nature genetics

Low copy number of the salivary amylase gene predisposes to obesity

Mario Falchi[1,39,40], Julia Sarah El-Sayed Moustafa[1,39], Petros Takousis[1], Francesco Pesce[1,2], Amélie Bonnefond[3–6], Johanna C Andersson-Assarsson[1,7,8], Peter H Sudmant[9], Rajkumar Dorajoo[1,10], Mashael Nedham Al-Shafai[1,11], Leonardo Bottolo[12], Erdal Ozdemir[1], Hon-Cheong So[13], Robert W Davies[14], Alexandre Patrice[6,15,16], Robert Dent[17], Massimo Mangino[18], Pirro G Hysi[18], Aurélie Dechaume[3,4,6], Marlène Huyvaert[3,4,6], Jane Skinner[19], Marie Pigeyre[4,6,15,16], Robert Caiazzo[4,6,15,16], Violeta Raverdy[6,15,16], Emmanuel Vaillant[3,4,6], Sarah Field[20], Beverley Balkau[21,22], Michel Marre[23,24], Sophie Visvikis-Siest[25], Jacques Weill[26], Odile Poulain-Godefroy[3,4,6], Peter Jacobson[7,8], Lars Sjostrom[7,8], Christopher J Hammond[18], Panos Deloukas[20,27,28], Pak Chung Sham[13], Ruth McPherson[29,30], Jeannette Lee[31], E Shyong Tai[31–33], Robert Sladek[34–36], Lena M S Carlsson[7,8], Andrew Walley[1,37], Evan E Eichler[9,38], Francois Pattou[4,6,15,16], Timothy D Spector[18,40] & Philippe Froguel[1,3–6,40]

Common multi-allelic copy number variants (CNVs) appear enriched for phenotypic associations compared to their biallelic counterparts[1–4]. Here we investigated the influence of gene dosage effects on adiposity through a CNV association study of gene expression levels in adipose tissue. We identified significant association of a multi-allelic CNV encompassing the salivary amylase gene (*AMY1*) with body mass index (BMI) and obesity, and we replicated this finding in 6,200 subjects. Increased *AMY1* copy number was positively associated with both amylase gene expression ($P = 2.31 \times 10^{-14}$) and serum enzyme levels ($P < 2.20 \times 10^{-16}$), whereas reduced *AMY1* copy number was associated with increased BMI (change in BMI per estimated copy = −0.15 (0.02) kg/m^2; $P = 6.93 \times 10^{-10}$) and obesity risk (odds ratio (OR) per estimated copy = 1.19, 95% confidence interval (CI) = 1.13–1.26; $P = 1.46 \times 10^{-10}$). The OR value of 1.19 per copy of *AMY1* translates into about an eightfold difference in risk of obesity between subjects in the top (copy number > 9) and bottom (copy number < 4) 10% of the copy number distribution. Our study provides a first genetic link between carbohydrate metabolism and BMI and demonstrates the power of integrated genomic approaches beyond genome-wide association studies.

We designed a gene-centric association study (GCAS) to identify common CNVs overlapping genes and inducing a dosage-dependent effect on gene expression, hypothesizing that these loci might be

[1]Department of Genomics of Common Disease, Imperial College London, London, UK. [2]Renal, Dialysis and Transplant Unit, Department of Emergency and Organ Transplantation (D.E.T.O.), University of Bari "Aldo Moro", Bari, Italy. [3]CNRS UMR 8199, Lille Pasteur Institute, Lille, France. [4]Lille 2 University, Lille, France. [5]Qatar Biomedical Research Institute (QBRI), Qatar Foundation, Doha, Qatar. [6]European Genomic Institute for Diabetes (EGID), Lille, France. [7]Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [8]Center for Cardiovascular and Metabolic Research, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [9]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [10]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore. [11]Research Division, Qatar Foundation, Doha, Qatar. [12]Department of Mathematics, Imperial College London, London, UK. [13]Department of Psychiatry, University of Hong Kong, Hong Kong, China. [14]Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [15]INSERM UMR 859, Lille, France. [16]Centre Hospitalier Régional Universitaire (CHRU) Lille, Lille, France. [17]Bariatric Program, Ottawa Hospital, Ottawa, Ontario, Canada. [18]Department of Twin Research & Genetic Epidemiology, King's College London, St. Thomas' Hospital Campus, London, UK. [19]Norwich Medical School, University of East Anglia, Norwich, UK. [20]Wellcome Trust Sanger Institute, Hinxton, UK. [21]Centre de Recherche en Epidémiologie et Santé des Populations, INSERM U1018, Epidemiology of Diabetes, Obesity and Chronic Kidney Disease over the Life Course, Villejuif, France. [22]Université Paris–Sud 11, UMRS 1018, Villejuif, France. [23]Department of Endocrinology, Diabetology and Nutrition, Bichat-Claude Bernard University Hospital, Assistance Publique–Hôpitaux de Paris (AP-HP), Paris, France. [24]INSERM U695, Paris 7 University, Paris, France. [25]INSERM UMR 1122 Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire, Université de Lorraine, Nancy, France. [26]Paediatric Endocrine Unit, Lille Teaching Hospital, Lille, France. [27]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [28]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [29]Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [30]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [31]Saw Swee Hock School of Public Health, National University of Singapore, National University Hospital System, Singapore. [32]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Hospital System, Singapore. [33]Duke–National University of Singapore Graduate Medical School, Singapore. [34]Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, Quebec, Canada. [35]Department of Medicine, Faculty of Medicine, McGill University, Montreal, Quebec, Canada. [36]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [37]Section of Genomic Medicine, National Heart and Lung Institute, Imperial College London, London, UK. [38]Howard Hughes Medical Institute, Seattle, Washington, USA. [39]These authors contributed equally to this work. [40]These authors jointly directed this work. Correspondence should be addressed to M.F. (m.falchi@imperial.ac.uk) or P.F. (p.froguel@imperial.ac.uk).

