



King's Research Portal

DOI:

[10.1186/s13148-016-0177-6](https://doi.org/10.1186/s13148-016-0177-6)

Document Version

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Al Muftah, W. A., Al-Shafai, M., Zaghlool, S. B., Visconti, A., Tsai, P-C., Kumar, P., Spector, T., Bell, J., Falchi, M., & Suhre, K. (2016). Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clinical Epigenetics*, 8, [13]. <https://doi.org/10.1186/s13148-016-0177-6>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

1 **Epigenetic Associations of Type 2 Diabetes and BMI in an Arab Population**

2 *Re-submission to Clinical Epigenetics, 10 January 2016*

3 Wadha A. Al Muftah^{1, 2, 3*}, Mashaël Al-Shafai^{1, 2, 3*}, Shaza B. Zaghlool¹, Alessia Visconti⁴, Pei-
4 Chien Tsai⁴, Pankaj Kumar¹, Tim Spector⁴, Jordana Bell⁴, Mario Falchi^{2, 4, +}, and Karsten Suhre^{1,}
5 4, 5+

6
7 ¹Bioinformatics Core, Weill Cornell Medical College in Qatar, Education City, PO Box 24144,
8 Doha, Qatar

9 ²Department of Genomics of Common Disease, Imperial College London, London, UK

10 ³Research Division, Qatar Science Leadership Program, Qatar Foundation, Doha, Qatar

11 ⁴Department of Twin Research & Genetic Epidemiology, King's College London, London SE1
12 7EH, UK

13 ⁵Helmholtz Zentrum München, Germany, Research Center for Environmental Health, 85764,
14 Neuherberg Germany

15
16 * Wadha A. Al Muftah and Mashaël Al-Shafai contributed equally to this work

17 + Karsten Suhre and Mario Falchi shared senior authorship.

18 Correspondence to: Karsten Suhre, PhD, Weill Cornell Medical College in Qatar, Qatar
19 Foundation – Education City, PO Box 24144, Doha, Qatar. Email: karsten@suhre.fr

23 **ABSTRACT**

24 *Background:* The prevalence of type 2 diabetes (T2D) and obesity has dramatically increased
25 within a few generations, reaching epidemic levels. In addition to genetic risk factors, epigenetic
26 mechanisms triggered by changing environment are investigated for their role in the
27 pathogenesis of these complex diseases. Epigenome-wide association studies (EWASs) have
28 revealed significant associations of T2D, obesity, and BMI with DNA methylation. However,
29 populations from the Middle East, where T2D and obesity rates are highest worldwide, have not
30 been investigated so far.

31 *Methods:* We performed the first EWAS in an Arab population with T2D and BMI and attempted
32 to replicate 47 EWAS associations previously reported in Caucasians. We used the Illumina
33 Infinium HumanMethylation450 BeadChip to quantify DNA methylation in whole blood DNA from
34 123 subjects of fifteen multigenerational families from Qatar. To investigate the effect of differing
35 genetic background and environment on the epigenetic associations, we further assessed the
36 effect of replicated loci in 810 twins from UK.

37 *Results:* Our EWAS suggested a novel association between T2D and cg06721411 (*DQX1*; p-
38 value= 1.18×10^{-9}). We replicated in the Qatari population seven CpG associations with BMI
39 (*SOCS3*, p-value= 3.99×10^{-6} ; *SREBF1*, p-value= 4.33×10^{-5} ; *SBNO2*, p-value= 5.87×10^{-5} ; *CPT1A*,
40 p-value= 7.99×10^{-5} ; *PRR5L*, p-value= 1.85×10^{-4} ; cg03078551 -intergenic region on chromosome
41 17; p-value= 1.00×10^{-3} ; *LY6G6E*, p-value= 1.10×10^{-3}) and one with T2D (*TXNIP*, p-
42 value= 2.46×10^{-5}). All the associations were further confirmed in the UK cohort for both BMI and
43 T2D. Meta-analysis increased the significance of the observed associations and revealed strong
44 heterogeneity of the effect sizes (apart from *CPT1A*), although associations at these loci
45 showed concordant direction in the two populations.

46 *Conclusion:* Our study replicated eight known CpG associations with T2D or BMI in an Arab
47 population. Heterogeneity of the effects at all loci except *CPT1A* between the Qatari and UK
48 studies suggests that the underlying mechanisms might depend on the genetic background and
49 environmental pressure. Our EWAS results provide a basis for comparison with other
50 ethnicities.

51

52

53

54 INTRODUCTION

55 The Qatari population is one of the under-studied Arab populations in T2D and obesity
56 research, despite the high prevalence of these diseases among Qataris, with an estimated
57 prevalence of ~23% for T2D (Scully 2012) and ~42% for obesity (World Health Organization
58 2015). The increased prevalence of T2D and obesity have occurred during a short period of
59 time (2-3 generations), suggesting an important contribution to the disease risk by changing
60 environment and life style factors, whose effects are potentially mediated by the epigenome.
61 Epigenetic modifications are changes that do not alter the primary DNA sequence itself and
62 include DNA methylation, histone modifications and other changes in chromatin structure that
63 may affect regulation of gene expression (Jaenisch and Bird 2003; Razin and Cedar 1977;
64 Riggs 1975). These epigenetic modifications are thought to provide a link between
65 environmental exposures and clinical phenotypes, and are also suspected to contribute to the
66 unexplained heritability of complex diseases (Manolio et al. 2009).

67 The recent development of genome-wide DNA methylation arrays, and of sequencing
68 technologies coupled with bisulfide treatment present novel opportunities to investigate the role
69 of DNA methylation in complex diseases through epigenome-wide association studies (EWAS).
70 Only a small number of EWASs have been published so far on T2D (Barres et al. 2009; Bell et
71 al. 2010; Chambers et al. 2015; Dayeh et al. 2014; Kulkarni et al. 2015; Toperoff et al. 2012;
72 Volkmar et al. 2012), and on obesity or BMI (Almen et al. 2012; Almen et al. 2014; Demerath et
73 al. 2015; Dick et al. 2014; Feinberg et al. 2010; Sun.D. et al. 2014; Wang et al. 2010). These
74 studies identified new potential T2D- and obesity- associated genes. Furthermore, efforts have
75 been made to study the effect of DNA methylation on gene expression and on metabolic profiles
76 in order to provide better understanding of disease mechanisms (Petersen et al. 2014; Riggs
77 1975). Most EWASs with T2D and obesity conducted so far were focused on Caucasian
78 populations. Little is known about whether their findings translate to other ethnicities and genetic
79 backgrounds. We performed here the first EWAS for T2D and BMI in an Arab population using
80 123 individuals from fifteen Qatari families.

81 Previous genetic studies have shown that translatability between populations does not always
82 hold. For example, genetic variants at the *PPARY* locus that associate with T2D in individuals of
83 European descent seem not to exert any effect on T2D risk in the Qatar population (Badii et al.
84 2008). More discrepancies might be expected for epigenetic risk factors, which are additionally
85 under strong environmental influence. To address this question, we then investigated whether
86 the effect of the replicated loci was homogeneous between Caucasians and the Qatari

87 population by using data from 810 twins from UK, therefore having different genetic background
88 and under different life style/environmental pressure.

89

90 MATERIAL AND METHODS

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

97 **Qatari family study.** The methylation data used in this study was obtained from whole blood
98 (the only easily accessible type of sample), and has been previously described in (Zaghloul et
99 al. 2015). Details on whole genome sequencing can be found in (Kumar et al. 2014). Briefly, the
100 study group consisted of 123 subjects of Qatari descent from 15 families of various sizes and
101 structures. The dataset included 72 females with mean age 39 (\pm 16.9) years and 51 males with
102 mean age 36.3 (\pm 17.2) years. The average BMI of the females was 28.3 (\pm 6.2) kg/m² and of
103 the males was 29.2 (\pm 7.2) kg/m². A total of 30 individuals consisting of 19 females and 11
104 males were previously diagnosed with T2D, ascertained by the diabetes clinic at Hamad
105 Medical Corporation. T2D subjects were receiving treatment for diabetes, and no other major
106 diseases were reported. DNA methylation profiling was performed by Illumina using the Infinium
107 HumanMethylation450K BeadChip platform and reported in the form of Beta-values. After
108 quality control and exclusion of individual probes containing SNPs within a region of \pm 110 base
109 pairs of the CpG site, based on 40x whole genome sequencing data (Kumar et al. 2014), a total
110 of 468,472 probes were selected for this study. Normalization was carried out using the
111 Lumi:QN+BMIQ pipeline, using the *smoothQuantileNormalization* method (**Supplementary**
112 **Figure 1**). Blood cell type coefficients of monocytes, granulocytes, NK-cells, B-cells, CD8⁺-T-
113 cells, and CD4⁺-T-cells, were estimated from the methylation data using the method described
114 by Houseman *et al.* (Houseman et al. 2012).

115 **TwinsUK cohort.** The TwinsUK cohort was established in 1992 to recruit monozygotic and
116 dizygotic twins (Moayyeri et al. 2013). More than 80% of participants are healthy female
117 Caucasians (age range from 16 to 98 years old). The cohort includes more than 13,000 twin

118 participants from all regions across the United Kingdom and many have had multiple visits over
119 the years. The TwinsUK cohort has been used in many epidemiological studies, and is
120 representative of the general UK population for a wide range of diseases and traits (Andrew et
121 al. 2001). DNA methylation was measured for 877 individuals randomly selected from the
122 TwinsUK cohort, 810 of who had both BMI and T2D information. All 810 subjects were female
123 Caucasians. The average BMI was 27.8 (\pm 5.2) kg/m² and 32 individuals were previously
124 diagnosed with T2D. The Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego,
125 CA) was used to measure DNA methylation. Details of experimental approaches have been
126 previously described (Tsaprouni et al. 2014). Normalization was carried out using the 'minfi' R
127 package (Aryee et al. 2014), a procedure equivalent to the Lumi:QN+BMIQ pipeline. DNA
128 methylation probes that mapped incorrectly or to multiple locations in the reference sequence,
129 and probes with detection p-value>0.05 or missing values were removed, resulting in 452,874
130 probes. Blood cell type coefficients were estimated from the methylation data using the method
131 described by Houseman *et al* (Houseman et al. 2012).

