Supplementary Information

Low copy number of the salivary amylase gene predisposes to obesity

Mario Falchi[1,39‡], Julia Sarah El-Sayed Moustafa[1,39], Petros Takousis[1,40], Francesco Pesce[1,2,40], Amélie Bonnefond[3,4,5,6,40], Johanna C. Andersson-Assarsson[7,8,1,40], 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[15,16,6], 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[15,16,4,6], Robert Caiazzo[15,16,4,6], Violeta Raverdy[15,16,6], 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[31,32,33], Robert Sladek[34,35,36], Lena M.S. Carlsson[7,8], Andrew Walley[1,37], Evan E. Eichler[9,38], Francois Pattou[15,16,6,4], Timothy D. Spector[18,41], Philippe Froguel[1,3,6,4,5,41‡]

Supplementary Tables and Figures: 2-37

Supplementary Note: 38-61

Supplementary Tables

Supplementary Table 1: Summary results from the GCAS with gene expression in adipose tissue in the Swedish discordant sib-pair dataset at different FDR thresholds.

FDR	Associations	Transcripts*	Probes*	Ensembl Genes*
10%	171	79	131	84
5%	149	66	115	73
1%	105	41	76	40

* Multiple associations may be observed for the same probe with different transcripts, some of which can be transcribed from the same gene. *P* values for FDR were corrected for the observed inflation (lambda = 1.12).

Supplementary Table 2: Association of copy-number signal at the *AMY1* and *AMY2* loci, inferred through qPCR, in subjects with complete BMI and fat mass data in the Swedish study of sib-pairs discordant for BMI.

Gene	N	Trait	β (SE)	*P*
AMY1	481	BMI	-0.34 (0.13)	8.08×10^{-3}
AMY2	478	BMI	-0.02 (0.59)	0.97
AMY1	481	Fat mass	-0.69 (0.26)	8.53×10^{-3}
AMY2	478	Fat mass	-0.11 (1.19)	0.92

Supplementary Table 3: Coordinates of genotyped loci in Build 37 of the human reference genome and the associate number of k-mer informative sites in each of these loci on chromosome 1.

Start	End	SUNKs	Paralog
104,085,264	104,131,162	15,298	*AMY2B*
104,140,218	104,173,355	2,150	*AMY2A*
104,186,809	104,209,356	109	*AMY1A1*
104,218,850	104,243,770	25	*AMY1A2*
104,278,480	104,307,257	432	*AMY1A3*
104,249,110	104,268,394	150	*AMYP*

Supplementary Table 4: Inferred diploid copy-number frequency at *AMY1* in the TwinsUK and DESIR samples. Estimates in TwinsUK were calculated using a random unrelated subsample.

Estimated *AMY1*	TwinsUK		DESIR	
copy-number	*N*	Frequency	*N*	Frequency
1	0	0.000	6	0.003
2	25	0.026	60	0.028
3	70	0.074	168	0.079
4	160	0.169	286	0.134
5	186	0.196	427	0.200
6	182	0.192	393	0.184
7	103	0.109	229	0.107
8	77	0.081	187	0.088
9	56	0.059	121	0.057
10	38	0.040	109	0.051
11	18	0.019	54	0.025
12	13	0.014	44	0.021
13	13	0.014	20	0.009
14	4	0.004	15	0.007
15	2	0.002	10	0.005
16	1	0.001	7	0.003
17	1	0.001	1	<0.001

Supplementary Table 5: *AMY1* estimated copy-number allele frequencies in the TwinsUK and DESIR samples. Estimates in TwinsUK were calculated using a random unrelated subsample. Allele frequencies were estimated through the EM algorithm implemented in CoNVEM[26]. (http://apps.biocompute.org.uk/convem/).

***AMY1* allele**	**Frequency TwinsUK**	**Frequency DESIR**
0	0.000	0.014
1	0.161	0.133
2	0.230	0.239
3	0.359	0.339
4	0.096	0.097
5	0.057	0.075
6	0.041	0.025
7	0.038	0.049
8	0.004	0.007
9	0.000	0.015
10	0.014	0.002
11	0.000	0.001
12	0.000	0.003
13	0.000	<0.001

Supplementary Table 6: Average BMI per inferred copy-number state in the TwinsUK cohort. Observations with less than 30 subjects are not reported.

Estimated *AMY1* copy-number	***N***	**BMI**	***SE***
2	44	26.73	0.87
3	124	26.72	0.41
4	210	26.32	0.34
5	310	26.50	0.28
6	296	26.05	0.30
7	142	25.02	0.33
8	121	25.78	0.34
9	84	25.16	0.43
10	60	25.28	0.59
11	37	25.02	0.60

Supplementary Table 7: Average estimated copy-number at the *AMY1* gene in the TwinsUK cohort per BMI category, presented according to the WHO International Classification of adult underweight, overweight and obesity.

Classification	***N***	**Mean *AMY1* copy-number**	***SE***
Underweight	14	7.07	0.79
Normal range	697	6.19	0.09
Pre obese	517	6.17	0.11
Obese class I	169	5.85	0.19
Obese class II	64	5.41	0.29
Obese class III	18	5.11	0.31

Supplementary Table 8. Average BMI per estimated copy-number state in the DESIR cohort. Observations with less than 30 subjects are not reported.

Estimated *AMY1* copy-number	***N***	**BMI***	***SE***
2	60	25.06	0.54
3	168	25.06	0.24
4	286	24.92	0.21
5	427	24.85	0.17
6	393	24.32	0.17
7	229	24.54	0.21
8	187	24.50	0.26
9	121	24.34	0.32
10	109	23.98	0.31
11	54	23.73	0.41
12	44	23.21	0.38

*BMI is the averaged value among the four time points.

Supplementary Table 9. Average estimated copy-number at the *AMY1* gene in the DESIR cohort per BMI category, presented according to the WHO International Classification of adult underweight, overweight and obesity.

Classification	***N***	**Mean *AMY1* copy-number**	***SE***
Underweight	28	6.89	0.51
Normal range	1239	6.49	0.08
Pre obese	733	6.04	0.09
Obese class I	115	5.63	0.21
Obese class II	20	5.00	0.48

Supplementary Table 10: Association of copy-number estimates in the extreme deciles of the *AMY1* copy-number distribution, measured by qPCR, with obesity risk in the AOB case control study and the case-control samples derived from the TwinsUK and DESIR population studies. Meta-analysis was assessed with the software METAL[1].

	Cases		Controls					
Sample	***N***	**Age***	***N***	**Age***	***β (SE)***	***P***	**OR (95%CI)**	
TwinsUK	53	52 (48-60)	137	51 (42-59)	-2.14 (1.02)	3.60×10^{-2}	8.49 (1.15-62.80)	
DESIR	26	51 (44-60)	301	46 (38-53)	-1.82 (0.51)	4.07×10^{-4}	6.18 (2.25-16.95)	
AOB	40	34 (29-40)	76	36 (33-39)	-2.22 (0.51)	1.33×10^{-5}	9.25 (3.40-24.19)	
	N		***N***		***β (SE)***	***P***	**OR (95%CI)**	***Het P*****
Meta-analysis	119		514		-2.04 (0.34)	2.52×10^{-9}	7.67 (3.92-14.99)	0.85

*median (1st, 3rd quartiles) ** Heterogeneity *P* value

Supplementary Table 11: Association results for the estimation of variance of obesity explained by estimated *AMY1* copy-number. The effect of *AMY1* on obesity was first assessed by comparing high (≥8) versus medium (≥5 and ≤7) and low (≤4) copy-number carriers in AOB and the TwinsUK- and DESIR-derived case-control studies (by selecting cases and controls from the respective population sample). Meta-analysis was performed using the software METAL[1]. The presented combined effect sizes were then used to calculate the proportion of variance of obesity explained by copy-number variation at the salivary amylase gene using the method of So *et al*[16].

Sample	***AMY1* copy-number**	***N***	***β* (SE)**	***P***
AOB	≤4	147	-	-
	≥5 ≤7	281	0.58(0.23)	1.12×10^{-2}
	≥8	135	0.80(0.26)	2.21×10^{-3}
TUK	≤4	246	-	-
	≥5 ≤7	498	0.75 (0.45)	9.30×10^{-2}
	≥8	218	2.40 (0.89)	7.00×10^{-3}
DESIR	≤4	325	-	-
	≥5 ≤7	686	0.51 (0.25)	3.93×10^{-2}
	≥8	393	0.91 (0.27)	7.68×10^{-4}

Meta analysis	***AMY1* copy-number**	***N***	***β* (SE)**	***P***	**Het *P****	**OR (95% CI)**
	≤4	718	-	-		1
	≥5 ≤7	1465	0.57 (0.16)	2.68×10^{-4}	0.90	1.78 (1.30-2.42)
	≥8	746	0.92 (0.18)	5.85×10^{-7}	0.23	2.51 (1.75-3.59)

Het P =heterogeneity P value*

Supplementary Table 12: Summary statistics of salivary amylase enzyme levels in serum categorised by copy-numbers of the *AMY1* gene estimated by qPCR in the ABOS sample. Only copy-numbers present in >10 subjects are reported.

Gene	*Copy-number*	*N*	Mean salivary amylase (IU/L)	*SE*
AMY1	2	16	4.06	0.98
	3	53	12.42	0.81
	4	51	15.41	1.06
	5	139	20.73	0.74
	6	65	23.00	1.11
	7	65	23.86	1.21
	8	39	27.33	1.60
	9	17	27.65	2.72
AMY2	1	56	18.68	1.29
	2	347	20.31	0.55
	3	56	24.00	1.37

Supplementary Table 13: LD analysis in the *AMY1* region in the TwinsUK sample. Please refer to Section 3.11 of the Supplementary Information for further details of this analysis.

		SNP-SNP		SNP-*AMY1* CN (please see text for 1,2,3 definitions)		
SNP	Position (bp)	r^2 (Haplotype Frequency)	r^2 (Corr)	*AMY1*(1)	*AMY1*(2)	*AMY1*(3)
rs10785777	103916755	-	-	0.01	0.02	0.02
rs6696797	104017778	0.96	0.96	0.02	0.02	0.02
rs7524694	104019631	0.07	0.07	0.02	0.01	0.01
rs12125927	104043185	0.01	0.01	0.00	0.00	0.00
rs2050592	104103057	0.01	0.01	0.01	0.02	0.02
rs10159229	104109958	0.06	0.06	0.02	0.02	0.02
rs12085609	104122556	1.00	1.00	0.02	0.02	0.02
rs4847151	104123662	1.00	1.00	0.02	0.02	0.02
rs6661612	104128618	1.00	1.00	0.02	0.02	0.02
rs1999478	104311712	0.13	0.13	0.02	0.01	0.01
rs504877	104315497	0.15	0.14	0.00	0.00	0.00
rs7515323	104318414	1.00	1.00	0.00	0.00	0.00
rs550437	104319480	1.00	1.00	0.00	0.00	0.00
rs11185098	104321349	0.20	0.23	0.05	0.02	0.04
rs1930213	104324720	0.20	0.23	0.00	0.00	0.00
rs1930212	104324819	0.14	0.12	0.03	0.03	0.03
rs1566154	104329877	0.05	0.05	0.02	0.02	0.01
rs1930211	104342898	0.38	0.32	0.00	0.00	0.00
rs10881476	104350777	0.99	0.98	0.00	0.01	0.00
rs9628996	104350971	1.00	1.00	0.00	0.01	0.00
rs12136807	104401489	0.09	0.10	0.00	0.00	0.00
rs1579026	104405855	1.00	1.00	0.00	0.00	0.00
rs1930192	104408821	1.00	1.00	0.00	0.00	0.00
rs1930194	104409499	1.00	1.00	0.00	0.00	0.00
s1930195	104414641	1.00	1.00	0.00	0.00	0.00
rs1999480	104419830	1.00	1.00	0.00	0.00	0.00
rs1182598	104429480	1.00	1.00	0.00	0.00	0.00
rs12143917	104433984	0.75	0.74	0.00	0.00	0.00
rs1182591	104435039	0.75	0.75	0.00	0.00	0.00
rs7531107	104436329	0.92	0.92	0.00	0.00	0.00
rs1182589	104436874	0.34	0.33	0.01	0.01	0.01
rs1182587	104437902	0.33	0.33	0.00	0.00	0.00
rs1930208	104445316	0.26	0.29	0.01	0.00	0.01
rs1111700	104454507	0.23	0.28	0.00	0.00	0.00
rs1475190	104470622	0.57	0.57	0.00	0.00	0.00
rs551284	104484407	0.89	0.91	0.00	0.00	0.00
rs619879	104487148	1.00	1.00	0.00	0.00	0.00
rs1330388	104492281	0.43	0.41	0.00	0.00	0.00
rs11185345	104502333	0.60	0.62	0.00	0.01	0.01
rs6694128	104505231	0.88	0.89	0.00	0.01	0.00
rs2208963	104529736	0.74	0.73	0.00	0.01	0.00
rs519975	104536508	0.87	0.87	0.00	0.01	0.01
rs508464	104540638	0.88	0.88	0.00	0.01	0.00
rs750236	104544187	0.96	0.97	0.00	0.00	0.00
rs750237	104544302	1.00	1.00	0.00	0.00	0.00
rs11185353	104545275	0.97	0.98	0.00	0.01	0.01
rs7546635	104545772	1.00	1.00	0.00	0.01	0.00
rs2181446	104546362	1.00	1.00	0.00	0.01	0.00
rs10881526	104546516	0.96	0.98	0.00	0.00	0.00
rs1330408	104546820	0.96	0.97	0.00	0.01	0.01
rs1330407	104546881	1.00	1.00	0.00	0.01	0.01
rs11185354	104547208	1.00	1.00	0.00	0.01	0.01
rs4141899	104547617	0.97	0.98	0.00	0.01	0.00
rs11185362	104561488	0.88	0.87	0.00	0.01	0.01

Supplementary Table 14: Association in the TwinsUK sample between BMI and eleven SNPs at the *FTO* gene included in the haploblock on chromosome 16q12.2 within 53,797kb-53,850kb. For comparison, association with estimated *AMY1* copy-number is reported for the same subset of subjects using a dichotomous variable for estimated *AMY1* copy-numbers <6 and ≥6.

