implications for disease genetics studies in the region - for complex diseases such as T2DM, the associated genetic variants in Arab populations would not *a priori* be expected to be the same (or to exert similar effect) as those discovered in European populations. For example, although rs7903146 and rs12255372 variants in *TCF7L2* gene show strong associations with T2DM in most populations, they show weak or no association in Arab populations⁸.

Furthermore, there exists a high level of inbreeding in the Arab region due to the practice of consanguineous marriages that are often between first cousins. This has led to excess of homozygosity for autosomal recessiveness^{9,10}. It is estimated that nearly two-thirds of genetically transmitted diseases in Arab patients follow an

Dasman Diabetes Institute, P.O. Box 1180, Dasman 15462, Kuwait. [†]These authors contributed equally to this work. *These authors jointly supervised this work. Correspondence and requests for materials should be addressed to T.A.T. (email: alphonse.thangavel@dasmaninstitute.org) or O.A. (email: oalsmadi@gmail.com)

autosomal recessive mode of inheritance¹¹. Consanguinity can exert an influence on the etiology of complex disorders if rare autosomal recessive alleles (as opposed to alleles that are common in the gene pool) are causally implicated¹². Therefore, it is possible to hypothesize that inbreeding helps in aggregating risk alleles and thereby in inducing possible cumulative recessive effects; such effects can lead to the elevation of risk factors, and subsequently to an increase in T2DM risk. Gosadi *et al.*¹³ reported a statistically significant inverse association between inbreeding coefficients and age at diagnosis of T2DM, and further observed that such an association is stronger in patients where both parents were afflicted with T2DM. For reasons mentioned above, transfer of knowledge on T2DM genetic markers derived by global genome surveys (on Euro-centric populations) to the Arab population can be complicated by differences in disease prevalence, risk factor profiles, genetic risk allele frequencies, gene-environment interactions, and increased homozygosity for autosomal recessive genes.

In this study, we explore the genome-wide genotypes from a cohort of Arab individuals from Kuwait for gene loci associated with 13 metabolic traits - namely height, weight, waist circumference (WC), waist circumference to height ratio (WcHtR), body mass index (BMI), glycated hemoglobin (HbA1c), fasting plasma glucose (FPG), triglyceride (TGL), high density lipoprotein (HDL), low density lipoprotein (LDL), total cholesterol (TC), systolic blood pressure (SBP), and diastolic blood pressure (DBP). For this purpose, we choose Cardio-MetaboChip genotyping array, which has successfully been employed for fine-mapping of established loci in populations such as Han Chinese¹⁴ and Africans¹⁵. We scrutinize about 45,793 variants related to metabolic traits in 1,965 Arab ethnic individuals. We report a number of risk variants that are associated with metabolic traits at varying significant levels of p-values. Furthermore, we examine the broader context in which these genetic variants operate to alter metabolic disease states. The results of this study that map disease alleles in the indigenous Arab population can provide important perspectives on the pathogenesis and diagnosis of metabolic trait associated diseases.

Results

Sample and marker sets. Upon performing sample-related quality control steps (such as identification of duplicate samples [leading to removal of $n = 1$ sample], assessment of samples for call rates $< 97\%$ [$n = 232$], assessment for gender mismatch [$n = 74$], identification of cryptic relatedness through IBD analysis [$n = 151$], and estimation of ancestry admixture to identify samples with ancestry mismatch [$n = 39$]), the sample size got reduced from 2,440 to 1,965. By performing marker-related quality control steps (such as $MAF < 0.01$, $HWE < 0.001$, genotype error < 0.1 , $mind < 0.1$ [leading to removal of $n = 68,582$ markers], LD-pruning [$n = 70,533$], and considering only those markers that possess a minimum of 5% minor allele frequency in the study population [$n = 24,740$]), we narrowed down the number of MetaboChip markers from 196,725 to 45,793. We used these filtered sample and marker sets to perform association tests between markers and quantitative phenotype traits.

The scatter plot representing the first two principal components of the sample set is presented in Supplementary Figure S1. The plot depicts the three genetic substructures of Kuwaiti population that we established in our previous study⁷.

Characteristics of study subjects. The demographic and biochemical characteristics of the study participants from the filtered set of 1,965 samples are shown in Table 1. The study population consisted largely of middle adulthood individuals (mean age = 44.88 ± 12.35 years) with an almost equal proportion of males to females (52%:48%). With a mean BMI of 31.53 ± 6.60 Kg/m², the population could largely be categorized as Class I obese; the observed high mean value for waist circumference (at 100.82 ± 14.41 cm) further reflected the extent of obesity. As regards the disease status, 27% of the participants were diabetic, 26% were hypertensive, and 9% suffered from cardiovascular disorders. The mean value for fasting plasma glucose (FPG) measurements for the study population at 6.37 ± 2.90 mmol/l indicated prediabetes; the mean value for HbA1c at 6.28 ± 1.68 mmol/l indicated increased risk of diabetes; the mean values for LDL at 3.25 ± 0.095 mmol/l, HDL at 1.13 ± 0.035 mmol/l, triglyceride (TGL) at 1.58 ± 1.04 mmol/l, and total cholesterol (TC) at 5.07 ± 1.07 mmol/l indicated normal or near optimal cholesterol levels. Mean values for systolic blood pressure (SBP) at 127.31 ± 17.06 mmHg and diastolic blood pressure (DBP) at 77.77 ± 10.73 mmHG indicated prehypertension.

Markers associated with phenotype traits either at 'close to genome-wide significant' p-values or at 'nominally significant' p-values. Results from tests of association pointed to only one risk variant (namely Chr15:40531386-rs12440118 from *ZNF106*) associated with HbA1c emerging at 'close to genome-wide significant' p-value of $3.41E-08$ (with Benjamini-Hochberg FDR p-value of 0.0032) (Table 2). Five other risk variants emerged as 'nominally' associated at the 'lenient' p-value threshold of $5.45E-07$ (and Benjamini-Hochberg FDR p-value ≤ 0.05); and they were seen associated with FPG (rs7144734/[*OTX2-AS1,RPL3P3*]) and TGL (rs17501809/*PLGRKT*, rs11143005/*LOC105376072*, rs900543/[*THSD4,NR2E3*], and Chr12:101494770/*IGF1*) (Table 2). The markers continued to be associated at significant p-values when the models were adjusted for medication status (Table 2) or for diabetes status or for obesity status (Table 3) - e.g. under adjustment for obesity status, the p-values remained significant at $\leq 8.8E-07$. All these 6 markers were identified by recessive models.

The Manhattan plots of the interrogated SNPs for the phenotype traits of HbA1c, FPG, and TGL are presented in Fig. 1. The Quantile-Quantile plots are shown in Supplementary Figure S2. The genomic-control inflation factors (which compare observed association statistics against the expected distributions) corresponding to the 13 tested traits are presented in Supplementary Table S1; values of ≤ 1.04 were obtained for these factors and hence it was not felt necessary to perform corrections for genomic-control inflation on association statistics. Regional plots, depicting trait association statistics for all the SNPs (typed in the MetaboChip) from a region of 1MB, for the 6 identified markers are shown in Supplementary Figure S3.

Results from power calculations suggested that the sample size of 1965 used in this study has the potential to detect up to 2% quantitative trait variance with 80% power for the SNPs with $MAF \geq 5\%$ at the claimed genome-wide significant p-value of $3.41E-08$. The sample sizes for the different R^2_G values are as follows (R^2_G ,

Characteristics	Study Participants
Sex (M:F)	1028:937
Age (in years \pm SD)	44.88 \pm 12.35
Height (cm \pm SD)	165.74 \pm 9.20
Weight (Kg \pm SD)	86.59 \pm 19.24
BMI (Kg/m ² \pm SD)	31.53 \pm 6.60
Waist circumference, WC (cm \pm SD)	100.82 \pm 14.41
Waist circumference to Height ratio, WcHtR (\pm SD)	0.602 \pm 0.111
Fasting plasma glucose, FPG (mmol/l \pm SD)	6.37 \pm 2.90
Glycated hemoglobin, HbA1c (mmol/l \pm SD)	6.28 \pm 1.68
Low-density lipoprotein, LDL (mmol/l \pm SD)	3.25 \pm 0.95
High-density lipoprotein, HDL (mmol/l \pm SD)	1.13 \pm 0.35
Triglyceride, TGL (mmol/l \pm SD)	1.58 \pm 1.04
Total Cholesterol, TC (mmol/l \pm SD)	5.07 \pm 1.07
Systolic blood pressure, SBP (mmHg \pm SD)	127.31 \pm 17.06
Diastolic blood pressure, DBP (mmHg \pm SD)	77.77 \pm 10.73
Obesity (Class I and more – i.e. BMI \geq 30 Kg/M ²) (case:control)	1062:903
Type 2 Diabetes (case:control)	536:1429
Hypertension (case:control)	513:1452
Cardiovascular disease, CVD (case:control)	185:1780
Lipid lowering medication [®] (Yes:No)	103:1862
Diabetes medication [®] (Yes:No)	249:1716
Hypertension medication [®] (Yes:No)	230:1735

Table 1. The demographic and biochemical characteristics of the study participants. [®]It reflects whether medication is administered around the time of participant registration; previous records of medication were not checked.

sample size): (0.015, 2677), (0.016, 2509), (0.017, 2360), (0.018, 2228), (0.019, 2109), (0.020, 2003), (0.021, 1907), (0.022, 1819), (0.023, 1739), (0.024, 1666), (0.025, 1598). The sample size of 1965 used in this study has the potential to pick up an effect size of <2.44 for HbA1c with the identified rs12440118 variant at 6% MAF (observed beta value = 2.00), <2.09 for FPG with the rs7144734 variant at 20% MAF (observed beta value = 1.465), <2.45 for TGL with the rs10860880 variant at MAF 6% (observed beta value = 1.59), <0.54 for TGL with the rs11143005 variant at 28% MAF (observed beta value = 0.4196), <1.625 for TGL with the rs900543 variant at 9% MAF (observed beta value = 1.625) at genome-wide significant p-value.

