

# Multiple genetic variants explain measurable variance in type 2 diabetes-related traits in Pakistanis

M. Islam · T. H. Jafar · A. R. Wood · N. M. G. De Silva ·  
M. Caulfield · N. Chaturvedi · T. M. Frayling

Received: 16 November 2011 / Accepted: 5 March 2012 / Published online: 28 April 2012  
© Springer-Verlag 2012

## Abstract

**Aims/hypothesis** Multiple genetic variants are associated with type 2 diabetes-related traits in Europeans, but their role in South Asian populations needs further study. We hypothesised that genetic variants associated with diabetes-related traits in Europeans would explain a similar proportion of phenotypic variance in a Pakistani population and could be used in Mendelian randomisation analyses.

**Methods** We used data from 2,131 individuals from the Control of Blood Pressure and Risk Attenuation Trial (COBRA) in Karachi, Pakistan. Individuals were aged 40 years or older.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-012-2560-y) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

M. Islam · T. H. Jafar (✉)  
Department of Community Health Sciences, Aga Khan University,  
Stadium Road, PO Box 3500, Karachi 74800, Pakistan  
e-mail: tazeen.jafar@aku.edu

T. H. Jafar  
Department of Medicine, Aga Khan University,  
Karachi, Pakistan

A. R. Wood · N. M. G. De Silva · T. M. Frayling  
Genetics of Complex Traits, Peninsula College of Medicine and  
Dentistry, University of Exeter,  
Exeter, UK

M. Caulfield  
Clinical Pharmacology, Barts and the London Genome Centre,  
Queen Mary University of London,  
London, UK

M. Caulfield  
William Harvey Research Institute, Barts and the London School  
of Medicine, Queen Mary University of London,  
London, UK

N. Chaturvedi  
International Centre for Circulatory Health,  
National Heart and Lung Institute, Imperial College London,  
London, UK

**Results** Combining information from multiple genetic variants showed that fasting glucose, BMI, triacylglycerol, and systolic and diastolic blood pressure variants explained 2.9%, 0.7%, 5.5%, 1.2% and 1.8% of the variance in those traits respectively. Genetic risk scores of fasting glucose, triacylglycerol, BMI, systolic blood pressure and diastolic blood pressure variants were associated with these traits, with per allele SD effects of 0.057 (95% CI 0.041, 0.074),  $p=3.44 \times 10^{-12}$ , 0.130 (95% CI 0.105, 0.155),  $p=2.9 \times 10^{-21}$ , 0.04 (95% CI 0.014, 0.072),  $p=0.004$ , 0.031 (95% CI 0.016, 0.047),  $p=7.9 \times 10^{-5}$ , 0.028 (95% CI 0.015, 0.042),  $p=5.5 \times 10^{-5}$ , respectively. These effects are consistent with those observed in Europeans, except that the effect of triacylglycerol variants in South Asians was slightly lower. Mendelian randomisation provided evidence that genetically influenced, raised triacylglycerol levels do not causally affect type 2 diabetes risk to the extent predicted from observational data ( $p=0.0003$  for difference between observed and instrumental variables correlations).

**Conclusions/interpretation** Genetic variants identified in Europeans are associated with type 2 diabetes-related traits in Pakistanis, with comparable effect sizes. Larger studies are needed to perform adequately powered Mendelian randomisation and help dissect the relationships between type 2 diabetes-related traits in diverse South Asian subgroups.

**Keywords** Fasting glucose · Genetic risk scores · Instrumental variable analysis · Mendelian randomisation · Triacylglycerol · Type 2 diabetes

## Abbreviations

COBRA Control of Blood Pressure and Risk Attenuation Trial  
GWAS Genome-wide association studies  
IV Instrumental variable  
SNP Single nucleotide polymorphism

## Introduction

The risk of type 2 diabetes is particularly elevated in South Asians compared with Western populations and is associated with greater morbidity and mortality rates [1]. About 20% of persons aged 40 years and older have diabetes in Pakistan, making it the seventh highest ranking country in terms of the global burden of diabetes. Interestingly, the interrelationships between intermediate metabolic traits such as BMI, blood pressure, lipids and glucose are inconsistent within different ethnic subgroups of South Asian populations compared with those observed in European-origin populations. For example, we have shown that Pakistani ethnic subgroups with the greatest burden of hypertension have the lowest relative prevalence of diabetes [2].

Genetic variants may be useful in helping us understand the complex relationships between metabolic traits and why these relationships may differ across populations. There are several reasons for further testing the role of genetic variants in type 2 diabetes-related traits in South Asian populations. First, genetic variants are randomly sorted in meiosis and are independent of non-genetic factors, including environmental risk factors, confounding factors or disease processes. This principle of Mendelian randomisation has been used to assess the causal relationship between several related metabolic phenotypes, e.g. C-reactive protein and heart disease [3], and sex hormone-binding globulin levels and diabetes [4]. Most recently, studies have combined multiple risk variants in a ‘genetic risk score’ to provide more power compared with individual variants. These studies have assessed the role of glucose in subclinical atherosclerosis [5] and triacylglycerol in type 2 diabetes [6]. Second, large meta-analyses of genome-wide association studies (GWAS) have recently increased the number of common variants robustly associated with type 2 diabetes-related quantitative traits to more than 100. These also include 95, 32 and 16 variants associated with lipid measures, BMI and fasting glucose, respectively, and most recently 29 variants associated with blood pressure [7–10]. Some of these large [10] meta-analyses showed that common variants identified in Europeans are associated with similar effect sizes in other ethnic groups [11, 12], but these studies tended to focus on single traits. Third, while individual common variants identified by GWAS have small effects with regard to the associated phenotype, they can, when combined, explain a more substantial proportion of phenotypic variance. For example, genetic loci associated with triacylglycerol or fasting glucose levels explain ~7.4% and ~3% of the variance in triacylglycerol [9] and fasting glucose [7] levels in European populations, respectively. These effects are similar in magnitude to well established risk factors, such as the effect of BMI on fasting glucose levels [13].

Testing variants against diabetes-related traits can be useful to help understand the relationship between those traits and how those relationships differ across ethnic groups. Thus, with regard to the *FTO* variant known to be associated with BMI in Europeans, we have recently shown that its association with type 2 diabetes differs between Europeans and South Asians, probably because BMI is a poorer surrogate of adiposity in South Asians [14].

The objectives of this study were: (1) to determine the extent to which genetic risk scores based on common gene variants detected in European-origin populations explain phenotypic variance in type 2 diabetes-related traits in a South Asian context; and (2) to use these genetic risk scores to detect any potential causal relationships of substantial magnitude in Mendelian randomisation analyses in South Asians. We used individual data from the Control of Blood Pressure and Risk Attenuation Trial (COBRA) cohort study in Karachi, a population-based study with measures of multiple metabolic traits and consisting of 2,073 (72%) participants of Muhajir ethnicity. We focused on three traits strongly associated with type 2 diabetes: fasting glucose levels, BMI and triacylglycerol levels. We also used the most recently identified blood pressure variants [12], because of the strong association between these traits in different ethnic subgroups in Pakistan [2].

## Methods

Details of participants are given in Table 1 and Fig. 1. Ethical approval was obtained from the Ethics Review Committee at the Aga Khan University, Karachi, Pakistan.