132 **Statistical Analysis.** Associations of T2D and BMI with DNA methylation levels in the Qatari
133 family study and TwinsUK cohort were carried out within a variance component framework to
134 model the resemblance among family members. Specifically, the association between the
135 phenotypic traits and DNA methylation levels was evaluated by using a linear mixed model
136 where the total phenotypic variance was partitioned into polygenic and environmental variance,
137 the latter including also measurement errors. DNA methylation levels were modeled as fixed
138 effects, whilst the polygenic and environmental effects were modeled as random components.
139 The phenotypic covariance matrix between subjects was modeled using the matrix of the
140 expected proportion of alleles shared IBD over the genome between each pair of individuals.
141 The significance of the associations was evaluated by comparing the likelihood of a full model
142 including the methylation status in the fixed effect, and the likelihood of a null model where the
143 effect of DNA methylation values was constrained to zero. Age, sex (only for the Qatari dataset),
144 smoking status, and the six Houseman blood cell type coefficients (for B cells, granulocytes,
145 monocytes, Natural Killer cells, CD8⁺-T-cells, and CD4⁺-T-cells) were included in the association
146 model. Additionally, BMI association analysis included T2D status as confounder, and vice
147 versa. BMI values were standardized to have zero mean and one standard deviation. Given the
148 limited sample size, and to avoid potential inflation of the association statistics by directly
149 carrying out the study on selected probes, we preliminarily performed an EWAS in the Qatari

150 family sample using the whole set of probes. False discovery rate was evaluated using Storey's
151 method (Storey 2002).

152 **Selection of CpG sites for replication.** At the time this study was conducted, two large
153 EWASs for T2D (Chambers et al. 2015; Kulkarni et al. 2015) and two for obesity and BMI
154 (Demerath et al. 2015; Dick et al. 2014; Sun.D. et al. 2014) were available. We attempted to
155 replicate the most significant CpG probe for each reported locus that reached genome-wide
156 significance in at least one of these four studies, resulting in eight CpG probes for T2D (**Table 1**)
157 and 39 for obesity and BMI (**Table 2**). Conservative Bonferroni method was used to correct for
158 multiple testing, considering an association replicated with T2D if its p-value was lower than
159 6.25×10^{-3} (0.05/8) and with BMI if its p-value was lower than 1.28×10^{-3} (0.05/39).

160 **Meta-analysis.** Meta-analyses between the Qatari and UK samples were carried out using the
161 GWAMA (Genome-Wide Association Meta-Analysis) software (Magi and Morris 2010).
162 Specifically, we used a fixed-effect model with inverse variance to combine the regression
163 coefficients of each study and their standard errors. Inter-study heterogeneity was estimated by
164 using the Cochran's Q-test and by measuring the proportion of variability that is explained by
165 between-trial heterogeneity (I^2 estimates, (Higgins et al. 2003)), both implemented in GWAMA.

166 RESULTS

167 Our EWAS (**Supplementary Figures 2 and 3**) identified one CpG association with T2D that
168 reached genome-wide Bonferroni significance (p-value $< 1.07 \times 10^{-7}$) (cg06721411 at *DQX1*; p-
169 value= 1.18×10^{-9}). No methylation probes were significantly associated with BMI after Bonferroni
170 correction for multiple testing, the strongest association being at cg17501210 (*RPS6KA2*; p-
171 value= 4.90×10^{-7}). The full EWAS association data is available as **Supplementary Files 1 and**
172 **2**. Q-Q plots (**Supplementary Figure 3**) of the EWASs for BMI and T2D showed mild inflation of
173 the p-value statistics (the genomic inflation factor was 1.10 for T2D and 1.09 for BMI). We also
174 replicated the association of our top T2D CpG cg06721411 (*DQX1*) in the TwinsUK cohort (p-
175 value= 9.00×10^{-3}).

176 We calculated the heritability of DNA methylation at these probes in the Qatari families. At 1%
177 FDR 41,374 (about 10%) methylation levels showed segregation between family members
178 (median heritability=0.70; 1st-3rd quartile=0.31–1.00). The replicated loci showed heritability
179 between 0.46-0.96, apart from cg11024682 (*SREBF1*) and cg07573872 (*SBNO2*) which were
180 not significant at 1% FDR.

181 We replicated the associations in the Qatari family study between T2D and cg19693031
182 (*TXNIP*; p-value= 2.46×10^{-5}) (**Table 1**) and between BMI and CpG sites cg18181703 (*SOCS3*; p-
183 value= 3.99×10^{-6}), cg11024682 (*SREBF1*; p-value= 4.33×10^{-5}), cg07573872 (*SBNO2*; p-
184 value= 5.87×10^{-5}), cg00574958 (*CPT1A*; p-value= 7.99×10^{-5}), cg07136133 (*PRR5L*; p-
185 value= 1.85×10^{-4}), cg03078551 (intergenic region on chromosome 17; p-value= 1.00×10^{-3}),
186 cg13123009 (*LY6G6E*; p-value= 1.10×10^{-3}) (**Table 2**). Boxplots and scatterplots of these
187 associations are in **Figures 1** and **2**. The distributions of the methylation values for these eight
188 CpG sites are in **Supplementary Figure 4**. Although we decided to adopt Bonferroni correction
189 for the replication study, 12 additional associations with BMI showed nominal level of
190 significance and same direction of the associations as the original EWASs (**Table 2**).

191 The eight replicated associations were analyzed in the TwinsUK cohort, and effects were
192 combined in meta-analyses. The meta-analysis of BMI with the TwinsUK results indicated a
193 moderate presence of heterogeneity between the two studies for cg00574958 (*CPT1A*; I^2
194 =56.8%; Cochran's heterogeneity statistic's p-value > 0.05). The meta-analysis increased the
195 significance of this replicated association to p-value= 7.32×10^{-14} . On the other hand, a
196 considerable presence of heterogeneity between the two studies was identified for all the other
197 associations (**Table 3**; Cochran's heterogeneity statistic's p-value < 0.05), despite association at
198 these loci was significant in both populations and with concordant direction (**Table 3**).

199 **DISCUSSION**

200 The high prevalence of T2D and obesity in Qatar motivated the initiation of genetic and
201 epigenetic research in this country. To the best of our knowledge, this is the first association
202 study of CpG methylation with T2D and BMI in an Arab population. We conducted a full EWAS,
203 and attempted to replicate in the Qatari population eight CpG sites associated with T2D (**Table**
204 **1**) and 39 CpG sites associated with BMI (**Table 2**) in Caucasians. Our EWAS with T2D
205 (**Supplementary Figures 2 and 3**) identified one significantly associated CpG site at
206 cg06721411 (*DQX1*; p-value= 1.18×10^{-9}), while the strongest association with BMI at
207 cg17501210 (*RPS6KA2*; p-value= 4.90×10^{-7}) did not reach genome-wide significance. The
208 inflation shown in the QQ-plots is possibly due to hidden confounders, including potentially
209 reduced folate levels in diabetic subjects. However, the observed inflation is only moderate and
210 does not substantially affect our conclusions. As only cg06721411 (*DQX1*) in the T2D EWAS
211 satisfies Bonferroni significance in our discovery cohort, we also replicated this locus in the
212 TwinsUK cohort (p-value= 9.00×10^{-3}). The effect was in the same direction. *DQX1* (DEAQ Box

213 RNA-Dependent ATPase 1) is a protein coding gene located on chromosome 2 and is classified
214 as a member of the DEAD/H family. The highest expression of the *DQX1* is found in the muscle
215 and liver (Ji et al. 2001).

216 Using conservative Bonferroni correction we replicated eight of the 47 associations: *SOCS3*,
217 *SREBF1*, *SBNO2*, *CPT1A*, *PRR5L*, an intergenic region on chromosome 17, and *LY6G6E* with
218 BMI; and *TXNIP1* with T2D, while nominal significance was reached for a further 12 loci
219 associated with BMI (**Table 2**). Despite association with methylation at *SOCS3* being previously
220 reported for both T2D and BMI (Chambers et al. 2015; Kulkarni et al. 2015), only the association
221 with BMI was replicated in our study. However, the association between *SOCS3* and T2D was
222 not significant in the study of Chambers and colleagues after adjustment for BMI, suggesting
223 that the observed association with T2D in their study may be driven by differences in adiposity
224 between their T2D cases and controls.

225 Although the mechanisms linking DNA methylation of *SOCS3*, *SREBF1*, *SBNO2*, *CPT1A*,
226 *PRR5L*, and *LY6GGE* with BMI and *TXNIP* to T2D are not fully established yet, some of these
227 genes have been already functionally linked to metabolic phenotypes. For instance, *TXNIP* is a
228 pro-apoptotic beta-cell factor and encodes for a protein that acts as a regulator of metabolism
229 and an inhibitor of the antioxidant thioredoxins. A recent study showed that *TXNIP* is involved
230 in glucose regulation by controlling insulin sensitivity in the periphery of the human body, and its
231 expression is elevated in the skeletal muscles in T2D patients (Parikh et al. 2007) indicating a
232 linkage to phenotype. We observed concordant results in this study: individuals diagnosed with
233 T2D show lower levels of *TXNIP* methylation, thus suggesting higher *TXNIP* expression. Also,
234 *SOCS3* belongs to the *SOCS* protein family, which is rapidly induced by cytokines, and acts as
235 an inhibitor of various cytokine signaling pathways. Previous studies have shown that *SOCS3* is
236 linked to phenotype by being a negative regulator of leptin (Bjorbaek et al. 1998; Bjorbaek et al.
237 1999) and insulin signaling (Emanuelli et al. 2000; Rui et al. 2002; Shi et al. 2004). In addition,
238 there is evidence for association between variants located near *SOCS3* with glucose
239 homeostasis, BMI and other obesity traits (Talbert et al. 2009; Tang et al. 2011).

