Epigenetic associations of type 2 diabetes and BMI in an Arab population

Wadha A. Al Muftah, Mashael Al-Shafa, Shaza B. Zaghloul, Alessia Visconti, Pei-Chien Tsai, Pankaj Kumar, Tim Spector, Jordana Bell, Mario Falchi and Karsten Suhre

Abstract

Background: The prevalence of type 2 diabetes (T2D) and obesity has dramatically increased within a few generations, reaching epidemic levels. In addition to genetic risk factors, epigenetic mechanisms triggered by changing environment and life style factors, whose effects are potentially mediated by the epigenome. Epigenetic modifications are changes that do not alter the primary DNA sequence itself and include DNA methylation, histone modifications, and other changes in chromatin structure that may affect the regulation of gene expression. These epigenetic modifications are thought to provide a link between environmental exposures and clinical phenotypes and are also suspected to contribute to the unexplained heritability of complex diseases.

Background

The Qatari population is one of the understudied Arab populations in type 2 diabetes (T2D) and obesity research, despite the high prevalence of these diseases among Qataris, with an estimated prevalence of ~23 % for T2D [44] and ~42 % for obesity (World Health Organization 2015) [58]. The increased prevalence of T2D and obesity has occurred during a short period of time (2–3 generations), suggesting an important contribution to the disease risk by changing environment and life style factors, whose effects are potentially mediated by the epigenome. Epigenetic modifications are changes that do not alter the primary DNA sequence itself and include DNA methylation, histone modifications, and other changes in chromatin structure that may affect the regulation of gene expression [24, 39, 40]. These epigenetic modifications are thought to provide a link between environmental exposures and clinical phenotypes and are also suspected to contribute to the unexplained heritability of complex diseases [30].

The recent development of genome-wide DNA methylation arrays, and of sequencing technologies, coupled
with bisulfide treatment presents novel opportunities to investigate the role of DNA methylation in complex diseases through epigenome-wide association studies (EWASs). Only a small number of EWASs have been published so far on T2D [6, 7, 12, 14, 26, 52, 56] and on obesity or BMI [1, 2, 15, 16, 18]; Sun.D. et al. [49]; [57]). These studies identified new potential T2D- and obesity-associated genes. Furthermore, efforts have been made to study the effect of DNA methylation on gene expression and on metabolic profiles in order to provide better understanding of disease mechanisms [38, 40].

Most EWASs with T2D and obesity conducted so far were focused on Caucasian populations, and little is known about whether their findings translate to other ethnicities and genetic backgrounds. We performed here the first EWAS for T2D and BMI in an Arab population using 123 individuals from 15 Qatari families.

Previous genetic studies have shown that translatability between populations does not always hold. For example, genetic variants at the PPARδ locus that associate with T2D in individuals of European descent seem not to exert any effect on T2D risk in the Qatari population [5]. More discrepancies might be expected for epigenetic risk factors, which are additionally under strong environmental influence. To address this question, we investigated whether the effect of the replicated loci was homogeneous between Caucasians and the Qatari population by using data from 810 twins from UK, therefore having different genetic backgrounds and under different life style/environmental pressures.

Methods
The studies were conducted in concordance with the Helsinki Declaration of ethical principles for medical research involving human subjects. The studies were approved by the relevant institutional review boards in Qatar (Institutional Review Board of Weill Cornell Medical College in Qatar, ethical approval numbers 2012–003 and 2012–0025) and in the UK (Guy’s and St. Thomas’ Hospital Ethics Committee). Written informed consent was obtained from every participant in each study.

Qatari family study
The methylation data used in this study was obtained from whole blood (the only easily accessible type of sample) and has been previously described in [59]. Details on whole-genome sequencing can be found in [27]. Briefly, the study group consisted of 123 subjects of Qatari descent from 15 families of various sizes and structures. The dataset included 72 females with mean age 39 (±16.9) years and 51 males with mean age 36.3 (±7.2) years. The average BMI of the females was 28.3 (±6.2) kg/m² and of the males was 29.2 (±7.2) kg/m². A total of 30 individuals consisting of 19 females and 11 males were previously diagnosed with T2D, ascertained by the diabetes clinic at Hamad Medical Corporation. T2D subjects were receiving treatment for diabetes, and no other major diseases were reported. DNA methylation profiling was performed by Illumina using the Infinium HumanMethylation450K BeadChip platform and reported in the form of beta values. After quality control and exclusion of individual probes containing SNPs within a region of ±110 base pairs of the CpG site, based on 40× whole-genome sequencing data [27], a total of 468,472 probes were selected for this study. Normalization was carried out using the Lumi:QN+BMIQ pipeline, using the smoothQuantileNormalization method (Additional file 1: Figure S1). Blood cell type coefficients of monocytes, granulocytes, NK-cells, B cells, CD8⁺T-cells, and CD4⁺T-cells were estimated from the methylation data using the method described by Houseman et al. [22].