### Study design

A population-based sample representative of adults in Karachi was recruited in a cluster randomised trial of strategies for control of hypertension (COBRA, ClinicalTrials.gov ID no. NCT00327574). The sampling details have been described previously [15]. Briefly, a multi-stage cluster random sampling design was used to randomly select 12 geographical clusters in Karachi, the largest metropolitan city in Pakistan. A census was done and a listing made of all individuals from all households in the selected areas. All participants aged 40 years or above and residing in the same household were invited to participate in the study by trained community health workers. All participants were evaluated after obtaining informed consent. A range of anthropometric and biochemical data were collected from all consenting participants, with and without hypertension, who had been screened for eligibility for the trial. The enrolled group was categorised according to the five main Pakistani ethnic subgroups into Muhajirs, Punjabis, Sindhis, Baluchis, Pashtun and others (Table 1)

**Table 1** Baseline characteristics of the enrolled participants

Characteristics <i>n</i>	Total 2,880	Type 2 diabetic participant 617	Non-diabetic participant 2,263	<i>p</i> value <sup>a</sup>
Age (years)	51.5±10.7	53.2±10.4	51.1±10.8	9.03×10 <sup>-6</sup>
Muhajir, <i>n</i> (%)	2,073 (72.0)	482 (78.1)	1,591 (70.3)	
Sindhi, <i>n</i> (%)	236 (8.2)	34 (5.5)	202 (8.9)	3.23×10 <sup>-4</sup>
Punjabi, <i>n</i> (%)	90 (3.1)	22 (3.6)	68 (3.0)	
Others, <i>n</i> (%)	481 (16.7)	79 (12.8)	402 (17.8)	
BMI (kg/m <sup>2</sup> )	25.75±5.47	27.07±5.43	25.39±5.42	1.12×10 <sup>-11</sup>
Waist circumference (cm)	89.67±12.44	94.11±11.65	88.45±12.38	9.03×10 <sup>-25</sup>
Fasting glucose (mmol/l)	6.41±2.82	10.42±3.96	5.34±0.57	1.46×10 <sup>-132</sup>
Systolic BP (mmHg)	136.86±23.65	143.66±24.63	135.00±23.03	1.20×10 <sup>-14</sup>
Diastolic BP (mmHg)	86.02±12.79	88.97±12.43	85.48±12.84	1.82×10 <sup>-5</sup>
Triacylglycerol (mmol/l)	1.78±1.07	2.24±1.39	1.66±0.93	1.83×10 <sup>-21</sup>
Serum cholesterol (mmol/l)	4.86±1.02	5.02±1.13	4.82±0.98	3.98×10 <sup>-5</sup>
Men, <i>n</i> (%)	1,376 (47.8)	264 (42.8)	1,112 (49.1)	0.005
Hypertension <sup>b</sup> , <i>n</i> (%)	1,270 (44.1)	362 (58.7)	908 (40.1)	1.94×10 <sup>-16</sup>

Data are mean±SD for all continuous quantitative traits

<sup>a</sup> For comparison of the characteristics between diabetic and non-diabetic individuals

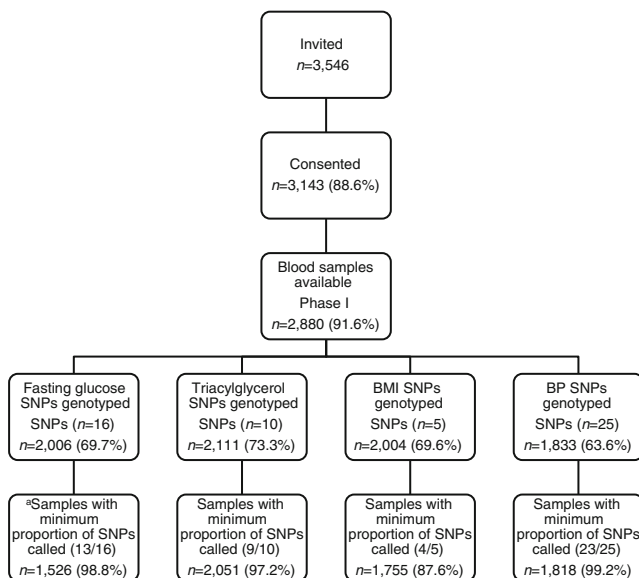
<sup>b</sup> Defined as patient taking antihypertensive drugs, or as mean systolic BP ≥140 mmHg or mean diastolic BP ≥90 mmHg

[16, 17]. The numbers of individuals involved differed slightly by trait, as genotyping success rates differed slightly and individuals with diabetes had to be excluded from the fasting glucose analysis.

### SNP selection

We only selected single nucleotide polymorphisms (SNPs) from GWAS studies of European individuals because these

studies include the majority of GWAS findings. We aimed to test the effects of SNPs identified through European GWAS in Pakistanis. We selected the ten SNPs most strongly associated with triacylglycerol levels [8, 9] (excluding *FADS1* rs174547 and *GCKR* rs1260326 because they are strongly associated with several other quantitative traits relevant to diabetes), the five SNPs most strongly associated with BMI [10], 16 SNPs associated with fasting glucose [7] and 29 SNPs associated with blood pressure [12].



**Fig. 1** Flow diagram of study procedures. <sup>a</sup>Participants with type 2 diabetes excluded *n*=462 (23%), where per cent is of *n*=2,006

### Genotyping

We genotyped all SNPs except the five BMI SNPs, using a modified Taqman assay, KASPAR assay ([www.kbioscience.co.uk](http://www.kbioscience.co.uk)). The five BMI SNPs were genotyped using a pre-designed Taqman SNP genotyping assay from Applied Biosystems (Warrington, UK), followed by genotype clustering using Klustercaller software (Kbiosciences, Hoddesdon, UK). The final number of analysed individuals differed by trait because of different genotype success rates (we required individuals to have >85% of SNPs successfully genotyped for any one trait typed) and, for fasting glucose, because we excluded individuals with diabetes.

**Fasting glucose SNPs** We excluded participants with diabetes as defined by use of glucose-lowering medication or fasting glucose ≥7 mmol/l, since these individuals would be on glucose-lowering agents, making their fasting glucose levels uninformative. Appropriate samples were available from 1,544 individuals, of which 18 (1%) failed four or

more of the 16 attempted SNPs in the batch; these samples were excluded. The call rates were therefore generated from a subset of 1,526 samples, for which at least 13 of 16 SNPs were called (Fig. 1). Call rates by SNP in these 1,526 samples ranged from 96.81% to 99.4%.

**Triacylglycerol SNPs** We used 2,111 participants, including those with diabetes. Of these, 60 (3%) failed two or more of the ten attempted SNPs in the batch and we excluded these samples. The call rates were therefore generated from a subset of 2,051 samples, for which at least nine of ten SNPs were called (Fig. 1). Call rates by SNP ranged from 98.5% to 100%.

**BMI SNPs** We used 2,004 participants, including those with diabetes. Of these, 249 (12%) failed two or more of the five attempted SNPs in the batch and were excluded. The call rates were therefore generated from a subset of 1,755 samples, for which at least four of the five SNPs were called (Fig. 1). Call rates by SNP ranged from 88.7% to 99.4%.

**Blood pressure SNPs** We used 1,833 samples. Of these, 15 (0.8%) failed 22 or more of the 25 attempted SNPs in the batch and were excluded. The call rates were therefore generated from a subset of 1,818 samples, for which at least 23 of 25 SNPs were called (Fig. 1). Call rates by SNP ranged from 94.2% to 98.6%.