240 The replicated methylation sites are within proximity of known genes suggesting a regulatory
241 role of the methylation. However, because expression data is not available for this population,
242 we used data from prior studies to confirm the functional relevance of methylation to the
243 expression of these genes. For instance, in (Chambers et al. 2015; Kulkarni et al. 2015), it was
244 shown that expression of *SREBF1* was reduced in adipose and skeletal muscle of diabetic

245 subjects. *SREBF1* was also shown to regulate carbohydrate metabolism and synthesis in an
246 animal model of obesity and T2D (Ruiz et al. 2014). In another study, qPCR experiments
247 showed that *CPT1A* expression is correlated with the methylation status of *CPT1A* gene with p-
248 value= 4.1×10^{-14} and replicated in the Framingham Heart Study with p-value= 3.1×10^{-13} (Irvin et
249 al. 2014). Differential methylation at *CPT1A* was also found to influence gene expression in Dick
250 et al. (Demerath et al. 2015; Dick et al. 2014; Sun.D. et al. 2014). Also, (Ueki, Kondo, Kahn
251 2004) showed an increase of the suppressor of the cytokine signaling proteins including *SOCS3*
252 in liver, muscle, and fat, in obesity. *SOCS3* over-expression in the fat cells was accompanied by
253 glycogen synthesis and activation of glucose uptake. We also used a recently available
254 database (Bonder et al. 2015) to look up our T2D EWAS hit (*DQX1*) and the 8 replicated loci
255 (*TXNIP*, *SOCS3*, *SREBR1*, *SBNO2*, *CPT1A*, *PRR5L*, cg03078551, and *LY6G6E*). We found
256 that all 9 queried methylation sites had more than 1 association with expression (cis-meQTL,
257 cis-eQTM, and/or trans-meQTL) with p-value of at least 1.24×10^{-7} and $FDR < 0.05$. This provides
258 further evidence for the functional relevance of methylation to the expression of these genes.
259 Therefore, differential methylation may suggest regulatory roles in these different cases.

260 Interestingly, two CpG sites replicated by our study (*CPT1A* and *TXNIP*) and a third CpG we
261 attempted to replicate (*ABCG1*) were also the only probes significantly associated with alpha-
262 hydroxybutyrate (AHB) in our recent EWAS with blood serum metabolomics traits (cg00574958
263 in *CPT1A*, p-value= 1.3×10^{-10} ; cg06500161 in *ABCG1*, p-value= 7.8×10^{-6} ; cg19693031 in *TXNIP*,
264 p-value= 7.2×10^{-8}) (Petersen et al. 2014). AHB is a sub-product of the ketones metabolism;
265 elevated AHB levels indicate potential insulin resistance (Gall et al. 2010) and this biomarker is
266 part of the metabolomics-based pre-diabetes test Quantose™ (Tripathy et al. 2015).

267 Other biomarkers of T2D and pre-diabetes have been previously associated with methylation at
268 *ABCG1*, *TXNIP* and *CPT1A*, including 3-methyl-2-oxovalerate (Menni et al. 2013), glycine
269 (Ferrannini et al. 2013), several lipid traits, including phosphatidylcholines (PCs) (Suhre et al.
270 2010), chylomicrons and their remnants, VLDL and IDL cholesterol particles (Petersen et al.
271 2014). They all showed diabetes related effect directions that are in agreement with the effect
272 directions observed in this study. Most interestingly, the list of metabolic traits associated with
273 CpGs (Petersen et al. 2014) also includes the product of *CPT1A* itself, palmitoylcarnitine.
274 Furthermore, higher levels of cg00574958 (*CPT1A*) methylation were also associated with
275 higher levels of related long-chain fatty acids in the EWAS reported by Petersen et al., including
276 palmitate (16:0), stearate (18:0), and oleate (18:1n9) [see Suppl. Table 5 of (Petersen et al.
277 2014). Higher levels of cg06500161 (*ABCG1*) methylation were also associated in a recent

278 EWAS with higher levels of chylomicrons and VLDL-cholesterol [see Supplementary Table 5 of
279 (Petersen et al. 2014) and **Table 4**].

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

287 Replicated associations identified in this study were also confirmed in the TwinsUK cohort
288 (**Table 3**). Meta-analysis increased the significance of the associations, but highlighted
289 heterogeneity of the effect sizes for all loci but *CPT1A*. Some heterogeneity of effects between
290 our results and what was reported in the original papers might be expected, as they could be
291 driven by potential differences in the normalization pipeline of the array data, or by the
292 correction of the methylation values using different confounders. However, despite there were
293 no differences in the normalization pipeline or in the use of confounders between the Qatari
294 family sample and the TwinsUK cohort, we still observed significant effect heterogeneity. This
295 heterogeneity may partly be explained by the different environmental pressures. While the
296 standardized BMI distribution was not different between the two samples (Kolmogorov–Smirnov
297 $P>0.05$) the distribution of six out of eight methylation values at the tested probes were different
298 in either location or shape (Kolmogorov–Smirnov $P<0.05$; **Supplementary Figure 4**).

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

308 Second, DNA methylation was measured using DNA extracted from whole blood that was the
309 only accessible type of sample and may not be representative of more disease-relevant tissues

310 for the diseases under study, such as pancreatic cells and adipose tissues. Studies of
311 methylation in obesity or T2D based on disease-relevant tissues such as skeletal muscle,
312 adipose tissue, or pancreatic islets are interesting, but only exist for relatively small studies

313 (Benton et al. 2015; Dayeh and Ling 2015; Ling et al. 2008; Maples et al. 2015a; Maples et al.
314 2015b; Nilsson et al. 2014; Nilsson et al. 2015; Ronn et al. 2015; Sjostrom et al. 2009). Since
315 DNA methylation can be strongly tissue dependent, and as our data was obtained from blood
316 samples, for consistency, we only selected methylation probes for replication from EWAS
317 studies that were also performed in blood. In addition, blood consists of various cell types
318 (including B cells, granulocytes, monocytes, Natural Killer cells, and T cells subset) that may
319 bias methylation estimates. Estimation of cell type coefficients from the methylation data using a
320 method described by (Houseman et al. 2012), and correction for these coefficients in the
321 association model is common practice and it is also applied here.

322 **CONCLUSION**

323 Given the early state of the epigenome-wide technologies, the number of published EWASs on
324 T2D and obesity so far is small. However, the availability of the technology along with the
325 availability of novel computational tools is expected to accelerate increase in the number of
326 studies conducted in this field. To the best of our knowledge, this study is the first EWAS of T2D
327 and obesity in an Arab population. Our EWAS identified one new CpG association with T2D at
328 *DQX1* that reached genome-wide Bonferroni significance. We also replicated 8 previously
329 reported T2D and BMI associations, although they were not genome-wide significant, confirming
330 the relevance of these CpG sites to these phenotypes.

331

332 **COMPETING INTERESTS**

333 The authors have no conflict of interest to disclose.

334

335 **AUTHORS' CONTRIBUTIONS**

336 WAAM, MAS, MF, and KS designed the study. MAS and WAAM collected the samples and
337 generated the data. SBZ, AV, and PK analyzed the data. WAAM, MAS, SBZ, AV, MF, and KS
338 wrote the manuscript. All authors read and approved the final manuscript.

339 **ACKNOWLEDGEMENTS**

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

342 **FUNDING**

343 This work was supported by the Qatar Science Leadership Program funds at the Research
344 Division, a program funded by the Qatar Foundation. This work was also supported by
345 Biomedical Research Program funds at Weill Cornell Medical College in Qatar, a program
346 funded by the Qatar Foundation. MF is in receipt of support from Qatar Foundation grant
347 GEQATDIAB. AV is funded by the British Skin Foundation, grant 5044i. The data coming from
348 the TwinsUK study was funded by the Wellcome Trust; European Community's Seventh
349 Framework Programme (FP7/2007-2013). The study also receives support from the National
350 Institute for Health Research (NIHR) - funded BioResource, Clinical Research Facility and
351 Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in
352 partnership with King's College London. The statements made herein are solely the
353 responsibility of the authors.

354 **References**

- 355 Almen MS, Nilsson EK, Jacobsson JA, Kalnina I, Klovins J, Fredriksson R, Schioth HB. 2014.
356 Genome-wide analysis reveals DNA methylation markers that vary with both age and
357 obesity. *Gene* 548(1):61-7.
- 358 Almen MS, Jacobsson JA, Moschonis G, Benedict C, Chrousos GP, Fredriksson R, Schioth HB.
359 2012. Genome wide analysis reveals association of a FTO gene variant with epigenetic
360 changes. *Genomics* 99(3):132-7.
- 361 Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. 2001. Are twins and
362 singletons comparable? A study of disease-related and lifestyle characteristics in adult
363 women. *Twin Res* 4(6):464-77.
- 364 Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA.
365 2014. Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium
366 DNA methylation microarrays. *Bioinformatics* 30(10):1363-9.
- 367 Badii R, Bener A, Zirie M, Al-Rikabi A, Simsek M, Al-Hamaq AO, Ghousaini M, Froguel P,
368 Wareham NJ. 2008. Lack of association between the Pro12Ala polymorphism of the PPAR-
369 gamma 2 gene and type 2 diabetes mellitus in the qatari consanguineous population. *Acta*
370 *Diabetol* 45(1):15-21.

371 Barres R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, Krook A, Zierath JR. 2009. Non-CpG
372 methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density.
373 *Cell Metab* 10(3):189-98.

374 Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, Akan P, Stupka E,
375 Down TA, Prokopenko I, et al. 2010. Integrated genetic and epigenetic analysis identifies
376 haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus.
377 *PLoS One* 5(11):e14040.

378 Benton MC, Johnstone A, Eccles D, Harmon B, Hayes MT, Lea RA, Griffiths L, Hoffman EP,
379 Stubbs RS, Macartney-Coxson D. 2015. An analysis of DNA methylation in human adipose
380 tissue reveals differential modification of obesity genes before and after gastric bypass and
381 weight loss. *Genome Biol* 16:8,014-0569-x.

382 Bjorbaek C, El-Haschimi K, Frantz JD, Flier JS. 1999. The role of SOCS-3 in leptin signaling
383 and leptin resistance. *J Biol Chem* 274(42):30059-65.

384 Bjorbaek C, Elmquist JK, Frantz JD, Shoelson SE, Flier JS. 1998. Identification of SOCS-3 as a
385 potential mediator of central leptin resistance. *Mol Cell* 1(4):619-25.