			FTO			***AMY1***		
SNP	**Position**	**Subjects**	**Beta**	**SE**	***P***	**Beta**	**SE**	***P***
rs1075440	53790906	1388	-0.41	0.19	2.95×10^{-2}	-0.86	0.25	5.52×10^{-4}
rs1558902	53803574	1324	0.39	0.19	3.47×10^{-2}	-0.89	0.25	4.14×10^{-4}
rs17817449	53813367	1368	0.42	0.18	2.32×10^{-2}	-0.85	0.25	6.76×10^{-4}
rs8043757	53813450	1368	0.42	0.18	2.32×10^{-2}	-0.85	0.25	6.76×10^{-4}
rs8050136	53816275	1388	0.42	0.18	2.12×10^{-2}	-0.86	0.25	5.52×10^{-4}
rs8051591	53816752	1382	0.44	0.18	1.62×10^{-2}	-0.86	0.25	5.76×10^{-4}
rs9935401	53816838	1382	0.44	0.18	1.62×10^{-2}	-0.86	0.25	5.76×10^{-4}
rs3751812	53818460	1388	0.45	0.18	1.44×10^{-2}	-0.86	0.25	5.52×10^{-4}
rs7190492	53828752	1388	-0.07	0.18	7.23×10^{-1}	-0.86	0.25	5.52×10^{-4}
rs8044769	53839135	1388	-0.27	0.18	1.31×10^{-1}	-0.86	0.25	5.52×10^{-4}
rs1421090	53850170	1388	-0.03	0.21	8.86×10^{-1}	-0.86	0.25	5.52×10^{-4}

Supplementary Table 15: Association in the DESIR sample between BMI and 104 SNPs of the *FTO* gene included in the haploblock on chromosome 16q12.2 within 53,797kb-53,850kb. For comparison, association with estimated *AMY1* copy-number is reported for the same subset of subjects using a dichotomous variable for *AMY1* estimated copy-numbers <6 and ≥6.

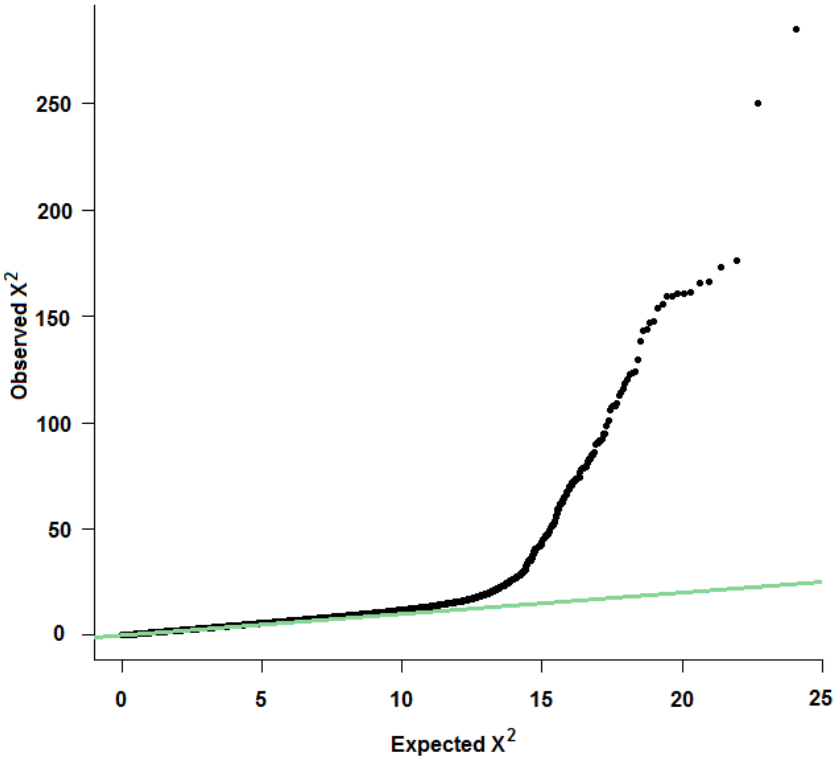
			FTO			***AMY1***		
SNP	**Position**	**Subjects**	**Beta**	**SE**	***P***	**Beta**	**SE**	***P***
rs60386982	53797452	2126	-0.02	0.16	8.98×10^{-1}	-0.64	0.14	5.94×10^{-6}
rs6499643	53797518	2125	-0.02	0.13	8.74×10^{-1}	-0.64	0.14	6.87×10^{-6}
rs4784323	53797565	2125	-0.23	0.11	3.82×10^{-2}	-0.64	0.14	6.55×10^{-6}
rs7206790	53797908	2128	0.18	0.10	7.70×10^{-2}	-0.65	0.14	5.46×10^{-6}
rs8047395	53798523	2124	-0.16	0.10	1.07×10^{-1}	-0.64	0.14	7.81×10^{-6}
rs8047587	53798622	2118	0.24	0.10	1.84×10^{-2}	-0.64	0.14	7.43×10^{-6}
rs9937053	53799507	2123	0.25	0.10	1.41×10^{-2}	-0.65	0.14	5.53×10^{-6}
rs9937354	53799847	2122	0.25	0.10	1.29×10^{-2}	-0.65	0.14	5.12×10^{-6}
rs9928094	53799905	2122	0.25	0.10	1.38×10^{-2}	-0.64	0.14	6.33×10^{-6}
rs9930333	53799977	2107	0.25	0.10	1.32×10^{-2}	-0.64	0.14	8.42×10^{-6}
rs9930397	53799985	2123	0.24	0.10	1.62×10^{-2}	-0.65	0.14	4.93×10^{-6}
rs9939973	53800568	2124	0.24	0.10	1.58×10^{-2}	-0.64	0.14	6.74×10^{-6}
rs9940646	53800629	2123	0.25	0.10	1.42×10^{-2}	-0.65	0.14	5.50×10^{-6}
rs9940128	53800754	2124	0.24	0.10	1.97×10^{-2}	-0.64	0.14	7.38×10^{-6}
rs1421085	53800954	2128	0.24	0.10	1.89×10^{-2}	-0.64	0.14	6.55×10^{-6}
rs9923147	53801549	2129	0.25	0.10	1.23×10^{-2}	-0.65	0.14	5.75×10^{-6}
rs9923544	53801985	2121	0.25	0.10	1.23×10^{-2}	-0.62	0.14	1.14×10^{-5}
rs11642015	53802494	2123	0.24	0.10	1.98×10^{-2}	-0.64	0.14	6.89×10^{-6}
rs8055197	53803156	2127	-0.14	0.10	1.44×10^{-1}	-0.65	0.14	5.33×10^{-6}
rs1558901	53803187	2124	0.24	0.10	1.76×10^{-2}	-0.65	0.14	5.70×10^{-6}
rs62048402	53803223	2127	0.24	0.10	2.02×10^{-2}	-0.64	0.14	6.15×10^{-6}
rs1558902	53803574	2115	0.25	0.10	1.56×10^{-2}	-0.65	0.14	5.76×10^{-6}
rs1861866	53804340	2127	-0.14	0.10	1.47×10^{-1}	-0.65	0.14	5.43×10^{-6}
rs10852521	53804965	2124	-0.14	0.10	1.50×10^{-1}	-0.64	0.14	6.99×10^{-6}
rs11075985	53805207	2122	0.24	0.10	1.67×10^{-2}	-0.64	0.14	6.87×10^{-6}
rs11075986	53805344	2125	-0.34	0.18	5.49×10^{-2}	-0.63	0.14	8.18×10^{-6}
rs2058908	53806145	2129	-0.21	0.12	7.57×10^{-2}	-0.64	0.14	7.31×10^{-6}
rs9922047	53806280	2120	-0.15	0.10	1.43×10^{-1}	-0.64	0.14	7.53×10^{-6}
rs56094641	53806453	2125	0.24	0.10	1.56×10^{-2}	-0.65	0.14	5.71×10^{-6}
rs17817288	53807764	2127	-0.13	0.10	2.05×10^{-1}	-0.64	0.14	7.61×10^{-6}
rs53808258	53808258	2126	-0.14	0.10	1.72×10^{-1}	-0.63	0.14	8.53×10^{-6}
rs55872725	53809123	2121	0.23	0.10	2.25×10^{-2}	-0.63	0.14	9.27×10^{-6}
rs1121980	53809247	2126	0.24	0.10	1.56×10^{-2}	-0.64	0.14	6.58×10^{-6}
rs73612011	53809861	2130	-0.34	0.20	8.45×10^{-2}	-0.64	0.14	6.56×10^{-6}
rs7187250	53810546	2125	0.21	0.10	3.93×10^{-2}	-0.65	0.14	5.90×10^{-6}
rs7193144	53810686	2109	0.20	0.10	5.29×10^{-2}	-0.61	0.14	2.09×10^{-5}
rs8063057	53812433	2129	0.23	0.10	2.37×10^{-2}	-0.64	0.14	6.26×10^{-6}
rs16945088	53812524	2130	-0.24	0.18	1.88×10^{-1}	-0.64	0.14	6.56×10^{-6}
rs8057044	53812614	2127	0.13	0.10	1.94×10^{-1}	-0.64	0.14	6.38×10^{-6}
rs17817449	53813367	2121	0.20	0.10	4.76×10^{-2}	-0.65	0.14	4.32×10^{-6}
rs9972653	53814363	2120	0.25	0.10	1.65×10^{-2}	-0.63	0.14	8.75×10^{-6}
rs11075987	53815161	2125	-0.12	0.10	2.17×10^{-1}	-0.64	0.14	7.57×10^{-6}
rs17817497	53815435	2123	0.23	0.10	2.53×10^{-2}	-0.65	0.14	5.35×10^{-6}
rs8050136	53816275	2124	0.22	0.10	3.05×10^{-2}	-0.64	0.14	6.38×10^{-6}
rs4783819	53816647	2123	-0.14	0.10	1.72×10^{-1}	-0.64	0.14	7.04×10^{-6}
rs8051591	53816752	2124	0.23	0.10	2.44×10^{-2}	-0.63	0.14	8.40×10^{-6}
rs9935401	53816838	2120	0.22	0.10	3.13×10^{-2}	-0.64	0.14	7.02×10^{-6}
rs9933509	53818167	2121	0.24	0.10	2.05×10^{-2}	-0.64	0.14	8.05×10^{-6}
rs3751812	53818460	2127	0.23	0.10	2.44×10^{-2}	-0.65	0.14	5.79×10^{-6}
rs3751813	53818708	2129	-0.26	0.10	1.12×10^{-2}	-0.64	0.14	6.38×10^{-6}
rs3751814	53818724	2129	0.23	0.10	2.62×10^{-2}	-0.65	0.14	4.74×10^{-6}
rs9931900	53818813	2125	0.24	0.10	1.86×10^{-2}	-0.64	0.14	7.10×10^{-6}