#### Statistical analyses

**Association of SNPs with phenotypes in South Asians** We  $\log_{10}$ -transformed triacylglycerol levels, and systolic and diastolic blood pressure. Fasting glucose did not require transformation. We generated within-study  $z$  scores for fasting glucose and  $\log_{10}$ -transformed triacylglycerol using the mean and SD of the samples. We used individual SNPs as independent variables and outcome measures as dependent variables in a per-allele test, with age, sex, BMI, ethnicity and clustering of household member. We coded ethnicity as a non-ordinal value of 1, 2, 3 and 4 for the Muhajir, Sindhi, Punjabi and other ethnicities respectively. For individual SNP associations, we did not have adequate power to replicate the effects observed in European samples, and we did not correct single SNP analyses for multiple testing. Instead, we report the number of SNPs associated in the same direction as European associations at  $p < 0.05$  and compare this with the one of 20 SNPs expected by chance. We also tested the significance of genetic risk scores that combine information from all known variants. For these tests we had sufficient power, and based on the combined variants explaining  $>0.7\%$  of the phenotypic variance and a minimal sample size of 1,500 individuals, we had  $>89\%$  power to detect effects at  $p = 0.05$ .

We created weighted allele scores and accounted for the varying, previously reported effect sizes of each SNP using Eq. (1), where  $w$  is the  $\beta$ -coefficient from the individual regressions of SNP genotype against the outcome.

$$\text{Weighted score} = (w_1 \times \text{SNP}_1) + (w_2 \times \text{SNP}_2) + \dots + (w_n \times \text{SNP}_n) \quad (1)$$

We rescaled the weighted score to reflect the number of available SNPs using Eq. (2) as described in Lin et al. [18]:

$$\text{Allele score} = \frac{\text{Weighted score} \times \text{Number of SNPs available}}{\text{Sum weight of the available SNPs}} \quad (2)$$

We used this allele score as the independent variable, and fasting glucose, triacylglycerol, BMI and systolic or diastolic blood pressure as the dependent variables, with age, sex, ethnicity, clustering of household member and BMI (except for the BMI analysis) as covariates in linear regression analyses.

For all traits, we also performed sensitivity analysis by computing the allele score based on only those alleles associated with the dependent variables at  $p < 0.05$ , and assessed their significance and the variability that was explained for each model.

We also evaluated the association between SNPs and secondary phenotypes. For this, we used the weighted allele score as independent variable and type 2 diabetes status (defined as use of glucose-lowering medications or a fasting blood glucose level of  $\geq 7.0$  mmol/l), hypertension status (defined as mean systolic blood pressure  $\geq 140$  mmHg or mean diastolic blood pressure  $\geq 90$  mmHg or taking anti-hypertensive medication) or hypertriacylglycerolaemia (defined as serum triacylglycerol  $> 1.7$  mmol/l) as the dependent variables in logistic regression analyses, with age and sex as covariates.

**Mendelian randomisation and instrumental variables analysis** We performed a series of instrumental variables (IV)-based Mendelian randomisation analyses on several pairs of traits as described below. We used the weighted allele score associated with the primary trait as the ‘instrument’ to test the relationship between each pair of primary and secondary phenotypes. An IV analysis relates the variation in the potentially causal risk factor of interest (here one of the quantitative traits) that is influenced by the ‘instrument’ (here the weighted allele score) to the outcome (here one of the quantitative traits or type 2 diabetes). This method assumes that the IV (1) is not associated with measured or unmeasured confounders (likely to be true for genetic variants) [11] and (2) is only related to the outcome via its effect on the risk factor. This produces an estimate of the causal effect in a similar way to an intention to treat analysis in a randomised controlled trial [11]. To ensure our tests were as close as possible to meeting these

assumptions, we excluded two SNPs (in or near *GCKR* and *FADS1*) known to have multiple effects on metabolic phenotypes.

When the outcome was a continuously distributed trait, we performed the IV estimation for each outcome in each study using the two-stage least squares estimator, implemented in the Stata command ‘ivreg2’. We tested for a difference between the IV and observational estimates using the Durbin–Wu–Hausman test of endogeneity.

When the outcome was a dichotomous trait (type 2 diabetes case–control status) we performed IV analysis using a logistic control function estimator. The analysis was performed in two stages. In the first stage, we assessed the observational association between the weighted allele score and the triacylglycerol *z* score. We saved the predicted values and residuals from this regression model. In the second stage, we used the predicted values from stage 1 as the independent variable (reflecting an unconfounded estimate of triacylglycerol levels due to these genotypes) and diabetes status as the dependent variable in a logistic regression analysis. The residuals from stage 1 were included as a covariate, representing residual variation in triacylglycerol levels that is not due to these genotypes. We then used a Wald test to assess the evidence of any difference between the predicted values coefficient (IV estimate of the causal effect of triacylglycerol levels on type 2 diabetes) and the residuals coefficient as a test of endogeneity.

The sensitivity analysis for the IV analysis was performed using the weighted allele score from the model only with alleles significantly associated with the outcomes.

Given a minimum sample size of 1,500, we had 80% power, at  $p < 0.05$ , to detect combinations of genetic variants that explain  $>0.5\%$  of phenotypic variance. It is difficult to estimate the statistical power of Mendelian randomisation tests, but an approximate method is to take the product of the SNP–primary trait and primary trait–secondary trait effects. For example, an effect of  $0.5\%$  variance would require a SNP–primary trait association explaining  $5\%$  variance and a primary trait–secondary trait association explaining  $10\%$  variance. Most associations in our study were weaker than this and power therefore lower than  $80\%$  in most tests. However, given that these calculations are very approximate, we present the results of the tests, which in themselves provide better power calculations for further study [19, 20]. All statistical analyses were performed in Stata/IC v.12.1 for Windows (Stata, College Station, TX, USA).

## Results

### Study procedures

Figure 1 illustrates the study flow diagram. Of the 3,546 invited participants aged 40 years or above, 3,143 (88.6%)

consented to enrol in the population-based study. We attempted genotyping of 2,131 individuals from whom DNA was available at the time of the genetic study.

### Fasting glucose SNPs

Results from the individual 16 SNPs and the weighted allele score are shown in Table 2. Of the 16 SNPs, seven were associated with fasting glucose at  $p < 0.05$  (all with the same allele direction as reported in Europeans,  $p = 0.00001$ , binomial, compared with the number expected by chance), and the weighted allele score was very strongly associated with fasting glucose ( $p = 8.9 \times 10^{-9}$ ). Together the SNPs explained  $2.9\%$  (95% CI 1.2, 4.6%) of the variance in fasting glucose or  $3.4\%$  in sensitivity analysis consisting of only the seven SNPs at  $p < 0.05$  as shown in electronic supplementary material (ESM) Table 1, a finding that is similar to the  $3\%$  (95% CI 2.7, 3.3%) reported in Europeans [7]. The association of the weighted allele score with fasting glucose is shown in Fig. 2a.

### Triacylglycerol SNPs

Results from the individual ten SNPs and the weighted allele score are shown in Table 3. Of the ten SNPs, six were associated with triacylglycerol at  $p < 0.05$  (all with the same allele direction as reported in Europeans,  $p = 0.000005$ , binomial, compared with the number expected by chance), and the weighted allele score was very strongly associated with triacylglycerol ( $p = 1.3 \times 10^{-25}$ ). Together the SNPs explained  $5.5\%$  (95% CI 3.5, 7.5%) of the variance in triacylglycerol or  $5.4\%$  in sensitivity analysis consisting of only the six SNPs at  $p < 0.05$  (ESM Table 1), which is slightly less, based on 95% CIs, than the  $7.4\%$  (95% CI 6.7, 8.1%) reported in Europeans [9]. The association of the weighted allele score with triacylglycerol levels is shown in Fig. 2b.