Received 15 August 2013; accepted 6 March 2014; published online 30 March 2014; doi:10.1038/ng.2939

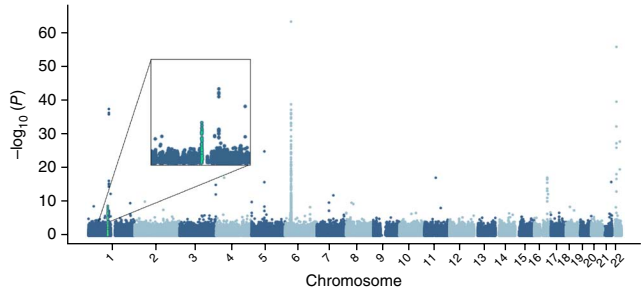

Figure 1 Manhattan plot of gene-centric CNV association study (GCAS) results with gene expression levels in subcutaneous adipose tissue using data from the Swedish family study. The chromosomal location for each probe is given on the *x* axis for each of the 22 autosomes, and $-\log_{10}(P)$ of the association between probe signal intensity and gene expression levels is shown on the *y* axis. The inset depicts the amylase region. The probes tested against amylase gene transcriptional levels are shown in green.

enriched for physiologically relevant CNVs. To achieve this aim, we conducted a family-based association analysis of signal intensity data from DNA arrays (log R ratio and B-allele frequency) with transcriptomic data from adipose tissue using famCNV[5] in 149 Swedish families ascertained through the identification of siblings who were discordant for obesity[6] (**Fig. 1**, **Table 1** and **Supplementary Fig. 1**). A total of 76 probes that hybridize within putative CNVs showed a dosage-dependent effect on gene expression at a false discovery rate (FDR) of 1% (**Supplementary Table 1**). Of these probes, only cnvi0020639, mapping to a CNV overlapping the amylase gene cluster (including the *AMY1* salivary amylase gene and the *AMY2* pancreatic amylase gene, expression probe set 208498_s_at; FDR = 6.88×10^{-3}), was also associated with adiposity (both BMI ($P = 3.86 \times 10^{-4}$) and fat mass ($P = 3.11 \times 10^{-4}$)) (**Supplementary Figs. 2–4**). Salivary amylase is encoded by the three distinct but closely related genes, *AMY1A*, *AMY1B* and *AMY1C*, which are referred to collectively here as *AMY1*. Analogously, except where otherwise noted, *AMY2* refers to the pancreatic amylase gene cluster encompassing the genes *AMY2A* and *AMY2B*. Reduced signal intensity at this probe, indicative of reduced copy number at this locus, was associated with increased adiposity (**Fig. 2**, **Table 2** and **Supplementary Figs. 2–4**).

This inverse association between copy number in the amylase region and BMI was first replicated using signal intensity data from DNA arrays in 972 subjects from TwinsUK (**Table 1**)[7]. The strongest association was observed at cnvi0022844 ($P = 1.13 \times 10^{-3}$; **Table 2**), which showed significant association with BMI after Bonferroni correction. When multiple probes were considered through principal-component analysis, the BMI association signal actually extended over a region between the hybridization sites for cnvi0022844 and cnvi0016754 ($P = 1.32 \times 10^{-3}$), which overlapped the cnvi0020639 probe site associated with adiposity in the Swedish discovery families. These results, although supportive of the association in the amylase region, did not permit us to distinguish which of the salivary or pancreatic amylase genes was driving the association with adiposity, necessitating the use of a non-array-based method of copy number measurement.

Table 1 Summary information on subjects included in this study

Sample	Total	Male	Female	Median age in years (1st–3rd quartiles)	Median BMI in kg/m^2 (1st–3rd quartiles)
Swedish	342	98	244	37 (33–43)	27.9 (22.6–36.5)
TwinsUK	1,479[a]	–	1,479	53 (45–60)	26.0 (22.8–28.4)
DESIR	2,137	942	1,195	52 (44–61)	24.6 (22.2–26.6)
AOB	563	160	403	35 (32–39)	–
Cases	205	39	166	36 (29–41)	46.2 (42.5–51.3)
Controls	358	121	237	35 (33–38)	21.6 (20.3–22.4)
SP2	658	237	421	46 (37–52)	–
Cases[b]	333	139	194	47 (40–54)	27.1 (25.9–28.9)
Controls	325	98	227	44 (34–51)	18.4 (17.5–19.1)
ABOS	468	122	346	43 (33–51)	46.2 (41.7–52.3)

[a]Consisting of 334 dizygotic and 193 monozygotic twin pairs and 425 singletons. [b]Including 136 obese (BMI ≥ 28 kg/m^2) and 197 overweight (23 kg/m^2 ≤ BMI < 28 kg/m^2) Singaporean Chinese subjects.

Consequently, we estimated copy number at *AMY1* and *AMY2* in 481 subjects from the Swedish families (**Table 1**) using quantitative PCR (qPCR). This approach generates a continuous intensity distribution from which integer copy numbers can be inferred by comparison to a reference sample of known copy number (**Supplementary Note**). Given the many technical challenges inherent in copy number measurement at multi-allelic loci[2,8–11], we treated these discretized measurements as relative estimates or surrogates correlated with the true underlying copy number state rather than absolute copy number genotypes.

