

1           **Regional fat depot masses are influenced by protein-coding gene variants**

2  
3

4   Matt J. Neville<sup>1,2¶</sup>, Laura BL. Wittemans<sup>3¶</sup>, Katherine E. Pinnick<sup>1</sup>, Marijana Todorčević<sup>1</sup>, Risto  
5   Kaksonen<sup>4</sup>, Kirsi H. Pietiläinen<sup>4,5</sup>, Jian'an Luan<sup>3</sup>, Robert A. Scott<sup>3</sup>, Nicholas J. Wareham<sup>3</sup>,  
6   Claudia Langenberg<sup>3</sup>, Fredrik Karpe<sup>1,2\*</sup>

7

8   <sup>1</sup>Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of  
9   Medicine, University of Oxford, Oxford

10   <sup>2</sup>Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford

11   <sup>3</sup>MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of  
12   Metabolic Science, Cambridge

13   <sup>4</sup>Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of  
14   Helsinki, Helsinki, Finland

15   <sup>5</sup>Abdominal Center, Endocrinology, Helsinki University Hospital, Helsinki, Finland

16

17   \*Corresponding Author:

18   [Fredrik.Karpe@ocdem.ox.ac.uk](mailto:Fredrik.Karpe@ocdem.ox.ac.uk)

19

20   ¶ These authors contributed equally to this work

21

## 22 **Abstract**

23 With the identification of a large number of genetic loci associated with human fat  
24 distribution and its importance for metabolic health, the question arises as to what  
25 the genetic drivers for discrete fat depot expansion might be. To date most studies  
26 have focussed on conventional anthropometric measures such as waist-to-hip ratio  
27 (WHR) adjusted for body mass index. We searched for genetic loci determining  
28 discrete fat depots mass size using an exome-wide approach in 3 large cohorts.  
29 Here we report an exome-wide analysis of non-synonymous genetic variants in  
30 17,212 participants in which regional fat masses were quantified using dual-energy  
31 X-ray absorptiometry. The missense variant *CCDC92*<sub>S70C</sub>, previously associated with  
32 WHR, is associated specifically with reduced visceral and increased leg fat masses.  
33 Allele-specific expression analysis shows that the deleterious minor allele carrying  
34 transcript also has a constitutively higher expression. In addition, we identify two  
35 variants associated with the transcriptionally distinct fat depot arm fat (*SPATA20*<sub>K422R</sub>  
36 and *UQCC1*<sub>R51Q</sub>). *SPATA20*<sub>K422R</sub>, a rare novel locus with a large effect size specific  
37 to arm, and *UQCC1*<sub>R51Q</sub>, a common variant exome-wide significant in arm but  
38 showing similar trends in other subcutaneous fat depots. In terms of the  
39 understanding of human fat distribution, these findings suggest distinct regulation of  
40 discrete fat depot expansion.

41

## 42 **Author summary**

43 Human fat storing tissues are heterogeneous and comprise functionally and  
44 structurally distinct regional fat depots, the relative size of which appear to have  
45 significant implications for health. Whilst it is known that inter-individual differences in  
46 fat distribution have genetic drivers, studies to date have focussed on crude

47 anthropometric approximations of region fat masses rather than precise measures.  
48 Here we describe an exome-wide analysis of a large collection of men and women  
49 who have undergone body scanning using dual-energy X-ray absorptiometry (DXA)  
50 to better define regional fat masses and identify new genetic drivers for human fat  
51 distribution. With this approach we identify three gene regions associated with  
52 distinct fat depots which can help to explain the variation in fat distribution between  
53 people and may lead to a better understanding of the depot specific fat  
54 tissue expansion.

## 56 **Introduction**

57

58 Beyond associations with chronic disease and overall obesity, as defined by body-  
59 mass index (BMI), it is becoming increasingly apparent that there is an even stronger  
60 relationship between body fat distribution and cardio-metabolic disease[1, 2]. For  
61 example, Yusuf *et al.*[1] showed that waist-to-hip ratio (WHR) is a stronger predictor of  
62 myocardial infarction than BMI. To date, the overwhelming majority genome and  
63 exome-wide association studies on fat distribution have focussed on waist and hip  
64 circumference and WHR[3, 4]. While these measures are easy and cheap to obtain  
65 on a large scale, they do not capture all variation in fat distribution. For example,  
66 WHR does not capture peripheral fat stored in the upper limbs and the distribution of  
67 overall central fat over the subcutaneous and visceral compartments, of which the  
68 latter have been suggested to have discordant effects on cardio-metabolic risk[5-8].  
69 Furthermore, circumference-based estimates of fat accumulation do not take into  
70 account differences in lean mass and bone structure and mass. Therefore, additional  
71 genetic association studies that call upon direct measures of regional fat mass would  
72 help unpick mechanisms underlying the expansion of distinct fat depots.  
73 Quantification of distinct fat depot masses requires imaging methods with post-image  
74 processing to derive delineation of tissues, such as magnetic resonance imaging or  
75 dual x-ray absorptiometry (DXA). We have therefore formed a large consortium with  
76 DXA-derived regional body fat measurements together with capability to pursue an  
77 exome chip discovery project of exonic gene coding variants relating to distinct fat  
78 depot size. We hypothesise that by identifying fat depot-specific genetic loci we may  
79 gain better insight into the site-specific role of adipose tissue to disease aetiology.

80

## 81 **Results and discussion**

82

83 We tested the associations of coding genetic variants covered on the Illumina  
84 Human Exome Bead chip with regional fat masses measured by DXA (GE Lunar  
85 iDXA). Our analyses included up to 17,212 participants of European ancestry from  
86 the Oxford Biobank[10], Fenland[11] and EPIC-Norfolk[12] cohorts (Table 1 and  
87 Table S1). We fitted within each cohort additive, recessive and dominant models for  
88 six DXA-derived adipose tissue regions, i.e., arm fat, leg fat, gynoid fat, total android  
89 fat, visceral abdominal fat and subcutaneous abdominal fat (Table S1), using  
90 RAREMETALWORKER[13]. The regional fat phenotypes were adjusted for the first 4  
91 principal components, age and total body fat percentage and the residuals were  
92 rank-based inverse normally transformed for men and women separately. Meta-  
93 analyses of the single variant association statistics were performed in  
94 RAREMETAL[14]. Only non-synonymous variants were considered and the cut-off  
95 for exome-wide statistical significance was  $p < 2E^{-7}$ . Three non-synonymous variants  
96 reached exome-wide significance (Fig1, Table 2 and S2 Table): rs11057401, a  
97 common missense variant in Coiled-Coil Domain Containing 92 (*CCDC92*<sub>S70C</sub>);  
98 rs62621401, a novel low-frequency missense variant in Spermatogenesis  
99 Associated 20 (*SPATA20*<sub>K422R</sub>) and rs4911494, a common missense variant in  
100 Ubiquinol-Cytochrome C Reductase Complex Assembly Factor 1 (*UQCC1*<sub>R51Q</sub>). An  
101 additional 30 non-synonymous variants reached suggestive significance across 38  
102 tests ( $p < 10^{-6}$ , S3 Table), including a large haplotype block on chromosome 17  
103 containing 8 missense variants across the *SPPL2C*, *MAPT*, *KANSL1* genes and  
104 *GDF5*<sub>S276A</sub> in LD with *UQCC1*<sub>R51Q</sub>.

105

106 Table 1. Study Cohorts

<b>Study name</b>	<b>Sample size (% men)</b>	<b>Genotyping array</b>	<b>QC Passed Variants. Total N<sup>a</sup></b>	<b>Polymorphic variants. N</b>
Fenland-ExomeChip	1145 (45.8%)	Illumina Exome BeadChip v1.0	240859	95739
Fenland-CoreExome	997 (45.2%)	Illumina Infinium Core Exome 24 v1 array	234179	85218
Fenland-Axiom	7363 (47.5%)	Affymetrix UK Biobank Axiom array	58240	57864
EPIC-Norfolk	3101 (45.0%)	Affymetrix UK Biobank Axiom array	56837	52020
Oxford Biobank Exome Chip	3281 (43.7%)	Illumina Exome BeadChip v1.0	245138	125912
Oxford Biobank Axiom	1325 (41.4%)	Affymetrix UK Biobank Axiom array	62732	56820

107 <sup>a</sup> Counts represent the number of variants in each dataset that overlap with the Illumina Exome  
108 Beadchip v1.0 content after standard QC metrics are applied.

109 Table 2. Primary Exome-Wide significant findings

rsID	Chr:Position (GRCH37)	Gene	Amino acid Change	Ref Allele	Alt Allele	DXA derived Fat Depot	N	Alternate Allele Frequency	Effect Size	Effect Size SD	Pvalue <sup>a</sup>	Haplotype region (GRCH37) <sup>b</sup>	Number of SNPs in LD with index SNP <sup>c</sup>
rs11057401	12:124427306	CCDC92	S70C	T	A	Total Android Fat	17184	0.321	-0.065	0.012	3.5E-08	chr12:124403769-124495203	99
						Visceral Fat	16967	0.321	-0.063	0.012	1.3E-07		
						Leg Fat	17184	0.321	0.075	0.012	4.9E-10		
rs62621401	17:48628160	SPATA20	K422R	A	G	Arm Fat	17204	0.016	-0.293	0.043	1.5E-11	chr17:48623825-48633043	7
rs4911494	20:33971914	UQCC1	R51Q	C	T	Arm Fat	17197	0.616	-0.063	0.012	1.3E-07	chr20:33887955-34194423	129

110 <sup>a</sup> Exome-wide significance was set to 2E<sup>-7</sup>

111 <sup>b</sup> The haplotype region is defined as the furthest 3' and 5' SNPs with a R<sup>2</sup> >0.9 with the index SNP

112 <sup>c</sup> SNP count is based on 1000 genome SNP data with SNPs in high LD (R<sup>2</sup>>0.9) with the index SNP