Upon searching the literature for evidences on the involvement of the identified gene loci (associated at either close to the stringent p-value threshold or at lenient p-value threshold) in metabolic trait related processes, the following affirmative observations were made:

Chr15:40531386 [ZNF106] with HbA1c. The variation leads to a non-synonymous amino acid change (W103R, aromatic to basic residue) in the encoded Zinc Finger Protein 106 (*ZNF106 alias ZFP106*). *ZNF106* is a product of *SIRM* gene and it shares the same initiation codon and differs in 3' UTR with another product, namely *SH3BP3*, of *SIRM*¹⁶; thus *ZNF106* is expected to share the same properties of *SIRM*. The *SIRM* protein has been described as a novel insulin-regulated SH3 binding protein that associates with *Grb2* and *FYN*¹⁷. It is further known that deregulation of *Slc25a20* gene, which is known to play an important role in insulin secretion, is accompanied by the upregulation of *ZNF106* in Type 1 diabetes¹⁸. These evidences shown in mice and cell lines surmise that *ZNF106* could be a potential causal gene in human too and could be implicated in human insulin receptor signaling pathway.

rs7144734 [OTX2-AS1, RPL3P3] with FPG. *OTX2 antisense RNA 1 (head to head)* is a transcription factor known to play an important role in controlling expression of *OTX2* gene. Although there is no direct association of *OTX* with glucose metabolism or diabetes, it is known to be involved in regulating gonadotrophin releasing hormone (*GnRH*) in hypogonadism. Interestingly, a study from the Middle East region found that 36.5% of 1089 men with T2DM had low serum testosterone levels; 17% of such T2DM patients with low serum testosterone levels had primary hypogonadism while the remaining had secondary hypogonadism¹⁹. A study on *OTX2* knockout in mice confirms that *OTX2* is indispensable for *GnRH* expression²⁰. The *GnRH* expression is found to be down-regulated among T2DM men²¹.

rs17501809 [PLGRKT] with TGL. The Plasminogen receptor with a C-terminal lysine is a transmembrane protein, known to regulate catecholamine release²². Catecholamine is known to be involved in the regulation of lipoproteins including triglyceride metabolism²³.

Phenotype	Variant	Gene	Model	MAF	Beta ^s	P-value ^s	Beta ^f	P-value ^f	Bonferroni p-value	Benjamini-Hochberg FDR p-value
A. Association at close to genome-wide significance of p-value $\leq 3.41E-08$										
HbA1c	chr15:40531386-rs12440118	<i>ZNF106</i> (W > R)	Recessive	0.1059	2.006	7.07E-08	1.498	2.70E-05	0.003	0.0032
B. Associations at nominal significance of p-value $\leq 5.45E-07$										
	rs7144734	<i>OTX2-ASI</i> (upstream)	Recessive	0.2036	1.465	2.82E-07	1.361	4.31E-07	0.012	0.012
TGL	rs17501809	<i>PLGRKT</i> (intronic)	Recessive	0.0593	1.807	1.04E-07	1.74	3.29E-07	0.0047	0.0016
TGL	rs11143005	<i>LOC105376072</i> (intronic)	Recessive	0.2829	0.4196	4.03E-07	0.4225	3.21E-07	0.018	0.0030
TGL	chr12:101494770/rs10860880	<i>IGF1</i> (downstream)	Recessive	0.05575	1.596	2.077-07	1.603	1.74E-07	0.0095	0.0019
TGL	rs900543	<i>THSD4, NR2E3</i> (intergenic/downstream to <i>NR2E3</i>)	Recessive	0.08696	1.625	1.27E-07	1.607	1.67E-07	0.0058	0.0016

Table 2. Markers that are shown to be associated with phenotype traits either at close to the claimed genome-wide significant p-value of $\leq 3.41 \times 10^{-08}$ or at nominally significant p-value of $\leq 5.45 \times 10^{-07}$.

^sTests for association were adjusted for age, sex, and the first 10 principal components. ^fTests for association with anthropometric and lipid traits were adjusted for age, sex, principal components, and lipid lowering medication status; tests for association with HbA1c and FGL were adjusted for age, sex, principal components, and diabetes medication status; tests for association with SBP and DBP were adjusted for age, sex, principal components, and hypertension medication status.

Marker and gene loci [@]	Metabolic trait	Statistics after regular adjustment (RA) for age, sex, and the first 10 principal components and further adjustments for Diabetes medication (DM), Diabetes status (DS), and BMI as a surrogate for Obesity Status (OB)							
		RA		RA + DM		RA + DS		RA + OB	
		Beta	p-value	Beta	p-value	Beta	p-value	Beta	p-value
chr15:40531386-rs12440118 <i>ZNF106</i> (W > R)	HbA1c	2.006	7.07E-08	1.50	2.705E-05	1.55	9.30E-05	1.96	1.17E-07
rs7144734 <i>OTX2-ASI</i> (upstream)	FPG	1.465	2.825E-07	1.36	4.312E-07	1.12	2.18E-06	1.51	8.8E-07
rs17501809 <i>PLGRKT</i> (intronic)	TGL	1.807	1.043E-07	1.74	3.297E-07	1.72	3.50E-07	1.75	1.95E-07
rs11143005 <i>LOC105376072</i> (intronic)	TGL	0.420	4.035E-07	0.42	3.218E-07	0.42	2.27E-07	0.41	5.83E-07
chr12:101494770/rs10860880 <i>IGF1</i> (downstream)	TGL	1.596	2.077-07	1.60	1.746E-07	1.52	6.21E-07	1.53	5.15E-07
rs900543 <i>THSD4, NR2E3</i> (intergenic/downstream to <i>NR2E3</i>)	TGL	1.625	1.27E-07	1.61	1.67E-07	1.58	2.26E-07	1.59	1.849E-07

Table 3. Effect of adjusting association tests for diabetes status and obesity status with markers identified at either close to the claimed genome-wide significant p-value of $\leq 3.41E-08$ or at nominally significant p-value of $\leq 5.45E-07$. [@]These markers are as listed in Table 2.

rs11143005 [LOC105376072] with TGL. This SNP is located in a long interspersed noncoding RNA (*lincRNA*) called *RP11-274B18.4-001* with no annotated function. In the gene expression data analysis (presented later in Results section), we find that this genotype upregulates the downstream *PGM5* gene which is known to be involved in glucose metabolism²⁴. *PGM* activity is essential in the formation of glucose-6-phosphate from galactose and glycogen, and in the formation of carbohydrates from glucose-6-phosphate.

rs900543 [THSD4, NR2E3] with TGL. The nuclear receptor subfamily 2 group E member 3 (*NR2E3*) gene is implicated in several eye-related disorders including enhanced S-cone syndrome, a form of retinitis pigmentosa, and Goldmann-Favre syndrome. *NR2E3*, like other nuclear receptors, is a target for retinoic acid which is important in many physiological functions and in the expression of genes that regulate energy metabolism²⁵. Till date there is no report on *NR2E3*'s direct role in lipid metabolism. However, discovery of *NR1D1* as an *NR2E3*-interacting protein²⁶ adds a new dimension: *NR1D1* is known to regulate human *ApoC3* gene promoter, a gene that plays an important role in plasma triglyceride and remnant lipoprotein metabolism²⁷.

Chr12:101494770 [IGF1, PAH] with TGL. Insulin-like growth factor 1 (*IGF1*) is similar to insulin in molecular structure. It plays an important role in cell growth and anabolism function. Recent study on *IGF1* deficient mice showed elevated levels of triglyceride²⁸, and another study on *IGF1* treatment in healthy individuals showed reduced levels of fasting and postprandial triglyceride²⁹. *IGF1* deficiency has been reported to be involved in the development of diabetic retinopathy in Thalassemia major patients from Italy³⁰ and in the development of metabolic syndrome in a British cohort³¹.

Additional susceptibility loci with suggestive evidence of association at p-values of $< 1.0E-05$. It is also important to consider the markers that reached p-values of $< 1.0E-05$ ('suggestive' evidence) in association tests. We identified a total of 30 such SNPs which were involved in 33 associations with 12 metabolic traits

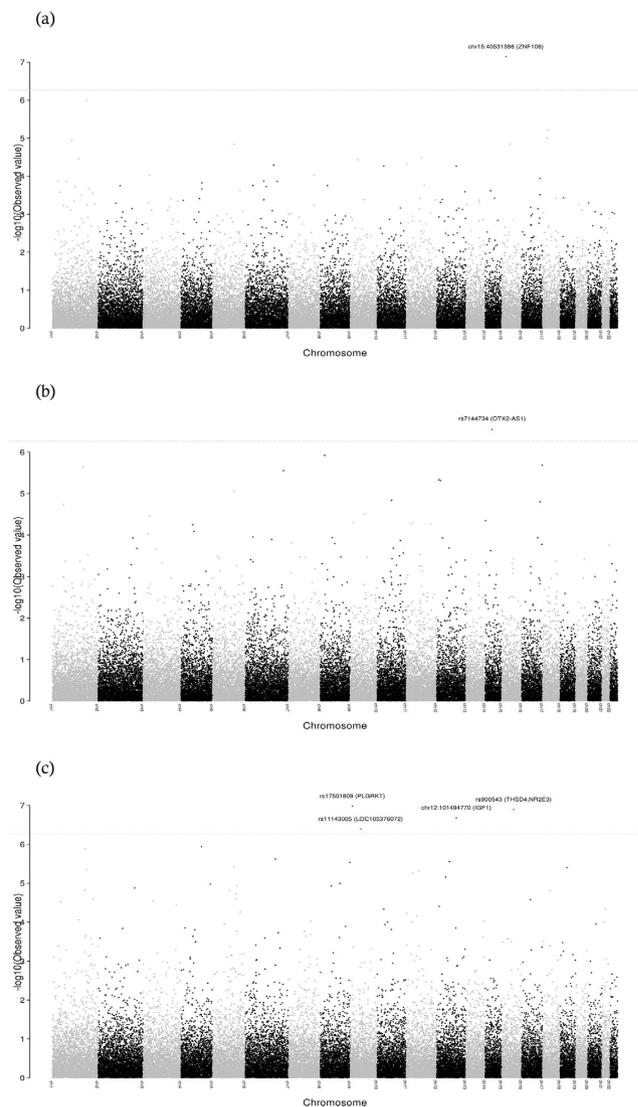


Figure 1. Manhattan plots of the 45793 tagSNPs employed to interrogate the genomes of the study population for associations with the phenotype traits of HbA1c (**1.a**), FPG (**1.b**), and TGL (**1.c**). The $-\log_{10}$ p-values were determined using linear regression methods adjusted for age, sex, first 10 principal components with respective quantitative trait. Labeled are those markers with a p-value of $\leq 5.45E-07$.

of height, weight, BMI, waist circumference (WC), waist circumference to height ratio (WcHtR), HbA1c, FPG, HDL, LDL, TGL, Total Cholesterol, and SBP (Table 4). 13 of these associations (involving 10 unique markers and 9 metabolic traits of weight, WC, WcHtR, BMI, HbA1c, LDL, HDL, TGL, and TC) were identified by models implementing additive mode of inheritance while 20 (involving 19 unique markers and 6 metabolic traits of weight, WC, HbA1c, FPG, TGL, and SBP) were identified by models implementing recessive mode of inheritance. All the associations identified under recessive mode of inheritance and 10 of the 13 associations identified under additive mode of inheritance were of risk effect.