Only three estimated copy number states (1–3) were detected for the pancreatic amylase gene (*AMY2*), and these were not associated with either BMI or fat mass (**Supplementary Table 2**). In contrast, copy number estimates at *AMY1* ranged from 2 to 14 and showed association with both BMI ($P = 8.08 \times 10^{-3}$) and fat mass ($P = 8.53 \times 10^{-3}$), confirming our previous DNA array–based analysis (**Supplementary Table 2**). We found greater correlation between signal intensity at cnvi0020639 and *AMY1* copy number ($r = 0.73$; $P < 2.20 \times 10^{-16}$) than with *AMY2* copy number ($r = 0.35$; $P = 1.25 \times 10^{-8}$), suggesting that the GCAS-identified association was mainly capturing copy number variation at *AMY1* rather than *AMY2* and justifying follow-up of the former. Furthermore, we validated the accuracy of the *AMY1* qPCR assay by using *AMY1* copy number estimates derived from whole-genome shotgun sequencing data for the 1000 Genomes Project[12] and observed a correlation of 0.94 ($P < 2.20 \times 10^{-16}$) between *AMY1* copy number estimates derived by qPCR and sequencing (**Supplementary Figs. 5–8** and **Supplementary Table 3**). To further validate the *AMY1* qPCR assay, we also compared copy number measured by qPCR in 96 samples from the French Data from the Epidemiological Study on the Insulin Resistance syndrome (DESIR) cohort[13] with *AMY1* copy number measured by digital PCR in the same samples, obtaining high correlation between the two methods ($r = 0.95$; $P < 2.20 \times 10^{-16}$; **Supplementary Fig. 9**). Analogously, high correlation ($r = 0.98$; $P < 2.20 \times 10^{-16}$; **Supplementary Fig. 9**) was also observed between copy numbers measured using the qPCR assay applied in this study and copy numbers obtained using a different qPCR assay on the same 96 DESIR samples.

To replicate the observed association in a larger sample, we next estimated *AMY1* copy number by qPCR in an additional sample of 1,479 female subjects from TwinsUK[14] and 2,137 male and female subjects from DESIR[13] (**Table 1**). The two population samples showed a similar copy number distribution (Wilcoxon test $P > 0.05$) with estimated median copy number of 6, ranging from 1 to 18 (**Supplementary Fig. 10** and **Supplementary Tables 4** and **5**). Meta-analysis of *AMY1* effects in TwinsUK and DESIR (total $n = 3{,}616$) showed significant association between reduced *AMY1* copy number and increased BMI (per-copy β (se) = −0.15 (0.02) kg/m^2; $P = 6.93 \times 10^{-10}$; **Fig. 3**, **Table 2** and **Supplementary Tables 6–9**). Results of associations assessed using both the qPCR intensity signal as a continuous measure as well as discretized using an unsupervised

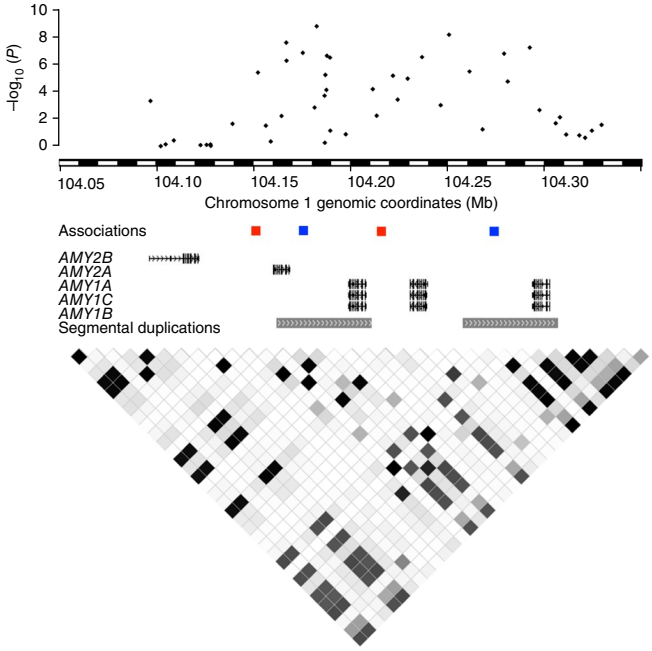

Figure 2 The amylase region in detail. Top, famCNV association results between signal intensity for probes mapping within 30 kb of the amylase cluster and amylase expression levels (probe set 208498_s_at) in adipose tissue in the Swedish family discovery study, with chromosomal coordinates given on the *x* axis and $-\log_{10}(P)$ on the *y* axis. Middle, locations of hybridization sites for probes showing association between signal intensity and BMI: cnvi0020639 (blue; Swedish family discovery study) and cnvi0022844 (red; TwinsUK). Because of the repetitive nature of this region, which contains six paralogs (including one pseudogenized copy) in the reference genome (**Supplementary Fig. 13**), the cnvi0020639 and cnvi0022844 probes were found to map to two locations each within the amylase gene cluster. Gene content in the amylase region, based on the human reference sequence (hg19; RefSeq), depicting *AMY2B*, *AMY2A* and the *AMY1A*, *AMY1B* and *AMY1C* genes, as well as two segmental duplications with high sequence similarity in the region. Bottom, LD between HapMap markers (release 23) calculated with Haploview[46] (darker shading corresponds to higher r^2 value).

clustering approach (*k* means) were concordant with those generated using integer copy numbers (**Supplementary Note**).

We then assessed the effect of *AMY1* copy number on obesity susceptibility by selecting obese cases (BMI ≥ 30 kg/m^2) and normal-weight controls (BMI < 25 kg/m^2) from TwinsUK and DESIR and by measuring *AMY1* copy number by qPCR in an additional 205 severely obese cases and 358 age-matched controls from the French Adult Obesity (AOB)[15] study (**Table 1** and **Supplementary Note**). In these European samples, subjects with lower estimated *AMY1* copy number showed significantly increased risk of obesity in each of the three samples (meta-analysis per estimated *AMY1* copy number: OR = 1.19, 95% CI = 1.13–1.26; $P = 1.46 \times 10^{-10}$; **Fig. 3** and **Table 2**). The *AMY1* copy number distribution in our sample ranged from 1 to 18 copies, with approximately 10% of subjects carrying fewer than 4 copies of *AMY1* and 10% of subjects carrying an *AMY1* copy number greater than 9 (**Table 2**). Given the multi-allelic nature of the *AMY1* CNV, this OR of 1.19 per copy of *AMY1* translates into about an eightfold difference in risk of obesity between subjects in the top (copy number > 9) and bottom (copy number < 4) deciles of the distribution of estimated *AMY1* copy numbers (OR = 7.67, 95% CI = 3.92–14.99; $P = 2.52 \times 10^{-9}$; **Supplementary Table 10**). Using a multifactorial liability threshold model[16], we estimated the proportion of total variance in obesity explained by estimated *AMY1* copy number to lie between 1.73 and 7.94% (95% CI; **Supplementary Table 11**). Therefore, on the basis of an estimated heritability of 40–70% for obesity[17,18], copy number variation at *AMY1* may account for 2.47–19.86% of the total genetic variation in obesity. Analogously, we estimated that between 0.66 and 4.40% of the proportion of genetic variance in BMI could be explained by inferred *AMY1* copy number in these European samples.