113 The rs11057401 *CCDC92*<sub>S70C</sub> variant (EAF=0.32) is predicted to cause a deleterious  
114 amino acid change as assessed by predictSNP[15]. The minor allele of rs11057401 shows  
115 significant opposing effects on android, specifically visceral fat mass, and lower body fat,  
116 in a sex-combined additive model (total abdominal fat mass=  $\beta$  -0.065, SD 0.0119,  $p=3.5E^{-8}$ ;  
117 visceral abdominal fat mass=  $\beta$  -0.063, SD 0.0119,  $p=1.3E^{-7}$ ; leg fat mass=  $\beta$  0.075, SD  
118 0.02,  $p=4.9E^{-10}$ ) (Table 2). These data extend WHR associations reported by Justice, *et*  
119 *al.*[3] and Lotta *et al.*[11] at this locus. Lotta *et al.*[11] describe the contribution of leg fat-  
120 mass; here we demonstrate an additional opposing effect specifically on abdominal  
121 visceral fat mass but not for abdominal subcutaneous fat mass, which would correspond to  
122 the already observed association with increased waist circumference but with the present  
123 analysis showing that the effect is confined to the intra-abdominal fat depot only. Found on  
124 chromosome 12q24, *CCDC92* is ubiquitously expressed with highest levels in adipose  
125 tissue, brain and testes. It is a nuclear protein interacting with the centriole-ciliary  
126 interface[11] and may also be involved in DNA repair[16]. The lead variant tags a large  
127 haplotype of at least 99 SNPs ( $r^2>0.9$ ) across a number of genes including the putative  
128 transcription factor Zinc Finger Protein 664 (*ZNF664*) and Dynein Axonemal Heavy chain  
129 10 (*DNAH10*) (Table 2). There is also strong evidence for multiple eQTL signals across  
130 this haplotype which includes three genes, i.e. *CCDC92*, *DNAH10* and *ZNF664*[17].  
131 Previous GWAS studies have also associated SNPs in this haplotype with a reduction in  
132 insulin resistance[11], improvements in metabolic syndrome[18], reduced WHRadjBMI[3,  
133 4], increased adiponectin levels[19] and with increased plasma HDL-cholesterol and  
134 reduced triglyceride concentrations[20-22]. Ablation of *CCDC92* and *DNAH10* in mouse  
135 OP9-K cells impairs adipogenesis and reduced lipid accumulation[11]. To further define  
136 the likely causative gene or genes in this complex region we undertook a number of gene  
137 expression studies in human regional adipocytes and whole adipose tissue. *CCDC92* and



138 *ZNF664* showed very similar expression profiles between abdominal subcutaneous  
139 (ASAT), gluteal subcutaneous (GSAT) and arm fat (Fig 2 and 3, Fig S1-3), whilst *DNAH10*  
140 expression could not be detected in cDNA from either of the diverse human adipose  
141 tissues or cultured primary human preadipocytes making it an unlikely target. Across a  
142 panel of 52 paired ASAT and GSAT cDNA samples (Fig S1), qPCR showed small  
143 differences in expression of *CCDC92* and *ZNF664* between ASAT and GSAT as well as  
144 between lean and obese individuals. In a cultured human primary preadipocyte  
145 differentiation time course experiment, both *CCDC92* and *ZNF664* showed a significant  
146 upregulation by day 4 of differentiation (Fig 2 A and B) but no difference in expression  
147 levels was observed between preadipocytes of ASAT and GSAT origin. Whilst this study  
148 focusses on exonic coding variants, recent studies have highlighted the need for caution  
149 when dissociating the analysis of such variants from surrounding eQTL signals[23]. To that  
150 end we also sought to investigate the reported eQTL signals at this locus, for both  
151 *CCDC92* and *ZNF664* (GTEx project[17] and[11] using allele-specific qPCR; a method that  
152 allows us to assess expressed allelic imbalance in heterozygous individuals and thus an  
153 eQTL. This showed a highly statistically significant increased expression of transcripts  
154 found on the minor allele haplotype for both genes (ASAT 5.8%, GSAT 4.9%, Fig 3 A and  
155 B). Of functional importance is that this allelic expression imbalance would result in the  
156 increased expression of the predicted deleterious serine-70-cysteine amino acid  
157 substitution in the *CCDC92* protein. Interestingly, zinc finger proteins such as *ZNF664*  
158 have been suggested to regulate the expression of near-by genes[18]. The observed co-  
159 regulatory expression pattern of *ZNF664* and *CCDC92* could then possibly be due to the  
160 eQTL acting on *ZNF664* which then upregulates the *CCDC92* deleterious variant. Further  
161 work needs to be done to investigate this.

162

163 The rs62621401 *SPATA20*<sub>K422R</sub> is a rare novel variant (EAF 0.016). This amino acid  
164 substitution is not predicted to be damaging and allele-specific qPCR on paired ASAT and  
165 GSAT cDNA samples of five Oxford Biobank participants heterozygous for rs62621401 did  
166 not reveal any suggestive eQTL (Fig 3C) either. This variant shows a large effect size and  
167 is the first locus to be associated with arm fat mass (Arm fat =  $\beta$  -0.29, SD 0.043, p-  
168 value=1.5E<sup>-11</sup>, Table 2 and Fig 1B) with an estimated per-allele effect size in the Oxford  
169 Biobank (n=4,606) of 125g less arm fat mass (-5.8%, CI -9.3% to -2.3%) in men and 67g (-  
170 2.6%, CI -5.3% to 0.4%) in women (approximate fat mass (grams) per allele after adjusting  
171 for covariates with % change in parentheses, Table S1). *SPATA20*, linked to  
172 spermatogenesis in mice[24], is highly expressed in human testes but also ubiquitously  
173 expressed, including in adipose tissue. *SPATA20* is a putative member of the thioredoxin  
174 family and other members of this family have been shown to be involved in preadipocyte  
175 proliferation[25] and pro-adipogenic *Wnt* signalling[26]. *SPATA20* expression was higher  
176 in men than women, although this was only significant for GSAT (p=0.01) (Fig S1).  
177 Expression of *SPATA20* during adipocyte differentiation showed an increase between  
178 days 0 to 7 of adipogenesis then a drop back to pre-differentiation levels between days 7  
179 and 14 (p=8.3E<sup>-8</sup>, Fig 2C) suggesting a role for this gene in adipocyte development.  
180 Surprisingly, despite the arm-specific association, expression of *SPATA20* was similar  
181 between arm fat, ASAT and GSAT (Fig S2 A). However, qPCR assessment of a number of  
182 developmental genes (Homeobox genes) in arm fat compared to ASAT and GSAT showed  
183 significant differences (Fig S2 B) indicating that arm fat is developmentally distinct from the  
184 other fat depots assessed. It is therefore possible that *SPATA20* is involved in an arm-  
185 specific developmental pathway.  
186

187 The rs4911494 *UQCC1*<sub>R51Q</sub> (EAF=0.62) variant was also associated with a loss of arm fat  
188 mass (Table 2 and Fig 1 C) but not predicted as damaging by predictSNP. Whilst exome-  
189 wide significance is only observed for arm fat, there is a trend towards less fat in all  
190 peripheral and subcutaneous fat depots in both genders (Fig 1 C and Table S1) for the  
191 effect allele as described here. Although it should be noted that the minor allele  
192 (MAF=0.38) would be associated with an increased fatmass. *UQCC1* is involved with  
193 mitochondrial respiratory chain complex III protein expression[27] and is structurally similar  
194 to the mouse *Bfzb* controlling mouse brown fat[28]. Previous associations at this locus  
195 include height[29], weight[30], WHRadjBMI[3] and osteoarthritis[31]. Another missense  
196 variant in the nearby canonical Wnt signalling gene *GDF5* (rs224331) is in LD with  
197 rs4911494 ( $r^2 > 0.9$ ) and reaches suggestive significance with arm fat in women ( $p = 8.97E^{-}$   
198  $07$ , SD 0.027,  $\beta$  0.13, Table S3). During adipogenesis, expression of *UQCC1* increases  
199 ( $p = 2.4E^{-22}$ , Fig 2D) but with no difference between ASAT and GSAT in cultured  
200 preadipocytes. Allele-specific qPCR showed that the minor allele (rs4911494) was  
201 associated with a small but statistically significant decrease in expression of *UQCC1* in  
202 ASAT (per allele percentage change in expression = -1.25%,  $p = 8.7E^{-4}$ ; Fig 3D). It is  
203 unclear whether this small change, if confirmed, would be biologically relevant. *GDF5*  
204 expression was not detected in adipose tissue cDNA samples. However, during  
205 adipogenesis *GDF5* showed transient expression at day 2, highlighting the possibility that  
206 *GDF5* is a regulator of early adipocyte differentiation.

207

## 208 **Conclusion**

209 This study represents the largest exome chip meta-analysis on DXA-derived discrete fat  
210 depot masses to date. The value of better-defined fat depot regions is illustrated at the  
211 *CCDC92* locus. The *CCDC92*<sub>S70C</sub> variant shows a clear signal for visceral fat mass, but

212 none for the adjacent subcutaneous abdominal fat mass and an opposing effect on lower-  
213 body fat mass most clearly observed in the whole leg. This is an important distinction from  
214 previous waist and WHR associations. Whether these opposing effects are because  
215 genetic variation at this locus has direct opposing adipose tissue mass effects in the  
216 depots or one depot is simply compensating to the mass change in the other is unclear  
217 and will require further investigation. A previous study investigating computerised  
218 tomography (CT) scan-derived visceral and subcutaneous fat mass also found  
219 associations at the *CCDC92* locus with *ZNF664* (rs1048497) and *DNAH10*  
220 (rs1316952)[32] but these SNPs are both in low LD with the index variant in this study ( $r^2$   
221 0.34 and 0.18, respectively) and may represent independent signals. No other loci  
222 associated with CT-derived visceral fat measures[32] were replicated to suggestive  
223 significance. These data and the depth of previous GWAS findings at the *CCDC92*-  
224 *ZNF664* locus highlight this as an important region in regulating adipose tissue distribution.  
225 In addition, we report two coding variants associated with arm fat: a novel low-frequency  
226 variant in *SPATA20* with a large effect size that seems to only effect arm fat and a  
227 common variant in *UQCC1* that additionally seem to have weak effects in other  
228 subcutaneous fat depots. Overall, when comparing these data to equivalent exome  
229 analysis using anthropometric measures[3] there were few replicated loci. Whilst this lack  
230 of replication is likely to be partly due to a lower power compared with significantly larger  
231 datasets using conventional anthropometric measures, we have identified a locus, not  
232 found for any traditional anthropometric traits, for arm fat and refined the tissue-specific  
233 association for another locus (*CCDC92*), highlighting the value of more defined regional fat  
234 measures.

235 Regional fat depots have distinct physiological regulation with an impact on the whole  
236 body metabolic homeostasis, with distinct transcriptomes demonstrating functional

237 differences and differences in origin[7]. This study provides genetic evidence for overall,  
238 distinct regulation of regional fat depot sizes.