rs9936385	53819169	2128	0.23	0.10	2.27×10^{-2}	-0.64	0.14	6.58×10^{-6}
rs9923233	53819198	2121	0.23	0.10	2.64×10^{-2}	-0.63	0.14	1.00×10^{-5}
rs9923312	53819367	2124	0.22	0.10	3.16×10^{-2}	-0.65	0.14	5.86×10^{-6}
rs11075989	53819877	2122	0.22	0.10	3.41×10^{-2}	-0.64	0.14	6.36×10^{-6}
rs11075990	53819893	2123	0.22	0.10	2.81×10^{-2}	-0.64	0.14	6.33×10^{-6}
rs11075991	53819937	2122	0.23	0.10	2.78×10^{-2}	-0.64	0.14	7.93×10^{-6}
rs11075992	53820066	2124	0.23	0.10	2.21×10^{-2}	-0.64	0.14	6.45×10^{-6}
rs9926289	53820503	2122	0.22	0.10	2.89×10^{-2}	-0.64	0.14	5.97×10^{-6}
rs9939609	53820527	2119	0.23	0.10	2.24×10^{-2}	-0.62	0.14	1.15×10^{-5}
rs17817712	53821125	2124	0.23	0.10	2.22×10^{-2}	-0.64	0.14	7.68×10^{-6}
rs7206410	53821297	2124	0.23	0.10	2.56×10^{-2}	-0.64	0.14	7.55×10^{-6}
rs7206629	53821413	2125	0.25	0.10	1.46×10^{-2}	-0.64	0.14	6.18×10^{-6}
rs7202116	53821615	2129	0.23	0.10	2.19×10^{-2}	-0.64	0.14	6.26×10^{-6}
rs7202296	53821690	2123	0.24	0.10	2.10×10^{-2}	-0.63	0.14	8.63×10^{-6}
rs7201850	53821862	2127	0.24	0.10	1.79×10^{-2}	-0.65	0.14	5.69×10^{-6}
rs66908032	53822142	2129	0.23	0.10	2.41×10^{-2}	-0.64	0.14	6.76×10^{-6}
rs72803697	53822183	2127	0.22	0.10	3.00×10^{-2}	-0.64	0.14	6.03×10^{-6}
rs7185735	53822651	2123	0.23	0.10	2.62×10^{-2}	-0.64	0.14	6.39×10^{-6}
rs62033406	53824226	2107	0.20	0.10	5.03×10^{-2}	-0.64	0.14	6.60×10^{-6}
rs9941349	53825488	2125	0.23	0.10	2.08×10^{-2}	-0.65	0.14	4.58×10^{-6}
rs7187961	53826034	2130	-0.26	0.13	4.39×10^{-2}	-0.63	0.14	8.40×10^{-6}
rs10468280	53827479	2120	0.22	0.10	2.83×10^{-2}	-0.64	0.14	7.21×10^{-6}
rs76442450	53827487	2129	-0.35	0.20	7.31×10^{-2}	-0.64	0.14	7.37×10^{-6}
rs62033408	53827962	2124	0.21	0.10	3.93×10^{-2}	-0.63	0.14	8.23×10^{-6}
rs73612051	53828037	2128	-0.39	0.20	4.57×10^{-2}	-0.63	0.14	9.19×10^{-6}
rs17817964	53828066	2127	0.22	0.10	3.41×10^{-2}	-0.64	0.14	6.24×10^{-6}
rs7190492	53828752	2127	-0.15	0.10	1.49×10^{-1}	-0.65	0.14	5.43×10^{-6}
rs73607075	53828968	2130	-0.34	0.20	8.69×10^{-2}	-0.63	0.14	8.83×10^{-6}
rs9930506	53830465	2124	0.23	0.10	2.56×10^{-2}	-0.65	0.14	5.61×10^{-6}
rs9932754	53830491	2121	0.23	0.10	2.55×10^{-2}	-0.65	0.14	5.59×10^{-6}
rs9933040	53830867	2120	0.22	0.10	2.99×10^{-2}	-0.65	0.14	5.99×10^{-6}
rs9922708	53831146	2110	0.23	0.10	2.45×10^{-2}	-0.64	0.14	6.29×10^{-6}
rs72805611	53831354	2122	0.20	0.10	5.01×10^{-2}	-0.64	0.14	6.72×10^{-6}
rs9922619	53831771	2126	0.22	0.10	2.90×10^{-2}	-0.65	0.14	5.21×10^{-6}
rs4783821	53834508	2129	-0.14	0.10	1.79×10^{-1}	-0.63	0.14	8.05×10^{-6}
rs114078316	53834763	2121	0.05	0.15	7.59×10^{-1}	-0.65	0.14	4.59×10^{-6}
rs11075993	53837144	2122	0.20	0.10	5.40×10^{-2}	-0.65	0.14	5.92×10^{-6}
rs72805613	53837342	2129	0.21	0.10	4.26×10^{-2}	-0.64	0.14	6.26×10^{-6}
rs62033415	53837369	2119	0.01	0.19	9.53×10^{-1}	-0.64	0.14	6.26×10^{-6}
rs73607083	53838295	2128	-0.32	0.20	1.01×10^{-1}	-0.64	0.14	6.47×10^{-6}
rs8044769	53839135	2125	-0.10	0.10	2.96×10^{-1}	-0.64	0.14	7.28×10^{-6}
rs12149182	53841177	2124	-0.13	0.10	2.19×10^{-1}	-0.64	0.14	6.70×10^{-6}
rs58239912	53842518	2127	-0.36	0.20	6.81×10^{-2}	-0.64	0.14	7.82×10^{-6}
rs111357538	53842712	2130	-0.38	0.23	9.23×10^{-2}	-0.64	0.14	7.48×10^{-6}
rs12149832	53842908	2123	0.20	0.10	4.78×10^{-2}	-0.65	0.14	5.09×10^{-6}
rs17218700	53844579	2124	0.05	0.15	7.51×10^{-1}	-0.63	0.14	8.77×10^{-6}
rs11642841	53845487	2119	0.14	0.10	1.68×10^{-1}	-0.64	0.14	6.38×10^{-6}
rs1861867	53848561	2114	-0.20	0.10	5.25×10^{-2}	-0.64	0.14	6.92×10^{-6}
rs11075994	53850079	2121	-0.18	0.11	1.03×10^{-1}	-0.63	0.14	9.62×10^{-6}
rs1421090	53850170	2128	-0.02	0.11	8.91×10^{-1}	-0.65	0.14	5.80×10^{-6}
rs5816908	53850412	2128	0.14	0.12	2.47×10^{-1}	-0.64	0.14	6.49×10^{-6}
rs9939811	53850868	2119	0.26	0.11	2.13×10^{-2}	-0.65	0.14	5.91×10^{-6}

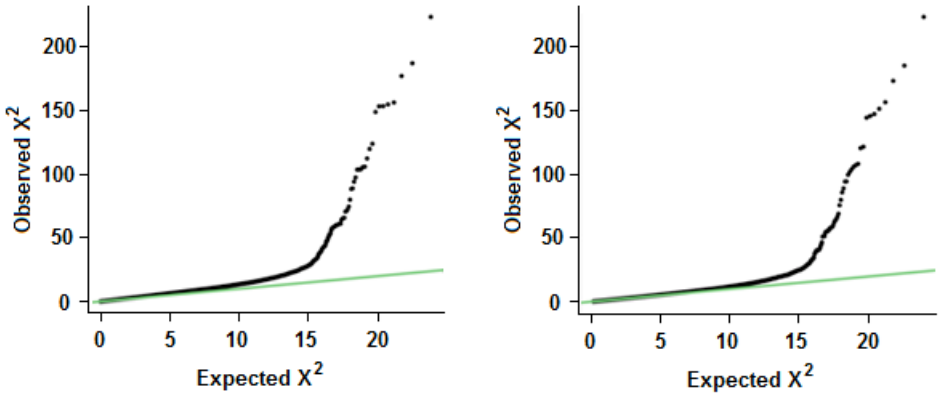
Supplementary Table 16: *AMY1* estimated copy-number effect by gender at different time points in the DESIR cohort, at recruitment (Year 0) and after 3, 6, and 9 years.

	Females		**Males**	
	β (SE)	***P***	***β (SE)***	***P***
Year 0	-0.14 (0.04)	4.76×10^{-4}	-0.10 (0.04)	5.55×10^{-3}
Year 3	-0.16 (0.04)	2.75×10^{-4}	-0.10 (0.04)	4.14×10^{-3}
Year 6	-0.19 (0.04)	3.25×10^{-5}	-0.12 (0.04)	8.99×10^{-4}
Year 9	-0.20 (0.05)	1.10×10^{-5}	-0.11 (0.04)	3.06×10^{-3}

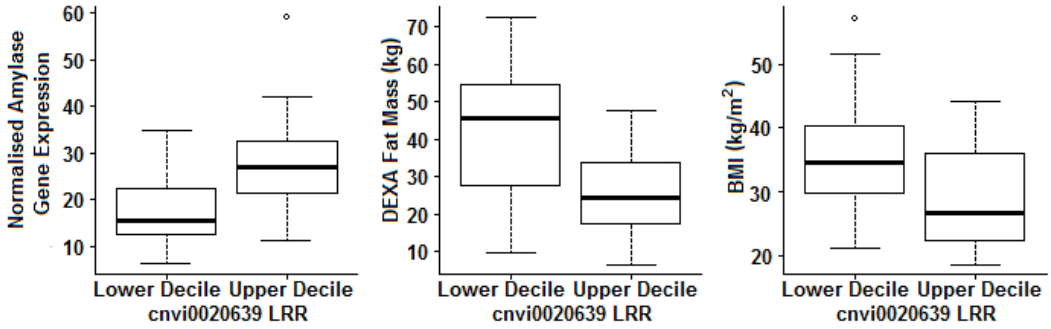
Supplementary Figures



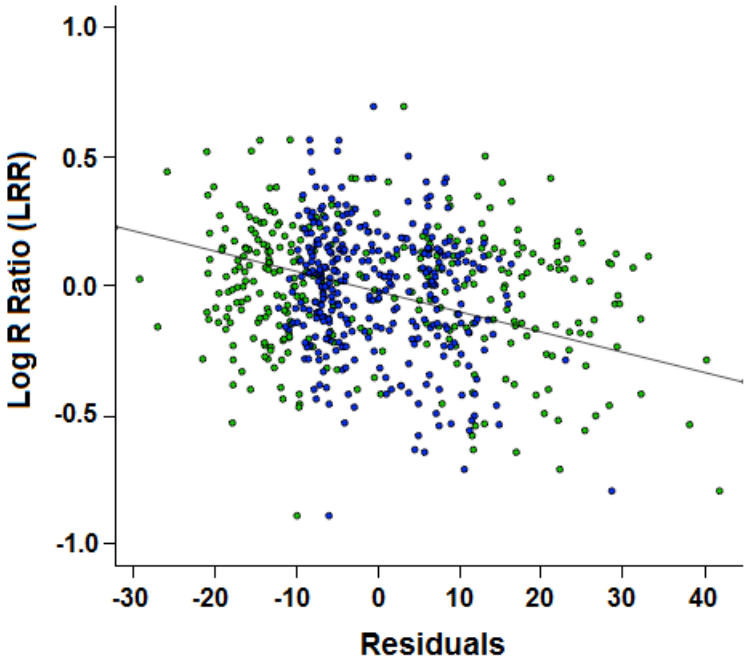
Supplementary Figure 1: Quantile-quantile (Q-Q) plot of famCNV test statistics for the GCAS in the Swedish discordant sib-pair sample set. GCAS was carried out between the transcripts and DNA-array probes lying within the transcript plus 30kb upstream and downstream, to encompass the coding regions and their internal and nearby regulatory regions. The Q-Q plot depicts the test statistics for all transcripts tested. The genomic inflation factor was 1.12, which was used for correction of the GCAS statistics by genomic control.



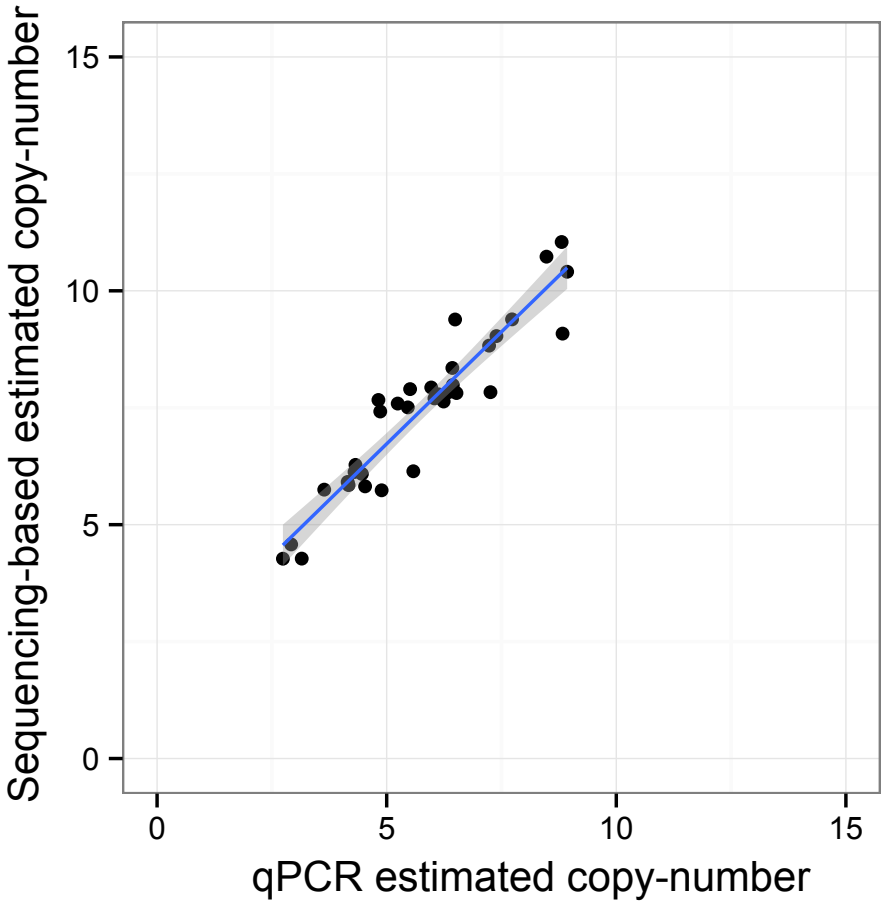
Supplementary Figure 2: Q-Q plots of the combined GCAS-BMI (left) and GCAS-fat mass (right) analyses. We assessed the significance of joint association for both gene expression and each of BMI and fat mass over the whole set of 348,150 Illumina probes used for the GCAS analysis. The cnvi0020639 probe identified in our study was ranked in the top 43 most significant hits for BMI (P = 5.50 x 10^{-10}) and within the 33 most significant for DEXA-derived fat mass (P = 1.85 x 10^{-10}). The strong associations observed between transcript levels and Illumina probes (**Supplementary Figure 1**) drove the majority of the highest associations in the joint association analyses. However, among the top hits only cnvi0020639 and one additional probe also showed a nominal univariate P value of association of less than 0.05 with both BMI and DEXA-derived fat mass.