As all the samples included in our analyses were of European origin, we reasoned that replication in a sample of different ancestry under differing environmental influences on obesity would provide greater support for its physiological role. We therefore selected a Singaporean

Table 2 Association of relative copy number in the amylase region with obesity and measures of adiposity

DNA array–based CNV analysis[a]	*N*	Trait	Associated probe			*P*		
Swedish families	342	BMI	cnvi0020639			3.86×10^{-4}		
Swedish families	331	Fat mass	cnvi0020639			3.11×10^{-4}		
TwinsUK	972	BMI	cnvi0022844			1.13×10^{-3}		
Population samples[b]	*N*	BMI[c]		Age (years)[c]	*β* (se)[d]	*P*		
TwinsUK	1,479	26.0 (22.8–28.4)		53 (45–60)	−0.18 (0.05)	5.91×10^{-4}		
DESIR	2,137	24.6 (22.2–26.6)		52 (44–61)	−0.14 (0.03)	2.49×10^{-7}		
	N				*β* (se)[d]	*P*		P^e_{het}
Meta-analysis	3,616				−0.15 (0.02)	6.93×10^{-10}		0.54
	Cases		Controls					
Obesity case-control samples[b]	*N*	Age (years)[c]	*N*	Age (years)[c]	*β* (se)[d]	*P*	OR (95 %CI)[d]	
TwinsUK	251	53 (47–60)	711	51 (44–59)	−0.26 (0.09)	3.61×10^{-3}	1.30 (1.08–1.55)	
DESIR	137	55 (47–64)	1,267	51 (42–59)	−0.16 (0.04)	7.47×10^{-5}	1.18 (1.09–1.27)	
AOB	205	36 (29–41)	358	35 (33–38)	−0.17 (0.04)	4.44×10^{-5}	1.19 (1.10–1.29)	
SP2	136	47 (37–54)	325	44 (34–51)	−0.15 (0.05)	3.73×10^{-3}	1.17 (1.05–1.29)	
	N		*N*		*β* (se)[d]	*P*	OR (95%CI)[d]	P^e_{het}
Meta-analysis[f]	593		2,336		−0.18 (0.03)	1.46×10^{-10}	1.19 (1.13–1.26)	0.62

Copy number estimates used in the association analyses for both the population samples and the obesity case-control samples were derived by qPCR. The listed numbers of samples are those that passed quality control and were used in the association analyses.
[a]Signal intensity data from Illumina SNP genotyping arrays. [b]Obesity case-control analyses in TwinsUK and DESIR were conducted using a subset (subjects with BMI < 25 kg/m^2 and those with BMI ≥ 30 kg/m^2) of the subjects included in the quantitative trait analysis in the population samples category. [c]Median (first and third quartiles). [d]Estimates calculated using integer *AMY1* copy numbers inferred from the underlying continuous distribution. [e]Heterogeneity *P* value. [f]The obesity case-control meta-analysis was limited to European samples.