239

240

## 241 **Materials and methods**

### 242 **Population cohorts**

243 **Oxford biobank.** The Oxford Biobank (OBB) cohort (<http://www.oxfordbiobank.org.uk>)  
244 consists of an age-stratified random sample of apparently healthy men and women (aged  
245 30 to 50 years) of European ancestry resident in Oxfordshire, UK, as described  
246 previously[10]. All participants gave written, informed consent to participate, and studies  
247 were approved by the Oxfordshire Research Ethics Committee (08/H0606/107+5). A total  
248 of 3,281 individuals from the Oxford Biobank had both measures of fat mass with GE  
249 Lunar iDXA[33] and Illumina Human Exome Beadchip genotypes after QC checks. An  
250 additional 1,325 individuals contained DXA data and Affymetrix UK Biobank Axiom array  
251 genotype data from which overlapping ExomeChip data was extracted for the purposes of  
252 this study (Table 1).

253

254 **Fenland.** The Fenland study is a population-based cohort study of participants without  
255 diabetes born between 1950 and 1975. Participants were recruited from general practice  
256 surgeries in Cambridge, Ely and Wisbech (UK) and underwent detailed metabolic  
257 phenotyping and genome-wide genotyping.  
258 A total of 1,145 individuals from the Fenland cohort had both measures of fatmass with GE  
259 Lunar iDXA[33] and Illumina Human Exome BeadChip genotypes after QC checks  
260 (Fenland-ExomeChip, Table 1), a further 997 had Illumina Infinium Core Exome data and  
261 7,363 had Affymetrix UK Biobank Axiom array genotype data from which overlapping  
262 ExomeChip data was extracted for the purpose of this study (Table 1).

263

264 **EPIC-Norfolk.** EPIC-Norfolk is a prospective cohort study of individuals aged between 40  
265 and 79 years and living in Norfolk county in the UK at the time of recruitment. EPIC-Norfolk  
266 is a constituent cohort of the European Prospective Investigation of Cancer (EPIC).  
267 A total of 3,101 individuals had Affymetrix UK Biobank Axiom array genotype data from  
268 which overlapping ExomeChip data was extracted for the purposes of this study (Table 1).

269

270

### 271 **DXA-derived depot-specific fat mass measures**

272 For all cohorts' depot-specific fat mass was quantified using GE Lunar iDXA (GE  
273 Healthcare, Bucks, UK). As previously described[33] these give high precision estimates of  
274 body composition. The standard setting of the Encore software (version 14.0; GE  
275 Healthcare, Bucks, UK) was used to automatically define regions of interest ensuring that  
276 boundaries were consistent between cohorts. The descriptives for the DXA measures  
277 used are presented in Supplementary Table 1. Visceral fat mass and android  
278 subcutaneous fat mass were not measured directly. Visceral fat mass was calculated  
279 using an algorithm within the Encore software as described elsewhere[33, 34] and the  
280 android subcutaneous fat mass was calculated by subtracting the visceral fatmass from  
281 total android fat mass. The DXA scanning was calibrated as per manufacturer's  
282 instructions.

283

### 284 **Exome-wide genotype analysis**

285

286 **Datasets.** Six data sets from three cohorts, Oxford Biobank[10], Fenland[35] and EPIC-  
287 Norfolk[12] (Table 1), equalling a total of 17,212 individuals of European ancestry were  
288 compiled for this analysis. The Illumina Exome BeadChip v1.0 genotype content was used  
289 as the base content. Where other genotype arrays were used (see Table 1) only the

290 content overlapping with the Illumina ExomeChip were selected. The breakdown of  
291 descriptives for each of the 6 datasets can be found in S1 Table. Standard quality control  
292 (QC) metrics were employed on each dataset separately and individuals and loci that  
293 failed QC removed before association analysis.

294

295 **Single-Variant analysis.** All DXA-derived phenotypes were log-transformed, adjusted for  
296 age, first 4 principal components (PCs) and percentage total fat mass (calculated as the  
297 percentage of total fat mass (grams) to total mass (grams)) and the residuals inverse  
298 normal transformed in the R statistical environment. Percentage total fat mass adjusted for  
299 age and PC1-4 was also included in the analysis to assess collider bias. Individual  
300 datasets were analysed separately in sex-combined and sex-specific analyses using  
301 RAREMETALWORKER[13] (<http://genome.sph.umich.edu/wiki/RAREMETALWORKER>).  
302 To account for cryptic relatedness, kinship matrices were first calculated and added into  
303 the analysis. Single-variant analysis was performed with, additive, recessive and dominant  
304 models.

305

306 **Meta-analysis.** Meta-analysis was carried out centrally using RAREMETAL[14]. Variants  
307 were excluded if they had a call rate <90%, Hardy-Weinberg equilibrium  $p$ -value <1E<sup>-7</sup>  
308 and markers on Y chromosome or mitochondrial genome. Exome-wide significance for the  
309 single-variant analysis was set, based on the full ExomeChip content, as  $p < 2E^{-7}$ . A  
310 suggestive significance was set to  $p < E^{-6}$ .

311 For this analysis we focussed on non-synonymous variants only, therefore all non-coding  
312 variants and synonymous variants were filtered out post meta-analysis. The exome-wide  
313 significant findings are presented in Fig 1 and S1 Table; the additional suggestive  
314 significant findings are presented in S3 Table.



315

## 316 **Additional informatics**

317 For the three exome-wide significant loci the amino-acid substitutions was assessed  
318 for functional significance using the predictSNP online consensus tool[15]  
319 (<https://loschmidt.chemi.muni.cz/predictsnp1/>). This allows for assessment across a  
320 number of different tools to generate a consensus assessment. For *CCDC92* the S70C  
321 missense variant was assessed; for *UQCC1* the R51Q was assessed and for *SPATA20*  
322 three different proteins as products of different splice variants were assessed (K422R,  
323 K406R and K362R).

324

325

326

## 327 **Adipose tissue gene expression panels**

328 Six genes found within the three index SNP LD boundaries (Table 2) were  
329 assessed for expression levels across a collection of human adipose tissue gene  
330 expression panels. Applied Biosystems Taqman assay-on-demand qPCR assays were  
331 selected for each gene that also avoid the index SNPs presented here, for *CCDC92* (ABI  
332 assay, hs01556139), *ZNF664* (ABI assay, hs00921074), *DNAH10* (ABI assay,  
333 hs1387352), *SPATA20* (ABI assay, hs00256188), *UQCC1* (ABI assay hs00921074) and  
334 *GDF5* (ABI assay, hs00167060).

335

336 For tissue panels, subcutaneous adipose tissue biopsies were collected by needle biopsy  
337 as previously described[36]. For cell-cultured human primary preadipocytes, of both  
338 abdominal subcutaneous fat (ASAT) and gluteofemoral fat (GSAT) origin, a differentiation  
339 time course (n=6) was performed as described in Todorčević, *et al.*[37]. All biopsies and

340 cells were homogenized in Tri-reagent (cat. no. T9424, Sigma-Aldrich, UK) and RNA was  
341 extracted with a standard Tri-reagent protocol. A total of 500ng RNA was used for cDNA  
342 synthesis following standard protocols and random hexamer primers using the cDNA  
343 Reverse Transcription Kit (Life Technologies, UK). Real-time PCR reactions were  
344 performed on a 1/40 cDNA dilution using Taqman Assays-on-Demand (Applied  
345 Biosystems) and Kapa Probe Fast Mastermix (Kapa Biosystems) in triplicate in a 6µl final  
346 volume and run on an Applied Biosystems 7900HT machine. Expression was assessed  
347 within each panel using a relative qPCR approach[38] and normalised using the previously  
348 assessed stably expressed endogenous control genes[36]. For the Lean/Obese Oxford  
349 Biobank panel (S1 Fig) the geometric mean of *PPIA*, *PGK1*, *PSMB6* and *IPO8* were used.  
350 *IPO8* was not used in a paired arm, ASAT and GSAT panel (S2 Fig) as it was not stably  
351 expressed between arm and the other depots. *PPIA* and *PGK1* were used as endogenous  
352 controls for primary cell culture experiments.

353 Neither *DNAH10* or *GDF5* could be detected above background in whole tissue cDNA  
354 panels. *GDF5* was however detected in a 14-day in vitro adipocyte differentiation time  
355 course.

356

357 Data for a panel of 52 paired ASAT and GSAT biopsy samples was used to assess  
358 expression between sexes, between ASAT and GSAT fat depots, and between lean and  
359 obese individuals. Descriptives for this panel are presented in S1 Fig. As both *SPATA20*  
360 rs62621401 and *UQCC1* rs4911494 were associated with arm fat mass their expression,  
361 along with *CCDC92* and *ZNF664* was assessed in a paired arm, ASAT and GSAT cDNA  
362 panel. As there is no published data on arm fat transcriptomics the additional *HOX* gene  
363 transcripts *HOXA5*, *HOXB8*, *HOXC8*, *HOXC9* and *HOXC11* were assessed as these are

364 known to be differentially expressed between ASAT, GSAT and visceral fat (These data  
365 are presented in S2 Fig).

366 The setup of a human primary adipocyte differentiation time course is described  
367 elsewhere[37]. Relative qPCR was run as above on the adipocyte panel for *CCDC92*,  
368 *ZNF664*, *SPATA20*, *UQCC1* and *GDF5*. Data is presented in Fig 2.

369

### 370 **Allele-specific qPCR**

371 Both the *CCDC92* and the *UQCC1* loci are associated with multiple eQTL signals.

372 Whilst we only consider non-synonymous variants in this analysis this does not discount  
373 that the coding locus is also under the influence of an eQTL. To assess the available data  
374 from resources such as the GTEx portal and to assess any eQTL effect between ASAT  
375 and GSAT fat depots we used the combined resources available within the Oxford  
376 Biobank.

377

378 Allele specific pPCR was run essentially as described in Fogarty *et al.*[39].

379 Taqman genotyping assays (Applied Biosystems) were selected to fall within the  
380 transcripts under investigation. For *CCDC92* the index SNP assay performed poorly so the  
381 Proxy SNP rs9863 (ABI assay, C\_206415\_30) was selected. To assess the nearby gene  
382 *ZNF664* a SNP in high LD with the *CCDC92* index SNP that fell within the *ZNF664*  
383 transcript, rs1054852 (ABI assay, C\_1169873\_10), was selected. For *SPATA20* the index  
384 SNP was used (rs62621401, ABI assay C\_25983779\_10) as was for *UQCC1* (rs4911494,  
385 ABI assay, C\_25472999\_10). As was previously stated neither *DNAH10*, nor *GDF5* could  
386 be detected in whole adipose tissue panels. Therefore, allele specific qPCR could not be  
387 assessed for these two genes.