### BMI SNPs

Results from the individual five SNPs and the weighted allele score are shown in Table 4. Of the five SNPs, two were associated with fasting glucose ( $p = 0.007$  and  $p = 0.059$ ) *FTO* and *GNPDA2*, respectively, with the same allele direction as reported in Europeans ( $p = 0.045$ , binomial, compared with the number expected by chance); the weighted allele score was associated with BMI ( $p = 0.00028$ ). Together the SNPs explained  $0.7\%$  (95% CI  $-0.2$ , 1.6%) (or  $0.9\%$  in sensitivity analysis consisting of the two most strongly associated SNPs [ESM Table 1]) of the variance in BMI, consistent with the  $1.4\%$  (95% CI 1.3, 1.5%) reported in Europeans for a wider set of 32 SNPs [10]. The association of the weighted allele score with BMI is shown in Fig. 2c.

**Table 2** Effects of 16 fasting glucose SNPs on fasting glucose in South Asian and European individuals

Gene			Reference study			South Asian			<i>p</i> value	
SNP	Nearest gene <sup>a</sup>	Allele (effect/other)	EAF	Effect size	95% CI	EAF	Effect size <sup>b</sup>	95% CI		
rs10830963	MTNR1B	G/C	0.30	0.067	(0.061, 0.073)	0.38	0.180 <sup>c</sup>	(0.111, 0.248)	2.85 × 10 <sup>-7</sup>	
rs560887	G6PC2	C/T	0.70	0.075	(0.069, 0.081)	0.89	0.204 <sup>c</sup>	(0.098, 0.309)	0.0002	
rs174547	FADS1	T/C	0.64	0.017	(0.011, 0.023)	0.85	0.112 <sup>c</sup>	(0.019, 0.205)	0.019	
rs7903146	TCF7L2	T/C	0.28	0.023	(0.015, 0.031)	0.30	0.093 <sup>c</sup>	(0.019, 0.167)	0.013	
rs2191349	DGKB– TMEM195 <sup>c</sup>	T/G	0.52	0.030	(0.024, 0.036)	0.61	0.082 <sup>c</sup>	(0.014, 0.150)	0.019	
rs7034200	GLIS3	A/C	0.49	0.018	(0.012, 0.024)	0.53	0.078 <sup>c</sup>	(0.010, 0.145)	0.024	
rs340874	PROX1	C/T	0.52	0.013	(0.007, 0.019)	0.53	0.073 <sup>c</sup>	(0.005, 0.141)	0.034	
rs179884	GCK (proxy rs4607517)	T/G	0.16	0.062	(0.054, 0.070)	0.26	0.061 <sup>c</sup>	(−0.015, 0.136)	0.116	
rs7944584	MADD	A/T	0.75	0.021	(0.015, 0.027)	0.82	0.086 <sup>c</sup>	(0.000, 0.172)	0.051	
rs780094	GCKR	C/T	0.62	0.029	(0.023, 0.035)	0.76	−0.038 <sup>c</sup>	(−0.116, 0.040)	0.334	
rs11920090	SLC2A2	T/A	0.13	0.020	(0.012, 0.028)	0.16	−0.037 <sup>c</sup>	(−0.130, 0.055)	0.429	
rs11605924	CRY2	A/C	0.49	0.015	(0.009, 0.021)	0.48	0.009 <sup>c</sup>	(−0.058, 0.076)	0.787	
rs11708067	ADCY5	A/G	0.78	0.027	(0.021, 0.033)	0.76	−0.015 <sup>c</sup>	(−0.095, 0.064)	0.706	
rs11071657	C2CD4B	A/G	0.63	0.008	(0.002, 0.014)	0.72	0.010 <sup>c</sup>	(−0.067, 0.087)	0.796	
rs10885122	ADRA2A	G/T	0.87	0.022	(0.014, 0.030)	0.78	−0.024 <sup>c</sup>	(−0.104, 0.057)	0.564	
rs11558471	SLC30A8	A/G	0.68	0.027	(0.019, 0.035)	0.74	0.047 <sup>c</sup>	(−0.031, 0.125)	0.239	
Weighted allele score								0.057 <sup>d</sup>	(0.041, 0.074)	3.44 × 10 <sup>-12</sup>

South Asians, *n*=1,526

EAF, Effect allele frequency

<sup>a</sup> Nearest gene information as in Dupuis et al [7]

<sup>b</sup> Adjusted for age, sex, BMI, ethnicity and clustering of household member

<sup>c</sup> Effect size for change in fasting glucose *z* score per fasting glucose-raising allele per individual SNPs

<sup>d</sup> Effect size for change in fasting glucose *z* score per unit increase in weighted allele score

<sup>e</sup> *TMEM195*, also known as *AGMO*

## Blood pressure SNPs

Results from the individual 25 SNPs and the weighted allele score are shown in Table 5. Of the 25 SNPs, four were associated with systolic and diastolic blood pressure at *p* < 0.05 (all with the same allele direction as reported in Europeans, *p*=0.068, binomial), and the weighted allele score was very strongly associated with systolic blood pressure. Together the SNPs explained 1.2% (95% CI 0.02, 2.38%) and 1.8% (95% CI 0.4, 3.2%) of the variance in systolic and diastolic blood pressure, respectively (0.9% and 1.3%, respectively, in sensitivity analysis [ESM Table 1]). These estimates of variance in systolic and diastolic blood pressure are similar to the 0.92% (95% CI 0.80, 1.04) and 0.94% (95% CI 0.82, 1.06) reported in Europeans [12]. The associations of the weighted allele score with systolic and diastolic blood pressure are shown in Fig. 2d, e.