386 Bonder MJ, Luijk R, Zhernakova D, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F,
387 van Galen M, Bot J, et al. 2015. Disease variants alter transcription factor levels and
388 methylation of their binding sites. *bioRxiv* .

389 Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, Wahl S, Elliott HR, Rota F, Scott
390 WR, et al. 2015. Epigenome-wide association of DNA methylation markers in peripheral
391 blood from indian asians and europeans with incident type 2 diabetes: A nested case-
392 control study. *Lancet Diabetes Endocrinol* 3(7):526-34.

393 Dayeh T and Ling C. 2015. Does epigenetic dysregulation of pancreatic islets contribute to
394 impaired insulin secretion and type 2 diabetes? *Biochem Cell Biol* :1-11.

395 Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB,
396 Eliasson L, Ronn T, et al. 2014. Genome-wide DNA methylation analysis of human
397 pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes
398 that influence insulin secretion. *PLoS Genet* 10(3):e1004160.

399 Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman AK,
400 Sandling JK, Li LA, Irvin MR, et al. 2015. Epigenome-wide association study (EWAS) of
401 BMI, BMI change and waist circumference in african american adults identifies multiple
402 replicated loci. *Hum Mol Genet* 24(15):4464-79.

403 Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, Meduri E, Morange PE,
404 Gagnon F, Grallert H, et al. 2014. DNA methylation and body-mass index: A genome-wide
405 analysis. *Lancet* 383(9933):1990-8.

406 Emanuelli B, Peraldi P, Filloux C, Sawka-Verhelle D, Hilton D, Van Obberghen E. 2000. SOCS-
407 3 is an insulin-induced negative regulator of insulin signaling. *J Biol Chem* 275(21):15985-
408 91.

409 Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris
410 TB, Launer L, Gudnason V, et al. 2010. Personalized epigenomic signatures that are stable
411 over time and covary with body mass index. *Sci Transl Med* 2(49):49ra67.

412 Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, Milburn MV, Kastenmuller
413 G, Adamski J, Tuomi T, et al. 2013. Early metabolic markers of the development of
414 dysglycemia and type 2 diabetes and their physiological significance. *Diabetes* 62(5):1730-
415 7.

416 Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, Ryals JA, Milburn MV,
417 Nannipieri M, Camastra S, et al. 2010. Alpha-hydroxybutyrate is an early biomarker of
418 insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One*
419 5(5):e10883.

420 Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-
421 analyses. *BMJ* 327(7414):557-60.

422 Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke
423 JK, Kelsey KT. 2012. DNA methylation arrays as surrogate measures of cell mixture
424 distribution. *BMC Bioinformatics* 13:86,2105-13-86.

425 Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibeault KS, Patel N,
426 Day K, Jones LW, et al. 2014. Epigenome-wide association study of fasting blood lipids in
427 the genetics of lipid-lowering drugs and diet network study. *Circulation* 130(7):565-72.

428 Jaenisch R and Bird A. 2003. Epigenetic regulation of gene expression: How the genome
429 integrates intrinsic and environmental signals. *Nat Genet* 33 Suppl:245-54.

430 Ji W, Chen F, Do T, Do A, Roe BA, Meisler MH. 2001. DQX1, an RNA-dependent ATPase
431 homolog with a novel DEAQ box: Expression pattern and genomic sequence comparison of
432 the human and mouse genes. *Mamm Genome* 12(6):456-61.

433 Kulkarni H, Kos MZ, Neary J, Dyer TD, Kent JW, Jr, Goring HH, Cole SA, Comuzzie AG, Almasy
434 L, Mahaney MC, et al. 2015. Novel epigenetic determinants of type 2 diabetes in mexican-
435 american families. *Hum Mol Genet* 24(18):5330-44.

436 Kumar P, Al-Shafai M, Al Muftah WA, Chalhoub N, Elsaid MF, Aleem AA, Suhre K. 2014.
437 Evaluation of SNP calling using single and multiple-sample calling algorithms by validation
438 against array base genotyping and mendelian inheritance. *BMC Res Notes* 7:747,0500-7-
439 747.

440 Ling C, Del Guerra S, Lupi R, Ronn T, Granhall C, Luthman H, Masiello P, Marchetti P, Groop
441 L, Del Prato S. 2008. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets
442 and effect on insulin secretion. *Diabetologia* 51(4):615-22.

443 Magi R and Morris AP. 2010. GWAMA: Software for genome-wide association meta-analysis.
444 *BMC Bioinformatics* 11:288,2105-11-288.

- 445 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos
446 EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex
447 diseases. *Nature* 461(7265):747-53.
- 448 Maples JM, Brault JJ, Witczak CA, Park S, Hubal MJ, Weber TM, Houmard JA, Shewchuk BM.
449 2015a. Differential epigenetic and transcriptional response of the skeletal muscle carnitine
450 palmitoyltransferase 1B (CPT1B) gene to lipid exposure with obesity. *Am J Physiol*
451 *Endocrinol Metab* 309(4):E345-56.
- 452 Maples JM, Brault JJ, Shewchuk BM, Witczak CA, Zou K, Rowland N, Hubal MJ, Weber TM,
453 Houmard JA. 2015b. Lipid exposure elicits differential responses in gene expression and
454 DNA methylation in primary human skeletal muscle cells from severely obese women.
455 *Physiol Genomics* 47(5):139-46.
- 456 Menni C, Fauman E, Erte I, Perry JR, Kastenmuller G, Shin SY, Petersen AK, Hyde C, Psatha
457 M, Ward KJ, et al. 2013. Biomarkers for type 2 diabetes and impaired fasting glucose using
458 a nontargeted metabolomics approach. *Diabetes* 62(12):4270-6.
- 459 Moayyeri A, Hammond CJ, Valdes AM, Spector TD. 2013. Cohort profile: TwinsUK and healthy
460 ageing twin study. *Int J Epidemiol* 42(1):76-85.
- 461 Nilsson E, Matte A, Perfilyev A, de Mello VD, Kakela P, Pihlajamaki J, Ling C. 2015. Epigenetic
462 alterations in human liver from subjects with type 2 diabetes in parallel with reduced folate
463 levels. *J Clin Endocrinol Metab* 100(11):E1491-501.
- 464 Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, Poulsen P, Ribel-
465 Madsen R, Pedersen NL, Almgren P, et al. 2014. Altered DNA methylation and differential
466 expression of genes influencing metabolism and inflammation in adipose tissue from
467 subjects with type 2 diabetes. *Diabetes* 63(9):2962-76.
- 468 Parikh H, Carlsson E, Chutkow WA, Johansson LE, Storgaard H, Poulsen P, Saxena R, Ladd
469 C, Schulze PC, Mazzini MJ, et al. 2007. TXNIP regulates peripheral glucose metabolism in
470 humans. *PLoS Med* 4(5):e158.
- 471 Petersen AK, Zeilinger S, Kastenmuller G, Romisch-Margl W, Brugger M, Peters A, Meisinger
472 C, Strauch K, Hengstenberg C, Pagel P, et al. 2014. Epigenetics meets metabolomics: An
473 epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet*
474 23(2):534-45.
- 475 Razin A and Cedar H. 1977. Distribution of 5-methylcytosine in chromatin. *Proc Natl Acad Sci U*
476 *S A* 74(7):2725-8.
- 477 Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet*
478 14(1):9-25.
- 479 Ronn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL, Jorgensen SW, Brons C,
480 Jansson PA, Eriksson KF, et al. 2015. Impact of age, BMI and HbA1c levels on the
481 genome-wide DNA methylation and mRNA expression patterns in human adipose tissue
482 and identification of epigenetic biomarkers in blood. *Hum Mol Genet* 24(13):3792-813.

483 Rui L, Yuan M, Frantz D, Shoelson S, White MF. 2002. SOCS-1 and SOCS-3 block insulin
484 signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *J Biol Chem* 277(44):42394-
485 8.

486 Ruiz R, Jideonwo V, Ahn M, Surendran S, Tagliabracci VS, Hou Y, Gamble A, Kerner J, Irimia-
487 Dominguez JM, Puchowicz MA, et al. 2014. Sterol regulatory element-binding protein-1
488 (SREBP-1) is required to regulate glycogen synthesis and gluconeogenic gene expression
489 in mouse liver. *J Biol Chem* 289(9):5510-7.

490 Scully T. 2012. Diabetes in numbers. *Nature Outlook* 485(7398):S2–S3.

491 Shi H, Tzamelis I, Bjorbaek C, Flier JS. 2004. Suppressor of cytokine signaling 3 is a
492 physiological regulator of adipocyte insulin signaling. *J Biol Chem* 279(33):34733-40.

493 Sjostrom L, Gummesson A, Sjostrom CD, Narbro K, Peltonen M, Wedel H, Bengtsson C,
494 Bouchard C, Carlsson B, Dahlgren S, et al. 2009. Effects of bariatric surgery on cancer
495 incidence in obese patients in sweden (swedish obese subjects study): A prospective,
496 controlled intervention trial. *Lancet Oncol* 10(7):653-62.

497 Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical*
498 *Society. Series B (Statistical Methodology)* 64(3):479-98.

499 Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall
500 WE, Weinberger KM, et al. 2010. Metabolic footprint of diabetes: A multiplatform
501 metabolomics study in an epidemiological setting. *PLoS One* 5(11):e13953.

502 Race-specific association between DNA methylation and body mass index: the Bogalusa Heart
503 Study [Internet]; c2014 [cited 2014 . Available from:
504 <http://www.ashg.org/2014meeting/abstracts/fulltext/f140122628.htm> .

505 Talbert ME, Langefeld CD, Ziegler J, Mychaleckyj JC, Haffner SM, Norris JM, Bowden DW.
506 2009. Polymorphisms near SOCS3 are associated with obesity and glucose homeostasis
507 traits in hispanic americans from the insulin resistance atherosclerosis family study. *Hum*
508 *Genet* 125(2):153-62.

509 Tang W, Zou JJ, Chen XF, Zheng JY, Zeng HZ, Liu ZM, Shi YQ. 2011. Association of two
510 polymorphisms within and near SOCS3 gene with obesity in three nationalities in xinjiang
511 province of china. *Acta Pharmacol Sin* 32(11):1381-6.