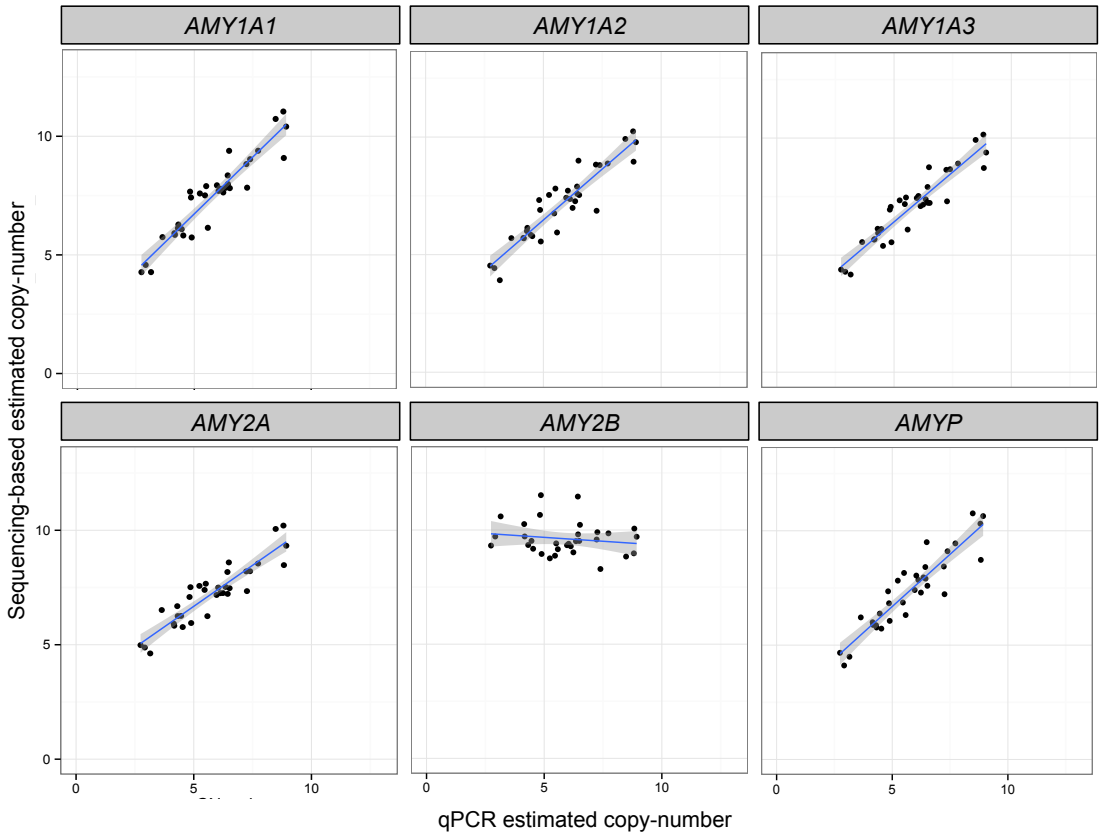
Supplementary Figure 3: Boxplot of (left to right) expression levels of the amylase genes in adipose tissue, DEXA-derived fat mass, and BMI, for subjects in the extreme upper and lower deciles of the cnvi0020639 Log R Ratio (LRR) distribution in the Swedish discovery sample.



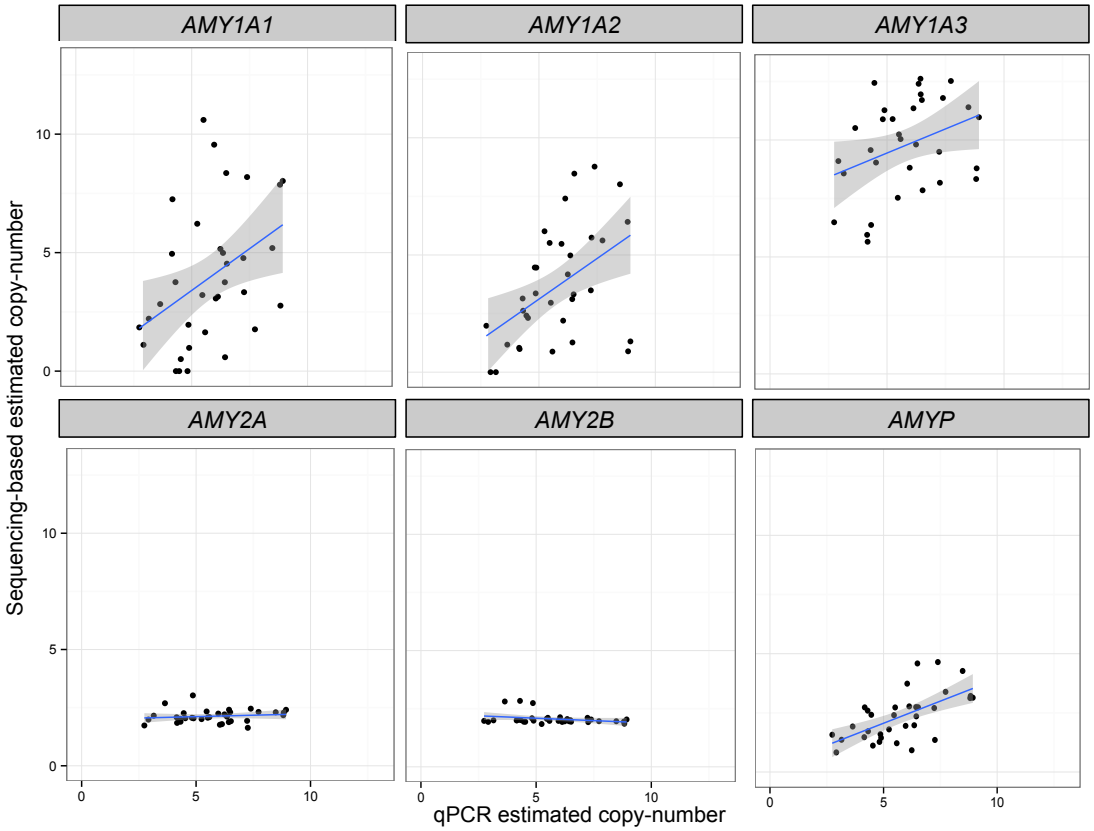
Supplementary Figure 4: Scatterplots of BMI (blue) and DEXA-derived fat mass (green) residuals, adjusted for sex and age, against Log R Ratio (LRR) at cnvi0020639 in the Swedish sample.



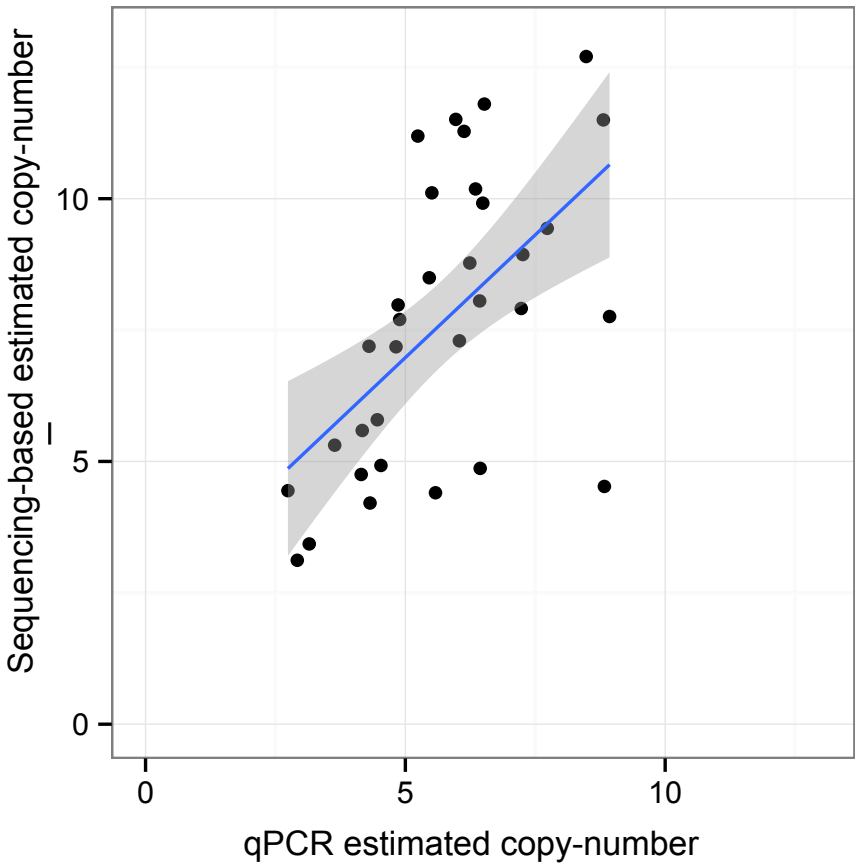
Supplementary Figure 5: Whole genome shotgun sequencing-based total copy-number estimates of the *AMY* genes versus copy-numbers estimated at *AMY1* by qPCR. The analyses using data from the 1000G project demonstrated a strong correlation between the two orthologous methods ($r = 0.94$, $P < 2 \times 10^{-16}$).



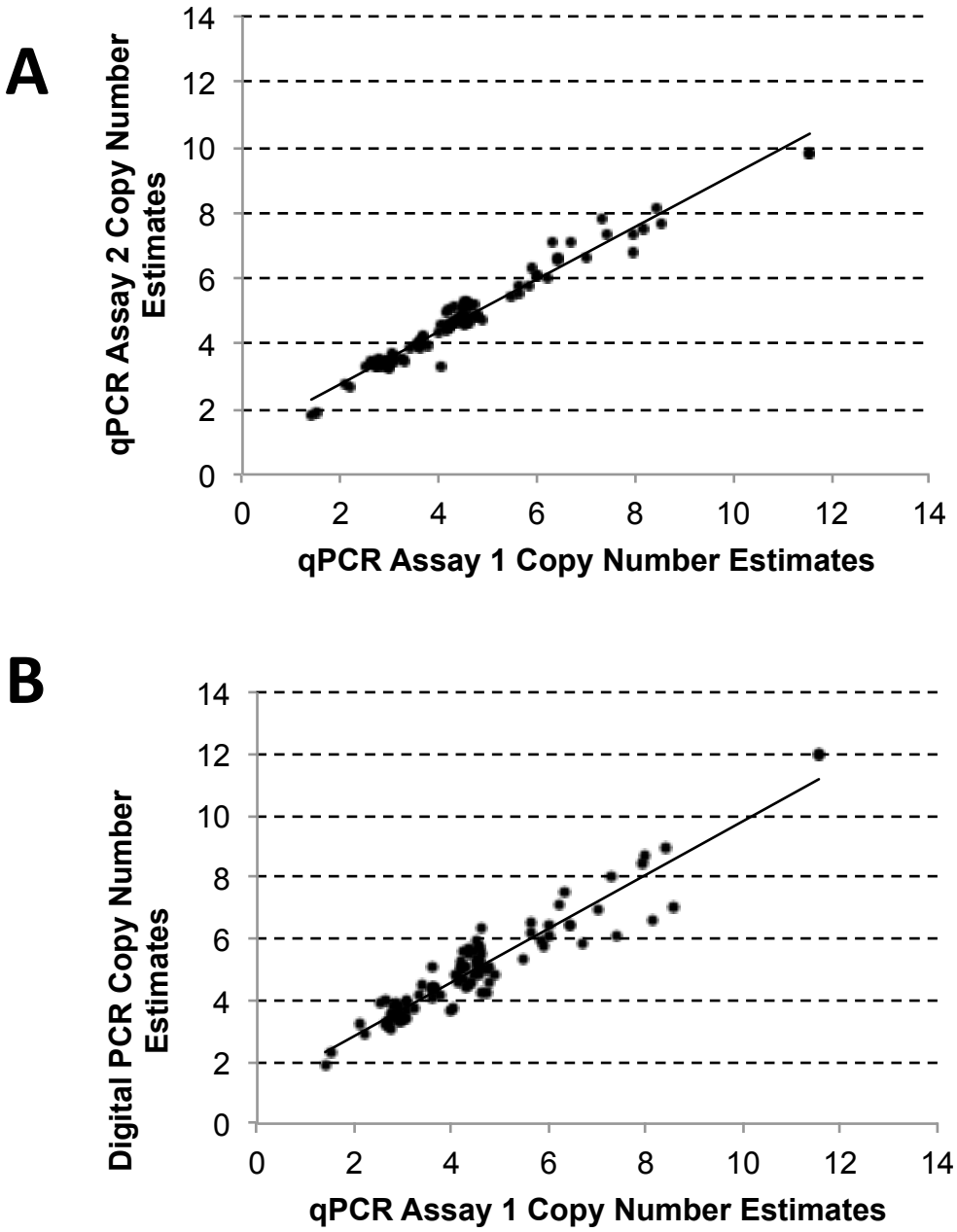
Supplementary Figure 6: 1000 Genomes whole genome shotgun-based total copy-number estimates *versus* qPCR based copy-number estimates for each of the amylase paralogs. Highly homologous *AMY1A1*, *AMY1A2*, *AMY1A3, AMY2A* and *AMYP* are strongly correlated with the qPCR based copy-number estimates while the more divergent *AMY2B* paralog is not.