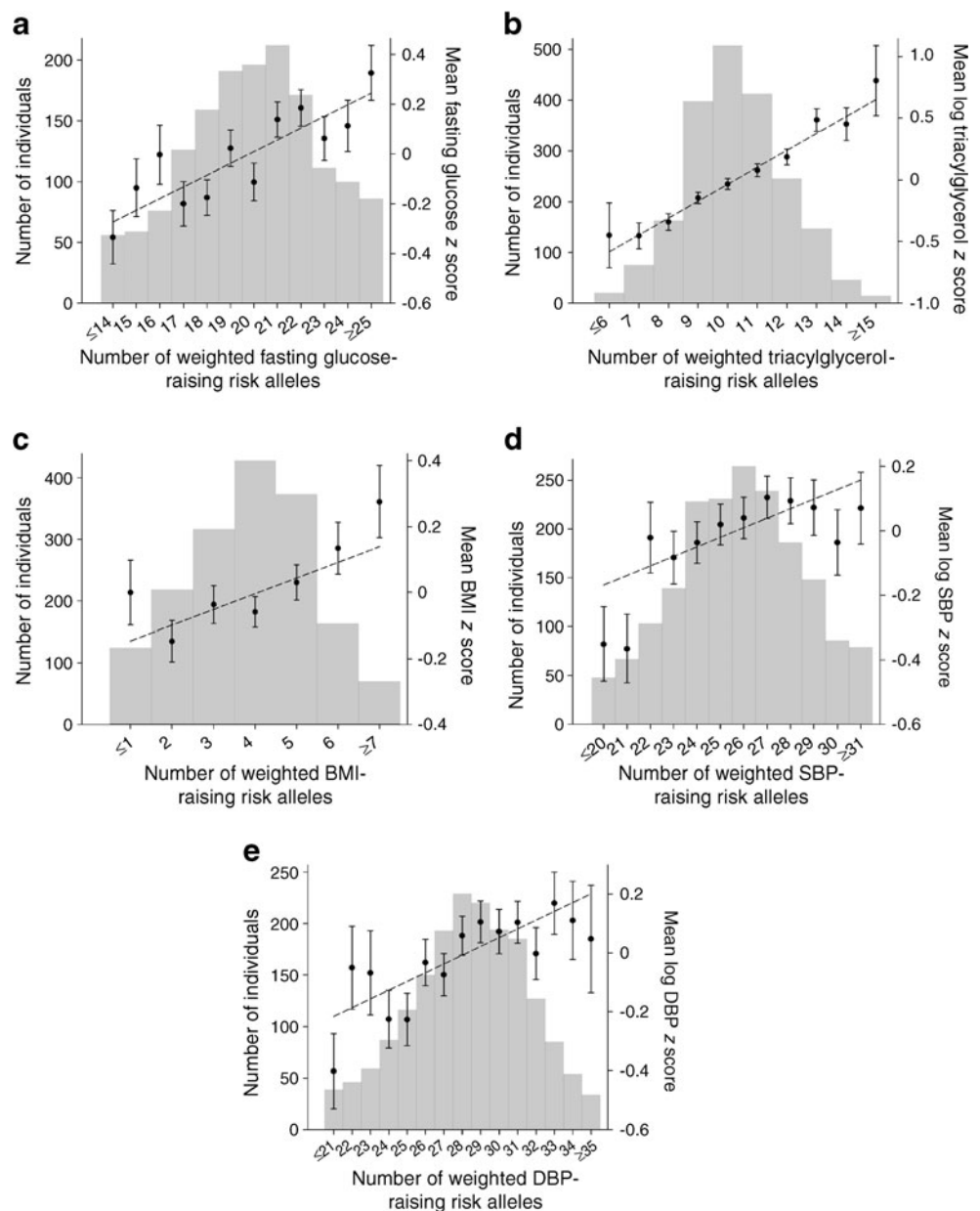
## Association of SNPs with secondary phenotypes

We observed a significant association between the weighted allele score of BMI and hypertriacylglycerolaemia (*p*=0.014). No other significant associations were detected (Table 6).

## Mendelian randomisation analysis

Significant observational associations were present between primary and secondary phenotypes as seen in Table 7, except between BMI and systolic blood pressure. For every 1 SD increase in log<sub>10</sub>-transformed triacylglycerol levels, the OR (95% CI) for type 2 diabetes was 1.77 (95% CI 1.59, 1.98), *p*=1.38 × 10<sup>-24</sup>. Each 1 SD increase in log<sub>10</sub>-transformed triacylglycerol levels was associated with a 0.039 SD (95% CI 0.029, 0.049) increase in fasting glucose levels. Every one unit increase in BMI was associated with a 0.353

**Fig. 2** The combined impact of the 16 SNPs associated with fasting glucose on fasting glucose levels in the control group **(a)**. **(b)** The combined impact, in all individuals, of SNPs associated with triacylglycerol (ten SNPs), **(c)** BMI (five SNPs), and **(d)** systolic BP (SBP) and **(e)** diastolic BP (DBP) (25 SNPs) on circulating triacylglycerol, BMI, and systolic and diastolic BP, respectively. Participants were grouped by the total number of trait-increasing alleles at all trait-associated SNPs. Grey bars, number of individuals in each trait-increasing allele group; circles, mean trait z score within each trait-increasing allele group; error bars, SEM of the observed mean trait; dashed line, fitted mean trait level



(95% CI 0.224, 0.482) mmHg increase of diastolic blood pressure, and 0.045 SD (95% CI 0.035, 0.054) and 0.017 SD (95% CI 0.014, 0.020) increases in triacylglycerol and fasting glucose, respectively (Table 7).

We found evidence against the hypothesis that genetically raised triacylglycerol levels are causally related to type 2 diabetes status, in keeping with recently reported results in Europeans [6]. We also found nominal evidence against the hypothesis that raised fasting glucose levels are causally related to triacylglycerol levels. In the triacylglycerol versus type 2 diabetes status and fasting glucose versus triacylglycerol analyses, the instrumental variable estimates of the associations were lower than the observed associations ( $p=0.0003$  and  $p=0.05$  for difference, respectively)

(Table 7). The sensitivity analysis revealed that findings were consistent (ESM Table 2).

**Discussion**

We have tested the role of multiple genetic variants in five metabolic phenotypes that are especially relevant to diabetes and related outcomes in a Pakistani population. Our results show that genetic risk scores containing multiple SNPs are associated with fasting glucose levels, triacylglycerol levels, BMI, and systolic and diastolic blood pressure in Pakistanis. These risk scores have similar effects on the total phenotypic variance explained in Pakistanis as they do in Europeans.

**Table 3** Effects of ten triacylglycerol SNPs in South Asian and European individuals on triacylglycerol

Gene			Reference study			South Asian			
SNP	Nearest gene <sup>a</sup>	Allele (effect/other)	EAF	Effect size	95% CI	EAF	Effect size <sup>b</sup>	95% CI	<i>p</i> value
rs7819412	XKR6– AMACIL2 <sup>c</sup>	A/G	0.52	0.04	(0.001, 0.079)	0.46	−0.025 <sup>c</sup>	(−0.084, 0.034)	0.399
rs7679	PLTP	C/T	0.19	0.07	(0.031, 0.109)	0.20	0.024 <sup>c</sup>	(−0.049, 0.097)	0.523
rs7557067	APOB	A/G	0.78	0.08	(0.041, 0.119)	0.50	0.018 <sup>c</sup>	(−0.04, 0.076)	0.536
rs714052	MLXIPL	A/G	0.88	0.16	(0.101, 0.219)	0.91	0.150 <sup>c</sup>	(0.048, 0.252)	0.004
rs662799	APOA5	G/A	0.06	0.25	(0.191, 0.309)	0.18	0.306 <sup>c</sup>	(0.231, 0.381)	1.77×10 <sup>−15</sup>
rs328	LPL	C/G	0.91	0.19	(0.151, 0.229)	0.91	0.131 <sup>c</sup>	(0.031, 0.232)	0.011
rs3135506	APOA5	C/G	0.06	0.24	(0.181, 0.299)	0.03	0.257 <sup>c</sup>	(0.091, 0.424)	0.002
rs2954029	TRIB1	A/T	0.66	0.11	(0.071, 0.149)	0.64	0.075 <sup>c</sup>	(0.014, 0.137)	0.016
rs17216525	CSPG3 <sup>f</sup> , CILP2, PBX4	C/T	0.93	0.11	(0.051, 0.169)	0.88	0.149 <sup>c</sup>	(0.059, 0.239)	0.001
rs10889353	ANGPTL3	A/C	0.67	0.05	(0.011, 0.089)	0.58	0.114 <sup>c</sup>	(0.055, 0.174)	0.0002
Weighted allele score							0.130 <sup>d</sup>	(0.105, 0.155)	2.9×10 <sup>−21</sup>

South Asians, *n*=2,051

EAF, Effect allele frequency

<sup>a</sup> Nearest gene information reported as in Kathiresan et al [9], except for rs328, which is from Kathiresan et al [8], and rs3135506 and rs662799 from De Silva et al [6]