388

389 From a panel of 200 paired ASAT and GSAT cDNA samples available from the Oxford  
390 Biobank, heterozygous individuals were selected. For *CCDC92* and *ZNF664* 28 paired  
391 ASAT and GSAT samples were selected, for *UQCC1* there were 34 and for *SPATA20*  
392 there were 5. Genomic DNA (gDNA) for these individuals were also retrieved and diluted  
393 to 1.5ng/μl. The gDNA is used as the control comparison to the cDNA samples as there is  
394 an equal quantity of both alleles in heterozygous gDNA samples. By comparing the ratio of  
395 the Ct values from each allele (the ratio of the genotype assay Vic or Fam fluorophore  
396 signals) between cDNA and gDNA any allelic expression differences observed in the  
397 cDNA samples can be resolved. This is particularly relevant as technical variation exists  
398 with each genotyping assay; particularly pronounced in *SPATA20* (Fig 3C).

399

400 Data are presented as the percentage of the minor allele Ct value compared to the major  
401 allele Ct. This is calculated by first generating a standard curve and regression statistic for  
402 each assay. A standard curve is generated from genomic DNA for individuals homozygous  
403 for the major allele (BB) and minor allele (bb). Genomic DNAs are diluted to 1.5ng/μl then  
404 BB and bb homozygotes are combined to ratios 80:20, 60:40, 50:50, 40:60, 80:20.

405 Following qPCR analysis using the dual-labelled TaqMan Genotyping assays the ratio of  
406 the B to b Ct values are calculated (Ct B minus Ct b) then plotted against the percentage  
407 of the minor allele in the dilution series. The linear regression statistic from this standard  
408 curve is then used to calculate the percentage minor allele expression of the unknown  
409 heterozygous individuals. The standard curves are presented in S3A-D Fig and allele-  
410 specific qPCR data for heterozygous individuals are presented in Fig 3.

411

412 For *CCDC92*, *ZNF664* and *UQCC1* there were sufficient cDNAs in the 200 panel, however  
413 for *SPATA20* there were only 5 individuals. Therefore, to improve the accuracy of the

414 *SPATA20* analysis, each sample was run in triplicate 4x across the assay plate and the  
415 average of all 4 sets of triplicates calculated. A single outlier in the *SPATA20* data was  
416 followed up in a second cDNA synthesis and persisted in both ASAT and GSAT samples.  
417 No phenotype differences were observed for this individual and no obvious genetic  
418 differences were found.

419

#### 420 **Statistical analysis**

421 Statistical significance was assessed for each experiment in SPSS v24. For  
422 estimates of per allele grams fat mass change, log phenotype data was analysed in a  
423 general linear model and adjusted for age, PC1-4 and total %fat mass then estimated  
424 marginal means were calculated (S1 Table).

425

## 426 **Acknowledgements**

427           We thank the volunteers from the Oxford Biobank ([www.oxfordbiobank.org.uk](http://www.oxfordbiobank.org.uk)) for  
428 their participation. The volunteer recruitment and recall process was funded by the  
429 National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC).  
430 The views expressed are those of the author(s) and not necessarily those of the NHS, the  
431 NIHR or the Department of Health recalling process of the volunteers.

432

433

434 **REFERENCES**

- 435 1. Yusuf S, Hawken S, Ounpuu S, Bautista L, Franzosi MG, Commerford P, et al. Obesity and  
436 the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control  
437 study. *Lancet*. 2005;366(9497):1640-9. doi: 10.1016/S0140-6736(05)67663-5.
- 438 2. Prentice AM, Jebb SA. Beyond body mass index. *Obes Rev*. 2001;2(3):141-7. doi:  
439 10.1046/j.1467-789x.2001.00031.x. PubMed PMID: 12120099.
- 440 3. Justice AE, Karaderi T, Highland H, Mahajan A, Gan W, Karpe F, et al. Protein-coding  
441 variants implicate novel genes related to lipid homeostasis contributing to body fat  
442 distribution. *Nature Genetics*. 2018; doi: 10.17863/CAM.20205.
- 443 4. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New  
444 genetic loci link adipose and insulin biology to body fat distribution. *Nature*.  
445 2015;518(7538):187-96. doi: 10.1038/nature14132.
- 446 5. Cantile M, Procino A, D'Armiento M, Cindolo L, Cillo C. HOX gene network is involved in  
447 the transcriptional regulation of in vivo human adipogenesis. *J Cell Physiol*.  
448 2003;194(2):225-36. doi: 10.1002/jcp.10210.
- 449 6. Min JL, Nicholson G, Halgrimsdottir I, Almstrup K, Petri A, Barrett A, et al. Coexpression  
450 network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for  
451 metabolic syndrome and related phenotypes. *PLoS genetics*. 2012;8(2). doi:  
452 10.1371/journal.pgen.1002505.
- 453 7. Karpe F, Pinnick KE. Biology of upper-body and lower-body adipose tissue--link to whole-  
454 body phenotypes. *Nat Rev Endocrinol*. 2015;11(2):90-100. Epub 2014/11/04. doi:  
455 10.1038/nrendo.2014.185. PubMed PMID: 25365922.
- 456 8. Hilton C, Karpe F, Pinnick KE. Role of developmental transcription factors in white, brown  
457 and beige adipose tissues. *Biochim Biophys Acta*. 2015;1851(5):686-96. Epub 2015/02/08.  
458 doi: 10.1016/j.bbali.2015.02.003. PubMed PMID: 25668679.
- 459 9. Wagner R, Machicao F, Fritsche A, Stefan N, Häring H-U, Staiger H. The genetic influence  
460 on body fat distribution. *Drug Discovery Today: Disease Mechanisms*. 2013;10(1-2):e5-  
461 e13. doi: <http://dx.doi.org/10.1016/j.ddmec.2013.05.003>.
- 462 10. Karpe F, Vasani SK, Humphreys SM, Miller J, Cheeseman J, Louise Dennis A, et al. Cohort  
463 Profile: The Oxford Biobank. *Int J Epidemiol*. 2017. doi: 10.1093/ije/dyx132.
- 464 11. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic  
465 analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human  
466 insulin resistance. *Nat Genet*. 2017;49(1):17-26. doi: 10.1038/ng.3714.
- 467 12. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, et al. EPIC-Norfolk: study  
468 design and characteristics of the cohort. *European Prospective Investigation of Cancer*. *Br J*  
469 *Cancer*. 1999;80 Suppl 1:95-103. doi: PMID:10466767. PubMed PMID: 10466767.
- 470 13. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-  
471 analysis for rare variants. *Bioinformatics*. 2014;30(19):2828-9. Epub 2014/06/03. doi:  
472 10.1093/bioinformatics/btu367. PubMed PMID: 24894501; PubMed Central PMCID:  
473 PMC4173011.
- 474 14. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of  
475 gene-level tests for rare variant association. *Nat Genet*. 2014;46(2):200-4. Epub 2013/12/15.  
476 doi: 10.1038/ng.2852. PubMed PMID: 24336170; PubMed Central PMCID:  
477 PMC3939031.
- 478 15. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust  
479 and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput*  
480 *Biol*. 2014;10(1):e1003440. doi: 10.1371/journal.pcbi.1003440.
- 481 16. Chaki M, Airik R, Ghosh AK, Giles RH, Chen R, Slaats GG, et al. Exome capture reveals  
482 ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response  
483 signaling. *Cell*. 2012;150(3):533-48. doi: 10.1016/j.cell.2012.06.028.

- 484 17. Ward MC, Gilad Y. Human genomics: Cracking the regulatory code. *Nature*.  
485 2017;550(7675):190-1. doi: 10.1038/550190a. PubMed PMID: 29022577.
- 486 18. Kraja AT, Chasman DI, North KE, Reiner AP, Yanek LR, Kilpeläinen TO, et al. Pleiotropic  
487 genes for metabolic syndrome and inflammation. *Mol Genet Metab*. 2014;112(4):317-38.  
488 doi: 10.1016/j.ymgme.2014.04.007.
- 489 19. Wu Y, Gao H, Li H, Tabara Y, Nakatochi M, Chiu Y-F, et al. A meta-analysis of genome-  
490 wide association studies for adiponectin levels in East Asians identifies a novel locus near  
491 WDR11-FGFR2. *Hum Mol Genet*. 2014;23(4):1108-19. doi: 10.1093/hmg/ddt488.
- 492 20. Chasman DI, Paré G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci  
493 associated with plasma lipoprotein size, concentration, and cholesterol content in genome-  
494 wide analysis. *PLoS genetics*. 2009;5(11):e1000730. doi: 10.1371/journal.pgen.1000730.
- 495 21. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide  
496 association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49(12):1758-  
497 66. doi: 10.1038/ng.3977.
- 498 22. Singaraja RR, Tietjen I, Hovingh GK, Franchini PL, Radomski C, Wong K, et al.  
499 Identification of four novel genes contributing to familial elevated plasma HDL cholesterol  
500 in humans. *J Lipid Res*. 2014;55(8):1693-701. doi: 10.1194/jlr.M048710.
- 501 23. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the  
502 accuracy of validated target identification through coding variant fine-mapping in type 2  
503 diabetes. *Nat Genet*. 2018;50(4):559-71. Epub 2018/04/09. doi: 10.1038/s41588-018-0084-  
504 1. PubMed PMID: 29632382; PubMed Central PMCID: PMC5898373.
- 505 24. Shi H-J, Wu AZ, Santos M, Feng Z-M, Huang L, Chen Y-M, et al. Cloning and  
506 characterization of rat spermatid protein SSP411: a thioredoxin-like protein. *J Androl*.  
507 2004;25(4):479-93. doi: 10.1002/j.1939-4640.2004.tb02819.x.
- 508 25. Song JS, Cho HH, Lee B-J, Bae YC, Jung JS. Role of thioredoxin 1 and thioredoxin 2 on  
509 proliferation of human adipose tissue-derived mesenchymal stem cells. *Stem cells and  
510 development*. 2011;20(9):1529-37. doi: 10.1089/scd.2010.0364.
- 511 26. Bahn YJ, Lee K-P, Lee S-M, Choi JY, Seo Y-S, Kwon K-S. Nucleoredoxin promotes  
512 adipogenic differentiation through regulation of Wnt/ $\beta$ -catenin signaling. *J Lipid Res*.  
513 2015;56(2):294-303. doi: 10.1194/jlr.M054056.
- 514 27. Tucker EJ, Wanschers BFJ, Szklarczyk R, Mountford HS, Wijeyeratne XW, van den Brand  
515 MAM, et al. Mutations in the UQCC1-Interacting Protein, UQCC2, Cause Human Complex  
516 III Deficiency Associated with Perturbed Cytochrome b Protein Expression. *PLoS genetics*.  
517 2013;9(12):e1004034. doi: 10.1371/journal.pgen.1004034.
- 518 28. Vetter K, Wurst W. Expression of a novel mouse gene 'mbFZb' in distinct regions of the  
519 developing nervous system and the adult brain. *Mechanisms of Development*.  
520 2001;100(1):123-6. doi: 10.1016/S0925-4773(00)00511-6.
- 521 29. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds  
522 of variants clustered in genomic loci and biological pathways affect human height. *Nature*.  
523 2010;467(7317):832-8. doi: 10.1038/nature09410.
- 524 30. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al. Sex-stratified  
525 Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism  
526 in Genetic Loci for Anthropometric Traits. *PLoS genetics*. 2013;9(6). doi:  
527 10.1371/journal.pgen.1003500.
- 528 31. Lettre G. The osteoarthritis and height GDF5 locus yields its secrets. *Nat Genet*.  
529 2017;49(8):1165-6. doi: 10.1038/ng.3924.
- 530 32. Fox CS, Liu Y, White CC, Feitosa M, Smith AV, Heard-Costa N, et al. Genome-wide  
531 association for abdominal subcutaneous and visceral adipose reveals a novel locus for  
532 visceral fat in women. *PLoS genetics*. 2012;8(5):e1002695. doi:  
533 10.1371/journal.pgen.1002695.