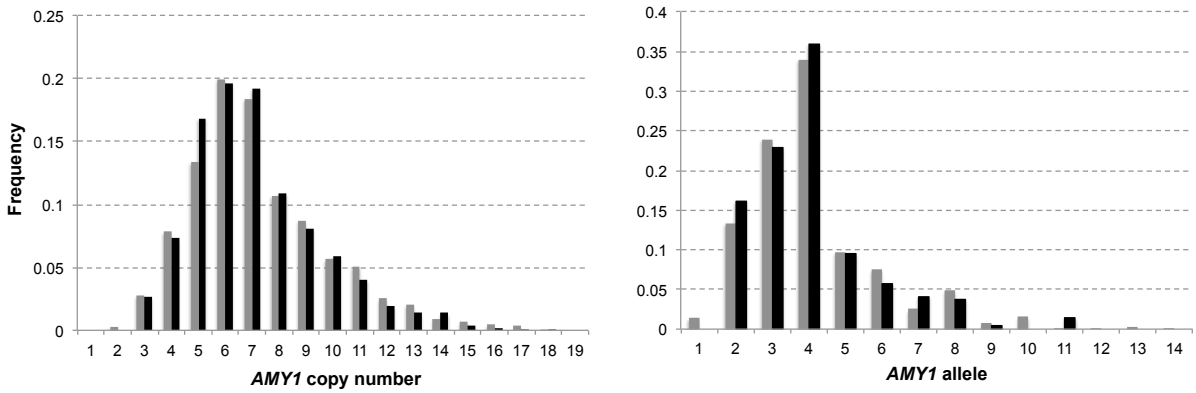
Supplementary Figure 7: 1000 Genomes whole genome shotgun-based paralog specific copy-number estimates for 34 individuals are plotted versus the qPCR based total *AMY1* copy-number estimates. As expected *AMY2A* and *AMY2B* exhibited little correlation with *AMY1* copy-number estimated by qPCR, while variation correlated with the total *AMY1* qPCR copy-number estimates was observed for all the *AMY1* paralogs. Within these 34 individuals *AMY1A2* was the most variable paralog.



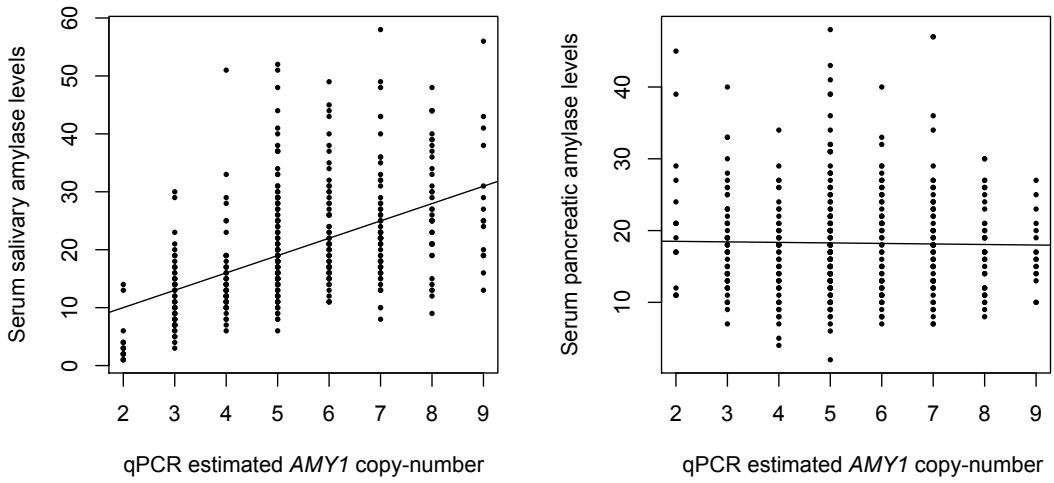
Supplementary Figure 8: The sum of read-depth based paralog specific copy-number estimates of each of the *AMY1* paralogs dervied from 1000 Genomes whole genome shotgun sequencing data is plotted versus the total *AMY1* copy-number estimated by qPCR (r=0.59, P=2.71 x 10^{-4}).



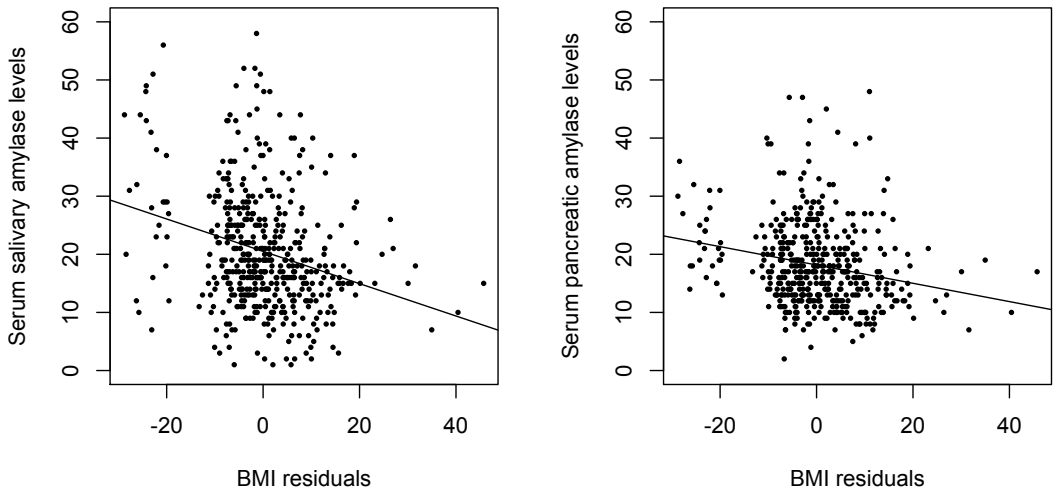
Supplementary Figure 9: A: total copy-number estimates of the *AMY1* gene assessed by qPCR using an alternative assay (assay 2) versus copy-numbers estimated at *AMY1* by the qPCR assay used in this study (assay 1) ($r = 0.98$, $P < 2.20 \times 10^{-16}$). B: total copy-number estimates of the *AMY1* gene assessed by digital PCR versus copy-numbers estimated at *AMY1* by the qPCR assay used in this study. The analyses show a strong correlation between the two methods ($r = 0.95$, $P < 2.20 \times 10^{-16}$).



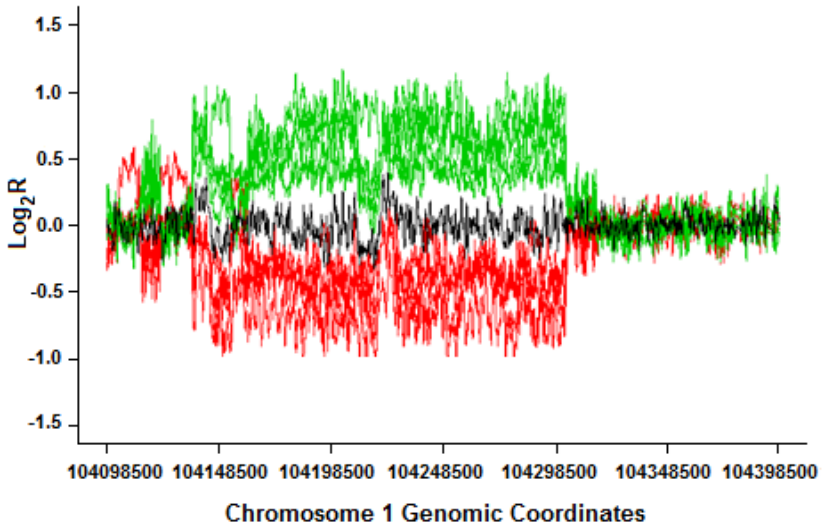
Supplementary Figure 10: Copy-number and allele frequency distributions at *AMY1* in TwinsUK (black bars) and DESIR (grey bars). Estimates in TwinsUK were calculated using a random unrelated subsample. Allele frequencies were estimated through the EM algorithm implemented in CoNVEM[26] (http://apps.biocompute.org.uk/convem/). See also **Table 2 and Supplementary Tables 4 and 5**.



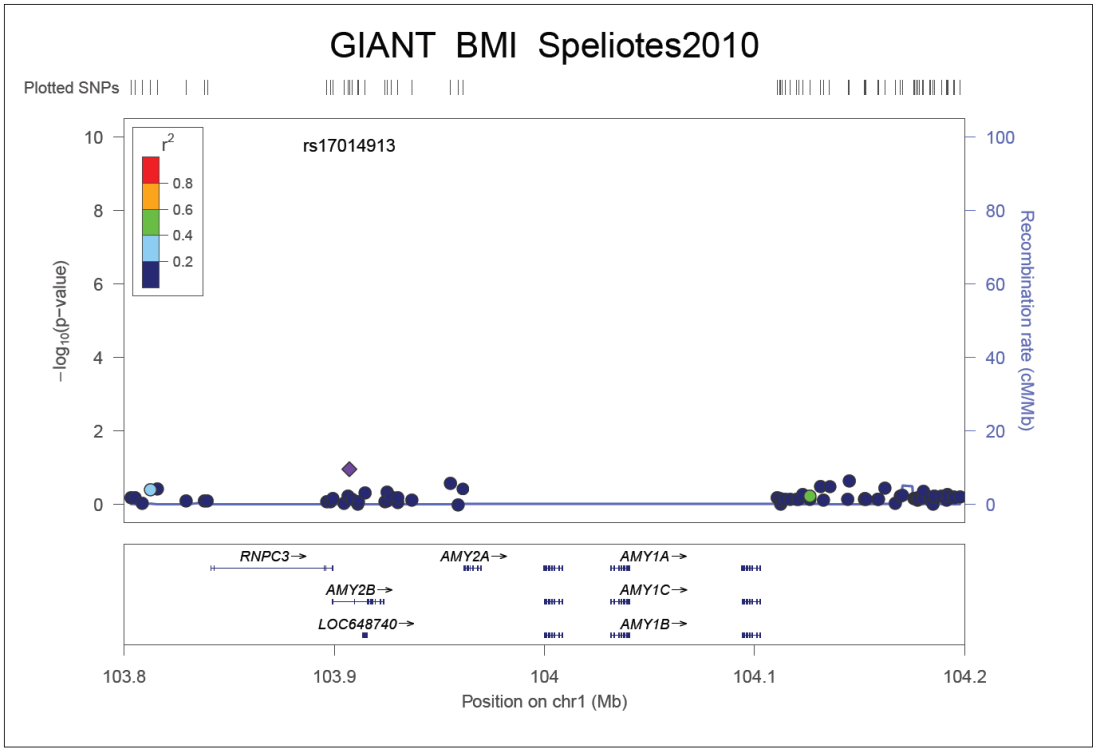
Supplementary Figure 11: Salivary (left) and pancreatic (right) serum amylase levels (IU/L) against estimated *AMY1* copy-number distribution inferred through qPCR for 468 subjects of the ABOS sample.



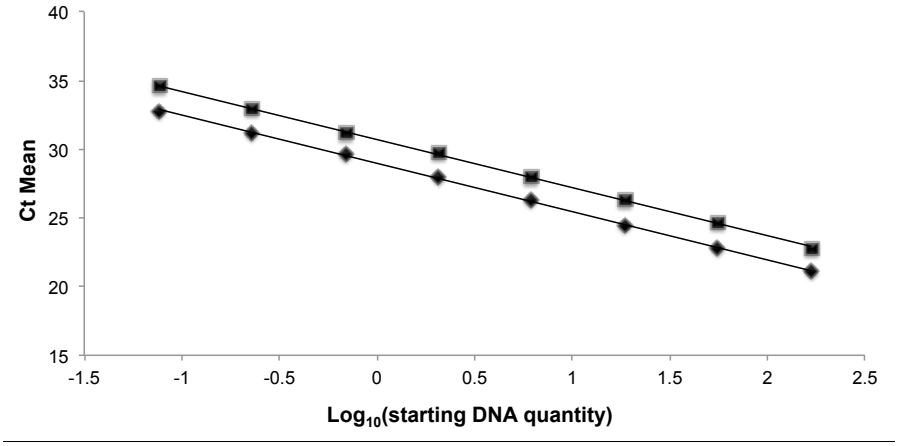
Supplementary Figure 12: Age and sex adjusted BMI residuals against salivary and pancreatic amylase levels in serum (IU/L) for 468 subjects of the ABOS sample.