<sup>b</sup> Adjusted for age, sex, BMI, ethnicity and clustering of household member

<sup>c</sup> Effect size for change in log<sub>10</sub>-triacylglycerol *z* score per triacylglycerol-raising allele for individual SNPs

<sup>d</sup> Effect size for change in log<sub>10</sub>-triacylglycerol *z* score per unit increase in weighted allele score

<sup>e</sup> *AMACIL2*, also known as *SLC35G5*

<sup>f</sup> *CSPG3*, also known as *NCAN*

The distribution of allele frequencies in the study population was also comparable to that reported in participants of European origin, indicating that the attributable risk from

these genes is likely to be similar across the two populations [7–10]. These results are consistent with data reported by recent large GWAS for lipids [21] and blood pressure [12],

**Table 4** Effects of five BMI SNPs in South Asian and European individuals on BMI

Gene			Reference study			South Asian			
SNP	Nearest gene <sup>a</sup>	Allele (effect/other)	EAF	Effect size	95% CI	EAF	Effect size <sup>b</sup>	95% CI	<i>p</i> value
rs9939609	FTO	A/T	0.42	0.390	(0.361, 0.429)	0.30	0.10 <sup>c</sup>	(0.027, 0.171)	0.007
rs10767664	BDNF	A/T	0.78	0.190	(0.131, 0.249)	0.71	−0.04 <sup>c</sup>	(−0.111, 0.026)	0.224
rs10938397	GNPDA2	G/A	0.40	0.180	(0.141, 0.219)	0.40	0.06 <sup>c</sup>	(−0.002, 0.123)	0.059
rs543874	SEC16B	G/A	0.19	0.220	(0.083, 0.357)	0.15	0.03 <sup>c</sup>	(−0.052, 0.121)	0.431
rs7359397	SH2B1	T/C	0.43	0.150	(0.111, 0.189)	0.22	0.01 <sup>c</sup>	(−0.071, 0.082)	0.885
Weighted allele score							0.04 <sup>d</sup>	(0.014, 0.072)	0.004

South Asians, *n*=1,755

EAF, Effect allele frequency

<sup>a</sup> Nearest gene information as previously reported [10]

<sup>b</sup> Adjusted for age, sex, ethnicity and clustering of household member

<sup>c</sup> Effect size for change in BMI *z* score per BMI-raising allele for individual SNPs

<sup>d</sup> Effect size for change in BMI *z* score per unit increase in weighted allele score



**Table 5** Effects of 25 systolic and diastolic blood pressure SNPs in South Asian and European individuals on systolic blood pressure

Gene	Reference study										South Asian									
	SBP					DBP					SBP					DBP				
	SNP	Nearest gene <sup>a</sup>	Allele (effect/other)	EAF	β-coefficient	95% CI	Effect size	95% CI	EAF	Effect size <sup>b</sup>	95% CI	p value	Effect size <sup>b</sup>	95% CI	p value					
rs11191548	CYP17A1–NT5C2	T/C	0.91	0.464	(0.337, 0.591)	1.095	(0.891, 1.299)	0.79	0.138 <sup>c</sup>	(0.063, 0.212)	0.0003	0.114 <sup>c</sup>	(0.036, 0.192)	0.004						
rs11737771	NPR3 (C5ORF23)	C/T	0.60	0.504	(0.384, 0.624)	0.261	(-0.051, 0.335)	0.61	0.012 <sup>c</sup>	(-0.051, 0.075)	0.718	0.024 <sup>c</sup>	(-0.042, 0.089)	0.477						
rs11953630	EBF1	C/T	0.63	0.412	(0.290, 0.534)	0.281	(-0.037, 0.357)	0.63	0.027 <sup>c</sup>	(-0.037, 0.091)	0.407	0.058 <sup>c</sup>	(-0.009, 0.124)	0.088						
rs12940887	ZNF652	T/C	0.38	0.362	(0.251, 0.473)	0.271	(-0.077, 0.340)	0.15	0.011 <sup>c</sup>	(-0.077, 0.098)	0.813	-0.007 <sup>c</sup>	(-0.098, 0.084)	0.877						
rs12946454	PLCD3	T/A	0.26	0.210	(0.106, 0.314)	0.447	(-0.096, 0.608)	0.28	-0.028 <sup>c</sup>	(-0.096, 0.041)	0.426	-0.023 <sup>c</sup>	(-0.094, 0.047)	0.519						
rs13082711	SLC4A7	C/T	0.22	0.315	(0.187, 0.443)	0.238	(-0.073, 0.316)	0.17	0.012 <sup>c</sup>	(-0.073, 0.096)	0.786	-0.026 <sup>c</sup>	(-0.113, 0.062)	0.562						
rs13107325	SLC39A8	C/T	0.50	0.981	(0.728, 1.234)	0.684	(-0.341, 0.843)	0.75	0.196 <sup>c</sup>	(-0.341, 0.734)	0.474	0.031 <sup>c</sup>	(-0.53, 0.591)	0.915						
rs13139571	GUCY1A3–GUCY1B3	C/A	0.76	0.321	(0.191, 0.451)	0.260	(-0.027, 0.340)	0.77	0.047 <sup>c</sup>	(-0.027, 0.122)	0.209	0.036 <sup>c</sup>	(-0.042, 0.113)	0.368						
rs1327235	JAG1	G/A	0.46	0.340	(0.221, 0.459)	0.302	(-0.073, 0.376)	0.45	-0.012 <sup>c</sup>	(-0.073, 0.05)	0.709	0.003 <sup>c</sup>	(-0.061, 0.067)	0.916						
rs1378942	CYP11A1–ULK3	G/T	0.65	0.416	(0.340, 0.492)	0.613	(-0.049, 0.735)	0.77	0.025 <sup>c</sup>	(-0.049, 0.099)	0.503	0.027 <sup>c</sup>	(-0.05, 0.103)	0.494						
rs17249754	ATP2B1	G/A	0.84	0.522	(0.389, 0.655)	0.928	(0.025, 1.136)	0.67	0.091 <sup>c</sup>	(0.025, 0.157)	0.007	0.063 <sup>c</sup>	(-0.006, 0.132)	0.072						
rs17367504	MTHFR–NPPB	A/G	0.85	0.547	(0.427, 0.667)	0.903	(0.028, 1.087)	0.80	0.104 <sup>c</sup>	(0.028, 0.18)	0.008	0.061 <sup>c</sup>	(-0.018, 0.141)	0.132						
rs17608766	GOSR2	C/T	0.14	0.556	(0.387, 0.725)	0.129	(-0.075, 0.235)	0.04	0.081 <sup>c</sup>	(-0.075, 0.237)	0.307	0.097 <sup>c</sup>	(-0.065, 0.259)	0.241						
rs198846	HHE	G/A	0.86	0.627	(0.447, 0.807)	0.457	(-0.060, 0.569)	0.91	0.046 <sup>c</sup>	(-0.06, 0.152)	0.396	-0.002 <sup>c</sup>	(-0.112, 0.108)	0.977						
rs2521501	FES	A/T	0.31	0.650	(0.507, 0.793)	0.359	(-0.036, 0.447)	0.24	0.036 <sup>c</sup>	(-0.036, 0.108)	0.33	0.066 <sup>c</sup>	(-0.01, 0.141)	0.088						
rs2932538	MOV10	C/T	0.75	0.388	(0.263, 0.513)	0.240	(-0.054, 0.316)	0.87	0.037 <sup>c</sup>	(-0.054, 0.127)	0.426	0.054 <sup>c</sup>	(-0.04, 0.147)	0.263						
rs3774372	ULK4	C/T	0.17	0.067	(-0.086, 0.220)	0.367	(-0.100, 0.463)	0.20	-0.024 <sup>c</sup>	(-0.1, 0.052)	0.536	0.020 <sup>c</sup>	(-0.058, 0.099)	0.614						
rs381815	PLEKHA7	T/C	0.26	0.349	(0.239, 0.459)	0.575	(-0.049, 0.747)	0.15	0.038 <sup>c</sup>	(-0.049, 0.124)	0.392	0.055 <sup>c</sup>	(-0.035, 0.144)	0.229						
rs419076	MECOM	T/C	0.47	0.409	(0.299, 0.519)	0.241	(-0.046, 0.308)	0.53	0.015 <sup>c</sup>	(-0.046, 0.077)	0.62	0.071 <sup>c</sup>	(0.007, 0.134)	0.029						
rs4373814	CACNB2(5')	G/C	0.55	0.373	(0.261, 0.485)	0.218	(-0.055, 0.287)	0.52	0.007 <sup>c</sup>	(-0.055, 0.068)	0.833	0.002 <sup>c</sup>	(-0.063, 0.066)	0.963						
rs6015450	GNAS–EDN3	G/A	0.12	0.896	(0.718, 1.074)	0.557	(-0.189, 0.667)	0.06	-0.059 <sup>c</sup>	(-0.189, 0.07)	0.369	-0.112 <sup>c</sup>	(-0.246, 0.023)	0.103						
rs653178	SH2B3	A/G	0.47	0.448	(0.364, 0.532)	0.598	(-0.154, 0.733)	0.88	-0.060 <sup>c</sup>	(-0.154, 0.034)	0.214	-0.085 <sup>c</sup>	(-0.183, 0.013)	0.09						
rs7129220	ADM	A/G	0.11	0.619	(0.445, 0.793)	0.299	(-0.185, 0.407)	0.20	-0.042 <sup>c</sup>	(-0.185, 0.101)	0.567	0.002 <sup>c</sup>	(-0.146, 0.15)	0.978						
rs805303	BAT2 <sup>e</sup> –BAT5 <sup>f</sup>	C/T	0.61	0.376	(0.266, 0.486)	0.228	(-0.088, 0.295)	0.69	-0.020 <sup>c</sup>	(-0.088, 0.047)	0.554	-0.040 <sup>c</sup>	(-0.11, 0.031)	0.267						
rs932764	PLCE1	G/A	0.44	0.484	(0.366, 0.602)	0.185	(0.013, 0.258)	0.45	0.076 <sup>c</sup>	(0.013, 0.138)	0.018	0.079 <sup>c</sup>	(0.014, 0.143)	0.018						
Weighted allele score																				
South Asians, <i>n</i> = 1,818																				