- 534 33. Vasan SK, Osmond C, Canoy D, Christodoulides C, Neville MJ, Di Gravio C, et al.  
535 Comparison of regional fat measurements by dual-energy X-ray absorptiometry and  
536 conventional anthropometry and their association with markers of diabetes and  
537 cardio:vascular disease risk. *Int J Obes (Lond)*. 2017;74:315. doi: 10.1038/ijo.2017.289.
- 538 34. Kaul S, Rothney MP, Peters DM, Wacker WK, Davis CE, Shapiro MD, et al. Dual-energy  
539 X-ray absorptiometry for quantification of visceral fat. *Obesity (Silver Spring)*.  
540 2012;20(6):1313-8. Epub 2012/01/26. doi: 10.1038/oby.2011.393. PubMed PMID:  
541 22282048; PubMed Central PMCID: PMC3361068.
- 542 35. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic  
543 analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human  
544 insulin resistance. *Nat Genet*. 2017;49(1):17-26. Epub 2016/11/14. doi: 10.1038/ng.3714.  
545 PubMed PMID: 27841877; PubMed Central PMCID: PMC5774584.
- 546 36. Neville MJ, Collins JM, Gloyn AL, McCarthy MI, Karpe F. Comprehensive Human  
547 Adipose Tissue mRNA and MicroRNA Endogenous Control Selection for Quantitative  
548 Real-Time-PCR Normalization. *Obesity*. 2010;19(4):888-92. doi: 10.1038/oby.2010.257.
- 549 37. Todorčević M, Hilton C, McNeil C, Christodoulides C, Hodson L, Karpe F, et al. A cellular  
550 model for the investigation of depot specific human adipocyte biology. *Adipocyte*.  
551 2017;6(1):40-55. doi: 10.1080/21623945.2016.1277052.
- 552 38. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR.  
553 *Nucleic Acids Res*. 2001;29(9):e45. doi: PMID: 11328886 PubMed Central PMCID:  
554 PMC55695.
- 555 39. Fogarty MP, Xiao R, Prokunina-Olsson L, Scott LJ, Mohlke KL. Allelic expression  
556 imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. *Hum Mol Genet*.  
557 2010;19(10):1921-9. doi: 10.1093/hmg/ddq067.  
558  
559

## 560 **Figure Captions**

561 **Fig 1. The effect size and direction of effect of Meta-analysis findings.** Effect size and  
562 direction of effect of the three exome-wide significant missense variants: **A.** rs11057401 in  
563 CCDC92 (EAF=0.32), **B.** rs62621401 in SPATA20 (EAF=0.016) and **C.** rs4911494 in  
564 UQCC1 (EAF=0.62). Data is presented for the 6 DXA measures under investigation and is  
565 presented as the beta value  $\pm$  SD. The meta-analysis significance level using an additive  
566 model for gender combined (All) as well as for gender stratified analysis, together with the  
567 N indicated to the right of the data in parentheses. DXA measures are Arm fat mass (Arm),  
568 Total android fat mass (Android), Subcutaneous android fat mass (Abdominal subcut),  
569 Visceral android fat mass (Abdominal visceral), Gluteal fat mass (Gynoid) and Leg fat  
570 mass (Leg). Exome-wide significant data ( $p < 2E^{-7}$ ) are in bold and underlined.

571

572 **Fig 2. Expression of candidate genes across a human primary adipocyte**  
573 **differentiation time course.** cDNA expression of *CCDC92* (**A**), *ZNF664* (**B**), *SPATA20*  
574 (**C**), *UQCC1* (**D**) and *GDF5* (**E**) was measured over a 14-day adipogenic differentiation  
575 time-course using primary preadipocytes from abdominal subcutaneous (ASAT) and  
576 gluteal subcutaneous (GSAT) fat depots[37]. Data are shown as DDCt values (normalized  
577 to *PPIA* and *PGK1*; n=6, mean  $\pm$  SEM). A multivariate general linear model was used to  
578 test for statistical significance between depots and time, and to assess depot x time  
579 interactions. *p*-values are presented in the shaded boxes, NS: non-significant.

580

581 **Fig 3. eQTL assessment of exome-wide significant loci by allele-specific qPCR**  
582 **expression.** Allelic expression was measured on 4 candidate transcripts in our three  
583 exome-wide significant regions using allele-specific qPCR. Data is presented as the % of  
584 the minor allele detected compared to the major allele, as described in the methods, with a

585 line indicating the mean and 95% CIs. To assess the rs11057401 eQTL haplotype the  
586 proxy SNP rs9863 was assessed for *CCDC92* (A) and the transcribed region proxy SNP  
587 rs1054852 for *ZNF664* (B). The index SNP rs62621401 was used to assess the *SPATA20*  
588 transcript (C) and the index SNP rs4911494 for *UQCC1*. Paired samples were compared  
589 between abdominal subcutaneous (ASAT) and gluteal subcutaneous (GSAT) and genomic  
590 DNA (gDNA). For each transcript ABI Taqman genotyping assays were selected that fall  
591 within the transcribed sequence. gDNA selected from the same individuals as the cDNAs  
592 acts as a paired control with presumed equal allele expression. Deviation from 50% for  
593 gDNA, particularly pronounced in *SPATA20* (C), represents inherent imbalance in assay  
594 technical performance and position of optimal Ct between Vic and Fam fluorescence. By  
595 using paired gDNAs to selected cDNAs allelic expression imbalance can be resolved by  
596 comparing cDNA to its paired gDNA. Significance was assessed with paired t-test in  
597 SPSSv24. Mean differences between comparisons and statistical significance is presented  
598 in shaded boxes. NS: Non-significant. The single outlier seen for *SPATA20* (C) was  
599 replicated in a second cDNA synthesis and both ASAT and GSAT. No phenotype  
600 differences were observed for this individual and no obvious genetic differences were  
601 observed.

602

## 603 **Supporting information**

604 **S1 Fig. mRNA expression of candidate genes across a lean/obese adipose tissue**  
605 **gene expression panel.** mRNA expression of the genes *CCDC92*, *ZNF664*, *UQCC1* and  
606 *SPATA20* across a panel of paired abdominal subcutaneous fat (ASAT) and gluteofemoral  
607 fat (GSAT) cDNA samples from the Oxford Biobank. The panel consisted of 25 male and  
608 29 female healthy individuals selected for either high or low BMI (Lean male, n=13, age  
609 44.5±0.9 yrs, BMI 22.7±0.3 kg/m<sup>2</sup>, fasting blood glucose 5.2±0.1 mmol/l; Obese males,

610 n=12, age 43.4±1.2 yrs, BMI 34.9±5.2 kg/m<sup>2</sup>, fasting blood glucose 5.6±0.1 mmol/l; Lean  
611 females, n=15, age 44±1.0 yrs, BMI 21.2±0.2 kg/m<sup>2</sup>, fasting blood glucose 4.8±0.1 mmol/l;  
612 Obese Females, n=14, age 44±1.0 yrs, BMI 33.6±0.6 kg/m<sup>2</sup>, fasting blood glucose 5.2±0.1  
613 mmol/l – data expressed as mean ±SEM). Data are shown as the mean ± SEM DDCT  
614 values (normalized to the geometric mean of the endogenous control genes *PPIA*, *PGK1*,  
615 *IPO8* and *PSMB6*) as described previously[36, 38]. A multivariate general linear model  
616 was used to test for statistical significance between gender, fat depots and obesity and to  
617 assess interactions. P-values are presented in the shaded box, NS: non-significant.  
618 There were small but significant differences in expression of *CCDC92*, *ZNF664* and  
619 *UQCC1* between fat depots in lean individuals but this difference was lost and expression  
620 was significantly reduced, in obese individuals. This is in keeping with a general quiescent  
621 state observed in transcripts associated with adipocyte metabolic activity in obesity.

622

623 **S2 Fig. mRNA expression of Candidate genes and Homeobox genes across a panel**  
624 **of 22 paired arm, abdominal subcutaneous adipose tissue (ASAT) and**  
625 **gluteofemoral adipose tissue (GSAT).** mRNA expression of the candidate genes **A:**  
626 *CCDC92*, *ZNF664*, *UQCC1* and *SPATA20* and a selection of developmental *HOX* genes  
627 **B:** *HOXA5*, *HOXB8*, *HOXC8*, *HOXC9* and *HOXC11* were determined by real-time qPCR.  
628 Data are shown as the mean ±SEM DDCT values (normalized to *PPIA*, *PGK1* and *PSMB6*;  
629 n=22). A univariate general linear model was used to test for statistical significance  
630 between depots. P-values for the *HOX* genes in **B** are presented in the shaded box.