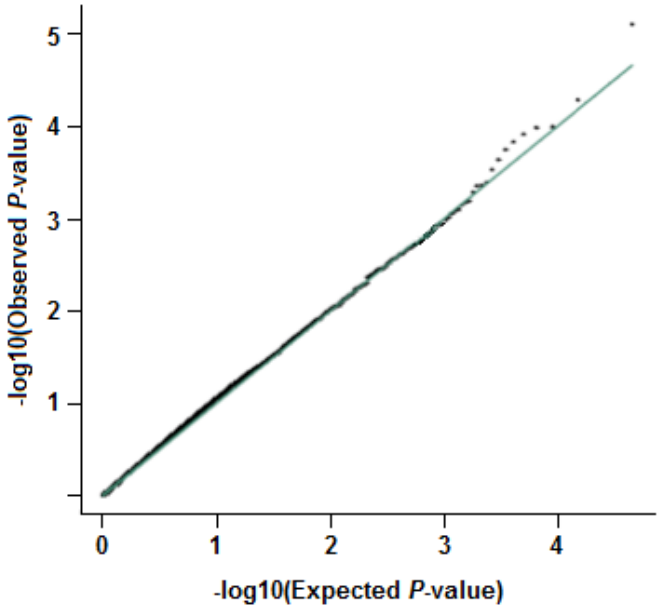
Supplementary Figure 13: The multi-allelic CNV region on chromosome 1 overlapping the amylase gene cluster in aCGH data from Conrad *et al* (2010)[17]. An eleven-point scrolling average is plotted for samples in the top and bottom 10% of the $\log_2$ ratio distribution. Samples with copy-number loss and gain relative to the reference sample are depicted in red and green, respectively. A sample showing no difference in copy-number from the reference sample is plotted in black. The boundaries of this region, as reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/) by the authors are chr1:104,136,714-104,317,018. The region encompasses *AMY2A*, and *AMY1A/B/C*.



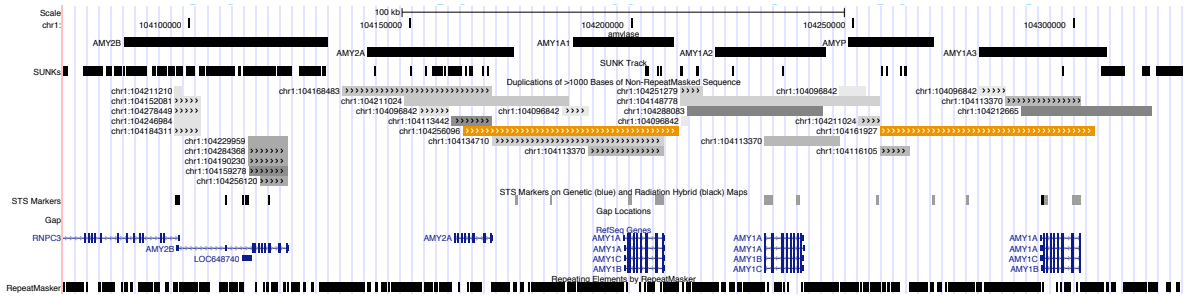
Supplementary Figure 14: Regional Manhattan plot of SNP association results in the amylase region from the latest BMI meta-analysis conducted by the GIANT consortium (Speliotes *et al*, 2010). X-axis: Chromosome 1 genomic coordinates. Y-axis: minus $\log_{10}$ (*P* value). Plot generated using LocusZoom[2].



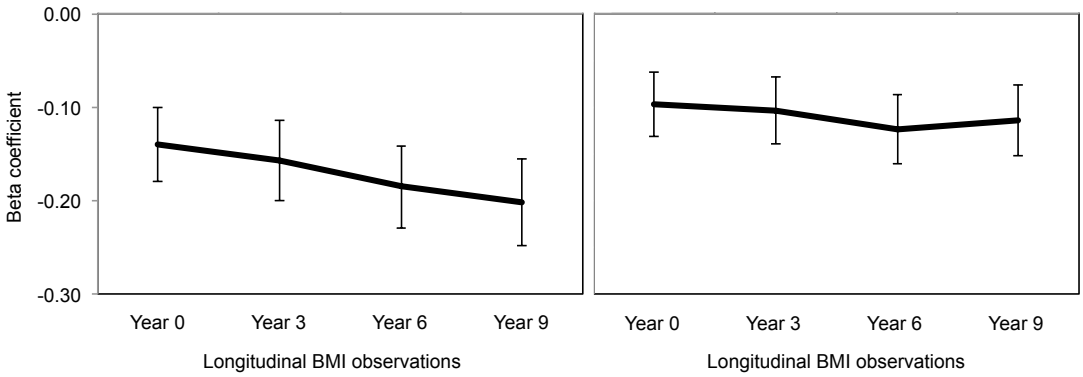
Supplementary Figure 15: PCR efficiency for the FAM (squares) copy-number target (*AMY1*) and VIC (diamonds) reference (*RNAse P*) assays. Assays showed comparable PCR efficiencies of 92.46% and 93.17%, respectively. X-axis: Log_{10}(starting DNA quantity) of the DNA template. Y-axis: Mean threshold cycle (Ct).



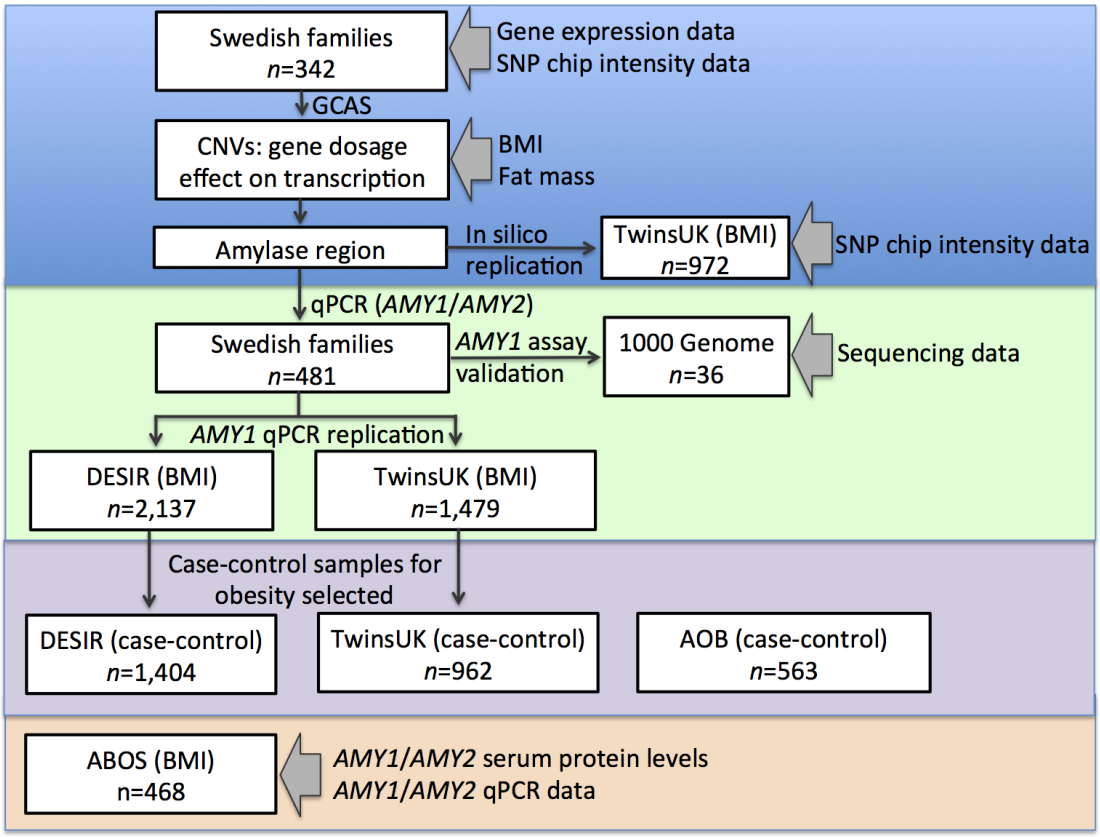
Supplementary Figure 16: Q-Q plots of the association results of BMI with 21,877 DNA-array CNV probes in the TwinsUK sample. The cnvi0022844 probe contained in the amylase region and found to be associated with BMI in this sample was ranked among the top 32 hits.



Supplementary Figure 17: Locations of paralogs in the amylase region. The loci chosen for the copy-number prediction from whole genome-shotgun sequencing in a panel of 36 individuals sequenced as part of the 1000 Genomes Project were selected to span each of the paralogs (on the top of the figure, from left to right: *AMY2B*, *AMY2A*, *AMY1A1*, *AMY1A2*, *AMYP*, *AMY1A3*). Regions were targeted to maximize the number of SUNKs.



Supplementary Figure 18: Effect of qPCR-inferred *AMY1* copy-number on BMI in females (left) and males (right) at different time points in the DESIR cohort, namely at recruitment (Year 0) and after 3, 6, and 9 years.



Supplementary Figure 19: Flow chart of experimental design

Supplementary Note

1. Study cohorts

Summary information for all samples included in this study is presented in **Table 1**. The studies were approved by the relevant institutional review boards in Sweden, UK, and France, and were conducted according to the principles of the Declaration of Helsinki. Written informed consent was obtained from every participant in each study.

1.1. Swedish families

The discovery sample included 154 Swedish families (651 subjects) ascertained through the presence of an obesity-discordant sib-pair in each family (body mass index [BMI] difference > 10 kg/m^2)[3]. Gene expression and DNA-array data were available for 149 families (342 subjects). Average family size was 4.34. Gene expression was measured in subcutaneous adipose tissue (SAT) using the Affymetrix Human Genome U133 Plus 2.0 microarrays. DNA from peripheral blood was genotyped in the same set of samples using the Illumina 610K-Quad array. Normalized intensity data were extracted using the Illumina GenomeStudio software, specifically the log R ratio (LRR), representing the total fluorescent intensity signals, and the B Allele frequency (BAF), representing the proportion of the total signal deriving from the B allele. Normalised intensity data from a total of 348,150 probes were included in our GCAS analysis. The single nucleotide polymorphisms (SNPs) in the amylase gene cluster region (n=12) were all in Hardy Weinberg equilibrium (HWE). Sample homogeneity was verified through Principal Component Analysis (PCA) using genotype data for 28,731 unlinked SNPs. Copy-numbers at *AMY1* and

AMY2 were estimated by quantitative real-time PCR (qPCR) for 481 subjects with complete data on both BMI and dual-energy X-ray absorptiometry (DEXA)-derived fat mass (**see Section 2**). Samples were plated in triplicate. To avoid spurious associations resulting from possible batch effects the samples were plated to ensure equal BMI distribution across plates (*P* value from Kruskal-Wallis test = 0.13).

1.2. TwinsUK

The St. Thomas' UK adult twin registry (TwinsUK) cohort is unselected for any disease or medical condition and is representative of the general UK population[4]. All subjects were recruited through national media campaigns and historically have included mostly females. To analyse signal intensity data from genome-wide SNP arrays in this sample, we selected the largest available uniform dataset of genotyping data from the TwinsUK cohort generated using the same genotyping platform, which consisted of 972 female subjects genotyped on the Illumina 610K-Quad array (Illumina)[5]. GWAS genotyping in the TwinsUK cohort was carried out in several stages over the years using different DNA-arrays, a number of which did not contain probes within the *AMY1* gene. Even for those chips containing probes within *AMY1*, while genotype calls can safely be combined from different GWAS batches, the same is not true for raw DNA-array intensity data. This is due to the fact that experimental batch effects between samples genotyped at different stages (and on different DNA-array chips) can be highly challenging to remove from such data, thus hampering CNV association studies. In order to avoid batch effects resulting from experimental differences both at the genotyping plate and at the experiment-level, we therefore focused our analysis of Illumina data in the TwinsUK cohort on the largest batch of data generated in a single genotyping experiment for this study sample (a total of 972

samples). The remaining genotyping experiments either included small numbers of subjects per-experiment and were genotyped on different chips, or were genotyped on the Illumina Hap300 chip which did not contain probes within the amylase region.

Intensity signals from 35 probes encompassing the amylase cluster region were included in our analyses. Normalised signal intensity data were extracted using the Illumina GenomeStudio software.

Replication of the association with BMI was carried out using 1,479 independent female samples with normal fasting glucose levels (<6.1 mmol/l) from the TwinsUK cohort.

DNA quality has repeatedly been shown to affect copy-number estimation from qPCR data[6-8], and we focused this experiment on those subjects collected after 2001, due to the fact that different extraction methods had been used in the previous years (Phenol-chloroform, salting out methods) and extraction was sometimes outsourced. Beginning in 2001, genomic DNA for the TwinsUK cohort was extracted in-house using a Nucleon BACC3 kit (Nucleon Biosciences, Coatbridge, UK). *AMY1* copy-number was inferred by qPCR as described in **Section 2**. Samples were plated in sextuplicate ensuring BMI homogeneity across plates, to avoid potential batch effects resulting from correlation between BMI and sample plate (*P*-value from Kruskal-Wallis test = 0.70).

1.3. DESIR

The Data from the Epidemiological Study on the Insulin Resistance syndrome (DESIR) study is a longitudinal study recruited through the French national social security system in western-central France between 1994-1996[9]. Sample homogeneity in DESIR was verified through PCA analysis based on genotyping data for 39,000

SNPs measured using the Illumina MetaboChip assay (data not shown). BMI association analysis in the DESIR sample was carried out using 2,137 normo-glycaemic subjects (fasting glucose levels <6.1 mmol/l) and unlikely to be affected by metabolic disorders (based on drug prescription records), representative of the general French population. To ensure the best DNA quality, we freshly extracted the DNA from the year-6 time-point of this longitudinal cohort. *AMY1* copy-number was inferred by qPCR as described in Section 2. In this sample, we opted to genotype each sample twice using four replicates each time (total = 8 measurements per sample). Copy-number concordance[10,11] between replicated plates was > 0.99 ($P < 2.0\ 10^{-16}$). We designed the plating taking care to avoid correlation between BMI and sample plate that could potentially generate spurious association (*P*-value from Kruskal-Wallis test = 0.26). Longitudinal BMI observations were available for 4 time points, at recruitment and after 3, 6, and 9 years.