TwinsUK cohort
The TwinsUK cohort was established in 1992 to recruit monozygotic and dizygotic twins [34]. More than 80 % of participants are healthy female Caucasians (age range from 16 to 98 years old). The cohort includes more than 13,000 twin participants from all regions across the United Kingdom, and many have had multiple visits over the years. The TwinsUK cohort has been used in many epidemiological studies and is representative of the general UK population for a wide range of diseases and traits [3]. DNA methylation was measured for 877 individuals randomly selected from the TwinsUK cohort, 810 of who had both BMI and T2D information. All 810 subjects were female Caucasians. The average BMI was 27.8 (±5.2) kg/m² and 32 individuals were previously diagnosed with T2D. The Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA, USA) was used to measure DNA methylation. Details of experimental approaches have been previously described [54]. Normalization was carried out using the “minfi” R package [4], a procedure equivalent to the Lumi:QN+BMIQ pipeline. DNA methylation probes that mapped incorrectly or to multiple locations in the reference sequence and probes with detection p value of >0.05 or missing values were removed, resulting in 452,874 probes. Blood cell type coefficients were estimated from the methylation data using the method described by Houseman et al. [22].

Statistical analysis
Associations of T2D and BMI with DNA methylation levels in the Qatari family study and in the TwinsUK cohort were carried out within a variance component framework to model the resemblance among family
members. Specifically, the association between the phenotypic traits and DNA methylation levels was evaluated by using a linear mixed model where the total phenotypic variance was partitioned into polygenic and environmental variance, the latter including also measurement errors. DNA methylation levels were modeled as fixed effects, whilst the polygenic and environmental effects were modeled as random components. The phenotypic covariance matrix between subjects was modeled using the matrix of the expected proportion of alleles shared IBD over the genome between each pair of individuals. The significance of the associations was evaluated by comparing the likelihood of a full model including the methylation status in the fixed effect, and the likelihood of a null model where the effect of DNA methylation values was constrained to zero. Age, sex (only for the Qatari dataset), smoking status, and the six Houseman blood cell type coefficients (for B cells, granulocytes, monocytes, natural killer cells, CD8^-T-cells, and CD4^-T-cells) were included in the association model. Additionally, BMI association analysis included T2D status as confounder and vice versa. BMI values were standardized to have zero mean and one standard deviation. Given the limited sample size, and to avoid potential inflation of the association statistics by directly carrying out the study on selected probes, we preliminarily performed an EWAS in the Qatari family sample using the whole set of probes. False discovery rate was evaluated using Storey’s method [47].

### Selection of CpG sites for replication

At the time this study was conducted, two large EWASs for T2D [12, 26] and two for obesity and BMI [15, 16] were available. We attempted to replicate the most significant CpG probe for each reported locus that reached genome-wide significance in at least one of these four studies, resulting in eight CpG probes for T2D (Table 1) and 39 for obesity and BMI (Table 2). Conservative Bonferroni method was used to correct for multiple testing, considering an association replicated with T2D if its p value was lower than $6.25 \times 10^{-3}$ (0.05/8) and with BMI if its p value was lower than $1.28 \times 10^{-3}$ (0.05/39).

### Meta-analysis

Meta-analyses between the Qatari and UK samples were carried out using the Genome-Wide Association Meta-Analysis (GWAMA) software [29]. Specifically, we used a fixed effects model with inverse variance to combine the regression coefficients of each study and their standard errors. Inter-study heterogeneity was estimated by using the Cochran’s Q test and by measuring the proportion of variability that is explained by between-trial heterogeneity ($I^2$ estimates, [21]), both implemented in GWAMA.

### Results

Our EWAS (Additional file 2: Figure S2 and Additional file 3: Figure S3) identified one CpG association with T2D that reached genome-wide Bonferroni significance ($p$ value $<1.07 \times 10^{-7}$) (cg06721411 at DQXI; $p$ value = $1.18 \times 10^{-6}$). No methylation probes were significantly associated with BMI after Bonferroni correction for multiple testing, the strongest association being at cg17501210 (RPS6KA2; $p$ value = $4.90 \times 10^{-7}$). The full EWAS association data is available as Additional files 4 and 5.

Q-Q plots (Additional file 3: Figure S3) of the EWASs for BMI and T2D showed mild inflation of the $p$ value statistics (the genomic inflation factor was 1.10 for T2D and 1.09 for BMI). We also replicated the association of our top T2D CpG, cg06721411 (DQXI), in the TwinsUK cohort ($p$ value = $9.00 \times 10^{-3}$).

We calculated the heritability of DNA methylation at these probes in the Qatari families. At 1 % FDR, 41,374 (about 10 %) methylation levels showed segregation between family members (median heritability = 0.70; first to third quartile = 0.31–1.00). The replicated loci showed

### Table 1 Replication of T2D-DNA methylation associations in the Qatari family study. The betas represent the slope of the regression model indicating the rate of change in the dependent variable (trait) as independent variable (methylation beta value) changes. Coordinates are in hg19

| Probe          | Chr | Position     | Gene symbol | Beta | SE  | $p$ value | Reference |
|----------------|-----|--------------|-------------|------|-----|----------|-----------|
| cg19693031     | 1   | 145441552    | TXNIP       | −2.41| 0.57| $2.46 \times 10^{-5}$ | [12, 26] |
| cg00574958     | 11  | 68607622     | CPT1A       | −3.77| 2.07| 0.068    | [12, 26] |
| cg11024682     | 17  | 17730094     | SREBF1      | 1.99 | 1.30| 0.124    | [12]     |
| cg09152259     | 2   | 128156114    | PROC        | −1.02| 0.71| 0.148    | [12]     |
| cg06500161     | 21  | 43656587     | ABCG1       | 2.17 | 1.52| 0.153    | [12]     |
| cg04999691     | 7   | 150027050    | C7orf29     | 1.87 | 1.40| 0.182    | [12]     |
| cg02650017     | 17  | 47301614     | PHOSPHO1    | −3.22| 2.58| 0.212    | [12]     |
| cg18181703     | 17  | 76354621     | SOCS3       | −0.66| 0.91| 0.465    | [12]     |
heritability between 0.46 and 0.96, apart from cg11024682 (SREBF1) and cg07573872 (SBNO2), which were not significant at 1% FDR.