For most of these suggestive associations, literature evidences for involvement of the gene loci in metabolic trait related processes were found (Table 5). The evidences are particularly prominent in the cases of [*TRA*, *TRD*] (WcHtR) and *C1orf106* (HbA1c), *GAPDHP56* (weight, BMI, and WC), *UST* (WC), [*MDGA1*, *ZFAND3*] (HbA1c), *CETP* (HDL), *BUD13* (TGL), *TMEM120B* (TGL), *TEX29* (weight), *KSR1* (HbA1c), *SLC28A3* (FPG), *PIK3C2G* (FPG), *LAMA4* (TGL), [*LY6D*, *GML*] (TGL), *MICAL2* (TGL), and *TTN* (SBP).

Examining the EBIGWAS Catalog for reports on associations between the identified markers and the exact metabolic traits. By way of examining the GWAS Catalog (the NHGRI-EBI Catalog of published genome-wide association studies available at <https://www.ebi.ac.uk/gwas/>), we checked whether previous GWA studies report the associations identified in our study. None of the six markers, that we identified as associated in the study population (either at close to genome-wide significance or at nominal p-value threshold), was reported with significant associations in the EBI GWAS Catalog (Supplementary Table S2). However, two of the markers with suggestive evidence of association in our study (namely rs1800775/*CETP* with HDL, and rs9326246/*BUD13*

Phenotype	Variant	Gene	Model	MAF	Beta [§]	P-value [§]	Beta [#]	P-Value [#]
WcHtR	rs17117722	<i>TRA, TRD</i> (3'UTR)	Additive	0.074	-0.0309	1.07E-06	-0.0303	2.316E-06
HbA1c	rs3767494	<i>C1orf106</i> (intronic)	Recessive	0.052	2.971	1.017E-06	2.503	1.557E-05
Weight	rs10005556	<i>GAPDHP56, LOC105377421</i> (intergenic)	Additive	0.050	6.63	1.24E-06	6.315	3.308E-06
Weight	rs4559034	<i>LOC102467224</i> (intronic)	Additive	0.311	2.901	5.24E-06	2.827	7.99E-06
BMI	rs10005556	<i>GAPDHP56, LOC105377421</i> (intergenic)	Additive	0.05	2.21	2.25E-06	2.278	5.03E-06
Waist Circumference	rs10005556	<i>GAPDHP56, LOC105377421</i> (intergenic)	Additive	0.05	4.851	1.29E-06	4.729	2.412E-06
Waist Circumference	rs9390649	<i>UST</i> (intronic)	Additive	0.36	2.138	1.72E-06	2.112	2.348E-06
HbA1c	rs17639988	<i>MDGA1, ZFAND3</i> (intergenic)	Additive	0.083	0.4722	2.21E-06	0.410	1.642E-05
HDL	rs1800775	<i>CETP</i> (downstream)	Additive	0.48	-0.0499	1.60E-06	-0.049	1.909E-06
LDL	chr2:43355981	<i>THADA</i> (intronic)	Additive	0.38	-0.1383	8.9E-06	-0.124	5.022E-05
TGL	rs9326246	<i>BUD13</i> (promoter)	Additive	0.11	0.237	5.19E-06	0.2315	9.853E-06
TGL	rs925530	<i>TMEM120B</i> (intronic)	Additive	0.37	0.1574	5.03E-06	0.159	3.668E-06
Total cholesterol	chr2:43355981	<i>THADA</i> (intronic)	Additive	0.38	-0.1657	1.37E-06	-0.1494	1.055E-05
Total cholesterol	rs10935794	<i>RPL32P9, LINC01213</i> (intergenic)	Additive	0.19	0.2037	3.65E-06	0.2056	2.447E-06
Height	rs3959929	<i>LOC105376567, LOC105376570</i> (intergenic)	Recessive	0.065	7.369	6.125E-06	NA [@]	NA [@]
Weight	rs1184476	<i>TEX29</i> (intergenic)	Recessive	0.055	36.52	1.12E-05	36.74	9.269E-06
HbA1c	chr1:160381119	<i>LOC105371466</i>	Recessive	0.053	2.332	3.48E-05	2.548	1.933E-06
HbA1c	rs17716285	<i>KSRI</i> (intronic)	Recessive	0.055	3.113	9.87E-06	3.25	1.19E-06
FPG	rs7729384	<i>LOC101927421</i> (intronic)	Recessive	0.122	2.169	8.855E-06	1.716	0.000202
FPG	rs3799125	<i>RGS17</i> (intronic)	Recessive	0.102	2.599	2.77E-06	2.372	5.906E-06
FPG	rs10247084	<i>LOC105375159, LOC105375160</i> (intergenic)	Recessive	0.059	3.865	0.000157	4.373	5.847E-06
FPG	rs883431	<i>SLC28A3</i> (upstream)	Recessive	0.184	1.287	3.187E-05	1.394	1.77E-06
FPG	rs747486	<i>ANKRD11</i> (intronic)	Recessive	0.076	3.714	2.055E-06	3.187	1.619E-05
FPG	rs4764409	<i>PIK3C2G</i> (intronic)	Recessive	0.085	3.314	4.852E-06	3.197	2.95E-06
TGL	rs12722856	<i>RPS3AP9, GAPDHP75</i> (intergenic)	Recessive	0.083	1.487	1.313E-06	1.428	3.599E-06
TGL	rs10014125	<i>LOC105377567</i> (intronic)	Recessive	0.176	0.6139	1.06E-05	0.6178	9.014E-06
TGL	rs17073574	<i>LAMA4</i> (intronic)	Recessive	0.137	0.7335	2.317E-06	0.7258	2.933E-06
TGL	rs11777524	<i>LY6D, GML</i> (intergenic)	Recessive	0.052	1.686	2.91E-06	1.692	2.619E-06
TGL	rs11602685	<i>MICAL2</i> (intronic)	Recessive	0.099	1.127	5.37E-06	1.125	5.437E-06
TGL	chr11:45779819	<i>DKFZp779M0652, SLC35C1</i> (intergenic)	Recessive	0.062	1.91	4.844E-06	1.884	6.319E-06
TGL	rs17569297	<i>LOC105369738, LOC105369739</i> (intergenic)	Recessive	0.128	0.773	6.963E-06	0.7732	6.7E-06
TGL	rs7342999	<i>LOC105372082, LOC105372084</i> (intergenic)	Recessive	0.129	0.8303	3.953E-06	0.8186	5.261E-06
SBP	rs10497520	<i>TTN</i> (coding: K > E)	Recessive	0.155	9.551	5.5E-06	9.339	8.558E-06

Table 4. Markers that achieved suggestive evidence of association with phenotype traits at p-values of $\lt; 1.0E-05$. [§]Tests for association were adjusted for age, sex, and the first 10 principal components. [#]Tests for association with anthropometric and lipid traits were adjusted for age, sex, principal components, and Lipid lowering medication status; tests for association with HbA1c and FPG were adjusted for age, sex, principal components, and diabetes medication status; tests for association with SBP and DBP were adjusted for age, sex, principal components, and hypertension medication status. [@]Tests of association for height were not adjusted for medication.

with TGL) were seen associated with respective phenotype traits at genome-wide significant p-values in GWAS Catalog (see Supplementary Table S2). In all other instances, the p-values in the GWAS Catalog failed to reach genome-wide significance. Thus, it is quite possible that many of the associations identified in our study are novel. We noticed that the effect sizes were not consistent between the data from Arab ethnicity presented in this study and the publicly available results in GWAS Catalog (except in the case of rs9326246/*BUD13* marker with beta values of 0.237 versus 0.2185). In most instances of identified associations, the beta values were close to 0 (ranging from -0.01 to + 0.01) in GWAS Catalog; as studies in the GWAS Catalog were carried out with large number of samples, it is not surprising to find values close to 0 for the effect size. However, it is to be noted that the effect sizes are not very informative when the p-values are not significant.

Marker/Gene	Phenotype trait	Comments
rs3767494/ <i>C1orf106</i>	HbA1c	A systematic analysis of genetic loci of cross autoimmune disorders implicated role of <i>c1orf106</i> gene in the etiology of Type 1 diabetes ⁷² .
rs17117722/[<i>TRA, TRD</i>]	WcHtR	Both the <i>TRA</i> and <i>TRD</i> genes are members of the class of T-Cell receptor proteins. T-cells play an important role in the immune adaptation to both obesity and malnutrition ⁷³ . T-cells play an early and critical role in inducing inflammation in obesity and in the accumulation of inflammatory macrophages in obese adipose tissue; both these two processes are known to promote insulin resistance.
rs10005556/ <i>GAPDHP56</i>	Weight/BMI/Waist Circumference	Genome region of <i>GAPDHP56</i> (chr4:131448001-131451000, chr4:131449501-131452500) is reported as associated with total fat mass in genome-wide rare variants analysis of UK10K Cohorts ⁵² .
rs1184476/ <i>TEX29</i>	Weight	Variants of <i>TEX29</i> (alias <i>C13orf16</i>) are associated with age at menarche among Chinese women. However, as of now there is no report of their role in Weight gain ⁵³ .
rs9390649/ <i>UST</i>	Waist circumference	A study using mice suggests an important role for Uronyl 2-sulfotransferase transfers in clearance of triglyceride-rich lipoprotein ⁷⁴ .
rs17639988/ <i>MDGA1, ZFAND3</i>	HbA1c	Intergenic SNPs between <i>MDGA1</i> and <i>ZFAND3</i> are implicated in T2DM in East Asian population ⁵⁴ .
rs17716285/ <i>KSRI</i>	HbA1c	Kinase suppressor of Ras 1 (<i>KSRI</i>) is required for Ras-mediated ERK pathway activation. This pathway is identified as modifiers of cellular insulin responsiveness. A study conducted on OLETF rat models for T2DM have shown that suppressing the <i>KSRI/ERK</i> complex brings down diabetic vascular complications ⁵¹ .
rs883431/ <i>SLC28A3</i>	FPG	A study in rat indicates upregulation of <i>SLC28A3</i> in glomerulus in diabetic condition. But till date no human study is reported ⁴⁹ .
rs4764409/ <i>PIK3C2G</i>	FPG	Variants from this gene are found to be associated with high HbA1c and low serum insulin levels in Japanese T2D population ⁵⁵ . <i>PI3K</i> orchestrates its action through <i>AKT2</i> pathway in the liver, as a main regulator of metabolism ⁷⁵ .
rs1800775/ <i>CETP</i>	HDL	rs1800775 is a well-known upstream (~2KB) variant of <i>CETP</i> gene shown to be replicated in European population with HDL ⁷⁶ .
rs9326246/ <i>BUD13</i>	TGL	Extended CARDIoGRAM study (involving participants of European ancestry) implicated rs9326246 marker as associated with TGL and as one of the important CAD susceptibility loci ⁷⁶ .
rs925530/ <i>TMEM120B</i>	TGL	The role of <i>TMEM120B</i> in fat metabolism has been established in knockdown cell line studies ⁵⁰ .
rs17073574/ <i>LAMA4</i>	TGL	Studies have shown association of <i>LAMA4</i> in obesity and diabetic nephropathy ^{77,78} .
rs11777524/ <i>LY6D, GML</i>	TGL	There is lack of literature reports on the role of <i>GML</i> (Glycosylphosphatidylinositol-anchored molecule like) in processes relating to triglyceride. But there is evidence for <i>GPIHBP1</i> (Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1) in causing hypertriglyceridemia. Protein sequence comparison analysis shows 28% identity between these two proteins. Furthermore, it is known that <i>GPIHBP1</i> evolved <i>LY6</i> -like genes ⁷⁹ .
rs11602685/ <i>MICAL2</i>	TGL	Its association with plasma uric acid in obese individuals is evident. However, it requires more literature support to establish its functional role in elevation of triglyceride levels ⁸⁰ .
rs10497520/ <i>TTN</i>	SBP	Involvement of the <i>TTN</i> gene in hypertension is substantiated by literature evidences. Rain <i>et al.</i> ⁸¹ found that phosphorylation of titin in right ventricular tissue was significantly reduced in pulmonary arterial hypertension patients (undergoing heart/lung transplantation) as compared with control subjects. Mutations in this gene are associated with familial hypertrophic cardiomyopathy ⁸² . Further, Ottenheijm <i>et al.</i> demonstrated that alternative splicing of the <i>TTN</i> gene was associated with diaphragm dysfunction in patients with mild to moderate chronic obstructive pulmonary disease ⁸³ .