1.4. French adult obesity cases (AOB)

The adult obesity study is fully described elsewhere[12]. Copy-numbers at *AMY1* were estimated by qPCR for 205 obese cases (BMI ≥ 30kg/m^2) and 358 age-matched controls (BMI < 25 kg/m^2) (**see Section 2**). Subjects were selected to have age not greater than 45 years and plasma fasting glucose levels less than 6.1 mmol/l. Samples were plated in triplicate. To avoid potential batch effects cases and controls were randomly mixed in each plate in equal proportion across all plates. Dixon's Q test was not significant for both the number of cases ($P = 0.36$) and controls ($P = 0.90$) and for the case/control ratio ($P = 0.21$) across plates.

1.5. ABOS

ABOS (Atlas Biologique de l'Obésité Sévère) is an on-going French cohort study (ClinicalTrials.gov Identifier: NCT01129297) conducted by the Département de Chirurgie Générale et Endocrinienne (Lille CHRU). The vast majority of the ABOS participants were severely obese and all were candidates for bariatric surgery. Copy-number inference at *AMY1* and *AMY2* through qPCR (**see Section 2**) was successful for 468 subjects, of whom 441 were obese with BMI $\geq$ 30 kg/m^2. Samples were plated in triplicate. We designed the plating taking care to avoid correlation between BMI and sample plate that could potentially result in spurious associations (*P*-value from Kruskal-Wallis test = 0.14).

Serum pancreatic and total amylase levels for 468 patients were measured by an enzymatic colorimetric assay with an autoanalyzer (CoBAS Icobas® 8000 modular analyser series; kits AMYL2-03183742122 and AMY-P-20766623322, Hoffman-La Roche Ltd). Serum salivary amylase levels were calculated by subtracting serum pancreatic amylase levels from total serum amylase levels.

1.6 Singapore Prospective Study Program (SP2)

SP2[13] is a population-based, cross-sectional study of 2,431 adult Singaporean subjects (Singaporean Chinese, Malays and Asian-Indians), aged between 24 and 95 years and this study that was specifically designed to examine the pathogenesis of cardiovascular and metabolic diseases. All individuals provided informed consent. Anthropometric measurements were performed by trained professionals at the clinics. Weight and height were measured in light clothing with no shoes. BMI was calculated as weight (in kg) over height (in m) squared. For this study we analysed 136 obese (BMI $\geq$ 28 kg/m^2), 197 overweight (23 $\leq$ BMI $<$ 28 kg/m^2), and 325

normal-weight controls with BMI < 23 kg/m^2 from the 2,431 adult Singaporean Chinese individuals included in SP2. *AMY1* copy number was measured by qPCR in all 658 subjects.

2. Copy-number estimation by quantitative real-time PCR

AMY1 and *AMY2A* gene copy-number were estimated by duplex quantitative real-time PCR (qPCR) on an Applied Biosystems 7900HT Real-Time PCR System, with Sequence Detection Software (SDS) version 2.3. Duplex reactions consisted of two assays, each with two primers and a TaqMan probe (Life Technologies); one specific for the target, *AMY1* (Hs07226362_cn) and *AMY2A* (Hs04204136_cn), and one specific for the reference (*RNase P*). The primers and probe of the Taqman qPCR assay employed in our analyses specifically target a region within exon 1 of the *AMY1* gene, which is absent in the *AMYP1* pseudogene, therefore ensuring specificity of the qPCR assay for *AMY1*. Values extracted following analysis with the SDS software were analysed with CopyCaller v1.0 software (Applied Biosystems), using analysis settings selected as recommended in the user guide. Diploid copy-numbers were estimated by ΔΔCt method using the reference DNA sample NA18972 (Coriell Cell Repositories, Camden, NJ), which carries 14 copies of the *AMY1* gene[14] and 2 copies of *AMY2*[15], and which was also used for between-experiment normalization. The sample NA18972 was specifically selected as a reference sample due to the fact that its copy number had been independently measured and confirmed in several published studies using methods including fiberFISH, qPCR and next-generation sequence analysis[14,15]. *AMY1* copy-number inference in both the TwinsUK and in the DESIR population sample was performed using an additional reference sample NA18956 which had been determined to carry 6 copies of the *AMY1* gene both by Perry *et al*[14]

and through the analysis of whole genome-shotgun sequencing data from the 1,000 Genomes Project[15] (**see Section 3.7**).

The ΔΔCt method assumes equal amplification efficiency for the target and the reference genes. PCR efficiency was evaluated for the target *AMY1* assay and the reference *RNase P* assay using a standard curve of serial dilutions of a DNA sample of known concentration. Log_{10} of the dilution factor was plotted against Ct mean values for each of the two assays. The slope of the line was then used to calculate PCR efficiency for each assay as follows:

$E = 10^{(-1/slope)}$

PCR Efficiency (%) = (E – 1) x 100

PCR efficiency for the target (*AMY1*) and reference (*RNase P*) assays was 92.46% and 93.17%, respectively (**Supplementary Figure 15**).

We further compared copy-number at the *AMY1* gene predicted using the qPCR assay to those predicted from whole-genome shotgun sequencing (**section 3.7** of this document), as well as by digital PCR and using an alternative qPCR assay (**section 3.8** of this document), showing strong correlation between the qPCR copy number measurements and those generated using each of the two other methods.

2.1. Quality control of large-scale population qPCR data

While quality control procedures have been standardized and are globally accepted for large genetic epidemiology studies such as GWAS, few large-scale qPCR-based CNV studies have been carried out to date. The quality control measures applied to qPCR data are aimed at identifying outliers for ΔCt values and their standard deviation among technical replicates (ΔCtSd) that can reflect differential DNA qualities or other possible technical issues. Of particular importance are those errors

affecting the calibrator sample, which in the ΔΔCt method is used to predict the copy-number state for all samples present in the same plate, and are thus likely to have an even stronger impact on the association study through their potential to induce batch effects in the data. Therefore, we designed our experiments to include two different calibrators on each plate on the two different tails of the copy-number distribution (see the previous section), both used for copy-number estimation to allow two additional quality control checks. At the plate level, we removed potential random fluctuations in the calibrators' ΔCt assignments (which may alter copy-number prediction for all samples in the same plate) by filtering out plates on the lowest 5% of the concordance correlation coefficient distribution calculated between copy-number states inferred by each of the two calibrators using the epiR R library[10,11]. At the sample level we then further removed those individual samples whose copy-number prediction was discordant for each of the two calibrators.

Since the DESIR experiment was designed using duplicate plates, the concordance was assessed between duplicate plates, and the predicted copy-number state was calculated by averaging the predictions between duplicate plates.

Furthermore, we assessed whether the use of integer predicted copy-numbers derived from non-integer calculated copy-numbers could in any way be resulting in spurious associations arising from rounding differences which could be cryptically associated with the phenotypic outcome. We tested each dataset (for both the BMI and obesity analyses) for association between the differences between exact and integer copy-numbers (referred to as residuals by Hollox *et al*[6]) and the phenotypic outcome (BMI or obesity). No association was observed between these copy-number 'residuals' and the phenotypic outcome in any of our datasets (association results are reported in each analysis section of the **Supplementary Note**).

3. Statistical analyses

Unless indicated otherwise, statistical analyses were carried out using R version 2.13.1[16]. All genomic coordinates are based on the February 2009 assembly of the human genome (hg19/GRCh37).

3.1. Gene-centric CNV association analysis

Gene-centric association (GCAS) was carried out using the linear mixed effects model implemented in famCNV, modelling gene expression levels in adipose tissue against genotype signal intensities[17]. famCNV, rather than inferring the discrete number of copies of a region, uses the continuous distribution of the intensity signals from SNP arrays to model the correlation with quantitative phenotypes in families[17]. As quantitative traits, we selected 29,546 transcripts (corresponding to 16,563 Ensembl genes) with known position in the genome and specificity > 70%. For each transcript, we focused the analyses on those probes lying within the transcript plus 30kb upstream and downstream, to encompass the coding regions and their internal and nearby regulatory regions (348,150 probes). The median (1st-3rd quartile) number of markers tested per transcript was 19 (12–33). Age and sex were included as covariates in the analyses. We applied the genomic control method to correct for inflation of the test statistics. Association *P* values were corrected for multiple testing using the False Discovery Rate (FDR) q-value method[18] in the gene-centric association analysis. See **Supplementary Table 1, Figure 1, Supplementary Figure 1**.

3.2. aCGH data exploration

High-resolution array comparative genome hybridization (aCGH) data for 40 individuals[19] were obtained from the Wellcome Trust Sanger Institute, U.K.: (http://www.sanger.ac.uk/cgi-bin/humgen/cnv/42mio/downloadBigDB.cgi).

An eleven-point scrolling average was calculated for $\log_2$ ratios for probes between 104.0Mb and 104.4Mb, encompassing a CNV region overlapping the amylase gene cluster. Samples in the top and bottom deciles of the $\log_2$ ratio distribution were plotted to visualize the CNV boundaries at this locus (**Supplementary Figure 13**).

3.3. CNV analysis of BMI and fat mass in the Swedish sample

Probes identified in the GCAS were tested in the Swedish sample against BMI and for a subgroup of 331 subjects with dual-energy X-ray absorptiometry (DEXA)-derived fat mass data using the mixed linear model implemented in famCNV. We selected, among the 131 SNP-chip probes showing significant association at 10% false discovery rate (FDR) in the GCAS (**Supplementary Table 1**), a subgroup of 81 probes located in genes where at least two associations between CNV and gene expression in adipose tissue were detected. Age and sex were included as covariates. Wilcoxon tests were assessed on BMI, fat mass, and amylase gene expression residuals, adjusted for age, sex and family clustering, between the top and bottom deciles of the LRR distributions for the associated probe cnvi0020639 (**Supplementary Figure 3**).

3.4. Joint analysis of the GCAS and each of BMI and fat mass

To assess the significance and ranking of our finding at the genome-wide scale, we performed joint association analysis for both gene expression levels in adipose tissue

and each of BMI and fat mass over the whole set of 348,150 Illumina probes used in the GCAS analysis. We first carried out CNV association analyses with famCNV for each of BMI and fat mass in the Swedish sample using the entire set of Illumina probes. We then combined the results from each of the two analyses with the GCAS results to assess the joint association at each probe with both expression levels of the overlapped gene and each of BMI and fat mass. We applied the method described by Brown[20], to account for the correlation between gene expression levels and each of the two phenotypes. Brown's method[20] provides an extension to Fisher's method when the variables, and therefore the tests, are not jointly independent. This method aggregates multiple dependent tests by assuming a multivariate normal distribution of the traits (and therefore of the tests) with a known covariance structure, and is appropriate in our study where one-sided likelihood ratio tests of association (such as those assessed by famCNV) are combined (**Supplementary Figure 2**).

3.5. BMI copy-number association analyses with LRR data in TwinsUK

Association analysis of LRR data from the Illumina platform with BMI was carried out in the TwinsUK cohort in a linear mixed effects model framework, including family structure and genotyping plate as random effects. LRR and BMI values were averaged within monozygous twin pairs to reduce measurement errors and environmental effects. Age and sex were included as covariates. Multiple testing was corrected for using Bonferroni correction. To better refine the CNV localization by weighting information provided by each probe, the association tests were also assessed using the first principal component calculated from the intensity data of consecutive probes in the region of interest using a sliding window of variable size ranging from a minimum of 2 probes to encompassing all probes across the region.

Difference in BMI between the top and bottom deciles of the LRR distributions for the associated cnvi0022844 probe was assessed using the Wilcoxon test. We investigated any potential bias of the association with the cnvi0022844 CNV probe by carrying out genome-wide association in the TwinsUK sample using all CNV probes on the Illumina 610k-Quad array (*n*=21,887) (**Supplementary Figure 16**).

3.6. Association of qPCR-estimated copy-number with quantitative traits in the Swedish sample

Estimated copy-number at *AMY1* and *AMY2* was inferred by qPCR for 481 subjects (478 for *AMY2*) from the Swedish families for whom complete data on both DEXA fat mass and BMI were available (236 parents and 242 siblings). A linear mixed effects model, implemented in the *lmekin* function from the R *kinship* package[21], was used to carry out association analysis between copy-numbers at each of *AMY1* and *AMY2* with each of BMI and fat mass, including the kinship matrix, calculated among all subjects, in the random effects to correct for relatedness among family members. Sex and age were included as covariates. Copy numbers included in the association analyses were restricted to those observed at a frequency above 1% in our study sample. Results are shown in **Supplementary Table 2**.

3.7. Amylase copy-number estimation by whole genome sequencing

To validate the amylase qPCR-based CNV assay, we compared total *AMY1* copy-number estimated by qPCR to those estimated from whole-genome shotgun sequencing in a panel of individuals sequenced as part of the 1000 Genomes Project Phase I[22]. We focused on a diverse panel of 36 individuals sequenced on the Illumina platform. Two individuals, namely NA18972 and NA18956 were used as calibrators.