We replicated in the Qatari family study the associations between T2D and cg19693031 (TXNIP; \( p \) value = \( 2.46 \times 10^{-5} \)) (Table 1) and between BMI and cg18181703

| Probe     | Chr | Position  | Gene symbol | Beta  | SE   | \( p \) value | Reference |
|-----------|-----|-----------|-------------|-------|------|--------------|-----------|
| cg18181703| 17  | 76354621  | SOCS3       | -10.78| 2.34 | \( 3.99 \times 10^{-6} \) | [12]      |
| cg11024682| 17  | 17730094  | SREBF1      | 14.56 | 3.56 | \( 4.33 \times 10^{-5} \) | [15]      |
| cg07573872| 11  | 41656298  | NA          | -19.23| 5.85 | \( 1.00 \times 10^{-3} \) | [15]      |
| cg13123009| 6   | 31681882  | LY6G6E      | 12.41 | 3.80 | \( 1.10 \times 10^{-3} \) | [15]      |
| cg09349128| 22  | 50372986  | NA          | -9.60 | 3.04 | \( 1.60 \times 10^{-3} \) | [15]      |
| cg08972190| 7   | 2138995   | MAD1L1      | 11.11 | 3.54 | \( 1.70 \times 10^{-3} \) | [15]      |

Table 2 Replication of BMI-DNA methylation associations in the Qatari family study. The betas represent the slope of the regression model indicating the rate of change in the dependent variable (trait) as independent variable (methylation beta value) changes. Coordinates are in hg19.
(SOCS3; \( p \) value = 3.99 \times 10^{-6}\), cg11024682 (SREBF1; \( p \) value = 4.33 \times 10^{-5}\), cg07573872 (SBNO2; \( p \) value = 5.87 \times 10^{-5}\), cg00574958 (CPT1A; \( p \) value = 7.99 \times 10^{-5}\), cg07136133 (PRRSL; \( p \) value = 1.85 \times 10^{-4}\), cg03078551 (intergenic region on chromosome 17; \( p \) value = 1.00 \times 10^{-3}\), and cg13123009 (LY6G6E; \( p \) value = 1.10 \times 10^{-3}\) (Table 2). Boxplots and scatter plots of these associations are in Figs. 1 and 2. The distributions of the methylation values for these eight CpG sites are in Additional file 6: Figure S4. Although we decided to adopt Bonferroni correction for the replication study, 12 additional associations with BMI showed nominal level of significance and the same direction of the associations as the original EWASs (Table 2).

The eight replicated associations were analyzed in the TwinsUK cohort, and effects were combined in meta-analyses. The meta-analysis of BMI with the TwinsUK results indicated a moderate presence of heterogeneity between the two studies for cg00574958 (CPT1A; \( I^2 = 56.8\%\); Cochran’s heterogeneity statistic’s \( p \) value >0.05) and increased the significance of this association to \( p \) value = 7.32 \times 10^{-14}\). On the other hand, a considerable presence of heterogeneity between the two studies was identified for all the other associations (Table 3; Cochran’s heterogeneity statistic’s \( p \) value <0.05), despite association at these loci was significant in both populations and with concordant direction (Table 3).

**Discussion**

The high prevalence of T2D and obesity in Qatar motivated the initiation of genetic and epigenetic research in this country. To the best of our knowledge, this is the first association study of CpG methylation with T2D and BMI in an Arab population. We conducted a full EWAS and attempted to replicate in the Qatari population eight CpG sites associated with T2D (Table 1) and 39 CpG sites associated with BMI (Table 2) in Caucasians. Our EWAS with T2D (Additional file 2: Figure S2 and Additional file 3: Figure S3) identified one significantly associated CpG site at cg06721411 (DQX1; \( p \) value = 1.18 \times 10^{-4}\), while the strongest association with BMI at cg17501210 (RP56KA2; \( p \) value = 4.90 \times 10^{-7}\) did not reach genome-wide significance. The inflation shown in the Q-Q plots is possibly due to hidden confounders, including potentially reduced folate levels in diabetic subjects. However, the observed inflation is only moderate and does not substantially affect our conclusions. As only cg06721411 (DQX1) in the T2D EWAS achieved Bonferroni significance criteria in our discovery cohort, we also replicated this locus in the TwinsUK cohort (\( p \) value = 9.00 \times 10^{-5}\). The effect was in the same direction. DQX1 (DEAQ box RNA-dependent ATPase 1) is a protein coding gene located on chromosome 2 and is classified as a member of the DEAD/H family. The highest expression of the DQX1 is found in the muscle and liver [25].