Table 5. Literature evidence for involvement of ‘suggestive’ markers (or gene loci) with p-values of <1.0E-05 in the associated phenotype traits.

Examining the EBI GWAS Catalog for reports on associations between the identified markers and related (not necessarily the exact) metabolic traits. Upon examining further the EBI GWAS Catalog for associations between the markers and related (not necessarily the exact associated trait in our study population) metabolic traits in other populations (see Online Supplement Data 2); we found that 17 of the 36 identified markers showed associations with related metabolic traits at p-values of $\leq E-03$ (highlighted in blue in Online Supplement Data 2). In the case of four of these markers, identified in our study as associated at nominal p-values (rs900543 with TGL, rs11143005 with TGL) or at suggestive evidence of p-values (rs17569297 with TGL, and rs10935794 with total cholesterol), associations at p-values of $< E-04$ were reported in other populations with related metabolic traits (rs900543 with fasting insulin at p-value = 9.40E-05, rs11143005 with 2 hour fasting glucose at p-value = 4.47E-05, rs17569297 with HDL at p-value = 1.51E-06, and rs10935794 with serum ratio of arabinosefructose at p-value = 9.840E-05) (Supplementary Table S3). Particularly interesting was the

rs17569297 marker that showed a p-value at suggestive evidence of association in both our study population and the European population.

Performance of most replicated exemplary markers relating to obesity and diabetes in our study population. Established susceptibility gene loci for traits relating to obesity and diabetes in European population have been tested for replication in other ethnic population groups; such exemplary loci that are known to replicate well in other ethnic population groups (such as Indians and Russians) include *PPARG*, *KCNJ11*, *TCF7L2*, *SLC30A*, *ABCC8*, *HHEX*, *CDKN2A*, *IGF2BP2*, *CDKALI*, and *FTO*^{32–34}. We examined the performance of markers from these gene loci in our MetaboChip data set (Supplementary Table S4). None of the markers from *ABCC8*, *HHEX*, *CDKN2A*, *IGF2BP2*, and *FTO* surfaced with at least a p-value of ≤ 0.05 for association in our data set. Markers from the other genes *PPARG*, *KCNJ11*, *TCF7L2*, *SLC30A*, and *CDKALI* did not reach significant association with any of the 13 tested metabolic traits; at best it reached a p-value of 0.001.

Gene expression regulation by identified markers. Examination of the Genotype-Tissue Expression (GTEx) data showed that seven of the 36 identified markers differentially regulate the expression of either the downstream or upstream genes (Table 6); such markers are Chr15:40531386 [*ZNF106*; W->R], rs11143005 [*LOC105376072*], rs1800775 [*CETP*], rs9326246 [*BUD13*], rs17716285 [*KSR1*], rs1177524 [*LY6D*, *GML*], and rs10497520 [*TTN*; K->E]. In addition, the rs11143005 marker from the above list was seen as regulating its own gene. It is interesting to note that in four instances the regulated up/downstream genes were also related to metabolic traits. Such instances are listed below:

rs11143005 [LOC105376072] with TGL. This SNP is located in a long interspersed noncoding RNA (*lincRNA*) called *RP11-274B18.4-001* of unknown function. As per GTEx data, this gene upregulates its downstream gene *PGM5* (phosphoglucomutase 5), which is involved in glucose metabolism²⁴. *PGM* activity is essential in the formation of carbohydrates from glucose-6-phosphate, and in the formation of glucose-6-phosphate from galactose and glycogen.

rs17716285 [KSR1] with HbA1c. The rs17716285 SNP from *KSR1* upregulates the expression of an upstream gene *NOS2* (Nitric oxide synthase 2). It has been strongly suggested that nitric oxide plays a crucial role in the regulation of blood pressure^{35,36}.

rs1800775 [CETP] with HDL. The rs1800775 SNP from *CETP* downregulates the upstream *NLRC5* gene. Charlesworth *et al.*, through combining results from association tests with genome-wide transcriptional profiling data for HDL, identified *NLRC5* as a novel loci associated with HDL³⁷.

rs10497520 [TTN] with SBP. The rs10497520 SNP from *TTN* gene downregulates the upstream *FKBP7* gene. Peter *et al.* report that the *FKBP7* gene showed a significant association with intensive lifestyle intervention response among overweight/obese diabetic individuals not receiving beta-blocker medications³⁸.

Pathway Analysis. In order to identify the pathways that are enriched with the reported gene loci, we mapped the genes onto canonical pathways. We found that five of the identified gene loci (*KSR1*, *PIK3C2G*, *MICAL2*, *CETP*, and *UST*) are involved in the following metabolic pathways: Ceramide signaling (*KSR1* and *PIK3C2G*; p-value for overlap = 9.56E-04), pregnenolone biosynthesis (*MICAL2*; p-value for overlap = 4.07E-03), ERK/MAPK signaling (*KSR1* and *PIK3C2G*; p-value for overlap = 5.08E-03), Histidine Degradation VI (*MICAL2*; p-value for overlap = 5.08E-03), and *LPS/IL-1* mediated inhibition of *RXR* function (*CETP* and *UST*; p-value for overlap = 6.97E-03). As enumerated below, these pathways relate to processes and disorders relating to metabolic traits:

Ceramide Signaling. This signaling pathway is implicated in the pathogenesis of insulin resistance and other obesity-associated metabolic diseases³⁹.

Pregnenolone biosynthesis. Pregnenolone is synthesized from cholesterol and is involved in causing obesity and insulin resistance⁴⁰.

ERK/MAPK signaling. This pathway is one of the well-studied pathways in the context of metabolic disorders. Cross-talk between the epidermal growth factor receptor (*EGFR*)-activated *MAPK/ERK* pathway and insulin signaling pathway in controlling glucose metabolism has been demonstrated in *Drosophila*⁴¹.

Histidine Degradation VI. It is well known that dysregulation of insulin leads to marked alteration in amino acid metabolism, and that histidine supplementation improves insulin resistance and reduces obesity⁴².

LPS/IL-1 Mediated Inhibition of RXR function. This pathway is known to be involved in causing diet-induced obesity, and noninsulin-dependent diabetes mellitus⁴³.

Figure 2 depicts gene interaction network showing highest overlapping score. The identified gene loci (namely *UST*, *CETP*, *MEM120B*, *ZNF106*, *KSR1*, *ANKRD11*, *LAMA4*, and *PIK3C2G*), that overlap with the above-mentioned pathways, were all seen interacting with one another (shown in red color) in the interaction network. The other interacting partner genes were also known to be involved in the etiology of complex metabolic disorders. For example: *PLG*, *PLAT*, *PLAUR*, *norepinephrine*, *TNF*, *MMP9*, *FYN* and *SHARPIN* are involved in

SNP	Gene mapped	Gene regulated	P-value	Effect Size	Regulation	Tissue
I. Markers associated at close to genome-wide significant p-value of $\leq 3.41E-08$						
Chr15:40531386-rs12440118	<i>ZNF106</i> (W > R)	<i>LRRC57</i> (downstream of <i>ZNF106</i>)	3.7E-09	-0.47	Down	Cells- transformed fibroblast
II. Markers with nominal evidence of association at p-value $\leq 5.45E-07$						
rs11143005	<i>LOC105376072</i> (intronic)	<i>LOC105376072</i> (intronic)	3.3E-08	0.23	Up	Whole Blood
rs11143005	<i>LOC105376072</i> (intronic)	<i>PGM5</i> (downstream of <i>LOC105376072</i>)	4.2E-10	0.50	Up	Whole Blood
III. Markers with suggestive evidence of association at p-value $< 1.0E-05$						
rs1800775	<i>CETP</i> (SNP located downstream of the gene)	<i>NLRC5</i> (Upstream of <i>CETP</i>)	4.3E-16	-0.35	Down	Cells- transformed fibroblast
rs9326246	<i>BUD13</i> (SNP located upstream of the gene)	<i>RP11-109L13.1</i> (downstream of <i>BUD13</i>)	4.8E-07	-0.80	Down	Skin- Sun exposed
rs17716285	<i>KSR1</i> (intronic)	<i>NOS2</i> (upstream of <i>KSR1</i>)	0.0000051	0.46	Up	Muscle skeletal
rs11777524	<i>LY6D, GML</i> (intergenic)	<i>LOC101928087</i> (upstream of <i>LY6D</i>)	0.0000041	-0.80	Down	Testis
rs10497520	<i>TTN</i> (K > E)	<i>FKBP7</i> (upstream of <i>TTN</i>)	0.000019	-0.19	Down	Cells- transformed fibroblast

Table 6. List of differentially regulated genes as determined by GTEx data analysis on variants associated with metabolic traits.

the etiology of dyslipidemia, hypertension, and cardiovascular disease (CVD); *APC* and *CCND1* are associated with obesity; *SUMO2*, *SMAD2*, *INS1*, *WNT1*, *TGFB1*, *TRAF2*, and *CCAR2* are associated with diabetes; whereas *CCND1*, *TNF*, *MMP9*, and *norepinephrine* are associated with comorbid conditions of metabolic disorders.