DBP, Diastolic BP; EAF, Effect allele frequency; SBP, Systolic BP

<sup>a</sup> Nearest gene information as previously reported in study on common polymorphisms affecting blood pressure [12]

<sup>b</sup> Adjusted for age, sex, BMI, ethnicity and clustering of household member

<sup>c</sup> Effect size for change in SBP or DBP z score per SBP- or DBP-raising allele for individual SNPs

<sup>d</sup> Effect size for change in SBP or DBP z score per unit increase in weighted allele score

<sup>e</sup> BAT2, also known as *PRRC2A*

<sup>f</sup> BAT5, also known as *ABHD16A*

**Table 6** Observed association of weighted allele scores for triacylglycerol, fasting glucose and BMI SNPs with secondary phenotypes

Weighted allele scores	Secondary phenotype	<i>n</i>	OR (95% CI)	<i>p</i> value
Triacylglycerol	Type 2 diabetes <sup>a</sup>	2,051	0.97 (0.91, 1.04)	0.410
Fasting glucose	Hypertriacylglycerolaemia <sup>b</sup>	1,526	0.99 (0.96, 1.03)	0.648
BMI	Hypertension <sup>c</sup>	1,329	0.98 (0.91, 1.06)	0.696
BMI	Hypertriacylglycerolaemia <sup>b</sup>	1,351	1.10 (1.02, 1.18)	0.014

<sup>a</sup>Type 2 diabetes status defined as use of glucose-lowering medications or fasting blood glucose level of  $\geq 7.0$  mmol/l

<sup>b</sup>Hypertriacylglycerolaemia defined as serum triacylglycerol  $>1.70$  mmol/l

<sup>c</sup>Hypertension defined as taking antihypertensive drugs, or mean systolic BP  $\geq 140$  mmHg or mean diastolic BP  $\geq 90$  mmHg

and provide further evidence that the majority of variants identified in GWAS have similar effects in South Asian populations to those in Europeans. This consistency between genetic effects in Pakistanis and Europeans probably reflects

generally similar patterns of linkage disequilibrium and consistent biological factors between Europeans and Asians, despite the frequently very different environmental components.

**Table 7** Mendelian randomisation analyses

Participants <i>n</i>	Phenotype		Change in secondary phenotype		IV estimate of change in secondary phenotype		Observed vs IV estimates <i>p</i> value for difference
	Primary	Secondary	Effect (95% CI)	<i>p</i> value	Effect (95% CI)	<i>p</i> value	
2051	Triacylglycerol	Type 2 diabetes <sup>b</sup>	1.769 <sup>a</sup> (1.586, 1.974)	$1.38 \times 10^{-24}$	0.795 <sup>a</sup> (0.498, 1.269)	0.336	0.0003
1526	Fasting glucose	Triacylglycerol <sup>c</sup>	0.186 (0.137, 0.236)	$1.02 \times 10^{-13}$	-0.123 (-0.458, 0.212)	0.47	0.05
1584	Triacylglycerol	Fasting glucose <sup>d</sup>	0.039 (0.029, 0.049)	$4.44 \times 10^{-15}$	0.007 (-0.036, 0.050)	0.75	0.13
1349	BMI	SBP <sup>e</sup>	-0.043 (-0.277, 0.190)	0.72			
1349	BMI	DBP <sup>f</sup>	0.353 (0.224, 0.482)	$8.11 \times 10^{-8}$	0.314 (-0.979, 1.607)	0.63	0.95
1349	BMI	Triacylglycerol <sup>g</sup>	0.045 (0.035, 0.054)	$1.54 \times 10^{-14}$	0.095 (-0.004, 0.194)	0.06	0.30
1349	BMI	Fasting glucose <sup>h</sup>	0.017 (0.014, 0.020)	$2.14 \times 10^{-25}$	0.011 (-0.020, 0.042)	0.50	0.70

Correlations between primary traits and secondary traits based on standard observed data and correlations calculated using an instrumental variable analysis, where the 'instrument' is the weighted allele score for the primary trait. If no association between traits was observed in standard analyses, we did not proceed with IV analysis

DBP, Diastolic blood pressure; SBP, Systolic blood pressure

<sup>a</sup>OR; type 2 diabetes patients in analysis,  $n=467$

<sup>b</sup>For every 1 SD increase in  $\log_{10}$ -triacylglycerol levels (millimoles per litre), the odds of type 2 diabetes was 1.77 (95% CI 1.59, 1.98); however, the weighted allele score of  $\log_{10}$ -triacylglycerol levels was not associated with type 2 diabetes in IV analysis ( $p=0.336$ )