Using a conservative Bonferroni correction, we replicated eight of the 47 associations: SOCS3, SREBF1, SBNO2, CPT1A, PRRSL, an intergenic region on chromosome 17, and LY6G6E with BMI; and TXNIP1 with T2D, while nominal significance was reached for a further 12 loci associated with BMI (Table 2). Despite association with methylation at SOCS3 being previously reported for both T2D and BMI [12, 26], only the association with BMI was replicated in our study. However, the association between SOCS3 and T2D was not significant in the study of Chambers and colleagues after adjustment for BMI, suggesting that the observed association with T2D in their study may be driven by differences in adiposity between their T2D cases and controls.

Although the mechanisms linking DNA methylation of SOCS3, SREBF1, SBNO2, CPT1A, PRRSL, and LY6G6E to BMI, and of TXNIP to T2D are not fully established yet, some of these genes have been already functionally linked to metabolic phenotypes. The replicated methylation sites are within proximity of known genes suggesting a regulatory role of the methylation. However, because expression data is not available for this population, we used data from prior studies to confirm the functional relevance of methylation to phenotypes and the expression of these genes. For instance, TXNIP is a

![Fig. 1 Boxplot of methylation beta values at cg19693031 (TXNIP) against diabetes state. The middle lines show the medians of the data while the boxes show the 25th to 75th percentiles (Q1 and Q3). The whiskers extend to include 99% of the data above while circles represent outliers. Beta value distributions at this probe in the diabetics and healthy individuals showed a difference in the levels of background methylation (Wilcoxon test \( p \) value = 2.10 \times 10^{-4}\) )


pro-apoptotic beta-cell factor and encodes for a protein that acts as a regulator of metabolism and an inhibitor of the antioxidant thioredoxins. A recent study showed that TXNIP is involved in glucose regulation by controlling insulin sensitivity in the periphery of the human body, and its expression is elevated in the skeletal muscles in T2D patients [37] indicating a linkage to phenotype. We observed concordant results in this study: individuals diagnosed with T2D show lower levels of TXNIP methylation, thus suggesting higher TXNIP
expression. Also, SOCS3 belongs to the SOCS protein family, which is rapidly induced by cytokines, and acts as an inhibitor of various cytokine signaling pathways. Previous studies have shown that SOCS3 is linked to phenotype by being a negative regulator of leptin [9, 10] and insulin signaling [17, 42, 45]. In addition, there is evidence for association between variants located near SOCS3 with glucose homeostasis, BMI, and other obesity traits [50, 51]. It was also shown that expression of SREBF1 was reduced in adipose and skeletal muscle of diabetic subjects [12, 26]. SREBF1 was shown to regulate carbohydrate metabolism and synthesis in an animal model of obesity and T2D [43]. In another study, qPCR experiments showed that CPT1A expression is correlated with the methylation status of CPT1A gene with \( p = 4.1 \times 10^{-14} \) and replicated in the Framingham Heart Study with \( p = 3.1 \times 10^{-13} \) [23]. Differential methylation at CPT1A was also found to influence gene expression in Dick et al. ([15, 16]; Sun.D. et al. [49]). Also, [55] showed an increase of the suppressor of the cytokine signaling proteins including SOCS3 in the liver, muscle, and fat, in obesity. SOCS3 overexpression in the fat cells was accompanied by glycogen synthesis and activation of glucose uptake. Finally, we also used a recently available database [11] to understand whether our T2D EWAS hit and the eight replicated were associated with any gene expression level. We found that all methylation sites had more than one association with expression (cis-meQTL, cis-eQTM, and/or trans-meQTL) at 5 % FDR. Therefore, differential methylation may suggest a regulatory role.

Interestingly, two CpG sites replicated by our study (CPT1A and TXNIP) and a third CpG we attempted to replicate (ABCG1) were also the only probes significantly associated with alpha-hydroxybutyrate (AHB) in our recent EWAS with blood serum metabolomics traits (cg00574958 in CPT1A, \( p = 1.3 \times 10^{-10} \); cg06500161 in ABCG1, \( p = 7.8 \times 10^{-6} \); cg19693031 in TXNIP, \( p = 7.2 \times 10^{-6} \)) [38]. AHB is a sub-product of the ketones metabolism; elevated AHB levels indicate potential insulin resistance [20], and this biomarker is part of the metabolomics-based pre-diabetes test Quantoset [53].

Other biomarkers of T2D and pre-diabetes have been previously associated with methylation at ABCG1, TXNIP, and CPT1A, including 3-methyl-2-oxovalerate [33], glycolenine [19], several lipid traits, including phosphatidylcholines (PCs) [48], chylomicrons, and their remnants, very-low-density lipoprotein (VLDL) and IDL cholesterol particles [38]. They all showed diabetes-related effect directions that are in agreement with the effect directions observed in this study. Most interestingly, the list of metabolic traits associated with CpGs [38] also includes the product of CPT1A itself, palmitoylcarnitine. Furthermore, higher levels of cg00574958 (CPT1A) methylation were also associated with higher levels of related long-chain fatty acids in the EWAS reported by Petersen et al., including palmitate (16:0), stearate (18:0), and oleate (18:1n9) (see Supplementary Table 5 of [38]). Higher levels of cg06500161 (ABCG1) methylation were also associated in a recent EWAS with higher levels of chylomicrons and VLDL-cholesterol (see Supplementary Table 5 of [38] and Table 4).