631

632

633 **S3 Fig. Allele specific qPCR standard curves and *CCDC92-ZNF664* regression**  
634 **analysis.** The standard curve and regression statistic used to calculate the percentage

635 minor allele expression with allele-specific qPCR is shown above for *CCDC92* (A),  
636 *ZNF664* (B), *SPATA20* (C) and *UQCC1* (D). To quantify any allelic expression imbalance  
637 for the four genes a standard curve was generated from genomic DNA for individuals  
638 homozygous for the Major allele (BB) and Minor allele (bb). Genomic DNAs are diluted to  
639 1.5ng/ $\mu$ l then BB and bb homozygotes were combined to ratios 80:20, 60:40,  
640 50:50,40:60,80:20 to generate a standard curve. Following qPCR analysis using dual  
641 labelled TaqMan Genotyping assays the ratio of the B to b allele Ct values are calculated  
642 (Ct B minus Ct b) then plotted against the percentage of the minor allele in the dilution  
643 series. The linear regression statistic from this (A, B, C and D above) is then used to  
644 calculate the percentage minor allele expression of our unknown individuals. For *CCDC92*  
645 (A), *ZNF664* (B) and *UQCC1* (D) three different pairs of homozygote individuals were used  
646 to generate each standard curve and a Mean  $\pm$  SEM plotted for each dilution (A, B and D).  
647 For *SPATA20* only one genomic DNA homozygote minor allele individual was available so  
648 an error bar cannot be displayed.

649 As discussed in the main text there was an observed co-regulatory pattern of expression  
650 between *CCDC92* and *ZNF664* across different cDNA panels. To assess any correlation  
651 between these two genes within the samples, the allele-specific qPCR paired data points  
652 were plotted and regression statistic calculated (Graphs E and F). For both ASAT (E) and  
653 GSAT (F) there was a significant correlation, further supporting the co-regulatory pattern of  
654 expression.

655

656 **S1 Table. Population cohort descriptives.**

657 **S2 Table. Exome-wide significant loci.** Detailed data on the three exome-wide  
658 significant loci described. DXA parameters are included for all measures and meta-  
659 analysis statistics for the additive model. DXA measures are arm fatmass (Arm), Total

660 android fat mass (Android), Subcutaneous android fat mass (Subcut), Visceral android fat  
661 mass (Visceral), Gluteal fat mass (Gluteal) and Leg fat mass (Leg). Effect size data for  
662 suggestive exome-wide significance ( $p \leq 10^{-6}$ ) is shown in bold. Exome-wide significant  
663 data ( $p < 2E^{-7}$ ) are in bold and underlined.

664 <sup>a</sup> The impact of missense variants were assessed using the PREDICTsnp online  
665 consensus tool[15] (<https://loschmidt.chemi.muni.cz/predictsnp1/>).

666 <sup>b</sup> Approximate fat mass (grams) changes per allele is shown where test reaches  
667 suggestive significance and were calculated as marginal means after adjusting for age,  
668 PCs1-4 and %fatmass as covariates in a general linear model, implemented in SPSS v24

669

670 **S3 Table. Exome-wide loci showing suggestive level of statistical significance.**

671 Additional non-synonymous loci where statistical tests did not reach exome-wide  
672 significance but did reach a suggestive significance cut off of  $p \leq 10^{-6}$  are included above.

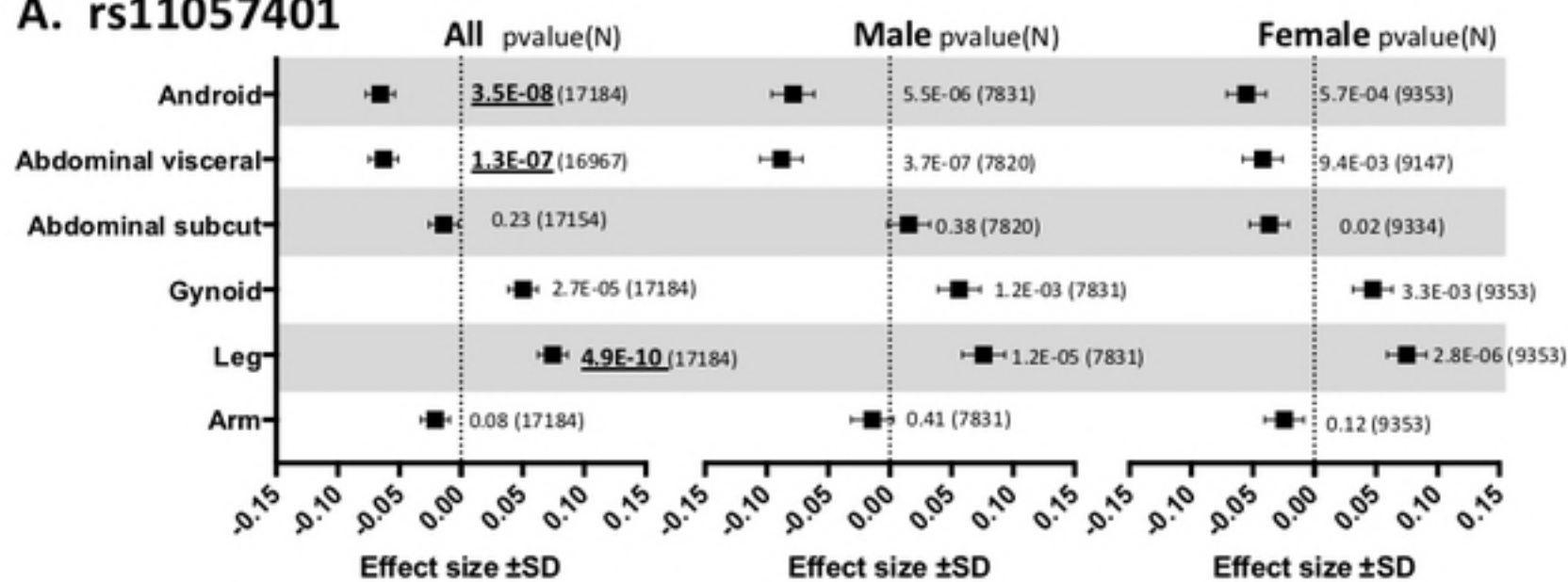
673 <sup>a</sup> Where it reaches suggestive significance the model is shown as Additive (add),  
674 Recessive(rec) or Dominant (Dom)

675 <sup>b</sup> The impact of missense variants were assessed using the predictSNP online consensus  
676 tool[15] (<https://loschmidt.chemi.muni.cz/predictsnp1/>).

677 <sup>c</sup> The cluster of 8 Missense SNPs found at the SPPL2C-MAPT-KANSL1 locus on  
678 chromosome 17 are part of a single haplotype that extends across ~400kb in this region  
679 containing >2300 SNPs ( $r^2 > 0.9$ ), rather than independent signals.

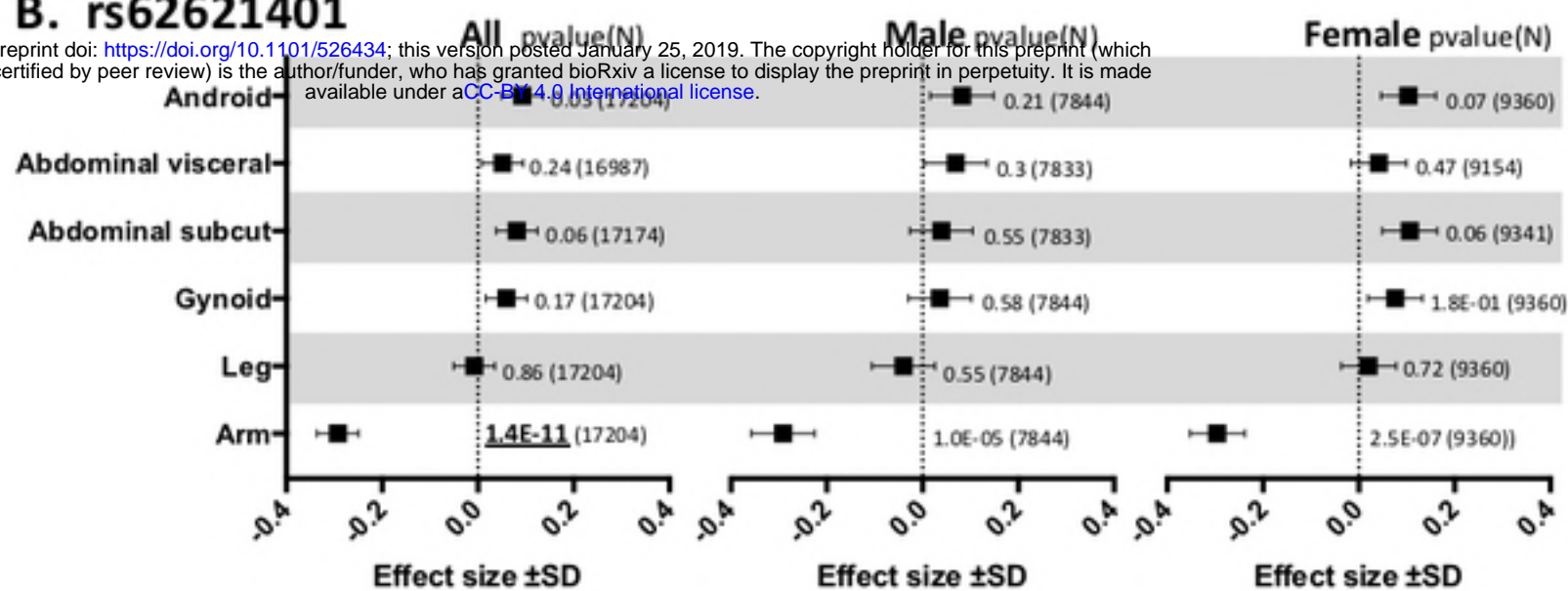
680

### A. rs11057401



### B. rs62621401

bioRxiv preprint doi: <https://doi.org/10.1101/526434>; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