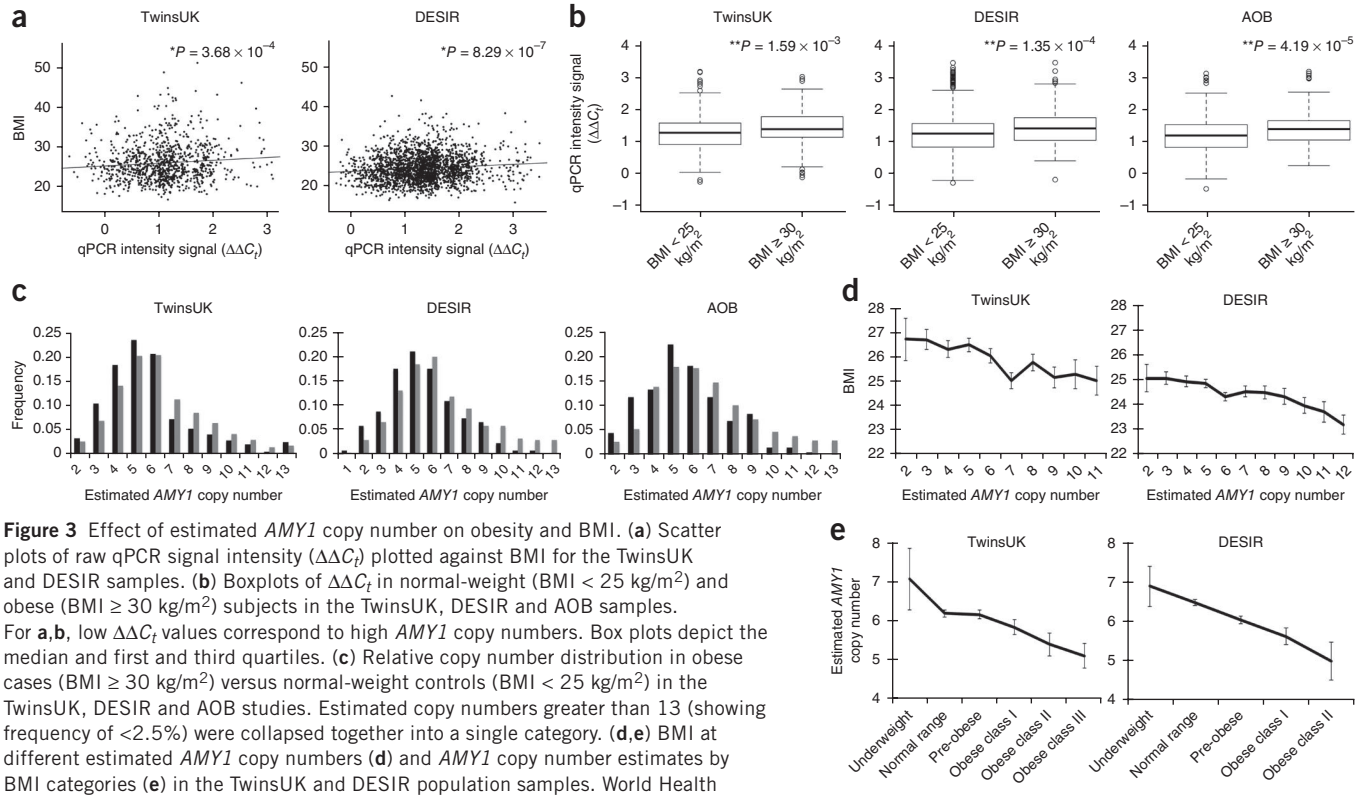

Figure 3 Effect of estimated *AMY1* copy number on obesity and BMI. (**a**) Scatter plots of raw qPCR signal intensity ($\Delta\Delta C_t$) plotted against BMI for the TwinsUK and DESIR samples. (**b**) Boxplots of $\Delta\Delta C_t$ in normal-weight (BMI < 25 kg/m^2) and obese (BMI ≥ 30 kg/m^2) subjects in the TwinsUK, DESIR and AOB samples. For **a**,**b**, low $\Delta\Delta C_t$ values correspond to high *AMY1* copy numbers. Box plots depict the median and first and third quartiles. (**c**) Relative copy number distribution in obese cases (BMI ≥ 30 kg/m^2) versus normal-weight controls (BMI < 25 kg/m^2) in the TwinsUK, DESIR and AOB studies. Estimated copy numbers greater than 13 (showing frequency of <2.5%) were collapsed together into a single category. (**d**,**e**) BMI at different estimated *AMY1* copy numbers (**d**) and *AMY1* copy number estimates by BMI categories (**e**) in the TwinsUK and DESIR population samples. World Health Organization BMI classification: underweight, <18.5 kg/m^2; normal range, 18.50–24.99 kg/m^2; pre-obese, 25.00–29.99 kg/m^2; obese class I, 30.00–34.99 kg/m^2; obese class II, 35.00–39.99 kg/m^2. All error bars, s.e.m. *, association between BMI and qPCR $\Delta\Delta C_t$ intensity signal, corrected for age, sex (DESIR), family (TwinsUK) and genotyping plate. **, Wilcoxon rank sum test.

Chinese case-control sample from the Singapore Prospective Study Program (SP2)[19]. A total of 136 obese and 197 overweight subjects were identified among the 2,431 Chinese subjects included in the SP2 cohort, with 325 matched lean Chinese SP2 normal-weight controls. *AMY1* copy number was measured by qPCR in all 658 subjects. Median copy number in SP2 normal-weight subjects was 6 (ranging from 2 to 16), similar to in our French DESIR and UK TwinsUK populations and in line with previous observations by Perry *et al.*[20]. Case-control association analysis in the Chinese sample showed that reduced estimated copy number for *AMY1* was associated with increased risk of obesity (per-copy OR = 1.17, 95% CI = 1.05–1.29; $P = 3.73 \times 10^{-3}$). Extending the case sample to include the 197 overweight subjects further confirmed the results (per-copy OR = 1.13, 95% CI = 1.06–1.21; $P = 3.52 \times 10^{-4}$).

To validate our *AMY1* genomic copy number data at the protein level, we investigated the effect of copy number variation at *AMY1* and *AMY2* on serum amylase enzyme levels and their relationship with BMI using 468 French morbidly obese subjects from the Atlas Biologique de l'Obésité Sévère (ABOS) study (**Table 1** and **Supplementary Table 12**). On average, salivary and pancreatic amylase proportions were approximately equal in serum (52% and 48%, respectively), and their levels showed close positive association with copy number variation at their respective genes ($P < 2.20 \times 10^{-16}$ and 1.04×10^{-11}, respectively; **Supplementary Fig. 11**). BMI was inversely associated with serum salivary amylase levels ($\beta = -0.23$ (0.04) kg/m^2; $P = 2.26 \times 10^{-7}$; **Supplementary Fig. 12**) and to a lesser extent with serum pancreatic amylase levels (β (se) = −0.23 (0.06) kg/m^2; $P = 2.29 \times 10^{-4}$; **Supplementary Fig. 12**), likely reflecting the physiological correlation between the levels of the two enzymes ($r = 0.21$; $P = 4.29 \times 10^{-6}$).

Salivary amylase catalyzes the hydrolysis of the α-1,4-glycosidic bonds of starch, initiating carbohydrate digestion in the oral cavity. Whereas individual salivary amylase levels vary in response to environmental factors, including psychological stress[21], they are genetically influenced by and directly correlate with the highly variable copy number at *AMY1* (refs. 20,22). Higher gene copy numbers at this locus are believed to have evolved in the human lineage as a consequence of a shift to a starch-rich diet[23]. Human populations traditionally consuming a high proportion of carbohydrates in their diet show higher copy numbers and salivary amylase activity than those consuming a low-starch diet[20,24]. The salivary glands and pancreas make similar contributions to the determination of overall levels of serum amylase[25], although enzyme activity is also detectable in other organs, including in adipose tissue[26,27]. Indeed, *AMY1* was among the 30% most highly expressed genes in adipose tissue in both our discovery sample and publicly available data from an unselected population sample, thus suggesting that this gene is actively expressed in adipose tissue (**Supplementary Note**). Whether adipose tissue is functionally involved in the link between *AMY1* copy number and obesity or whether this link implicates a different tissue in which *AMY1* is also actively transcribed warrants further investigation.