The direction of the associations for all metabolites at these three loci is coherent with the association of CPT1A and TXNIP being in one direction (lower methylation values associated with T2D or obesity) and that of ABCG1 in the opposite one (higher methylation values being associated with obesity). Taken together, these observations support the claim that lower methylation of the CPT1A and TXNIP loci and increased methylation of the ABCG1 locus associate with a well-defined diabetes-specific metabolic phenotype, which is mirrored by the association of

### Table 3 Meta-analyses of the replicated loci in the Qatari study with the TwinsUK results

| Trait Probe (gene) | Qatari cohort | TwinsUK cohort | Meta-analysis |
|-------------------|--------------|----------------|--------------|
|                   | Beta | SE | p value | Beta | SE | p value | Beta | SE | p value |
| T2D cg19693031 (TXNIP) | −2.41 | 0.57 | 2.46 × 10\(^{-3}\) | −0.34 | 0.12 | 6.74 × 10\(^{-3}\) | −0.43 (−0.66/−0.20) | 0.12 | 2.71 × 10\(^{-4}\) |
| BMII cg18181703 (SOCS3) | −10.78 | 2.34 | 3.99 × 10\(^{-6}\) | −2.90 | 0.71 | 5.45 × 10\(^{-5}\) | −3.56 (−4.90/−2.23) | 0.68 | 1.59 × 10\(^{-7}\) |
| CG cg11024682 (SREBF1) | 14.56 | 3.56 | 4.33 × 10\(^{-5}\) | 5.86 | 0.91 | 2.12 × 10\(^{-10}\) | 6.39 (4.67/8.12) | 0.88 | 4.28 × 10\(^{-13}\) |
| CG cg07538782 (SBNQ2) | −9.96 | 2.48 | 5.87 × 10\(^{-5}\) | −2.64 | 0.81 | 1.16 × 10\(^{-3}\) | −3.35 (−4.85/−1.84) | 0.77 | 1.41 × 10\(^{-5}\) |
| CG cg00574958 (CPT1A) | −21.04 | 5.33 | 7.99 × 10\(^{-5}\) | −12.43 | 1.90 | 1.08 × 10\(^{-10}\) | −13.40 (−16.91/−9.89) | 1.79 | 7.32 × 10\(^{-14}\) |
| CG cg07136133 (PRRS1) | −10.43 | 2.79 | 1.85 × 10\(^{-4}\) | −3.82 | 0.78 | 8.02 × 10\(^{-7}\) | −4.36 (−5.84/−2.89) | 0.75 | 6.43 × 10\(^{-9}\) |
| CG cg03078551 (NA) | −19.23 | 5.85 | 1.00 × 10\(^{-3}\) | −7.35 | 1.43 | 3.36 × 10\(^{-7}\) | −8.02 (−10.74/−5.30) | 1.39 | 7.97 × 10\(^{-9}\) |
| CG cg13123009 (LY6G6E) | 12.41 | 3.80 | 1.10 × 10\(^{-3}\) | 3.06 | 1.06 | 3.81 × 10\(^{-3}\) | 3.74 (1.73/5.74) | 1.02 | 2.56 × 10\(^{-4}\) |

The table reports the results we obtained using a fixed effects model with inverse variance to combine the regression coefficients of each study and their standard errors. \( p \) values, effect sizes (beta), and their standard errors (SE) are reported for both studies and for the meta-analysis results. For the meta-analysis, we also reported: upper/lower 95 % CI for beta (beta 95U/95L), and heterogeneity estimates (\( I^2 \)).
Table 4  
CpG–metabotype associations at the three replicated loci. Only associations with metabotypes that were significant at a $P$ value $<1.3 \times 10^{-5}$ (Bonferroni correction of testing multiple metabolic traits) are shown; effect size (beta$^\prime$), $P$ value of the linear model, and number of samples ($N$) (data from Supplementary Table 5 of [38]; for details on this dataset, see there).