Transethnic variation in allele frequencies. Frequencies of risk alleles for T2DM and metabolic traits often vary between populations, and thereby cause population differences in the risk due to particular genetic factors. The identified thirty six markers are common markers (*i.e.* MAF > 5%) in the study population. Examination of allele frequencies at these markers in 1000 Genomes Project Phase 3 populations (Table 7) revealed that as many as twenty four out of thirty six markers appear as low-frequency variants (MAF < 5%) in one or more global populations; and in as many as half of the instances, they appear as rare variants (at MAF < = 1%) in East Asian (15 times) and/or African (11 times) populations. Upon considering the 36 markers to establish Pearson correlation coefficients, we found that the risk variants from Arab population showed least correlation with East Asian populations at 0.412 [p-value = 0.00737, CI = 0.12 to 0.63] and highest correlation with South Asians at 0.738 [p-value = 3.604E-08, CI = 0.55 to 0.85]. The values for correlation coefficients with populations from the 1000 Genomes Project Phase 3 were intermediate, ranging from 0.588 to 0.647 (with AMR - Ad Mixed Americans, 0.588 [p-value = 5.19E-05; CI = 0.34–0.75], with EUR - Europeans, 0.647 [p-value = 4.84–06; CI = 0.42–0.79], with AFR - Africans, 0.654 [p-value = 3.43E-06; CI = 0.43–0.80]). These observations illustrate transethnic variations in the penetrance of risk alleles.

Discussion

Overall, this study identified (Chr15:40531386-rs12440118/*ZNF106*/W->R) as a risk variant associated with HbA1c at close to genome-wide significant p-value, and five other risk variants 'nominally' associated (p-value $\leq 5.45E-07$) with fasting plasma glucose (rs7144734/[*OTX2-AS1, RPL3P3*]) and triglyceride (rs17501809/*PLGRKT*, rs11143005/*LOC105376072*, rs900543/[*THSD4, NR2E3*], and Chr12:101494770/*IGF1*). Involvement of these gene loci in metabolic traits was amply supported by literature evidences. Furthermore, the study identified 30 additional variants with suggestive evidence of association with the phenotype traits of height, weight, BMI, waist circumference (WC), waist circumference to height ratio (WcHtR), HbA1c, FPG, HDL, LDL, TGL, total cholesterol, and SBP. Most of these gene loci with suggestive evidence of association - particularly [*TRA, TRD*] (WcHtR), *C1orf106* (HbA1c), *GAPDHP56* (weight/BMI/WC), *UST* (WC), [*MDGA1, ZFAND3*] (HbA1c), *CETP* (HDL), *BUD13* (TGL), *TMEM120B* (TGL), *TEX29* (Weight), *KSR1* (HbA1c), *SLC28A3* (FPG), *PIK3C2G* (FPG), *LAMA4* (TGL), [*LY6D, GML*] (TGL), *MICAL2* (TGL), and *TTN* (SBP) - were supported by literature evidences on their involvement in metabolic trait related processes.

Most of the reported variants are harbored in either intronic or intergenic regions. However, two of the identified markers lead to non-synonymous amino acid changes in the encoded proteins; these missense variants are: Chr15:40531386/*ZNF106*/W->R associated with HbA1c at close to genome-wide significant p-value, and rs10497520/*TTN*/K->E with suggestive evidence of association (p-value $< 1.0E-05$) with SBP. In both these instances, the amino acid substitution leads to change in the charge associated with the side chain. In the case of W->R (*trp*-> *arg*) substitution, while both the amino acids are hydrophobic, *trp* has non-charged side chain and *arg* has positively charged side chain. In the case of K->E (*lys*-> *glu*), while both the amino acids are hydrophilic, *lys* is positively charged and *glu* is negatively charged. Thus, though the tendency to get buried into protein core or to get exposed on the surface is not changed, the polarity is altered. Change in polarity may affect the interactions that the amino acid residue has with other residues from the protein or with non-protein molecules.

We find it striking that all the 6 associations identified either at close to genome-wide significance or at the lenient p-value threshold of $\leq 5.45E-07$, and as many as 20 of the 33 associations identified with suggestive

SNP	Minor allele frequencies (MAF) [@] in					
	AFR (African)	AMR (Ad Mixed American)	EAS (East Asian)	EUR (European)	SAS (South Asian)	Arab (Study population)
rs3767494	0.02	0.11	0.13	0.05	0.07	0.053
Chr15:40531386-rs12440118	0.01	0.1	0.01	0.14	0.07	0.106
rs7144734	0.12	0.2	0.07	0.29	0.17	0.204
rs17501809	0	0.04	0	0.08	0.04	0.059
rs11143005	0.42	0.14	0	0.19	0.18	0.283
rs900543	0.44	0.12	0.26	0.04	0.06	0.087
rs10860880	0.28	0.04	0.01	0.02	0.04	0.056
rs17117722	0.34	0.06	0	0.06	0.01	0.074
rs10005556	0.197	0.023	0	0.027	0.033	0.05
rs4559034	0.431	0.157	0.009	0.286	0.253	0.31
rs9390649	0.243	0.225	0.271	0.409	0.401	0.36
rs17639988	0.002	0.048	0.001	0.078	0.034	0.08
rs1800775	0.607	0.543	0.486	0.488	0.59	0.47
rs9326246	0.011	0.133	0.238	0.093	0.193	0.11
rs925530	0.616	0.427	0.451	0.328	0.367	0.37
rs10935794	0.061	0.207	0.326	0.282	0.26	0.19
rs3959929	0.004	0.069	0	0.11	0.015	0.065
rs1184476	0.29	0.035	0	0.038	0.027	0.054
chr1:160381119	0	0	0	0	0	0.053
rs17716285	0.007	0.055	0	0.093	0.019	0.055
rs7729384	0.066	0.216	0.118	0.242	0.05	0.12
rs3799125	0.029	0.086	0.105	0.196	0.098	0.1
rs10247084	0.132	0.085	0.107	0.072	0.077	0.06
rs11785103	0.003	0.027	0.001	0.044	0.039	0.0642
rs883431	0.402	0.442	0.497	0.172	0.191	0.18
rs4764409	0.036	0.411	0.322	0.196	0.194	0.085
rs747486	0.009	0.036	0.114	0.049	0.134	0.075
rs12722856	0.005	0.029	0	0.067	0.056	0.083
rs10014125	0.365	0.097	0.067	0.132	0.109	0.175
rs17073574	0.042	0.082	0.188	0.069	0.116	0.137
rs11777524	0.062	0.049	0.325	0.08	0.224	0.052
rs11602685	0.118	0.073	0.041	0.095	0.029	0.1
chr11:45779819	0	0	0.168*	0.033*	0	0.06
rs17569297	0.008	0.072	0.001	0.129	0.153	0.13
rs7342999	0.588	0.222	0.202	0.088	0.08	0.13
rs10497520	0.408	0.509	0.217	0.812	0.544	0.155

Table 7. 1000 Genomes Project Phase 3 allele frequencies for markers identified in this study. [@]MAF values less than 5% are shown in italics. *MAF is obtained from Illumina product files (http://support.illumina.com/downloads/cardio-metaboChip_product_support_files.html).

Furthermore, in the following two cases of identified suggestive markers, significant associations have been previously established with the respective metabolic traits in Euro-centric populations (Supplementary Table S2):

1. rs1800775/*CETP*/HDL - the EBI GWAS Catalog reports this variant as significantly associated with HDL in European population (with a p-value of $\leq 3.33\text{E-}644$) and in mixed populations (with a p-value of $\leq 2.05\text{E-}306$)^{56,57}.
2. rs9326246/*BUD13*/TGL - the EBI GWAS Catalog reports this variant as significantly associated with TGL in European population (with a p-value of $\leq 1.27\text{E-}229$), in populations included in the CARDIoGRAM consortium (with a p-value of $\leq 4.70\text{E-}124$) and in mixed populations (with a p-value of $\leq 4.79\text{E-}124$)⁵⁶⁻⁵⁹.

In addition, the rs17569297 identified in our study as associated with TGL at suggestive evidence of p-value has been reported as associated with a related metabolic trait of HDL in European population at similar p-values⁶⁰ (Supplementary Table S3).

It is revealed that seven of the identified genetic variants lead to differential expression of either the downstream or upstream genes (see Table 6). Interestingly, four of these seven genes (namely *PGM5* regulated by rs11143005, *NLRC5* by rs1800775, *NOS2* by rs17716285, *FKBP7* by rs10497520) have literature evidence for

associations with metabolic traits^{24,35–38} and thus these genes are potential candidate genes for further functional studies on processes relating to metabolic traits in this population.

The identified gene loci was seen to map to the canonical pathways of Ceramide signaling, Pregnenolone biosynthesis, *ERK/MAPK* signaling, Histidine Degradation, and *LPS/IL-1* Mediated inhibition of *RXR* function; all these five pathways relate to pathogenesis of obesity, insulin resistance and signaling. Key gene loci (apart from the loci identified in this study) that mapped to these pathways include *KSRI*, *PIK3C2G*, *MICAL2*, *CETP*, *UST*, *TMEM120B*, *ZNF106*, *ANKRD11*, and *LAMA4*. Gene interaction network analysis revealed extensive interactions between these genes and others that are involved in etiology of dyslipidemia/hypertension/CVD (such as *PLG*, *PLAT*, *PLAUR*, *norepinephrine*, *TNF*, *MMP9*, *FYN* and *SHARPIN*), obesity (such as *APC* and *CCND1*), T2DM (such as; *SUMO2*, *SMAD2*, *INS1*, *WNT1*, *TGFB1*, *TRAF2*, and *CCAR2*), and comorbid conditions of metabolic disorders (such as *CCND1*, *TNF*, *MMP9*, and *norepinephrine*). All these interacting genes warrant future targeted genotyping studies for metabolic traits in Arab population.