Lower blood amylase levels have been observed in both obese humans[28] and rats[29] and have recently been associated with increased risk of metabolic abnormalities[30,31] and reduced preabsorptive insulin release[32]. Furthermore, a recent study in mice fed a high-fat and high-sugar diet suggested association of the amylase locus with weight gain[33]. In these mice, this locus was also shown to be associated with the proportion of Enterobacteriaceae in the gut microbiota[33], which have previously been correlated with obesity in humans[34].

Rare CNVs have recently been implicated in highly penetrant forms of obesity[35,36] and severe thinness[37] through gene dosage–dependent effects. Common biallelic CNVs have also been associated with BMI[38–41]; however, as most of these CNVs are reliably tagged by surrounding SNPs[42], they share the same properties of small effect size and limited predictive value for obesity risk. In contrast, complex multi-allelic CNVs show lower linkage disequilibrium (LD) with surrounding SNPs (**Supplementary Table 13**) and are consequently less detectable by SNP-based genome-wide association study (GWAS)[43]. Surprisingly, *FTO* is the most replicated obesity susceptibility gene identified through GWAS[41], yet, in our analyses, estimated *AMY1* copy number appeared to show stronger association with BMI than *FTO* SNPs (**Supplementary Tables 14** and **15**). It is conceivable that high structural variability in the amylase region and subsequent low SNP coverage (**Supplementary Figs. 13–15**) might have hampered previous SNP-based GWAS attempts to detect association between the amylase cluster and adiposity. Indeed, examination of data from the most recent meta-analysis for BMI conducted by the Genetic Investigation of ANthropometric Traits (GIANT) Consortium[41] showed a large gap in SNP coverage across the locus encompassing the salivary amylase gene (**Supplementary Fig. 14**).

Present DNA high-throughput methods for CNV assessment, including array-, PCR- and sequencing-based approaches, are all affected by a wide number of variables, including DNA source, extraction method, quality and concentration, as well as experimental factors, inducing batch effects[10,11,44]. These factors complicate copy number measurement at multi-allelic CNVs and hinder the pooling of data from multiple centers. The observed association of *AMY1* with obesity may rekindle interest in the role of multi-allelic CNVs in common disease, driving the development of novel technological approaches for the accurate and high-throughput measurement of absolute copy number at such loci. These technological improvements will enable high-quality association analyses at such loci in larger sample sizes similar to those included in SNP association studies and are mandatory for disease risk assessment at the individual level, paving the way toward personalized medicine.

Our study provides a first genetic link between carbohydrate metabolism and obesity, with low copy number at *AMY1* resulting in decreased salivary amylase levels and a higher risk of obesity. This finding provides intriguing insight into some of the biological mechanisms underlying obesity, as well as a novel rationale for the investigation of innovative obesity treatments based on the manipulation of digestive enzyme levels.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene expression microarray data for the complete Swedish discordant sibling pair study sample have been deposited in the Gene Expression Omnibus (GEO) under accession GSE27916.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank the study participants in each of the studies, as well as the field investigators for the recruitment and examination of study subjects. The authors would also like to thank P. Maboudou and T. Brousseau for the measurement of serum amylase levels in the ABOS study subjects and P. Gele for the handling of ABOS samples. The authors thank the Genotyping Facility at the Wellcome Trust Sanger Institute for generating the TwinsUK SNP array data. The authors also thank E. Petretto, M. Manca, L. Coin, T. Andrew, I. Prokopenko, T. Norat and S. Richardson for helpful discussions, L. Arnalsteen, H. Verkindt, C. Eberle and M.F. Six for their contributions to the ABOS study and M. Deweider and F. Allegaert for laboratory processing of the DESIR samples, as well as S. Burbidge and M. Harvey of the Imperial College High-Performance Computing service for their assistance (http://www3.imperial.ac.uk/ict/services/hpc).

This work was supported by grants from the Wellcome Trust (grants 079534/z/06/z and 085555), the Medical Research Council (K2010-55X-11285-13), the Swedish Research Council, the Swedish Foundation for Strategic Research to the Sahlgrenska Center for Cardiovascular and Metabolic Research, the Swedish Diabetes Foundation and the Swedish federal government under the LUA/ALF agreement and by the European Community's Seventh Framework Programme (FP7/2007-2013) EUROCHIP project. M.F. is supported by the Medical Research Council (MR/K01353X/1), the British Skin Foundation (5044i), Qatar Foundation (GEQATDIAB) and the Commission of the European Communities (115005). P.F. is supported by the Medical Research Council (G1002084/1), the Imperial College Healthcare National Health Service (NHS) Trust Biomedical Research Centre (P46304 WMDI), the Commission of the European Communities (294785) and the Qatar Foundation (GEQATDIAB).

The DESIR study has been supported by INSERM contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis and by INSERM (Réseaux en Santé Publique, Interactions entre les Déterminants de la Santé), Cohortes Santé TGIR, Association Diabète Risque Vasculaire, Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon.

DESIR Study Group. INSERM U780: B.B., M.A. Charles, P. Ducimetière, E. Eschwège; INSERM U367: F. Alhenc-Gelas; CHU (Centre Hospitalier Universitaire) D'Angers: Y. Gallois, A. Girault; Bichat Hospital: F. Fumeron, M. Marre, R. Roussel; CHU Rennes: F. Bonnet; CNRS UMR 8090, Lille: S. Cauchi, P.F.; Centres d'Examens de Santé: Alençon, Angers, Blois, Caen, Chartres, Chateauroux, Cholet, Le Mans, Orleans-Tours; Institute de Recherche Médecine Générale: J. Cogneau; general practitioners of the region; Institute Inter-Regional pour la Santé: C. Born, E. Caces, M. Cailleau, N. Copin, O. Lantieri, J.G. Moreau, F. Rakotozafy, J. Tichet, S. Vol.