| Metabolic trait                  | cg06500161 (ABCG1) | cg00574958 (CPT1A) | cg19693031 (TXNIP) |
|---------------------------------|---------------------|--------------------|---------------------|
| 1-Oleoylglycerol (1-monoolein)  | 9.435 $^{*}$ 2.84 $\times 10^{-11}$ 1676 | $-0.983 4.37 \times 10^{-11}$ 1676 | $-0.841 1.02 \times 10^{-14}$ 1674 |
| Alpha-hydroxybutyrate (AHB)     | 1.834 7.80 $\times 10^{-6}$ 1749 | $-0.926 1.30 \times 10^{-10}$ 1749 | $-0.574 7.24 \times 10^{-8}$ 1747 |
| 3-methyl-2-oxovalerate          | 0.771 1.31 $\times 10^{-5}$ 1749 | $-0.605 4.50 \times 10^{-5}$ 1749 | $-0.472 6.81 \times 10^{-13}$ 1747 |
| Glycine                         | $-0.531 2.91 \times 10^{-5}$ 1744 | $4.064 7.58 \times 10^{-9}$ 1744 | $0.656 4.91 \times 10^{-6}$ 1742 |
| Palmitoylcarnitine              | 1.729 1.35 $\times 10^{-5}$ 1737 | $-0.849 2.28 \times 10^{-6}$ 1737 | $-0.579 3.11 \times 10^{-8}$ 1735 |
| PC aa C36:4                     | 2.549 2.82 $\times 10^{-5}$ 1781 | $-0.947 2.07 \times 10^{-8}$ 1781 | $-0.628 1.66 \times 10^{-6}$ 1779 |
| PC aa C42:0                     | $-0.883 3.91 \times 10^{-8}$ 1781 | $2.276 3.36 \times 10^{-6}$ 1781 | $6.181 1.11 \times 10^{-13}$ 1779 |
| PC aa C42:1                     | $-0.844 4.63 \times 10^{-7}$ 1781 | $12.688 4.18 \times 10^{-5}$ 1781 | $4.631 4.93 \times 10^{-12}$ 1779 |
| PC ae C44:6                     | $-0.918 8.52 \times 10^{-12}$ 1781 | $17.438 4.88 \times 10^{-6}$ 1781 | $5.075 5.48 \times 10^{-13}$ 1779 |
| Chylo-A (nM)                    | 594.7 8.44 $\times 10^{-14}$ 1773 | $-1.000 1.35 \times 10^{-13}$ 1773 | $-0.996 1.11 \times 10^{-21}$ 1771 |
| Chylo-B (nM)                    | 29.53 7.09 $\times 10^{-6}$ 1766 | $-1.000 3.24 \times 10^{-12}$ 1766 | $-0.988 8.30 \times 10^{-18}$ 1764 |
| Chylo-Rem (nM)                  | 416.1 1.34 $\times 10^{-11}$ 1772 | $-1.000 2.38 \times 10^{-12}$ 1772 | $-0.991 5.40 \times 10^{-15}$ 1770 |
| IDL (nM)                        | 30.86 1.32 $\times 10^{-9}$ 1773 | $-0.993 6.00 \times 10^{-7}$ 1773 | $-0.868 2.22 \times 10^{-7}$ 1771 |
| VLDL-A (nM)                     | 150.7 2.72 $\times 10^{-13}$ 1773 | $-1.000 9.23 \times 10^{-14}$ 1773 | $-0.985 5.73 \times 10^{-19}$ 1771 |
| VLDL-B (nM)                     | 166.0 4.22 $\times 10^{-13}$ 1773 | $-1.000 9.83 \times 10^{-12}$ 1773 | $-0.979 9.24 \times 10^{-16}$ 1771 |

The loci with the respective clinical phenotypes, obesity, and diabetes.

Replicated associations identified in this study were also confirmed in the TwinsUK cohort (Table 3). Meta-analysis increased the significance of the associations, but highlighted heterogeneity of the effect sizes for all loci but CPT1A. Some heterogeneity of effects between our results and what was reported in the original papers might be expected, as they could be driven by potential differences in the normalization pipeline of the array data or by the correction of the methylation values using different confounders. However, despite that there were no differences in the normalization pipeline or in the use of confounders between the Qatari family sample and the TwinsUK cohort, we still observed significant effect heterogeneity. This heterogeneity may partly be explained by the different environmental pressures. While the standardized BMI distribution was not different between the two samples (Kolmogorov–Smirnov $p$ value $>0.05$), the distribution of six out of eight methylation values at the tested probes was different in either location or shape (Kolmogorov–Smirnov $p$ value $<0.05$; Additional file 6: Figure S4).

There are some limitations to our present study that we are aware of. First, the use of Illumina Infinium HumanMethylation450K arrays targets only a subset of methylation sites across the human genome. Array-based technologies are sensitive to artifacts induced by genetic variants (SNPs) within probe binding sites. This problem is commonly addressed by excluding probes that contain known SNPs, based on the annotations given in the Illumina manifest. However, these annotations are based on SNP tagging technologies and might provide incomplete information, in particular, in less studied non-Caucasian populations. One of the strengths of our present study is the ability to fully remove such potentially confounding genetic effects, based on the availability of whole-genome sequencing data with deep coverage.

Second, DNA methylation was measured using DNA extracted from whole blood that was the only accessible type of sample and may not be representative of more disease-relevant tissues for the diseases under study, such as pancreatic cells and adipose tissues. Studies of methylation in obesity or T2D based on disease-relevant tissues such as the skeletal muscle, adipose tissue, or pancreatic islets are interesting but only exist for relatively small studies [8, 13, 28, 31, 32, 35, 36, 41, 46]. Since DNA methylation can be strongly tissue dependent and as our data was obtained from blood samples, for consistency, we only selected methylation probes for replication from EWAS studies that were also performed in the blood. In addition, the blood consists of various cell types (including B cells, granulocytes, monocytes, natural killer cells, and T cells subset) that may bias methylation estimates. Estimation of cell type coefficients from the methylation data was performed using a method described by [22], and correction for these coefficients in the association model is common practice and it is also applied here.
Conclusions
Given the early state of the epigenome-wide technologies, the number of published EWASs on T2D and obesity so far is small. However, the availability of the technology along with the availability of novel computational tools is expected to accelerate the increase in the number of studies conducted in this field. To the best of our knowledge, this study is the first EWAS of T2D and obesity in an Arab population. Our EWAS identified one new CpG association with T2D at DQX1 that reached genome-wide Bonferroni significance. We also replicated eight previously reported T2D and BMI associations, although they were not genome-wide significant, confirming the relevance of these CpG sites to these phenotypes.