512 Toperoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissan B, Wainstein J, Friedlander Y,
513 Levy-Lahad E, Glaser B, et al. 2012. Genome-wide survey reveals predisposing diabetes
514 type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet*
515 21(2):371-83.

516 Tripathy D, Cobb JE, Gall W, Adam KP, George T, Schwenke DC, Banerji M, Bray GA,
517 Buchanan TA, Clement SC, et al. 2015. A novel insulin resistance index to monitor
518 changes in insulin sensitivity and glucose tolerance: The ACT NOW study. *J Clin*
519 *Endocrinol Metab* 100(5):1855-62.

520 Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Vinuela A, Grundberg E, Nelson
521 CP, Meduri E, et al. 2014. Cigarette smoking reduces DNA methylation levels at multiple
522 genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 9(10):1382-
523 96.

524 Ueki K, Kondo T, Kahn CR. 2004. Suppressor of cytokine signaling 1 (SOCS-1) and SOCS-3
525 cause insulin resistance through inhibition of tyrosine phosphorylation of insulin receptor
526 substrate proteins by discrete mechanisms. *Mol Cell Biol* 24(12):5434-46.

527 Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, Calonne E,
528 Volkmar U, Igoillo-Esteve M, Naamane N, et al. 2012. DNA methylation profiling identifies
529 epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J*
530 31(6):1405-26.

531 Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, Maria BL, Dong Y, Treiber F, Gutin B,
532 et al. 2010. Obesity related methylation changes in DNA of peripheral blood leukocytes.
533 *BMC Med* 8:87,7015-8-87.

534 Global health observatory data repository [Internet]; c2015 [cited 2015 February 5]. Available
535 from: <http://apps.who.int/gho/data/node.main.A900A?lang=en> .

536 Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Falchi M, Suhre K. 2015. Association of
537 DNA methylation with age, gender, and smoking in an arab population. *Clin Epigenetics*
538 7(1):6,014-0040-6. eCollection 2015.

539

540

541

FIGURES AND TABLES

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	<i>TXNIP</i>	-2.41	0.57	2.46x10 ⁻⁵	(Chambers et al. 2015; Kulkarni et al. 2015)
cg00574958	11	68607622	<i>CPT1A</i>	-3.77	2.07	0.068	(Chambers et al. 2015; Kulkarni et al. 2015)
cg11024682	17	17730094	<i>SREBF1</i>	1.99	1.30	0.124	(Chambers et al. 2015)
cg09152259	2	128156114	<i>PROC</i>	-1.02	0.71	0.148	(Chambers et al. 2015)
cg06500161	21	43656587	<i>ABCG1</i>	2.17	1.52	0.153	(Chambers et al. 2015)
cg04999691	7	150027050	<i>C7orf29</i>	1.87	1.40	0.182	(Chambers et al. 2015)
cg02650017	17	47301614	<i>PHOSPHO1</i>	-3.22	2.58	0.212	(Chambers et al. 2015)
cg18181703	17	76354621	<i>SOCS3</i>	-0.66	0.91	0.465	(Chambers et al. 2015)

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.

Probe	Chr	Position	Gene Symbol	Beta	SE	p-value	Reference
cg18181703	17	76354621	<i>SOCS3</i>	-10.78	2.34	3.99x10 ⁻⁶	(Chambers et al. 2015)
cg11024682	17	17730094	<i>SREBF1</i>	14.56	3.56	4.33x10 ⁻⁵	(Demerath et al. 2015)
cg07573872	19	1126342	<i>SBNO2</i>	-9.96	2.48	5.87x10 ⁻⁵	(Demerath et al. 2015)
cg00574958	11	68607622	<i>CPT1A</i>	-21.04	5.33	7.99x10 ⁻⁵	(Demerath et al. 2015)
cg07136133	11	36422377	<i>PRR5L</i>	-10.43	2.79	1.85x10 ⁻⁴	(Demerath et al. 2015)
cg03078551	17	41656298	NA	-19.23	5.85	1.00x10 ⁻³	(Demerath et al. 2015)
cg13123009	6	31681882	<i>LY6G6E</i>	12.41	3.80	1.10x10 ⁻³	(Demerath et al. 2015)
cg09349128	22	50327986	NA	-9.60	3.04	1.60x10 ⁻³	(Demerath et al. 2015)
cg08972190	7	2138995	<i>MAD1L1</i>	11.11	3.54	1.70x10 ⁻³	(Demerath et al. 2015)
cg06192883	15	52554171	<i>MYO5C</i>	6.68	2.27	3.20x10 ⁻³	(Demerath et al. 2015)
cg06500161	21	43656587	<i>ABCG1</i>	11.90	4.17	4.30x10 ⁻³	(Chambers et al. 2015)
cg27243685	21	43642366	<i>ABCG1</i>	11.01	4.11	7.30x10 ⁻³	(Demerath et al. 2015)
cg06946797	16	11422409	NA	-6.24	2.36	8.10x10 ⁻³	(Demerath et al. 2015)
cg12992827	3	101901234	NA	-5.60	2.14	8.90x10 ⁻³	(Demerath et al. 2015)
cg23998749	1	154968781	NA	7.47	2.87	9.30x10 ⁻³	(Demerath et al. 2015)
cg26354221	22	24822802	<i>ADORA2A</i>	12.11	4.77	0.011	(Demerath et al. 2015)
cg11592786	15	89533581	NA	-28.45	12.46	0.022	(Demerath et al. 2015)
cg26403843	5	158634085	<i>RNF145</i>	3.63	1.61	0.024	(Chambers et al. 2015)
cg26033520	10	74004071	NA	4.92	2.20	0.025	(Demerath et al. 2015))
cg01844514	7	149557121	<i>ZNF862</i>	6.55	3.48	0.060	(Demerath et al. 2015)
cg14017402	2	86225602	NA	3.98	2.35	0.090	(Demerath et al. 2015)
cg22891070	19	46801642	<i>HIF3A</i>	1.64	1.05	0.120	(Dick et al. 2014)

							2015)
cg06876354	2	121020189	<i>RALB</i>	7.28	4.76	0.126	(Demerath et al. 2015)
cg15871086	18	56526595	NA	4.08	2.86	0.155	(Demerath et al. 2015)
cg07814318	15	31624584	<i>KLF13</i>	3.64	2.60	0.161	(Dick et al. 2014)
cg04816311	7	1066650	<i>C7orf50</i>	2.07	1.98	0.295	(Demerath et al. 2015)
cg04927537	17	76976091	<i>LGALS3BP</i>	1.95	1.87	0.297	(Demerath et al. 2015)
cg04869770	1	164561550	<i>PBX1</i>	2.46	2.50	0.325	(Demerath et al. 2015)
cg25178683	17	76976267	<i>LGALS3BP</i>	2.23	2.38	0.350	(Demerath et al. 2015)
cg20954977	2	232260116	<i>B3GNT7</i>	1.65	1.93	0.391	(Demerath et al. 2015)
cg18568872	15	90606494	<i>ZNF710</i>	2.41	3.55	0.497	(Demerath et al. 2015)
cg00863378	16	56549757	<i>BBS2</i>	1.37	2.51	0.584	(Demerath et al. 2015)
cg17560136	8	21915510	<i>EPB49</i>	1.18	2.64	0.654	(Demerath et al. 2015)
cg13708645	12	121974305	<i>KDM2B</i>	0.67	1.71	0.697	(Demerath et al. 2015)
cg15695155	12	121973871	<i>KDM2B</i>	0.81	3.44	0.813	(Demerath et al. 2015)
cg27614723	15	92399897	<i>SLCO3A1</i>	0.81	4.02	0.840	(Demerath et al. 2015)
cg09664445	17	2612406	<i>CLUH</i>	-0.86	5.28	0.871	(Dick et al. 2014)
cg18307303	5	158757456	<i>IL12B</i>	-0.24	3.39	0.943	(Demerath et al. 2015)

Table 3. Meta-analyses of the replicated loci in the Qatari study with the TwinsUK results. The table reports the results we obtained using a fixed-effect 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).

Trait	Probe (Gene)	Qatari Cohort			TwinsUK Cohort			Meta-analysis			
		Beta	Se	p-value	Beta	Se	p-value	Beta (95U/95L)	se	p-value	I^2
T2D	cg19693031 (<i>TXNIP</i>)	-2.41	0.57	2.46×10^{-5}	-0.34	0.12	6.74×10^{-3}	-0.43 (-0.66/-0.20)	0.12	2.71×10^{-4}	92.1%
BMI	cg18181703 (<i>SOCS3</i>)	-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}	90.4%
	cg11024682 (<i>SREBF1</i>)	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}	82.2%
	cg07573872 (<i>SBNO2</i>)	-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}	87.3%
	cg00574958 (<i>CPT1A</i>)	-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}	56.8%
	cg07136133 (<i>PRR5L</i>)	-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}	80.4%
	cg03078551 (<i>NA</i>)	-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}	74.3%
	cg13123009 (<i>LY6G6E</i>)	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}	82.2%

Table 4. CpG–metabotype associations at the three replicated loci. Only associations with metabolotypes that were significant at P-value < 1.3×10^{-5} (Bonferroni correction of testing multiple metabolic traits) are shown; effect size (Beta'), P-value of the linear model, and number of samples (N) [data from Suppl. Tab. 5 of (Petersen et al. 2014), for details on this dataset see there].