Upon viewing the results from pathway analysis and gene expression (GTEX) data analysis in the context of our previous studies on leptin-mediated hypertension^{61,62}, we can hypothesize that ceramide signaling is probably a key pathway connecting obesity-induced diabetes and obesity-induced hypertension in Arab population. Ceramide signaling plays a role in central control of feeding via regulating leptin levels⁶³. Two of the identified gene loci in our study namely *KSRI* and *PIK3C2G* (associated with HbA1c and FPG, respectively) are mapped to this pathway. Interestingly, in addition to this, the presented GTEX results revealed that rs17716285 harbored in *KSRI* upregulates its upstream *NOS2* (Nitric oxide synthase 2) gene. Experimental studies have shown that glucose releases endothelial nitric oxide (NO), which in turn contributes to renal hyperperfusion in models of diabetes; upon examining whether this translates into the human condition, Schneider *et al.*⁶⁴ reported that poor glycaemic control is related to higher NO activity which in turn leads to hypertension. Plenty of research reports exist suggesting that normal release of nitric oxide plays a crucial role in homeostasis of blood pressure, and that impairment of NO level can cause hypertension^{35,36}. However, models that implement the role of NO in the pathogenesis of the comorbidity (of diabetes and hypertension) are yet to be demonstrated. The model (incorporating the genes from the ceramide signaling) that we propose in this study requires further functional studies in this population.

Studies to identify genetic markers for the metabolic traits of HbA1c and FPG are usually performed on healthy individuals⁶⁵. Efforts to replicate the identified markers/loci in individuals with diabetes have not resulted in much success; it is yet to be established whether the low reproducibility is due to differences in study power or in ethnicity, or in environment (such as medication for diabetes)⁶⁵. The Kuwaiti population, which contributed to our study cohort, has a high prevalence of obesity and diabetes^{2,66}; as a result, 54% of our study participants are obese, 27% are diabetic, and 13% are under diabetes medication (Table 1). Upon introducing adjustment for diabetes status or for obesity in association tests, the p-values and beta values still remained significant (Table 3). In our future studies, we will focus on replicating these markers in entirely non-diabetic/non-obese cohort and in entirely diabetic/obese cohort.

A potential limitation of this study relates to absence of an independent sample set for replicating the results. However, as discussed so far, most of the identified associations benefit from literature evidence - involvement of some of the identified gene loci in metabolic processes have been demonstrated in mice or cell lines; markers from some of the identified gene loci have been established to be associated with metabolic traits in previous GWA studies; and two of the markers (rs1800775 [from *CETP* gene and associated with HDL]; rs9326246 [from *BUD13* gene and associated with TGL]) with suggestive evidence of association have been reported in other European populations as associated with the respective traits at genome-wide significant p-values. In spite of the absence of replication using an independent sample set, the study presents to the community valuable data on potential markers associated with metabolic traits in an ethnic population which has been under-represented in global genome-wide surveys.

Our study could identify only one marker at a p-value close to genome-wide significance; this might be because of small sample size. A mention about limitations in achieving a large target for participant recruitment is in order here. Establishment of Arab ethnicity as part of the recruitment process was rigorous - ethnicity of every participant was confirmed via a rigorous questionnaire that addressed parental lineages up to three generations. Participants, selected from the computerized register of the Public Authority of Civil Information (PACI) by way of random sampling, were invited to visit us if they wished to participate in the research project; the acceptance rate by the participants was low as there exists a cultural barrier for the local Arabs to disclose their tribe ancestry and to donate blood samples. In spite of the achieved small sample size, the reported loci are promising and they deserve to be examined in large-sized Arab cohorts.

In conclusion, results based on our study with Arab population from Kuwait pinpoint crucial genes that take part in metabolism and metabolic disorders. The presented Pathway analysis and Genotype Tissue Expression data analysis reveal the broader context in which genetic variants regulate metabolic processes and suggests further gene loci for future studies on metabolic traits in Arab population. The study also demonstrates the association of recessive signatures and their pleiotropic effects with the etiology of metabolic disorder. Culturally and geographically distinct ethnic populations such as those of Arabian Peninsula are under-represented in global genome survey studies; the presented study bridges this gap.

Methods

Study participants. A total of 2,440 participants from Kuwait were randomly recruited under protocols approved by the scientific and ethics advisory boards at Dasman Diabetes Institute. The participants include: (i) A random representative sample of adults (>18 years of age) of Arab ethnicity across the six governorates of the State of Kuwait. A stratified random sampling technique was used for the selection of participants from the computerized register of the Public Authority of Civil Information (PACI). PACI is a government body which

maintains all records of personal information of both Kuwaiti citizens as well as expatriates (that includes citizens of other Arab countries from the region); and (ii) Patients with diabetes or prediabetes seeking tertiary medical care in clinics at our institute. At the time of participant recruitment, the nationality was confirmed and ethnicity of every participant was confirmed via a rigorous questionnaire that addressed parental lineages up to three generations; data on age, sex, illness (e.g. diabetes and cardiovascular complications), and medication were recorded; and furthermore, vital signs such as height, weight, waist circumference (WC), and blood pressure readings were recorded. Informed consent was obtained from each of the participants. Participant recruitment for this study was carried out as part of approved research projects^{7,66} at Dasman Diabetes Institute.

Sample collection and genotyping. Upon signing written consent forms, blood samples were collected after confirming that the participants were fasting overnight. The methods to collect blood samples and to measure vital signs were carried out in accordance with guidelines laid in place by the institutional Ethical Review Committee. Lipid profile and glucose measurements were recorded. Genra Puregene[®] kit (Qiagen, Valencia, CA, USA) was used to extract DNA as per manufacturer's protocols. DNA was quantified, with the requirement that the A260/A280 ratio is in the range of 1.8–2.1, using both Quant-iT[™] PicoGreen[®] dsDNA Assay Kit (Life Technologies, NY, USA) and Epoch Microplate Spectrophotometer. Frozen DNA stocks were diluted to a working solution of concentration at 50 ng/μl as recommended by Illumina (Illumina, CA, USA). Samples were genotyped using the Illumina HumanCardio-MetaboChip array utilizing Infinium HD Assay Ultra genotyping assay methods. Genotyping assay includes whole genome amplification, fragmentation, hybridization, staining and imaging of Cardio-MetaboChips using the Illumina iSCAN system. We distributed the 2,440 samples onto 27 batches (each homogenized for sex and diabetes status) and then carried out genotyping. The MetaboChip is a custom genotyping array based on the Illumina iSelect platform and is designed for use in replication of top association signals from the largest available GWAS meta-analysis for metabolic-related traits; it fine mapped 257 genome-wide significant association signals for 15 of these traits⁶⁷. The MetaboChip contains ~200,000 SNP markers that include rare variants identified in the 1000 Genomes Project as well.

Power calculation. We performed power calculation using Quanto software version 1.2.4 (University of Southern California, <http://biostats.usc.edu/Quanto.html>). Power calculation was performed separately for additive and recessive genetic models. We considered 'Gene only' hypothesis in independent individuals with quantitative traits. Marginal genetic effect estimate (R_G^2) was set to range from 0.015 to 0.025 by incrementing by 0.001 (so as to detect a genetic effect that can explain at least 1.5% to 2.5% of the trait variance). For each of the phenotype traits, population mean \pm SD value of the respective quantitative trait (Table 1) was used. Power for the analysis was set at 80% and type 1 error at genome-wide significant p-value was considered.

Quality control and statistical analysis. Raw intensity data from all of the 27 batches were pooled and genotypes were called using GenCall algorithm implemented in GenomeStudio software. Samples with call rates <97% were removed. A high quality set of SNPs was derived by applying a series of quality metric thresholds: SNPs with call rate of <98%, of low intensity (AB R Mean \leq 0.25), poor cluster separation (Cluster Sep < 0.3), with heterozygote clusters too close to homozygotes (AB T Mean \leq 0.2 or \geq 0.8), excess of heterozygotes (Het Excess \geq 0.2), and those with fewer than expected heterozygotes (Het Excess \leq -0.3) were removed. We used GenomeStudio software for gender estimation and removed those samples reporting gender mismatches. We also excluded duplicate samples.

We used PLINK⁶⁸ to perform quality control procedures on markers. Markers that failed to reach cut-off values for missingness per individual ($-mind$ 0.1), allele frequency ($-MAF$ 0.01), missingness per marker ($-geno$ 0.1), and Hardy-Weinberg equilibrium (HWE < 10^{-3}) were ignored. These procedures reduced the size of marker set to 128,143 SNPs. Since the markers mapped in MetaboChip are from gene loci known to be associated with metabolic disorders and related traits, the association tests would pick up a high number of markers that are in linkage disequilibrium (LD) with one another. Hence, we pruned LD with ' $-indep-pairwise$ ' option in PLINK with an R^2 value > 0.3 with any other SNP within a 50-SNP sliding window (advanced by 5 SNPs each time); this stringent LD pruning step ensures that no two SNPs are in strong LD with one another.

We examined the sample set for relatedness using PLINK's ' $-genome$ ' feature (PI_HAT > 0.125, *i.e.* up to third degree relatives) and randomly removed one sample per pair of related individuals. We used ADMIXTURE software⁶⁹ for ancestry estimation and excluded samples with improper ethnicity.

We used EIGENSTRAT⁷⁰ for performing Principal component (PC) analysis with the following choices for parameters: number of eigenvectors to output (numoutvec = 10), turn off outlier removal (numoutlieriter = 0), number of principal components along which to remove outliers during each iteration (numoutlierevec = 10), and number of standard deviations which an individual must exceed along one of the top (numoutlierevec) principal components in order for that individual to be removed/classified as an outlier (outliersigmathresh = 6.0). All the 10 principal components were used as covariates in quantitative trait association tests to adjust for population stratification.

Quantitative trait association analysis was performed, using linear regression method available in PLINK, for each of the 13 traits namely height, weight, waist circumference (WC), waist circumference to height ratio (WcHtR), body mass index (BMI), glycosylated hemoglobin (HbA1c), fasting plasma glucose (FPG), triglyceride (TGL), high density lipoprotein (HDL), low density lipoprotein (LDL), total cholesterol (TC), systolic blood pressure (SBP), and diastolic blood pressure (DBP). Additive mode of inheritance is the most common method used in global GWA studies. However, considering the genetic background of this population, we also used models based on recessive mode of inheritance. Tests of association were adjusted for age, sex, and the first 10 principal components (denoted as R, Regular Adjustments). The tests of associations were repeated with further adjustment for medication status (denoted as R + M, regular and medication adjustments) - all the anthropometric and

lipid traits were adjusted for lipid lowering medication; FPG and HbA1c were adjusted for diabetes medication and hypertension traits were adjusted for hypertension medication.