Additional files

Additional file 1: Figure S1. DNA methylation for the 123 samples presented as boxplots. Circles represent outliers, and the red and green boxes represent the effect of the quality control on the data: (a) before preprocessing, (b) after color bias adjustment, and (c) after quantile normalization (see [59] for details) 548 KB

Additional file 2: Figure S2. Manhattan plots for epigenome-wide association of CpG methylation sites with (A) BMI (B) T2D Coordinates are in hg19. The red line indicates a conservative Bonferroni significance threshold of p value = 1.07 × 10−7, 277 KB

Additional file 3: Figure S3. Q-Q plots of the EWAS results with (A) BMI (genomic inflation factor = 1.09), and (B) T2D (genomic inflation factor = 1.10). The red line shows the expected p values. 49.7 KB

Additional file 4: Whole list of BMI associations with CpG methylation sites from the EWAS in the Qatari family sample (provided in csv-separated format). Coordinates are in hg19. 31.5 MB

Additional file 5: Whole list of T2D associations with CpG methylation sites from the EWAS in the Qatari family sample (provided in csv-separated format). Coordinates are in hg19. 31.2 MB

Additional file 6: Figure S4. Distribution of methylation beta value for the Qatari family study (red) and the TwinsUK cohort (blue). While the standardized BMI distribution was not different between the two samples (Kolmogorov-Smirnov p value=0.05), the distribution of six out of eight methylation values at the tested probes were different in either location or shape (Kolmogorov-Smirnov p value<0.05) 309 KB

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
WAAM, MAS, MF, and KS designed the study. MAS, WAAM, JB, TS, and PCT collected the samples and generated the data. SBZ, AV, and PK analyzed the data. WAAM, MAS, SBZ, AV, MF, and KS wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank the Qatar Diabetes Association (QDA) and Cindy McKeon for their help in the sample recruitment. We are grateful to all study participants for their contribution to this research study.

Funding
This work was supported by the Qatar Science Leadership Program funds at the Research Division, a program funded by the Qatar Foundation. This work was also supported by Biomedical Research Program funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. MF is in receipt of support from Qatar Foundation grant GEQATDIAB. AV is funded by the British Skin Foundation, grant 504Hi. The data coming from the TwinsUK study was funded by the Wellcome Trust; European Community’s Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR)—funded BioResource and the Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. The statements made herein are solely the responsibility of the authors.