Metabolic trait	cg06500161 (ABCG1)			cg00574958 (CPT1A)			cg19693031 (TXNIP)		
	Beta'	P-value	N	Beta'	P-value	N	Beta'	P-value	N
1-oleoylglycerol (1-monoolein)	9.435	2.84×10^{-11}	1676	-0.983	4.37×10^{-11}	1676	-0.841	1.02×10^{-14}	1674
alpha-hydroxybutyrate (AHB)	1.834	7.80×10^{-6}	1749	-0.926	1.30×10^{-10}	1749	-0.574	7.24×10^{-8}	1747
3-methyl-2-oxovalerate	0.771	1.31×10^{-5}	1749	-0.605	4.50×10^{-5}	1749	-0.472	6.81×10^{-13}	1747
Glycine	-0.531	2.91×10^{-6}	1744	4.064	7.58×10^{-9}	1744	0.656	4.91×10^{-6}	1742
Palmitoylcarnitine	1.729	1.35×10^{-5}	1737	-0.849	2.28×10^{-6}	1737	-0.579	3.11×10^{-8}	1735
PC aa C36:4	2.549	2.82×10^{-5}	1781	-0.947	2.07×10^{-8}	1781	-0.628	1.66×10^{-6}	1779
PC aa C42:0	-0.883	3.91×10^{-8}	1781	22.276	3.36×10^{-6}	1781	6.181	1.11×10^{-13}	1779
PC aa C42:1	-0.844	4.63×10^{-7}	1781	12.688	4.18×10^{-5}	1781	4.631	4.93×10^{-12}	1779
PC ae C44:6	-0.918	8.52×10^{-12}	1781	17.438	4.88×10^{-6}	1781	5.075	5.48×10^{-13}	1779
Chylo-A (nM)	594.7	8.44×10^{-14}	1773	-1.000	1.35×10^{-13}	1773	-0.996	1.11×10^{-21}	1771
Chylo-B (nM)	29.53	7.09×10^{-6}	1766	-1.000	3.24×10^{-12}	1766	-0.988	8.30×10^{-18}	1764
Chylo-Rem (nM)	416.1	1.34×10^{-11}	1772	-1.000	2.38×10^{-12}	1772	-0.991	5.40×10^{-15}	1770
IDL (nM)	30.86	1.32×10^{-9}	1773	-0.993	6.00×10^{-7}	1773	-0.868	2.22×10^{-7}	1771
VLDL-A (nM)	150.7	2.72×10^{-13}	1773	-1.000	9.23×10^{-14}	1773	-0.985	5.73×10^{-19}	1771
VLDL-B (nM)	166.0	4.22×10^{-13}	1773	-1.000	9.83×10^{-12}	1773	-0.979	9.24×10^{-16}	1771

FIGURES

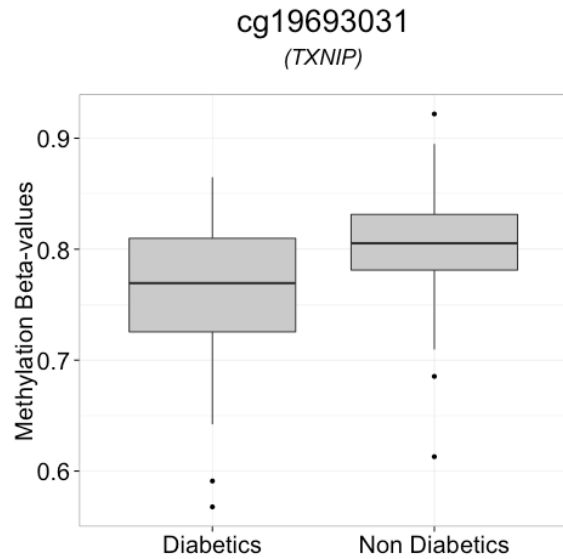
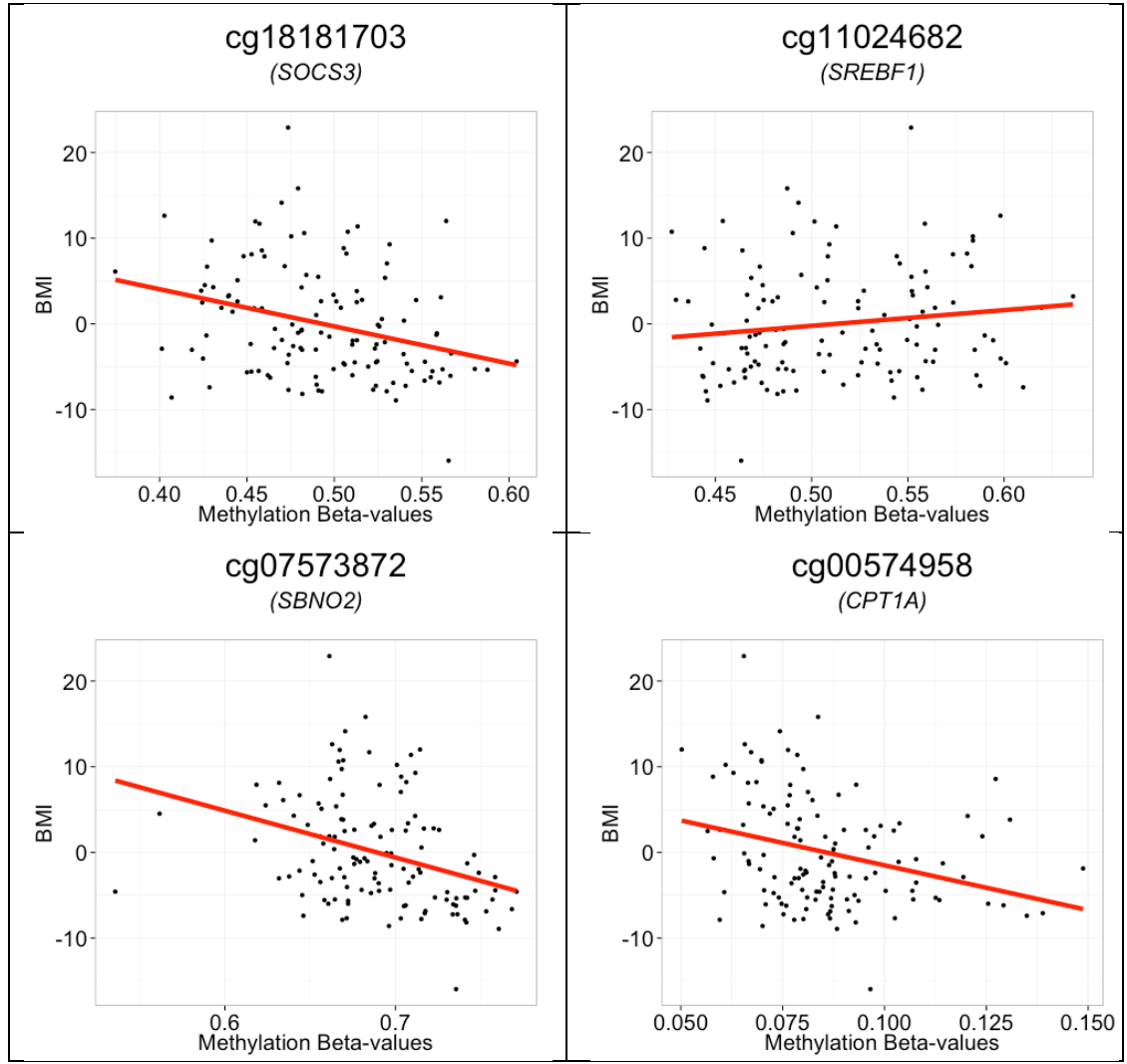


Figure 1. Boxplot of methylation Beta-values at cg19693031 (TXNIP) against state. The middle lines show the medians of the data while the boxes show the percentiles (Q1 and Q3). The whiskers extend to include 99% of the data above which represent outliers. Beta-value distributions at this probe in the diabetics and healthy showed a difference in the levels of background methylation (Wilcoxon test p-value=2.