Regional, Q-Q, Manhattan and PCA plots. In order to generate regional association plot for a SNP-trait association, all the SNPs (typed in MetaboChip) from a region of around 1MB centered on the SNP were tested for association with the trait; the resultant statistics and the SNPs were displayed in the regional association plot. Region-plot tool (<https://github.com/pgxcentre/region-plot>) was used to produce regional plots. Q-Q plots, Manhattan plots and PCA plots were generated using R scripts (<https://cran.r-project.org>).

P-value thresholds for assessing the statistical significance of genotype-phenotype associations. *P-value threshold for association at genome-wide significance.* The p-value threshold for genome-wide significance needs to be calibrated for the 13 tested traits, for the 2 tests pertaining to adjustments for medication status, and for the two genetic models of recessive and additive modes of inheritance; the p-value threshold derived using such a stringent criteria is $2.10E-08$ ($=0.05/[45,793*2*13*2]$). However, the 13 traits are probably correlated to some extent and hence the effective number of independent traits can be lesser than 13. Upon performing Spearman correlation analysis between the 13 traits and then calculating effective number of independent traits using matSpD tool⁷¹ (<http://neurogenetics.qimrberghofer.edu.au/matSpD/>), it turned out that the effective number of independent traits is 8.04. By way of considering the number of markers tested, number of genetics models, number of tests considering medication correction, and effective number of independent traits, we found $3.41E-08$ ($=0.05/[45793*2*8*2]$) as the 'stringent' p-value threshold for genome-wide significance. We also required that the associations passed through Bonferroni correction and Benjamini–Hochberg FDR correction procedures at p-values ≤ 0.05 .

P-value threshold for association at nominal significance. We further defined a 'lenient' p-value threshold of $5.45E-07$ ($=0.05/[45,793*2]$) calculated on the basis that the tests for multiple traits are independent and are not concurrent. We termed the associations whose p-values passed through such a lenient p-value threshold as 'nominal' associations.

P-value threshold for associations with suggestive level of significance. We further defined a third category of markers as showing "suggestive evidence of association" if the observed p-values were $< 1.0E-05$.

Expression data, pathway, and allele frequency analysis. To get insights into the manifestation of the effects that the identified variants might exert on regulation of gene expression in human tissues, we examined the database of expression quantitative trait loci using the GTEx (Genotype-Tissue Expression) portal available at <http://www.gtexportal.org>. Pathway analysis for the identified gene loci from this study was performed using Ingenuity Pathway Analysis software (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity). Transethnic frequency analysis on the identified markers was performed by correlating 1000 Genomes Project Phase 3 allele frequencies for the African (AFR), Ad Mixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS) populations to those in the study population using Pearson-correlation tests.

References

- Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 766–781, doi: 10.1016/S0140-6736(14)60460-8 (2014).
- Channanath, A. M., Farran, B., Behbehani, K. & Thanaraj, T. A. State of diabetes, hypertension, and comorbidity in Kuwait: showcasing the trends as seen in native versus expatriate populations. *Diabetes Care* **36**, e75, doi: 36/6/e75 [pii]10.2337/dc12-2451 (2013).
- Al-Daghri, N. M. *et al.* Assessing the contribution of 38 genetic loci to the risk of type 2 diabetes in the Saudi Arabian Population. *Clin Endocrinol (Oxf)* **80**, 532–537, doi: 10.1111/cen.12187 (2014).
- Ghassibe-Sabbagh, M. *et al.* T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. *Sci Rep* **4**, 7351, doi: 10.1038/srep07351 (2014).
- Al Safar, H. S. *et al.* A genome-wide search for type 2 diabetes susceptibility genes in an extended Arab family. *Ann Hum Genet* **77**, 488–503, doi: 10.1111/ahg.12036 (2013).
- Rodriguez-Flores, J. L. *et al.* Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Res* **26**, 151–162, doi: 10.1101/gr.191478.115 (2016).
- Alsmadi, O. *et al.* Genetic substructure of Kuwaiti population reveals migration history. *PLoS One* **8**, e74913, doi: 10.1371/journal.pone.0074913 (2013).
- Alsmadi, O. *et al.* Weak or no association of TCF7L2 variants with Type 2 diabetes risk in an Arab population. *BMC Med Genet* **9**, 72, doi: 10.1186/1471-2350-9-72 (2008).
- Teebi, A. S. Autosomal recessive disorders among Arabs: an overview from Kuwait. *J Med Genet* **31**, 224–233 (1994).
- Al-Awadi, S. A. *et al.* Consanguinity among the Kuwaiti population. *Clin Genet* **27**, 483–486 (1985).
- Tadmouri, G. O., Sastry, K. S. & Chouchane, L. Arab gene geography: From population diversities to personalized medical genomics. *Glob Cardiol Sci Pract* **2014**, 394–408, doi: 10.5339/gcsp.2014.54 (2014).
- Bittles, A. H. & Black, M. L. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci USA* **107** Suppl 1, 1779–1786, doi: 10.1073/pnas.0906079106 (2010).
- Gosadi, I. M., Goyder, E. C. & Teare, M. D. Investigating the potential effect of consanguinity on type 2 diabetes susceptibility in a Saudi population. *Hum Hered* **77**, 197–206, doi: 10.1159/000362447 (2014).
- Kuo, J. Z. *et al.* Trans-ethnic fine mapping identifies a novel independent locus at the 3' end of CDKAL1 and novel variants of several susceptibility loci for type 2 diabetes in a Han Chinese population. *Diabetologia* **56**, 2619–2628, doi: 10.1007/s00125-013-3047-1 (2013).
- Buyske, S. *et al.* Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS One* **7**, e35651, doi: 10.1371/journal.pone.0035651 (2012).
- Zuberi, A. R., Christianson, G. J., Mendoza, L. M., Shastri, N. & Roopenian, D. C. Positional cloning and molecular characterization of an immunodominant cytotoxic determinant of the mouse H3 minor histocompatibility complex. *Immunity* **9**, 687–698 (1998).

17. Salvatore, P., Hanash, C. R., Kido, Y., Imai, Y. & Accili, D. Identification of sirm, a novel insulin-regulated SH3 binding protein that associates with Grb-2 and FYN. *J Biol Chem* **273**, 6989–6997 (1998).
18. Soni, M. S. *et al.* Downregulation of carnitine acyl-carnitine transferase by miRNAs 132 and 212 amplifies glucose-stimulated insulin secretion. *Diabetes* **63**, 3805–3814, doi: 10.2337/db13-1677 (2014).
19. Al Hayek, A. A. *et al.* Prevalence of low testosterone levels in men with type 2 diabetes mellitus: a cross-sectional study. *J Family Community Med* **20**, 179–186, doi: 10.4103/2230-8229.122006 (2013).
20. Diaczok, D. *et al.* Deletion of Otx2 in GnRH neurons results in a mouse model of hypogonadotropic hypogonadism. *Mol Endocrinol* **25**, 833–846, doi: 10.1210/me.2010-0271 (2011).
21. Baccetti, B. *et al.* Insulin-dependent diabetes in men is associated with hypothalamo-pituitary derangement and with impairment in semen quality. *Hum Reprod* **17**, 2673–2677 (2002).
22. Bai, H. *et al.* The novel plasminogen receptor, plasminogen receptor(KT) (Plg-R(KT)), regulates catecholamine release. *J Biol Chem* **286**, 33125–33133, doi: 10.1074/jbc.M111.218693 (2011).
23. Iriyama, K., Teranishi, T., Mori, H., Nishiwaki, H. & Kusaka, N. Effects of exogenous catecholamines on glucose and fat metabolism and on triglycerides in the rat liver during total parenteral nutrition. *JPEN J Parenter Enteral Nutr* **8**, 412–415 (1984).
24. Edwards, Y. H., Putt, W., Fox, M. & Ives, J. H. A novel human phosphoglucomutase (PGM5) maps to the centromeric region of chromosome 9. *Genomics* **30**, 350–353, doi: 10.1006/geno.1995.9866 (1995).
25. Bastholt, L. *et al.* Oral idarubicin in the treatment of advanced breast cancer. *Acta Oncol* **28**, 893–897 (1989).
26. Cheng, H. *et al.* Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. *Hum Mol Genet* **13**, 1563–1575, doi: 10.1093/hmg/ddh173 (2004).
27. Coste, H. & Rodriguez, J. C. Orphan nuclear hormone receptor Rev-erbalph α regulates the human apolipoprotein CIII promoter. *J Biol Chem* **277**, 27120–27129, doi: 10.1074/jbc.M203421200 (2002).
28. De Ita, J. R. *et al.* Altered liver expression of genes involved in lipid and glucose metabolism in mice with partial IGF-1 deficiency: an experimental approach to metabolic syndrome. *J Transl Med* **13**, 326, doi: 10.1186/s12967-015-0684-9 (2015).
29. Bianda, T., Zapf, J., Froesch, E. R. & Schmid, C. Insulin-like growth factor-1 lowers fasting and postprandial triglyceride levels without affecting chylomicron clearance in healthy men. *J Intern Med* **245**, 455–462 (1999).
30. De Sanctis, V. *et al.* Does Insulin Like Growth Factor-1 (IGF-1) Deficiency Have a “Protective” Role in the Development of Diabetic Retinopathy in Thalassamia Major Patients? *Mediterr J Hematol Infect Dis* **7**, e2015038, doi: 10.4084/MJHID.2015.038 (2015).
31. Hypponen, E., Boucher, B. J., Berry, D. J. & Power, C. 25-hydroxyvitamin D, IGF-1, and metabolic syndrome at 45 years of age: a cross-sectional study in the 1958 British Birth Cohort. *Diabetes* **57**, 298–305, doi: 10.2337/db07-1122 (2008).
32. Chauhan, G. *et al.* Impact of common variants of PPAR γ , KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5,164 Indians. *Diabetes* **59**, 2068–2074, doi: 10.2337/db09-1386 (2010).
33. Sokolova, E. A., Bondar, I. A., Shabelnikova, O. Y., Pyankova, O. V. & Filipenko, M. L. Replication of KCNJ11 (p.E23K) and ABCC8 (p.S1369A) Association in Russian Diabetes Mellitus 2 Type Cohort and Meta-Analysis. *PLoS One* **10**, e0124662, doi: 10.1371/journal.pone.0124662 (2015).
34. Chang, Y. C. *et al.* Common variation in the fat mass and obesity-associated (FTO) gene confers risk of obesity and modulates BMI in the Chinese population. *Diabetes* **57**, 2245–2252, doi: 10.2337/db08-0377 (2008).
35. Tsuda, K. Leptin and nitric oxide in blood pressure regulation in humans. *Am J Hypertens* **27**, 1428, doi: 10.1093/ajh/hpu156 (2014).
36. Kadekaro, M. & Summy-Long, J. Y. Centrally produced nitric oxide and the regulation of body fluid and blood pressure homeostases. *Clin Exp Pharmacol Physiol* **27**, 450–459 (2000).
37. Charlesworth, J. C. *et al.* Toward the identification of causal genes in complex diseases: a gene-centric joint test of significance combining genomic and transcriptomic data. *BMC Proc* **3** Suppl 7, S92 (2009).
38. Peter, I. *et al.* Genetic modifiers of cardiorespiratory fitness response to lifestyle intervention. *Med Sci Sports Exerc* **46**, 302–311, doi: 10.1249/MSS.0b013e3182a66155 (2014).
39. Chavez, J. A. & Summers, S. A. A ceramide-centric view of insulin resistance. *Cell Metab* **15**, 585–594, doi: 10.1016/j.cmet.2012.04.002 (2012).
40. He, J. *et al.* PXR ablation alleviates diet-induced and genetic obesity and insulin resistance in mice. *Diabetes* **62**, 1876–1887, doi: 10.2337/db12-1039 (2013).
41. Zhang, W., Thompson, B. J., Hietakangas, V. & Cohen, S. M. MAPK/ERK signaling regulates insulin sensitivity to control glucose metabolism in *Drosophila*. *PLoS Genet* **7**, e1002429, doi: 10.1371/journal.pgen.1002429 (2011).
42. Feng, R. N. *et al.* Histidine supplementation improves insulin resistance through suppressed inflammation in obese women with the metabolic syndrome: a randomised controlled trial. *Diabetologia* **56**, 985–994, doi: 10.1007/s00125-013-2839-7 (2013).
43. Yamauchi, T. *et al.* Inhibition of RXR and PPAR γ ameliorates diet-induced obesity and type 2 diabetes. *J Clin Invest* **108**, 1001–1013, doi: 10.1172/JCI12864 (2001).
44. Bashiri, H. *et al.* Celiac disease in type-1 diabetes mellitus: coexisting phenomenon. *J Res Med Sci* **16** Suppl 1, S401–406 (2011).
45. Al-Hussaini, A., Sulaiman, N., Al-Zahrani, M., Alenizi, A. & El Haj, I. High prevalence of celiac disease among Saudi children with type 1 diabetes: a prospective cross-sectional study. *BMC Gastroenterol* **12**, 180, doi: 10.1186/1471-230X-12-180 (2012).
46. Al Hayek, A. A., Khawaja, N. M., Khader, Y. S., Jaffal, S. K. & Ajlouni, K. M. The prevalence of Hypogonadism among diabetic and non-diabetic men in Jordan. *J Diabetes Complications* **28**, 135–140, doi: 10.1016/j.jdiacomp.2013.11.004 (2014).
47. Mirzaei, M. R., Amini, M. & Aminoroaya, A. The prevalence of hypogonadism in diabetic men in Isfahan Endocrine and Metabolism Research Center, Isfahan, Iran. *J Res Med Sci* **17**, 602–606 (2012).
48. Al-Adsani, A. & Gader, F. A. Combined occurrence of diabetes mellitus and retinitis pigmentosa. *Ann Saudi Med* **30**, 70–75, doi: 10.4103/0256-4947.59381 (2010).
49. Rodriguez-Mulero, S. *et al.* Expression of concentrative nucleoside transporters SLC28 (CNT1, CNT2, and CNT3) along the rat nephron: effect of diabetes. *Kidney Int* **68**, 665–672, doi: 10.1111/j.1523-1755.2005.00444.x (2005).
50. Batrakou, D. G., Heras de Las, J. L., Czapiewski, R., Mouras, R. & Schirmer, E. C. TMEM120A and B: Nuclear Envelope Transmembrane Proteins Important for Adipocyte Differentiation. *PLoS One* **10**, e0127712, doi: 10.1371/journal.pone.0127712 (2015).
51. Nemoto, S., Taguchi, K., Matsumoto, T., Kamata, K. & Kobayashi, T. Pravastatin normalizes ET-1-induced contraction in the aorta of type 2 diabetic OLETF rats by suppressing the KSR1/ERK complex. *Am J Physiol Heart Circ Physiol* **303**, H893–902, doi: 10.1152/ajpheart.01128.2011 (2012).
52. Consortium, U. K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90, doi: 10.1038/nature14962 (2015).
53. Delahanty, R. J. *et al.* Evaluation of GWAS-identified genetic variants for age at menarche among Chinese women. *Hum Reprod* **28**, 1135–1143, doi: 10.1093/humrep/det011 (2013).
54. Cho, Y. S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67–72, doi: 10.1038/ng.1019 (2011).
55. Daimon, M. *et al.* Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population. *Biochem Biophys Res Commun* **365**, 466–471, doi: 10.1016/j.bbrc.2007.10.180 (2008).
56. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274–1283, doi: 10.1038/ng.2797 (2013).
57. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713, doi: 10.1038/nature09270 (2010).

58. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* **41**, 56–65, doi: 10.1038/ng.291 (2009).
59. Consortium, C. A. D. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25–33, doi: 10.1038/ng.2480 (2013).
60. Heid, I. M. *et al.* Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* **1**, 10–20, doi: 10.1161/CIRCGENETICS.108.776708 (2008).
61. Alsmadi, O. *et al.* Leptin in association with common variants of MC3R mediates hypertension. *Am J Hypertens* **27**, 973–981, doi: 10.1093/ajh/hpt285 (2014).
62. Alsmadi, O. *et al.* Response to leptin and nitric oxide in blood pressure regulation in humans. *Am J Hypertens* **27**, 1429–1430, doi: 10.1093/ajh/hpu178 (2014).
63. Gao, S. *et al.* Important roles of brain-specific carnitine palmitoyltransferase and ceramide metabolism in leptin hypothalamic control of feeding. *Proc Natl Acad Sci USA* **108**, 9691–9696, doi: 10.1073/pnas.1103267108 (2011).
64. Schneider, M. P. *et al.* Poor glycemic control is related to increased nitric oxide activity within the renal circulation of patients with type 2 diabetes. *Diabetes Care* **36**, 4071–4075, doi: 10.2337/dc13-0806 (2013).
65. Soranzo, N. Genetic determinants of variability in glycosylated hemoglobin (HbA(1c)) in humans: review of recent progress and prospects for use in diabetes care. *Curr Diab Rep* **11**, 562–569, doi: 10.1007/s11892-011-0232-9 (2011).
66. Elkum, N., Al-Arouj, M., Sharifi, M., Shaltout, A. & Bennakhi, A. Prevalence of childhood obesity in the state of Kuwait. *Pediatr Obes*, doi: 10.1111/ijpo.12090 (2015).
67. Voight, B. F. *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* **8**, e1002793, doi: 10.1371/journal.pgen.1002793 (2012).
68. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575, doi: S0002-9297(07)61352-4 [pii]10.1086/519795 (2007).
69. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664, doi: gr.094052.109 [pii]10.1101/gr.094052.109 (2009).
70. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909, doi: 10.1038/ng1847 (2006).
71. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221–227, doi: 10.1038/sj.hdy.6800717 (2005).
72. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* **47**, 381–386, doi: 10.1038/ng.3245 (2015).
73. Gerriets, V. A. & MacIver, N. J. Role of T cells in malnutrition and obesity. *Front Immunol* **5**, 379, doi: 10.3389/fimmu.2014.00379 (2014).
74. Stanford, K. I. *et al.* Heparan sulfate 2-O-sulfotransferase is required for triglyceride-rich lipoprotein clearance. *J Biol Chem* **285**, 286–294, doi: 10.1074/jbc.M109.063701 (2010).
75. Braccini, L. *et al.* PI3K-C2gamma is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. *Nat Commun* **6**, 7400, doi: 10.1038/ncomms8400 (2015).
76. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189–197, doi: 10.1038/ng.75 (2008).
77. Ewens, K. G., George, R. A., Sharma, K., Ziyadeh, F. N. & Spielman, R. S. Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. *Diabetes* **54**, 3305–3318 (2005).
78. Vaicik, M. K. *et al.* Laminin alpha4 deficient mice exhibit decreased capacity for adipose tissue expansion and weight gain. *PLoS One* **9**, e109854, doi: 10.1371/journal.pone.0109854 (2014).
79. Holmes, R. S. & Cox, L. A. Comparative studies of glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1: evidence for a eutherian mammalian origin for the GPIHBP1 gene from an LY6-like gene. *3 Biotech* **2**, 37–52, doi: 10.1007/s13205-011-0026-4 (2012).
80. Li, W. D. *et al.* A genome wide association study of plasma uric acid levels in obese cases and never-overweight controls. *Obesity (Silver Spring)* **21**, E490–494, doi: 10.1002/oby.20303 (2013).
81. Rain, S. *et al.* Right ventricular diastolic impairment in patients with pulmonary arterial hypertension. *Circulation* **128**, 2016–2025, 2011–2010, doi: 10.1161/CIRCULATIONAHA.113.001873 (2013).
82. Herman, D. S. *et al.* Truncations of titin causing dilated cardiomyopathy. *N Engl J Med* **366**, 619–628, doi: 10.1056/NEJMoa1110186 (2012).
83. Ottenheijm, C. A. *et al.* Titin and diaphragm dysfunction in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **173**, 527–534, doi: 10.1164/rccm.200507-1056OC (2006).

Acknowledgements

The authors are extremely thankful to the Biostatistics and Epidemiology Department for their efforts and excellent work on recruiting the participants, conducting the interviews, collecting the samples, and managing the data. The authors further thank Motasem Melhem for DNA QC and help with recruiting participants, Maisa Mahmoud for help with recruiting participants, Daisy Thomas for help with participant recruitment and phenotyping. The Tissue Bank Core Facility is acknowledged for sample processing and DNA extraction. The Ethical Review Committee and the Scientific Advisory Board at Dasman Diabetes Institute are acknowledged for approving the study. The authors thank Prof. Jaakko Tuomilehto for his support and guidance. The authors acknowledge the Kuwait Foundation for Advancement of the Sciences (KFAS) for funding this study that meets the objectives of projects RA-2010-005 (PI: OA; Co-PI: TAT) and RA-2010-004 (PI: NE).

Author Contributions

T.A.T., O.A. and N.E. performed the study design for this work and directed the study components. T.A.T., P.H., O.A., and N.E. developed the manuscript. P.H. performed all the data analysis and participated significantly in interpreting the results. S.E.J. participated in data analysis. F.A. performed the experiments of genome-wide genotyping. P.H. and N.E. are considered as joint first authors.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hebbar, P. *et al.* Genetic risk variants for metabolic traits in Arab populations. *Sci. Rep.* **7**, 40988; doi: 10.1038/srep40988 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017