References
1. Almen MS, Nilsson EK, Jacobsson JA, Kalnina I, Klovins J, Fredriksson R, et al. Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. Gene. 2014;548(1):161-7.
2. Almen MS, Jacobsson JA, Moschonis G, Benedict C, Chrrousos GP, Fredriksson R, et al. Genome wide analysis reveals association of a FTO gene variant with epigenetic changes. Genomics. 2012;99(3):132–7.
3. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res. 2001;4(6):464–77.
4. Aytee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive bioconductor package for the analysis of Illumina DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9.
5. Badii R, Bener A, Zirie M, Al-Rikabi A, Simsek M, Al-Hamaq AO, et al. Lack of association between the Pro12Ala polymorphism of the PPAR-gamma 2 gene and type 2 diabetes mellitus in the Qatari consanguineous population. Acta Diabetol. 2008;45(1):15–21.
6. Barnes R, Oster ME, Yan J, Runa A, Fritz T, Caidahl K, et al. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab. 2009;10(3):189–98.
7. Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PLoS One. 2010;5(11), e14040.
8. Benton MC, Johnstone A, Eccles D, Harmon B, Hayes MT, Lea RA, et al. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genom Biol. 2015;16:5(14):0569-x.
9. Bjorbaek C, El-Hachem K, Frantz JD, Flier JS. The role of SOCS-3 in leptin signaling and leptin resistance. J Biol Chem. 1999;274(42):30059–65.
10. Bjorbaek C, Elmqvist JK, Frantz JD, Shoelson SE, Flier JS. Identification of SOCS-3 as a potential mediator of central leptin resistance. Mol Cell. 1998;114:619–25.
11. Bonder MJ, Luijk R, Zhernakova D, Moed M, Deelen P, Vermaat M, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference or shape (Kolmogorov-Smirnov p value<0.05) 309 KB
circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24(15):4464–79.
16. Dick KJ, Nelson CP, Taspinori L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body mass index: a genome-wide analysis. Lancet. 2014; 383(9923):1990–8.
17. Emanuelli B, Pesalá P, Pilloucos C, Sawkia-Verhelle D, Hilton D, Van Obberghen E. SOCS-3 is an insulin-induced negative regulator of insulin signaling. J Biol Chem. 2002;275(21):15985–91.
18. Feinberg AP, Irazny RA, Fradin D, Ayene MJ, Murakami P, Aspelund T, et al. Personalized epigenomic signatures that are se over time and covary with body mass index. Sci Transl Med. 2010;2(49):49a67.
19. Ferrannini E, Natali A, Camasta S, Nanipieri M, Mari A, Adam KP, et al. Early metabolic markers of the development of dysglycemia type 2 diabetes and their physiological significance. Diabetes. 2013;62(5):1730–7.
20. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhe PJ, et al. Alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. PLoS One. 2010;5(5), e10883.
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60.
22. Houseman EA, Acommodo WP, Keister DC, Christensen BC, Maris CT, Nelson HH, et al. DNA methylation arrayas as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:265-2105-13-265.
23. Ivin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. Circulation. 2014;130(7):565–72.
24. Jacobsen R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33(Suppl):245–54.
25. Ji W, Chen F, Do T, Do A, Roe BA, Meisler MH. DQX1, an RNADependent ATPase homolog with a novel DEAQ box: expression pattern and genomic sequence comparison of the human and mouse genes. Mamm Genome. 2001;12(6):456–66.
26. Kulkarni H, Kos MZ, Neary J, Dyer TD, Kent Jr JW, Goring HH, et al. Novel regulatory element-binding protein-1 (SREBP-1) is required to regulate insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. J Biol Chem. 2002;277(44):42394–8.
27. Kumar P, Al-Shafai M, Al Muftah WA, Chalhoub N, Elsaid MF, Aleem AA, et al. Carnitine palmitoyltransferase 1B (CPT1B) gene to lipid exposure with impact of age, BMI and HbA1C levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet. 2015;24(13):3792–813.
28. Rui L, Yuan M, Frantz D, Shoelson S, White MF. SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. J Biol Chem. 2002;277(44):42394–8.
29. Ruiz R, Jideonwo V, Ahn M, Surendran S, Tagliabracci VS, Hou Y, et al. Sterol regulatory element-binding protein-1 (SREBP-1) is required to regulate glycolysis and gluconeogenic gene expression in mouse liver. J Biol Chem. 2014;289(9):5510–7.
30. Scully T. Diabetes in numbers. Nature Outlook. 2012;485(7398):52–3.
31. Shi H, Tzameli I, Bjoerbaek C, Flier JS. Suppressor of cytokine signaling 3 is a physiological regulator of adipocyte insulin signaling. J Biol Chem. 2004; 279(37):34373–40.
32. Spoorst L, Gunnesson A, Sjostrom CD, Narbono K, Petlonen M, Wedel H, et al. Effects of bariatric surgery on cancer incidence in obese patients in Sweden (Swedish obese subjects study): a prospective, controlled intervention trial. Lancet Oncol. 2009;10(7):653–62.
33. Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B (Statistical Methodology). 2002;64(3):479–98.
34. Suhrke K, Meinert CL, Doring A, Altmairer E, Belcledi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplet biomarkers study in an epidemiological setting. PLoS One. 2010;5(11), e13953.
35. Sun D. Race-specific association between DNA methylation and body mass index: the Bogalusa Heart Study [Internet]. 2014 [cited 2014. Available from: http://www.ashg.org/2014meeting/abstracts/fulltext/f140122628.htm.
36. Talbott ME, Langefeld CD, Ziegler J, Mychaleckyj JC, Haffner SM, Norris JM, et al. Polymorphisms near SOCS3 are associated with obesity and glucose homeostasis traits in Hispanic americans from the insulin resistance atherosclerosis family study. Hum Mol Genet. 2009;12(5):153–62.
37. Tang W, Zou JJ, Chen XF, Zheng HZ, Liu ZM, et al. Association of two polymorphisms within and near SOCS3 gene with obesity in three nationalities in Xirijiang Province of China. Acta Pharmacol Sin. 2011;32(11):1381–6.
38. Toporoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissin B, et al. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Hum Mol Genet. 2012; 21(2):2371–83.
39. Tripathy D, Cobb JE, Gall W, Adam KP, George T, Schwennek DC, et al. A novel insulin resistance index to monitor changes in insulin sensitivity and glucose tolerance: the ACT NOW study. J Clin Endocrinol Metab. 2015;100(5):1855–62.
40. Taspinori LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics. 2014(9)10:3932–96.
41. Ueki K, Kondo T, Kahn CR. Suppressor of cytokine signaling 1 (SOCS-1) and SOCS-3 cause insulin resistance through inhibition of tyrosine phosphorylation of insulin receptor substrate proteins by discrete mechanisms. Mol Cell Biol. 2004;24(12):5344–56.
42. Volkmar M, Dedeunouwer S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, et al. Global health observatory data repository [Internet]; c2015 [cited 2015 February 5]. Available from: http://apps.who.int/gho/data/node.main.A900A?lang=en.
43. Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, et al. Obesity related metabolic markers of the development of dysglycemia and type 2 diabetes: evidence for a genetic basis. Diabetes. 2013;62(5):1730–1737.
44. Wouters S, Van der Schueren C, Grootx J, Baets L, Fransen A, Krom C, et al. Epigenetic biomarkers in blood predict cancer risk in an epidemiologic setting. PLoS One. 2010;5(11), e13953.
45. Xia Y, Shigemori A, Tsipourova R, Pan Y, Ambrosio R, Beatty B, et al. Epigenetic signatures to identify cancer susceptibility genes in human peripheral blood. Hum Mol Genet. 2015;24(18):5330–40.
46. Xie C, Xin D, Pekarek J, Fu G, Wang Z, Song X, et al. The role of methylated DNA in obesity and type 2 diabetes: a systematic review and meta-analysis. J Diabetes. 2015;7(2):210–21.