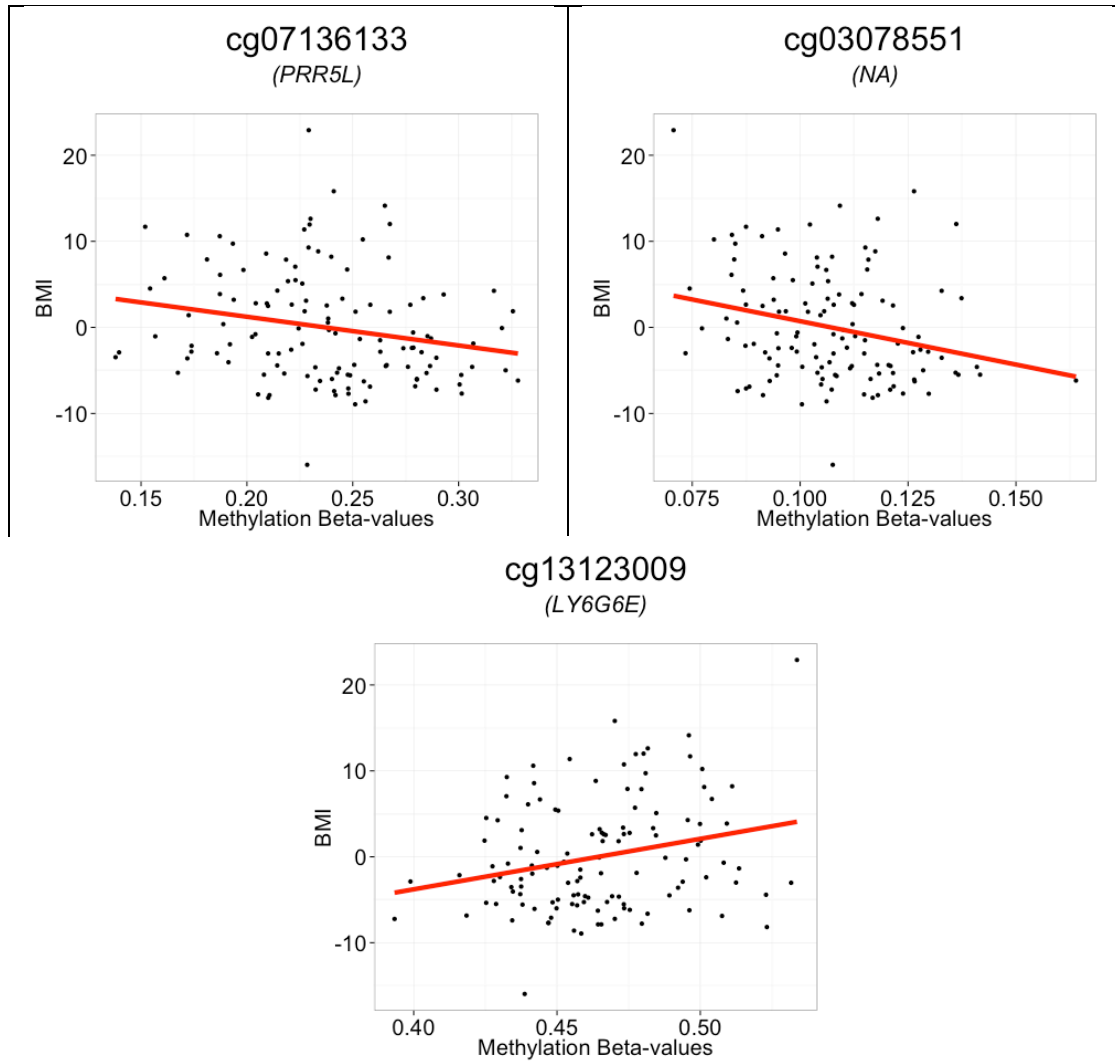
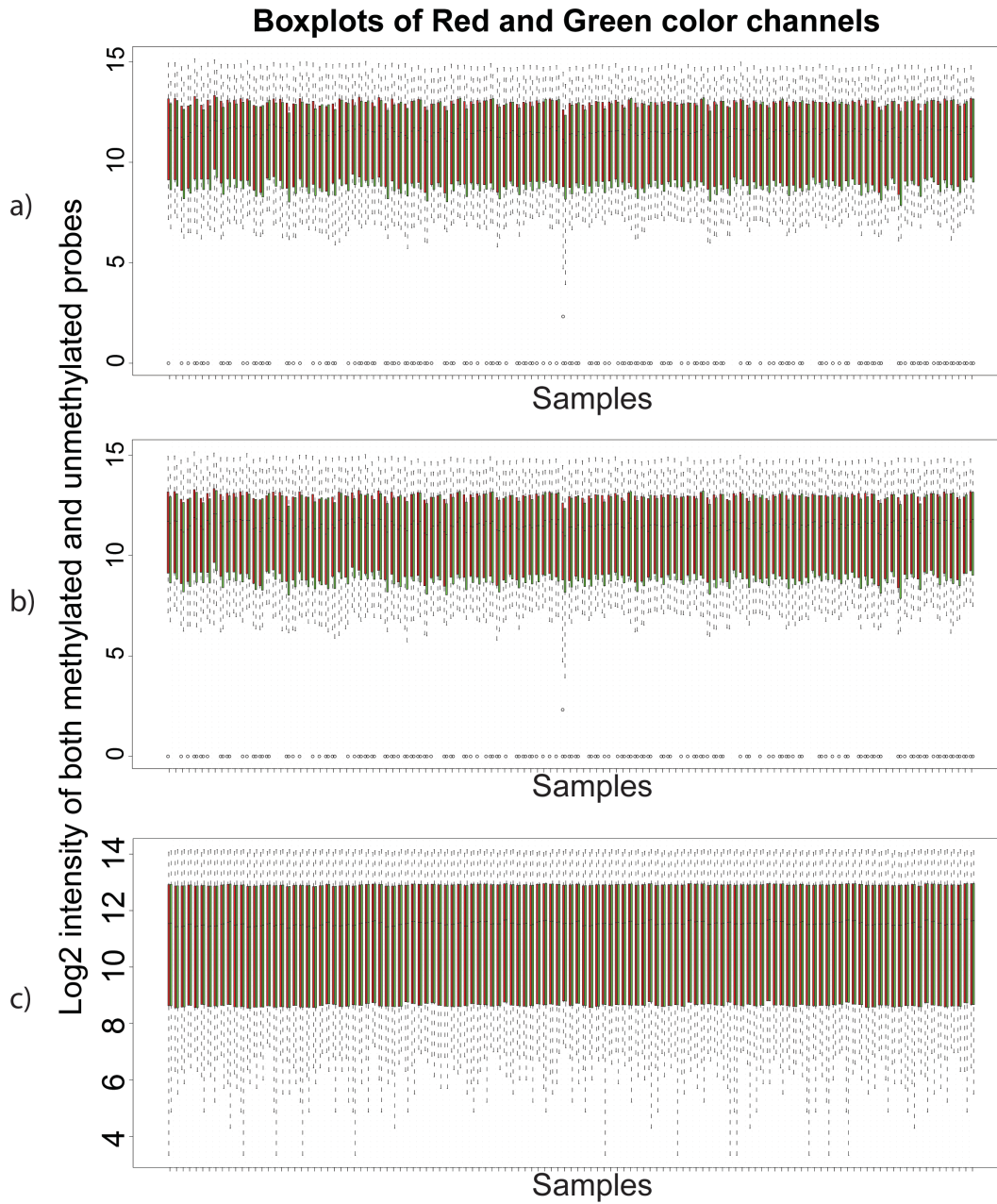
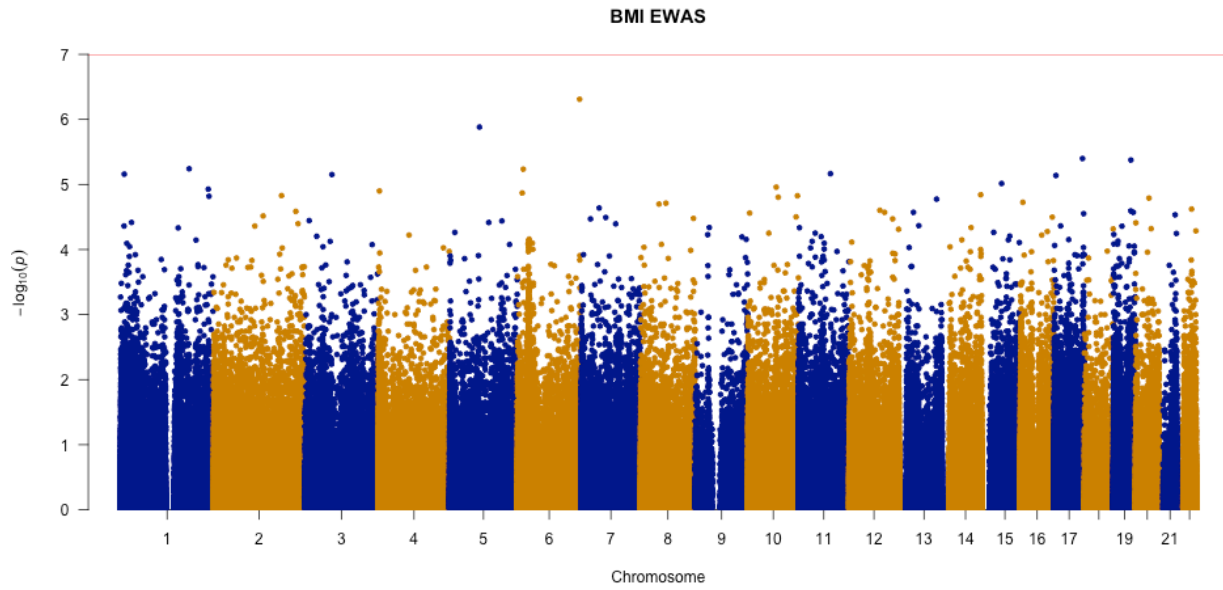


Figure 2. Scatterplots of BMI against methylation Beta-values. Red lines represent the slopes of the regression model. BMI values were corrected for age and sex.

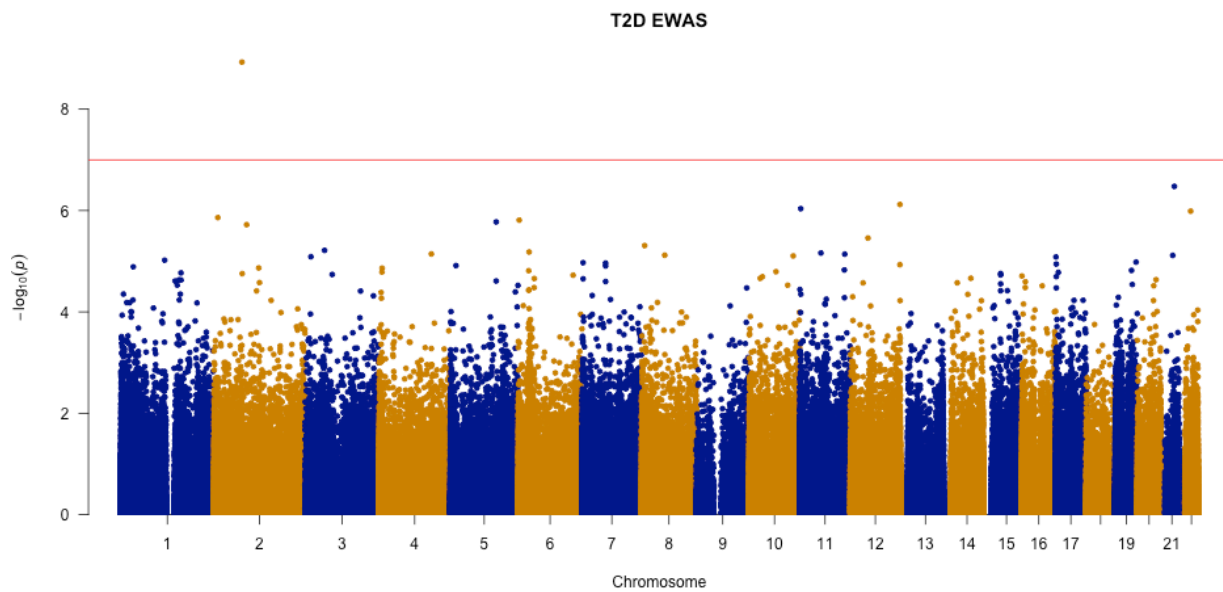
SUPPLEMENTARY FIGURES



Supplementary Figure 1. DNA methylation for the 123 samples presented as boxplots. Circles represent outliers and the red and green boxes represent the two color channels showing the effect of the quality control on the data: (a) before preprocessing, (b) after color bias adjustment, and (c) after quantile normalization (see (Zaghlool et al. 2015) for details).