### C. rs4911494

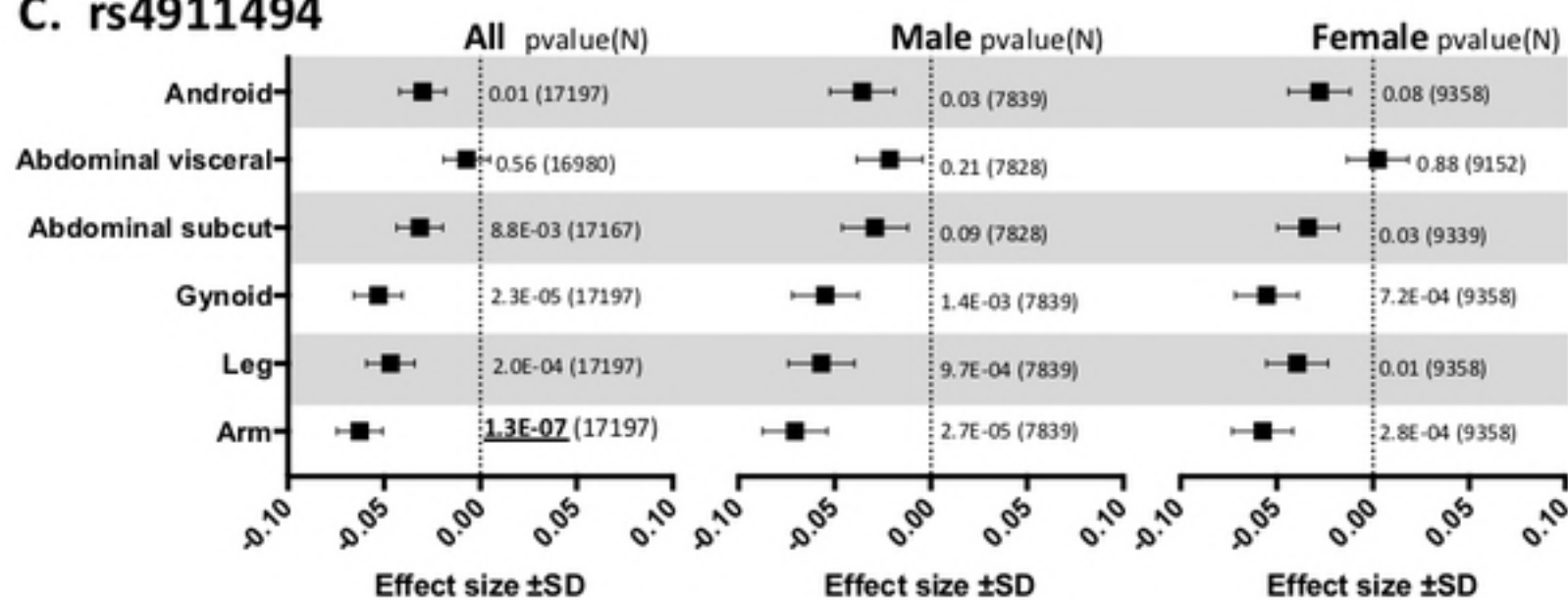


Figure 1

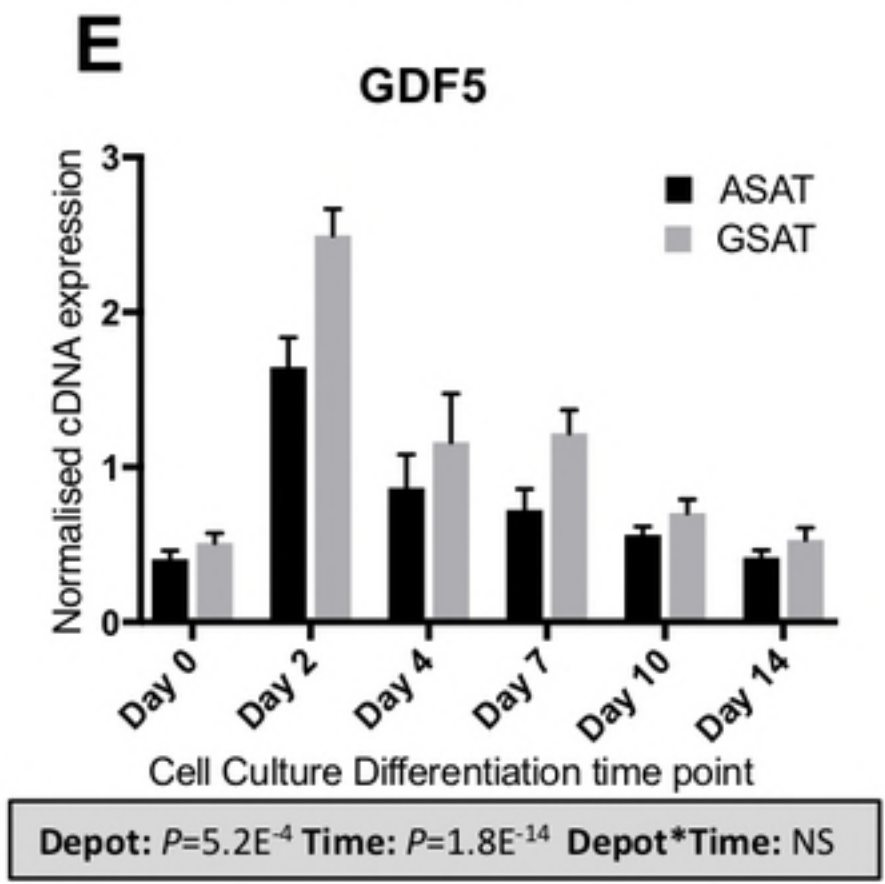
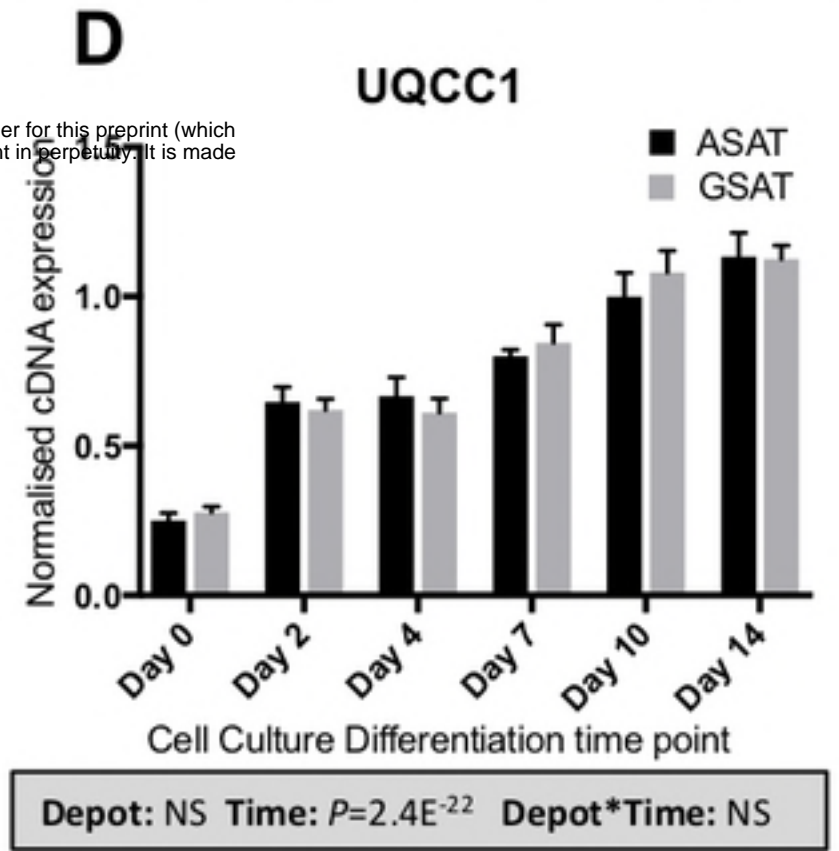
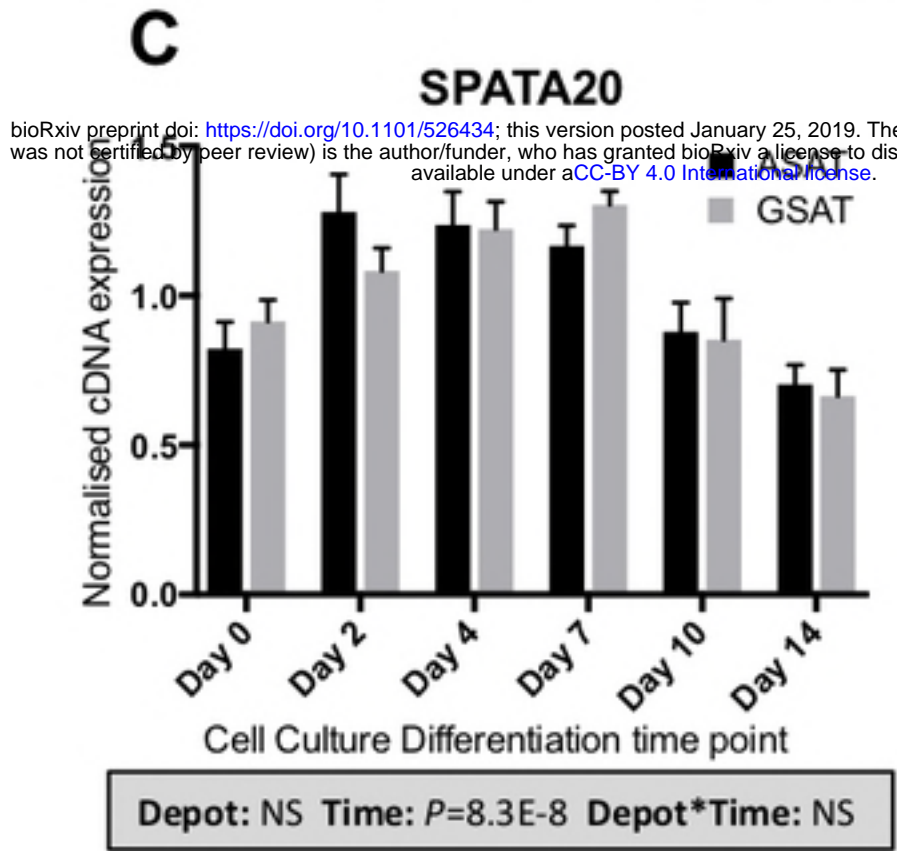
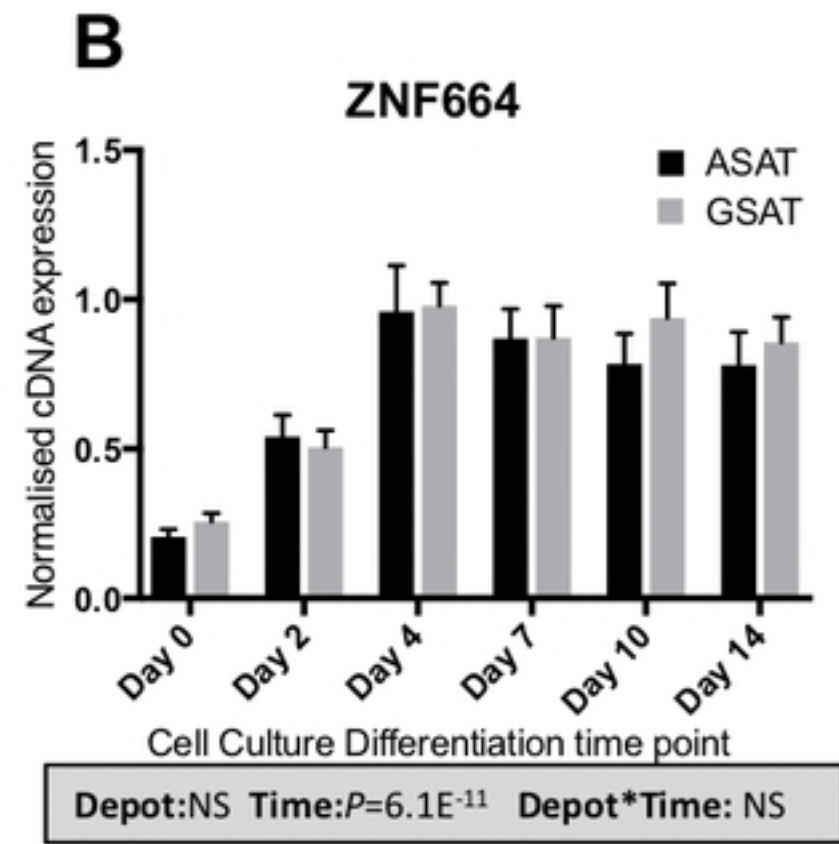
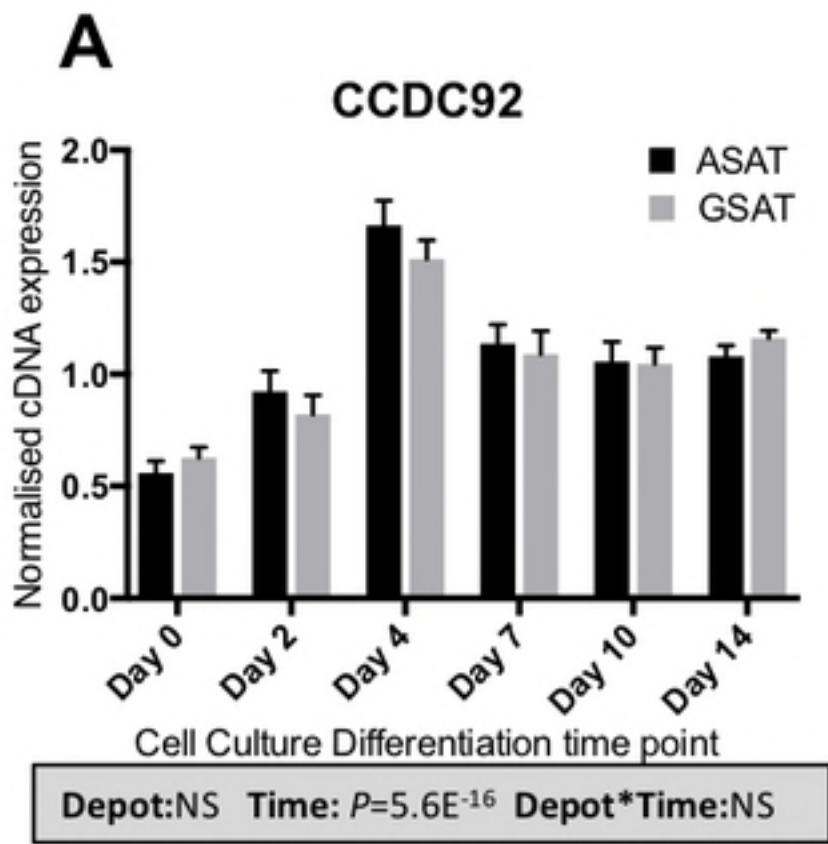
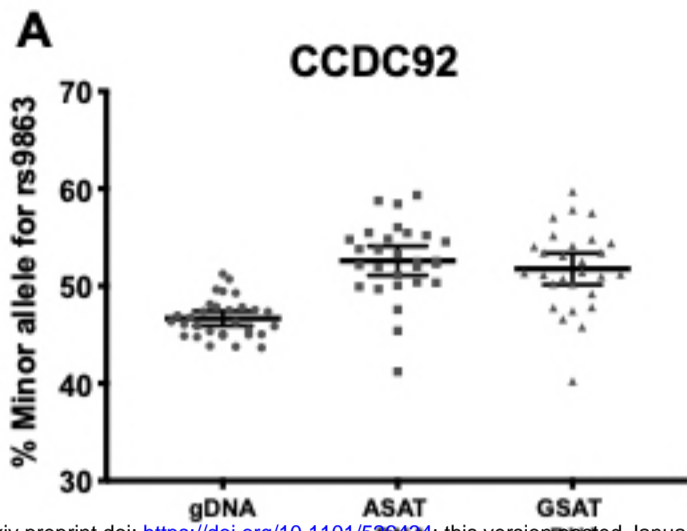