This study makes use of data generated by the Genome Structural Variation Consortium (principal investigators N. Carter, M. Hurles, C. Lee and S. Scherer) who we thank for pre-publication access to their CNV discovery (and/or) genotyping data, made available through the websites http://www.sanger.ac.uk/research/areas/humangenetics/cnv/ and http://dgv.tcag.ca/dgv/app/home as a resource to the community. Funding for the project was provided by the Wellcome Trust (grant 077006/Z/05/Z), the Canada Foundation of Innovation and the Ontario Innovation Trust, the Canadian Institutes of Health Research, the Genome Canada/Ontario Genomics Institute, the McLaughlin Centre for Molecular Medicine, the Ontario Ministry of Research and Innovation, the Hospital for Sick Children Foundation, the Department of Pathology at Brigham and Women's Hospital and US National Institutes of Health grants HG004221 and GM081533.

TwinsUK was funded by the Wellcome Trust; the authors acknowledge funding of the GWAS with the support of the Wellcome Trust Sanger Centre and the National Eye Institute via a US National Institutes of Health/Center for Inherited Disease Research (CIDR) genotyping project grant (principal investigator T. Young); also the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413. The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The project also received support from a Biotechnology and Biological Sciences Research Council (BBSRC) project grant (G20234). The ABOS study was partially supported by Lille University Hospital and INSERM (Centre d'Investigation Clinique de Lille) and by grants from Conseil Regional Nord-Pas de Calais (ARCIR Obésité).

F. Pesce was supported by a fellowship from the European Renal Association–European Dialysis and Transplant Association (ERA-EDTA; ALTF 72-2010). M.N.A.-S. is supported by a PhD scholarship from the Research Division of the Qatar Foundation. R.S. is a recipient of a Chercheur Boursier award from the Fonds de la Recherche en Santé du Québec and a New Investigator Award from the Canadian Institutes of Health Research. P.D. is supported by the Wellcome Trust. T.D.S. is in receipt of an NIHR Senior Investigator award and a European Research Council Senior Researcher award.

AUTHOR CONTRIBUTIONS

M.F. and P.F. conceived the study. M.F., P.F. and T.D.S. directed the project. M.F., J.S.E.-S.M. and P.F. wrote the manuscript. P.T., F. Pesce, A.B. and J.C.A.-A. contributed equally to this work. A.B., T.D.S., R.S., F. Pattou, H.-C.S., P.H.S., L.B.,

F. Pesce, P.T., R. Dorajoo, P.C.S. and E.E.E. edited the manuscript. J.S.E.-S.M., P.T., F. Pesce, J.C.A.-A., R. Dorajoo, M.N.A.-S., E.O., A.B., A.D. and M.H. performed the laboratory experiments. M.F., J.S.E.-S.M., J.C.A.-A., L.B., P.H.S., E.E.E., P.C.S. and H.-C.S. performed the statistical analyses. R.W.D., A.P., R. Dent, M. Mangino, P.G.H., J.S., M.P., R.C., V.R., E.V., S.F., B.B., M. Marre, S.V.-S., J.W., O.P.-G., P.J., L.S., C.J.H., P.D., R.M., J.L., E.S.T., L.M.S.C., A.W., F. Pattou, T.D.S. and P.F. provided samples, data and/or reagents. All authors commented on and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Gonzalez, E. *et al.* The influence of *CCL3L1* gene–containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
2. Hollox, E.J. *et al.* Psoriasis is associated with increased β-defensin genomic copy number. *Nat. Genet.* **40**, 23–25 (2008).
3. Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human β-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.* **79**, 439–448 (2006).
4. Aitman, T.J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
5. Eleftherohorinou, H. *et al.* famCNV: copy number variant association for quantitative traits in families. *Bioinformatics* **27**, 1873–1875 (2011).
6. Walley, A.J. *et al.* Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. *Int. J. Obes. (Lond.)* **36**, 137–147 (2012).
7. Hysi, P.G. *et al.* A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. *Nat. Genet.* **42**, 902–905 (2010).
8. Field, S.F. *et al.* Experimental aspects of copy number variant assays at *CCL3L1*. *Nat. Med.* **15**, 1115–1117 (2009).
9. Hollox, E.J. in *Genomic Structural Variants* Vol. 838 (ed. Feuk, L.) 187–207 (Springer New York, 2012).
10. Aldhous, M.C. *et al.* Measurement methods and accuracy in copy number variation: failure to replicate associations of β-defensin copy number with Crohn's disease. *Hum. Mol. Genet.* **19**, 4930–4938 (2010).
11. Carpenter, D., Walker, S., Prescott, N., Schalkwijk, J. & Armour, J.A. Accuracy and differential bias in copy number measurement of *CCL3L1* in association studies with three auto-immune disorders. *BMC Genomics* **12**, 418 (2011).
12. Sudmant, P.H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330**, 641–646 (2010).
13. Balkau, B., Eschwege, E., Tichet, J. & Marre, M. Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). *Diabetes Metab.* **23**, 428–434 (1997).
14. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906 (2006).
15. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* **41**, 157–159 (2009).
16. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
17. Maes, H.H., Neale, M.C. & Eaves, L.J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
18. Stunkard, A.J., Foch, T.T. & Hrubec, Z. A twin study of human obesity. *J. Am. Med. Assoc.* **256**, 51–54 (1986).
19. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* **44**, 307–311 (2012).
20. Perry, G.H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
21. Chatterton, R.T. Jr., Vogelsong, K.M., Lu, Y.C., Ellman, A.B. & Hudgens, G.A. Salivary α-amylase as a measure of endogenous adrenergic activity. *Clin. Physiol.* **16**, 433–448 (1996).
22. Park, K.S. Heritability of urine and plasma amylase activity. *Jinrui Idengaku Zasshi* **22**, 79–88 (1977).
23. Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009).
24. Squires, B.T. Human salivary amylase secretion in relation to diet. *J. Physiol. (Lond.)* **119**, 153–156 (1953).
25. Mifflin, T.E., Hortin, G. & Bruns, D.E. Electrophoretic assays of amylase isoenzymes and isoforms. *Clin. Lab. Med.* **6**, 583–599 (1986).
26. Mirski, A. Metabolism of adipose tissue *in vitro*. *Biochem. J.* **36**, 232–241 (1942).
27. Bingham, S.A. *et al.* Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr.* **4**, 847–858 (2001).
28. Kondo, T., Hayakawa, T., Shibata, T., Sato, Y. & Toda, Y. Serum levels of pancreatic enzymes in lean and obese subjects. *Int. J. Pancreatol.* **3**, 241–248 (1988).
29. Schneeman, B.O., Inman, M.D. & Stern, J.S. Pancreatic enzyme activity in obese and lean Zucker rats: a developmental study. *J. Nutr.* **113**, 921–925 (1983).
30. Nakajima, K. *et al.* Low serum amylase in association with metabolic syndrome and diabetes: a community-based study. *Cardiovasc. Diabetol.* **10**, 34 (2011).
31. Nakajima, K., Muneyuki, T., Munakata, H. & Kakei, M. Revisiting the cardiometabolic relevance of serum amylase. *BMC Res. Notes* **4**, 419 (2011).
32. Mandel, A.L. & Breslin, P.A. High endogenous salivary amylase activity is associated with improved glycemic homeostasis following starch ingestion in adults. *J. Nutr.* **142**, 853–858 (2012).
33. Parks, B.W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* **17**, 141–152 (2013).
34. Karlsson, C.L. *et al.* The microbiota of the gut in preschool children with normal and excessive body weight. *Obesity (Silver Spring)* **20**, 2257–2261 (2012).
35. Walters, R.G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (2010).
36. Bochukova, E.G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666–670 (2010).
37. Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102 (2011).
38. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
39. Jarick, I. *et al.* Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum. Mol. Genet.* **20**, 840–852 (2011).
40. Sha, B.Y. *et al.* Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J. Hum. Genet.* **54**, 199–202 (2009).
41. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
42. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
43. Locke, D.P. *et al.* Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.* **79**, 275–290 (2006).
44. Barnes, C. *et al.* A robust statistical method for case-control association testing with copy number variation. *Nat. Genet.* **40**, 1245–1252 (2008).