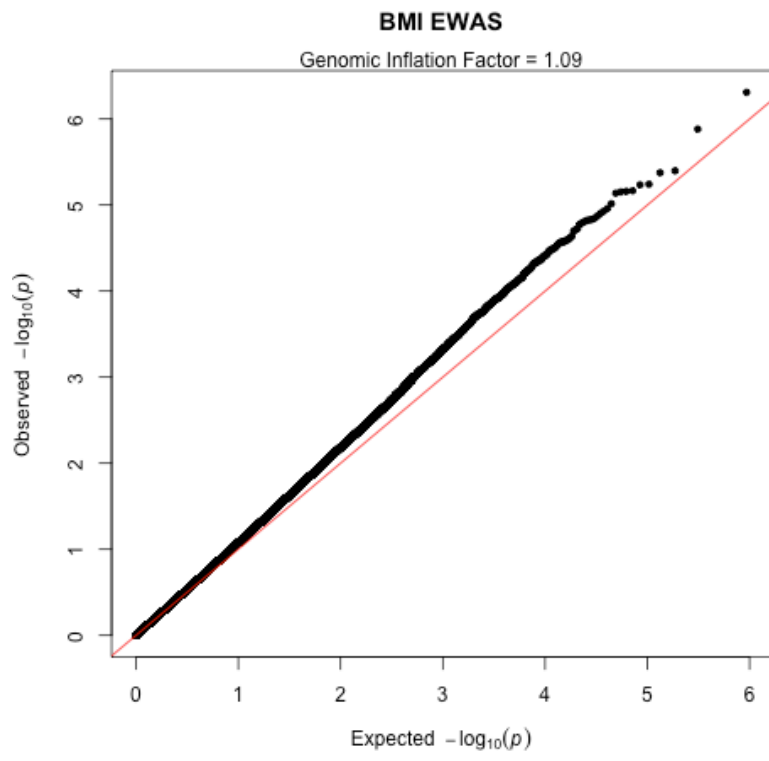


(A)

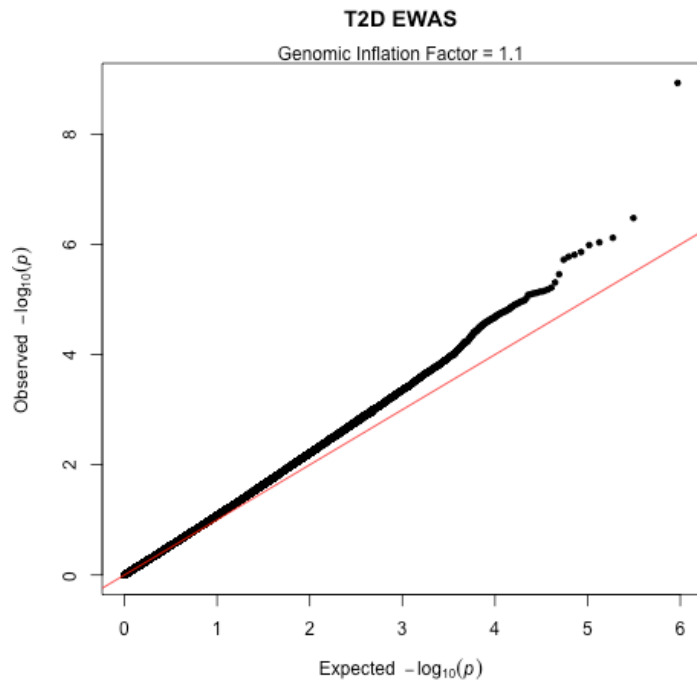


(B)

Supplementary Figure 2. 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 $p\text{-value}=1.07\times 10^{-7}$.



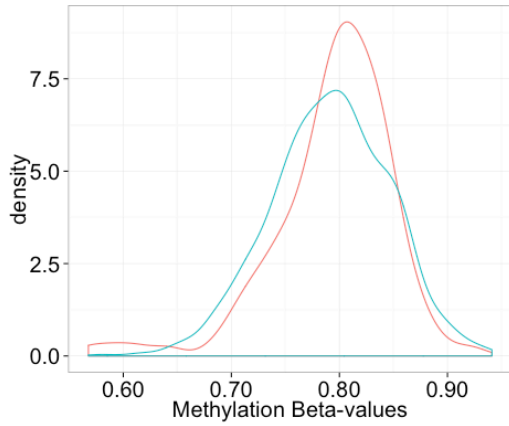
(A)



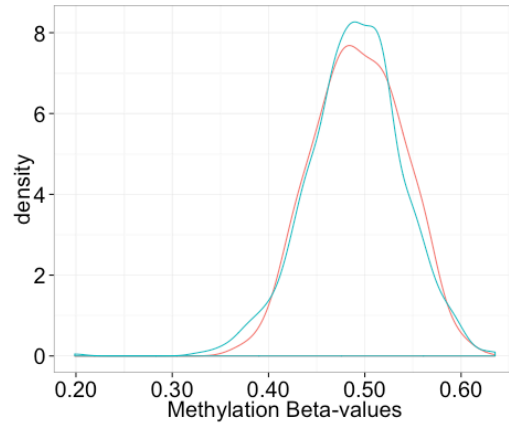
(B)

Supplementary Figure 3. 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.

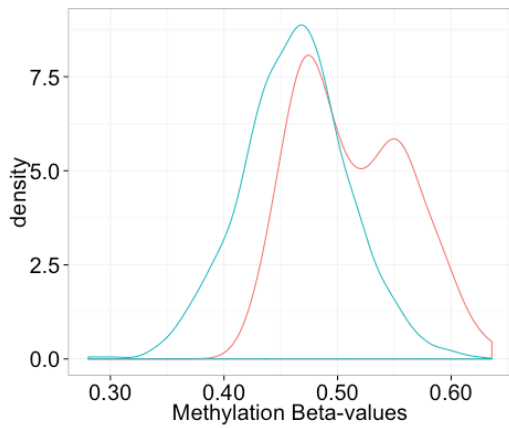
cg19693031
(*TXNIP*)



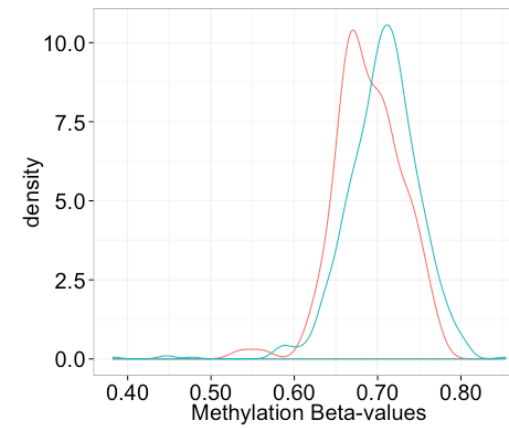
cg18181703
(*SOCS3*)



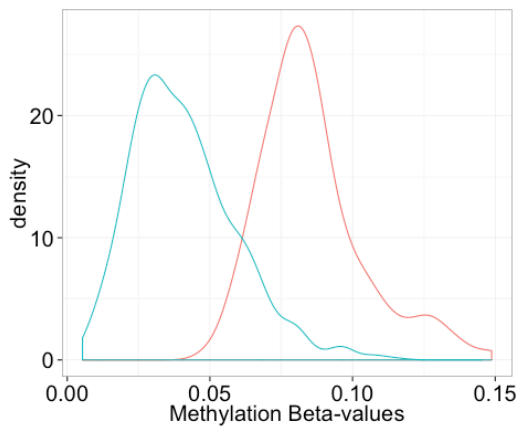
cg11024682
(*SREBF1*)



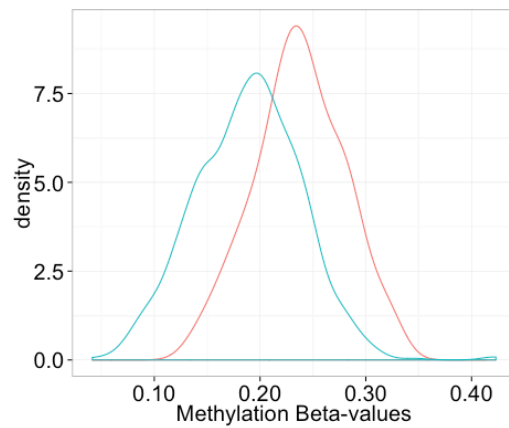
cg07573872
(*SBNO2*)

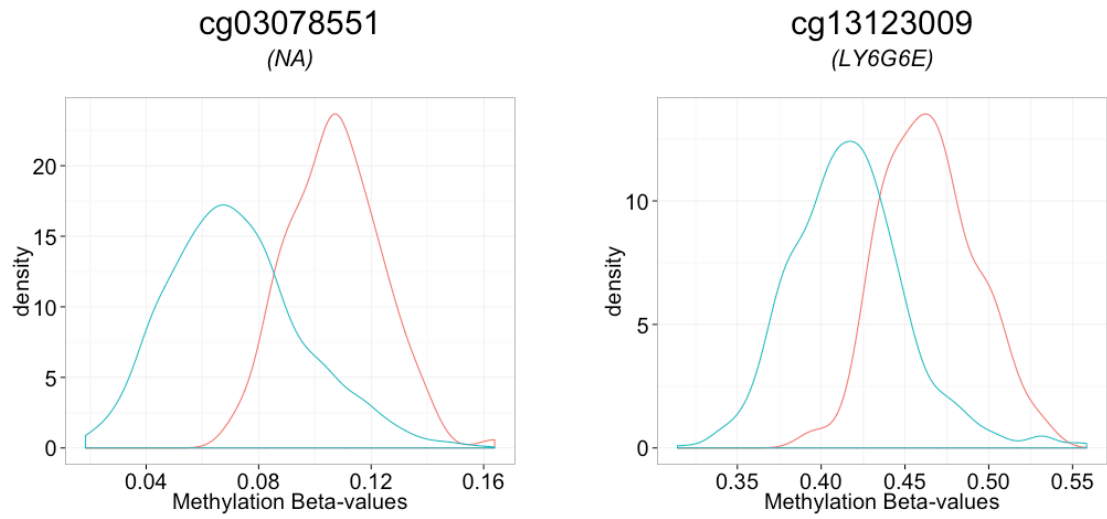


cg00574958
(*CPT1A*)



cg07136133
(*PRR5L*)





Supplementary Figure 4. Distribution of methylation Beta-value for the Qatari family study (red) and the TwinsUK cohort (blue).

SUPPLEMENTARY FILE

Supplementary File 1. 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.

Supplementary File 2. 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.