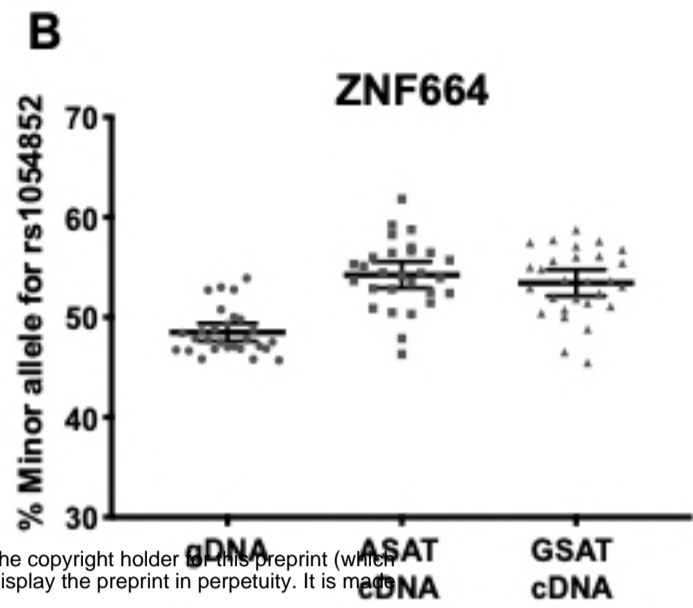
Figure2



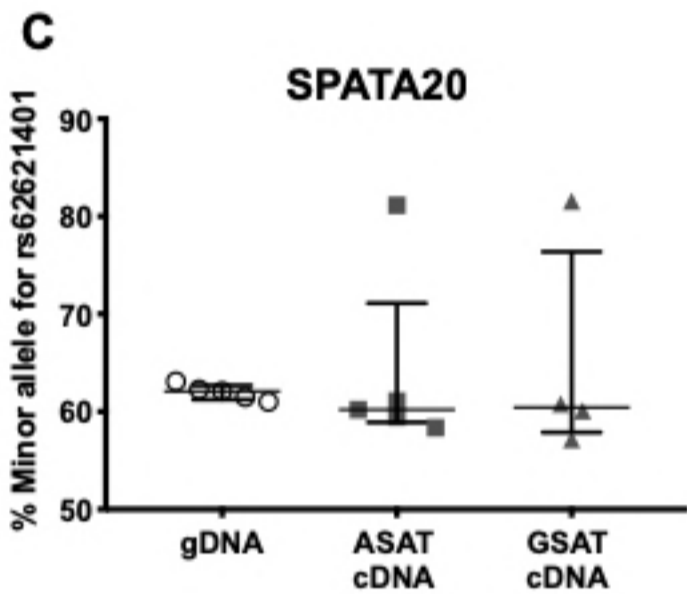


bioRxiv preprint doi: <https://doi.org/10.1101/529434>; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

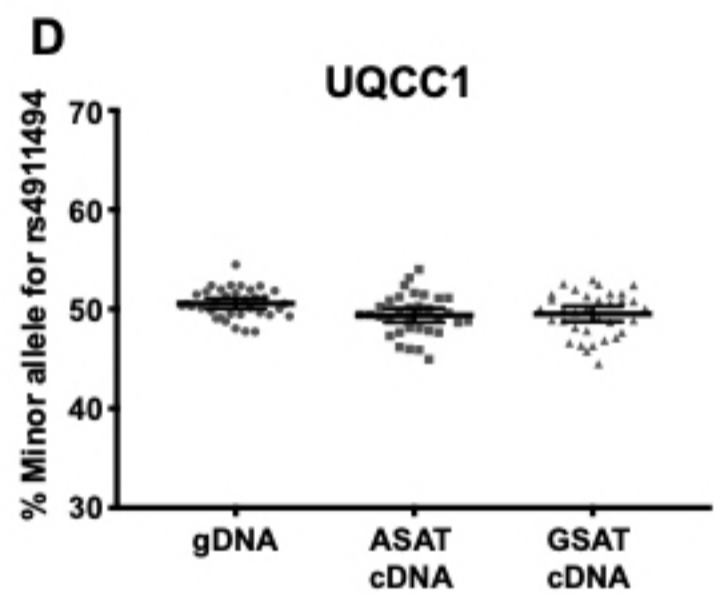
n=28	Comparison	Mean Difference	Pvalue
	gDNA-ASAT	-5.81±0.88	4.42E-07
	gDNA-GSAT	-4.97±0.88	5.52E-06
	ASAT-GSAT	0.84±0.38	0.038



n=28	Comparison	Mean Difference	Pvalue
	gDNA-ASAT	-5.74±0.71	1.20E-08
	gDNA-GSAT	-4.94±0.66	5.30E-08
	ASAT-GSAT	0.8±0.27	0.007



n=5 ASAT n=4 GSAT	Comparison	Mean Difference	Pvalue
	gDNA-ASAT	-2.06±4.05	NS
	gDNA-GSAT	-2.78±5.31	NS
	ASAT-GSAT	0.11±0.6	NS



n=34	Comparison	Mean Difference	Pvalue
	gDNA-ASAT	1.25±0.34	8.72E-04
	gDNA-GSAT	1.09±0.49	0.033
	ASAT-GSAT	-0.16±0.41	NS