<sup>c</sup>Each 1 SD increase in fasting plasma glucose (millimoles per litre) was associated with a 0.186 (0.137, 0.236) SD increase in  $\log_{10}$ -triacylglycerol (millimoles per litre) levels in linear regression analysis; however, the weighted allele score of fasting glucose was not associated with  $\log_{10}$ -triacylglycerol level in IV analysis ( $p=0.47$ )

<sup>d</sup>Each 1 SD increase in  $\log_{10}$ -triacylglycerol (millimoles per litre) was associated with a 0.039 (0.029, 0.049) SD increase in fasting glucose levels (millimoles per litre) in linear regression analysis; however, the weighted allele score of  $\log_{10}$ -triacylglycerol was not associated with fasting glucose levels in IV analysis ( $p=0.75$ )

<sup>e</sup>Each 1 unit increase in BMI (kilograms per square metre) was associated with a -0.043 (-0.277, 0.190) decrease in SBP (millimetre of mercury) in linear regression analysis

<sup>f</sup>Each 1 unit increase in BMI (kilograms per square metre) was associated with a 0.353 (0.224, 0.482) increase in DBP (millimetre of mercury) in linear regression analysis; however, the weighted allele score of BMI was not associated with DBP in IV analysis ( $p=0.63$ )

<sup>g</sup>Each 1 unit increase in BMI (kilograms per square metre) was associated with a 0.045 (0.035, 0.054) SD increase in  $\log_{10}$ -triacylglycerol (millimoles per litre) in linear regression analysis; however, the weighted allele score of BMI was not associated with  $\log_{10}$ -triacylglycerol in IV analysis ( $p=0.06$ )

<sup>h</sup>Each 1 unit increase in BMI (kilograms per square metre) was associated with a 0.017 (0.014, 0.020) SD increase in fasting glucose (millimoles per litre) in linear regression analysis; however, the weighted allele score of BMI was not associated with fasting glucose levels in IV analysis ( $p=0.50$ )

A second aim of our study was to use genetic risk scores to help understand possible major causal directions between metabolic traits. Further studies are needed to increase statistical power for these tests. However, we did observe evidence that genetically influenced, raised triacylglycerol and glucose levels do not causally affect type 2 diabetes risk and triacylglycerol levels, respectively. The lack of evidence that genetically influenced triacylglycerol levels causally affect type 2 diabetes risk is consistent with recent evidence from Europeans [6].

The strengths of our study include its population-based design, which means the effect size estimates and proportions of variance explained are likely to be representative of the situation in Pakistani individuals. We have also brought together information from several traits in one study, allowing a simple comparison of effects between traits, and between Europeans and South Asians.

There are some limitations to this study. First, we did not use all confirmed SNPs for all traits; for example, recent publications have identified additional variants associated with triacylglycerol levels and BMI [10, 21]. However, the SNPs we used represent most of those with the strongest effects on these traits and account for the majority of the phenotypic variance reported for known SNPs in Europeans [10, 21]. Second, our sample was representative of Karachi and thus predominantly consists of a single ethnic subgroup, i.e. Muhajirs (representative of North Indian descent); it remains to be evaluated whether our findings can be generalised to South Indian populations. Finally, the study was possibly underpowered to perform robust Mendelian randomisation analyses, although the estimates of our IV results (Table 7) do provide more meaningful power estimates for larger studies [2, 17].

It should also be emphasised that we did not study the impact of genetic risk score on BMI, fasting glucose, triacylglycerol and blood pressure levels in relation to other environmental risk factors (e.g. diet, physical activity) in this population. These metabolic risk factors could have population-specific familial clustering and could potentially interact with the genetic scores. This may be particularly relevant in native South Asian populations exposed to in-utero nutritional deficiencies with the subsequent enhanced susceptibility to metabolic disorders with weight gain [22].

In conclusion, our analyses provide valuable evidence that the relationships between genetic variants associated with important type 2 diabetes-related traits are similar in Pakistanis and Europeans. Our instrumental variable analysis findings, moreover, suggest that genetic risk scores may help disentangle the causal relationships between type 2 diabetes-related traits in larger sample sizes.

**Acknowledgements** We thank N. Poulter (Imperial College London, UK), and J. Hatcher (Aga Khan University, Pakistan) for advice and S. Hashmi (Aga Khan University, Pakistan) for project implementation and R. Bux (Aga Khan University, Pakistan) for data management support.

**Funding** The study was financially supported by a research award (080747/Z/06/Z) from the Wellcome Trust, UK. The design, conduct, analysis, interpretation and presentation of the data were the responsibility of the authors, with no involvement from the funder.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

**Contribution statement** MI, THJ and TMF analysed and interpreted the data, and drafted the manuscript. THJ, NC, and TMF contributed to the conception and design of the study. All authors contributed to the interpretation of data, critical review and approved the final version of manuscript to be published.

## References

1. Danaei G, Finucane MM, Lu Y et al (2011) National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet* 378:31–40
2. Jafar TH, Levey AS, Jafary FH et al (2003) Ethnic subgroup differences in hypertension in Pakistan. *J Hypertens* 21:905–912
3. Wensley F, Gao P, Burgess S et al (2011) Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* 342: d548
4. Perry JR, Weedon MN, Langenberg C et al (2009) Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. *Hum Mol Genet* 19:535–544
5. Zabrodsky S (1990) [Chronic candidosis]. *Prakt Zubn Lek* 38:115–122 (article in Czech)
6. De Silva NM, Freathy RM, Palmer TM et al (2011) Mendelian randomization studies do not support a role for raised circulating triglyceride levels influencing type 2 diabetes, glucose levels, or insulin resistance. *Diabetes* 60:1008–1018
7. Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42:105–116
8. Kathiresan S, Melander O, Guiducci C et al (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40:189–197
9. Kathiresan S, Willer CJ, Peloso GM et al (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41:56–65
10. Speliotes EK, Willer CJ, Berndt SI et al (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42:937–948
11. Lawlor DA, Bedford C, Taylor M, Ebrahim S (2003) Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *J Epidemiol Commun Health* 57:134–140
12. Ehret GB, Munroe PB, Rice KM et al (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478:103–109

13. Freathy RM, Timpson NJ, Lawlor DA et al (2008) Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* 57:1419–1426
14. Rees SD, Islam M, Hydrie MZ et al (2011) An FTO variant is associated with type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. *Diabet Med* 28:673–680
15. Jafar TH, Qadri Z, Chaturvedi N (2008) Coronary artery disease epidemic in Pakistan: more electrocardiographic evidence of ischaemia in women than in men. *Heart* 94:408–413
16. Hussain R (1999) Community perceptions of reasons for preference for consanguineous marriages in Pakistan. *J Biosoc Sci* 31:449–461
17. Jafar TH, Levey AS, White FM et al (2004) Ethnic differences and determinants of diabetes and central obesity among South Asians of Pakistan. *Diabet Med* 21:716–723
18. Lin X, Song K, Lim N et al (2009) Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score—the CoLaus Study. *Diabetologia* 52:600–608
19. Neter J, Wasserman W, Kutner MH (1990) *Applied linear statistical models: regression, analysis of variance, and experimental designs*. Irwin, Homewood
20. Pierce BL, Ahsan H, Vanderweele TJ (2011) Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 40:740–752
21. Teslovich TM, Musunuru K, Smith AV et al (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707–713
22. Bhargava SK, Sachdev HS, Fall CH et al (2004) Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *N Engl J Med* 350:865–875