ONLINE METHODS

Further detailed methods are provided in the **Supplementary Note**. The studies were approved by the relevant institutional review boards in Sweden, the UK, France and Singapore and were conducted according to the principles of the Declaration of Helsinki. Written informed consent was obtained from every participant in each study.

Associations were assessed using linear mixed-effects models, including plate as a random effect and family structure as an additional random effect where appropriate. Age and sex were included as covariates.

Discovery. The discovery sample included 149 Swedish families (342 subjects) ascertained through the identification of an obesity-discordant sibling pair (BMI difference >10 kg/m^2)[6]. Gene expression for 29,546 transcripts (16,563 Ensembl genes) was measured in subcutaneous adipose tissue using the Affymetrix Human Genome U133 Plus 2.0 microarray. GWAS signal intensity data from Illumina 610K-Quad arrays were available for 348,150 probes lying within each transcript plus 30 kb upstream and downstream to encompass the coding regions and their internal and nearby regulatory regions.

qPCR was carried out to infer relative copy number measurements reflecting the underlying copy number distribution at *AMY1* and *AMY2*, respectively, using the TaqMan assays Hs07226362_cn (qPCR efficiency is shown in **Supplementary Fig. 16**) and Hs04204136_cn on an Applied Biosystems 7900HT Real-Time PCR System. Association analyses were carried out for 481 subjects with complete data on BMI and dual-energy X-ray absorptiometry (DEXA)-derived fat mass.

Replication. *In silico* replication of the BMI association was conducted using 972 female subjects from the UK adult twin registry (TwinsUK) cohort[14] using intensity signals from Illumina 610K-Quad arrays[7]. Association with BMI and obesity was analyzed in 2 population samples using qPCR estimates of *AMY1* copy number for 1,479 female twins from TwinsUK[14] and 2,137 subjects from the French Data from the Epidemiological Study on the Insulin Resistance syndrome (DESIR)[13] cohort. This cohort, with data collected at four time points over the course of 9 years, was also used to explore longitudinal genetic effects and gene-by-sex interactions (**Supplementary Fig. 17**, **Supplementary Table 16** and **Supplementary Note**). Obesity association with qPCR data was also assessed in an additional case-control sample of 205 obese cases and 358 age-matched controls from the French Adult Obesity study (AOB)[15]. An additional case-control sample was extracted from the Singapore Prospective Study Program (SP2) cohort, a population-based study including 2,431 adult Chinese Singaporean subjects[19]. Obesity in the Chinese population was defined as BMI ≥ 28 kg/m^2, and normal weight was defined as BMI < 23 kg/m^2, on the basis of criteria set by the Working Group on Obesity in China[45] and the World Health Organization expert consultation for Asia[46]. Accordingly, a total of 136 obese and 197 overweight subjects were identified among the 2,431 Chinese subjects of the SP2 cohort, with 325 matched lean SP2 subjects selected as normal-weight controls.

To avoid any potential population stratification affecting our association analyses resulting from known differences in the *AMY1* copy number distribution across populations traditionally consuming high-starch versus low-starch diets[20], we carried out genotype principal-component analysis using genome-wide SNP array data to ensure that samples included in each analysis were of the same ancestry and genetic background. Furthermore, *AMY1* association analyses were conducted separately in each of the study populations and then combined by meta-analysis using METAL[47] rather than pooling.

Protein levels. Atlas Biologique de l'Obésité Sévère (ABOS) is a French cohort comprising candidates for bariatric surgery. Serum pancreatic and total amylase levels for 468 individuals were measured by an enzymatic colorimetric assay with an autoanalyzer (CoBAS Icobas 8000 modular analyzer series; kits AMYL2-03183742122 and AMY-P-20766623322, Hoffman–La Roche). Serum salivary amylase levels were calculated by subtracting serum pancreatic amylase levels from total serum amylase levels.

45. Zhou, B.F. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults—study on optimal cut-off points of body mass index and waist circumference in Chinese adults. *Biomed. Environ. Sci.* **15**, 83–96 (2002).
46. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* **363**, 157–163 (2004).
47. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

doi:10.1038/ng.2939