Specifically, we employed two approaches, total copy-number genotyping, which estimates the total copy-number of all of the genes within the family, and paralog-specific SUN (singly unique nucleotide) based genotyping, across the six *AMY* paralogs (including one pseudogenized copy) in the reference genome[15,23] (**Supplementary Table 3; Supplementary Figure 5**).

The total sequencing coverage of these individuals ranged from 2.6×-14.6× coverage (median, 5.23× coverage). Whole-genome Illumina shotgun sequenced reads from individuals were mapped to the repeat-masked human reference genome (Build hg19/GRCh37) using the *mrsFASTc* alignment software[24] and normalized for GC content. A calibration curve was then constructed over regions of known copy-number and was used to estimate copy-numbers at additional loci in each individual. More details can be found in the cited references.

Total copy-number estimation across the *AMY1A1* gene demonstrated that the qPCR and sequencing-based copy-numbers were highly concordant, with a correlation of r=0.94 between the two orthologous methods (**Supplementary Figure 5**). As the total copy-number genotyping method is estimating the aggregate copy-number of all members of the gene family, similarly high correlations were observed over each of the highly homologous paralogs *AMY1A2* and *AMY1A3* and *AMY2A* with r=0.95, r=0.96, r=0.94 respectively. *AMY2B*, which is a more ancient copy of amylase and further diverged from the other copies, showed no correlation with the qPCR assay (r=-0.19; **Supplementary Figure 6**). The concordance of these two independent methods demonstrates that each is able to estimate copy-number over a large dynamic range and provides validation for the *AMY1* qPCR assay.

To determine which of the individual paralogs was varying (**Supplementary Figure 17**), we next performed SUNK-based paralog-specific inference across each of the

amylase copies (**Supplementary Figure 7**). Paralog specificity was achieved by considering reads only mapping to singly unique k-mers (SUNKs), which are sequence k-mers that are unique within the genome and thus specifically tag a location. As the individual *AMY* paralogs are highly homologous, the number of SUNKs tagging the individual paralogs is few and thus the results are less precise. For example, we identify only 25 SUNKs for *AMY1A2*, limiting the number of reads that can be used to accurately infer copy number (**Supplementary Table 3**). However, while the qPCR-based assay is specific to *AMY1*, the SUNK based copy-number estimates were able to distinguish between each of *AMY1A1*, *AMY1A2* and *AMY1A3* in addition to the *AMYP* pseudogene. As expected, there was no correlation between either the *AMY2A* or the *AMY2B* paralogs with the *AMY1* qPCR assay (P>0.05) (**Supplementary Figure 7**). Additionally, little copy-number variation was observed among these 34 individuals in the *AMY2* paralogs. The correlation between the qPCR assay and the total copy-number of all the *AMY1* paralogs estimated using the SUNK-based method was 0.59 (P=2.71 x 10^{-3}) (**Supplementary Figure 8**), providing further validation for the qPCR assay. Additionally, the SUNK-based copy-number analysis demonstrated that among these 34 individuals *AMY1A2* was the most variable paralog.

3.8. Amylase copy-number estimation by digital PCR (dPCR) and alternative qPCR assay

Digital PCR analysis was performed on the BioMark System (Fluidigm) on 96 samples from the DESIR cohort, using the qdPCR 37K Digital Array (Fluidigm), according to manufacturer's recommendations. This digital array consists of 48 panels, each containing 770 individual reaction chambers. Briefly, the final reaction mix for a digital panel comprised 1.2µL DNA (at 5ng/µL; this concentration was measured using the Qubit dsDNA HS Assay kit [Invitrogen]), 2µL TaqMan Gene Expression Master Mix (Applied Biosystems), 0.4µL 20x GE Sample Loading Reagent (Fluidigm), 0.2µL *AMY1* TaqMan assay (Hs07226362_cn; Applied Biosystems) and 0.2µL *RNase P* human TaqMan Copy Number Reference (Applied Biosystems). A total of 10µL of 1× GE Sample Loading Reagent (Fluidigm) with 4µL sample mix were aliquoted into each sample inlet on the digital array and were distributed throughout the partitions within each panel using an automated IFC controller-MX (Fluidigm). Each DNA sample was analysed in quadruplicate, with single replicate outliers removed. Inferred copy-numbers at *AMY1* showed high correlation between qPCR and digital PCR (r=0.95; P<2.20 x 10^{-16}; **Supplementary Figure 9**). We additionally measured copy numbers at *AMY1* using an alternative qPCR assay (Hs07226361_cn; Applied Biosystems) targeting a region within exon 1 of the *AMY1* gene, which is absent in the *AMYP1* pseudogene, therefore ensuring specificity of the qPCR assay for *AMY1*. The entire set of 96 DESIR samples was analysed in duplicate plates, with each sample analysed in triplicate on each of the duplicate plates (total = 6 measurements per sample). Inferred copy-numbers at *AMY1* were again highly correlated between our qPCR assay and the qPCR results using the alternative assay (r=0.98; P<2.20 x 10^{-16}; **Supplementary Figure 9**).

3.9. qPCR copy-number association analyses in the TwinsUK and DESIR population samples

To finely quantify at the highest resolution the relationship between BMI and estimated copy-number at the *AMY1* gene in the continuous BMI distribution of this unselected population sample, we carried out detailed BMI analyses in 1,479 female twins from the TwinsUK cohort and 2,137 subjects from the DESIR cohort. The case-control obesity association studies in these samples were carried out on the subset of samples from these cohorts with BMI ≥30 and <25 kg/m^2. In both the TwinsUK and DESIR samples, copy-number was predicted using two qPCR calibrator samples, one at each end of the copy-number distribution (NA18956 CN = 6 and NA18972 CN = 14). Estimated copy-numbers higher than 13 (showing frequency < 2.5%) were collapsed together into a single category. In TwinsUK, the association between either BMI or obesity and estimated integer copy-numbers at *AMY1* was tested using a mixed effects model with age as a covariate and including family and genotyping plate as random effects. In DESIR the association between either BMI or obesity and estimated integer copy-numbers at *AMY1* was tested using a mixed effects model with age and sex as covariates and including genotyping plate as a random effect. Gender-stratified analyses were also conducted where association between either BMI or obesity was assessed separately in males and females using a mixed effects model with age as a covariate and including genotyping plate as a random effect. Results are reported in **Table 2**, **Supplementary Tables 6-9** and **Section 3.9**.

In both studies, we also calculated empirical *P* values through 10,000 random permutations of the copy-number state across subjects while retaining the phenotypic information and observing the number of times (*n*) the observed association was equated or exceeded by chance. The empirical *P* value, calculated as (*n* + 1)/10,001,

was significant for both TwinsUK ($P = 5.00 \times 10^{-4}$) and DESIR ($P = 2.00 \times 10^{-4}$).

To determine whether rounding of the estimated copy-number to the closest integer value could be resulting in any spurious association, we carried out associations with both the residuals (*integer value* - *calculated value*) and its absolute value against both BMI and obesity. Associations were not significant for either BMI in TwinsUK ($P = 0.20$; $P_{absolute} = 0.25$) and DESIR ($P = 0.60$; $P_{absolute} = 0.79$) or for obesity in TwinsUK ($P = 0.12$; $P_{absolute} = 0.56$) and DESIR ($P = 0.46$; $P_{absolute} = 0.59$).

Additionally, inspired by the approach described by Field *et al*[25], we also conducted the association analyses using the alternative approach of applying unsupervised clustering (k-means) to divide the raw ΔCt signal into discrete categories, obtaining comparable results - correlation between inferred copy-number and raw signal discretized by k-mean clustering in TwinsUK was $r > 0.98$; $P < 2.2 \times 10^{-16}$ and in DESIR was $r > 0.95$; $P < 2.2 \times 10^{-16}$.

In both TwinsUK and DESIR, the association with obesity between carriers of low and high copy-numbers was carried out using a binary variable to categorize subjects below and above the first (≤3 copy-numbers) and last (≥10 copy-numbers) decile of the estimated copy-number distribution, respectively (**Supplementary Table 10**).

3.10. Longitudinal Analysis and Sex-Specific Effects in DESIR

Interaction between age and copy-number at *AMY1* in DESIR was carried out by comparing the fit of the null mixed models which included a random effect for intercept and slope where the individual BMI at the 4 time points was determined by the additive effects of age and estimated copy-numbers (both transformed into z-scores), and the full model which also included the age × copy-number interaction term (see also **Supplementary Table 16; Supplementary Figure 18**). Significant

interaction was observed between *AMY1* estimated copy-number and subject age, showing an increased effect of *AMY1* estimated copy-number on the individual's BMI throughout life. The interaction between subject age and *AMY1* estimated copy-number was significant only in females ($P_{females} = 5.38 \times 10^{-3}$, $P_{males} = 0.21$).

The longitudinal study between *AMY1* estimated copy-number and BMI comprised four times more observations (4 observations per individual) than the cross-sectional analysis. Given the limited sample size, the latter analysis was therefore slightly underpowered to detect both main effects and interactions with this highly polymorphic locus. Comparison of a simpler null model (no interaction between estimated copy-number and sex) of longitudinal association between *AMY1* estimated copy-number and BMI including both males and females (with sex as a covariate) against an alternative model that includes also the interaction term between estimated copy-number and sex, against an alternative model that includes also the interaction term between estimated copy-number and sex was highly significant ($P<2.20 \times 10^{-16}$). The longitudinal interaction between sex and estimated copy-number at *AMY1* was indeed significant, with $P = 1.71 \times 10^{-3}$. In summary, while a baseline effect was detected for BMI in males (cross-sectional $P = 1.92 \times 10^{-3}$), as also showed in **Supplementary Figure 18** BMI in females appeared to be more strongly affected by estimated copy-number at *AMY1* throughout life.

In our cross-sectional analyses, although there was no significant difference in the CNV distribution between male and female subjects ($P = 0.96$), when the meta-analysis of the DESIR and TwinsUK cohorts was carried out only in the female subsample (n=2,674), we found a stronger effect of estimated *AMY1* copy-number on BMI (for each additional copy of *AMY1*: β = -0.18 [0.03] kg/m²; $P = 9.38 \times 10^{-8}$; **Table 1)**.

Analogously, meta-analysis of the three TwinsUK, DESIR and AOB case-control studies (593 cases; 2,336 controls) suggested the existence of a stronger effect of estimated *AMY1* copy-number on obesity risk in females (per *AMY1* copy-number: OR = 1.22 [1.15-1.30]; $P = 8.87 \times 10^{-10}$ in females; OR = 1.12 [1.02-1.24]; $P = 1.86 \times 10^{-2}$ in males).

While it is essential to note that our cross-sectional study samples were underpowered to detect interaction effects between estimated *AMY1* copy-number and sex, our analyses suggest that potential sex-specific effects between estimated *AMY1* copy-number and adiposity may exist, and may be worthy of further assessment in larger replication samples.

3.11. CNV-SNP LD Analysis in the *AMY1* Region in the TwinsUK Sample

To better understand why the *AMY1* locus has not been identified by large scale BMI and obesity SNP meta-analyses, we explored in the TwinsUK sample whether variability in *AMY1* copy-number could be captured by surrounding SNPs located within an extended region surrounding the salivary amylase gene cluster.

In the TwinsUK cohort, genotype data for 54 genotyped and imputed SNPs located in a window of 250kb either side of the *AMY1* region (no SNPs located within the *AMY1* region itself) were available for a subset of 1,388 subjects.

As previously-described, total diploid copy-number measurement at *AMY1* does not distinguish how many copies are carried on each of the two chromosomes. We therefore estimated the r^2 between each SNP and *AMY1* by squaring the correlation coefficients between the SNP genotypes (recoded as 0, 1, 2) with (1) the predicted copy-numbers at *AMY1*, (2) genotypes at *AMY1* recoded as 0, 1 corresponding to copy-numbers above (≥) and below the median, and (3) genotypes at *AMY1* recoded

as 0, 1, 2 including the top 25%, median 50%, and lower 25% of the copy-number distribution, respectively. We showed LD between consecutive SNPs in this region calculated using both haplotype frequencies and correlation coefficients to be highly concordant (r>0.99) (**Supplementary Table 13**). The maximum r^2 observed between SNPs and copy-number at *AMY1* (either using the entire distribution or recoding it in a smaller number of classes) was 0.05, indicating negligible LD between SNPs and *AMY1* copy-numbers. Together with the paucity of SNPs in the amylase region included in SNP meta-analyses, this negligible LD between *AMY1* copy-number states and surrounding SNPs has likely prevented detection of association of the *AMY1* locus with adiposity in previous GWAS.

3.12. *FTO* association study

We aimed to contrast the effect of *AMY1* estimated copy-number on BMI variation with that of *FTO*, which is the locus showing the strongest, most replicated genetic association with BMI. We carried out BMI association analyses using all SNPs included in the haploblock on chromosome 16q12.2 (chr16: 53,797kb-53,850kb), where previously-reported associations with *FTO* have largely been identified[22]. Eleven genotyped and imputed SNPs were included in the TwinsUK cohort for a subset of 1,388 subjects genotyped on the Illumina Hap300 and 610K-Quad arrays[22]. The DESIR cohort has recently been genotyped with the Illumina MetaboChip and data for 104 SNPs were available for 2,132 subjects from this study.

For comparison, association between estimated copy-numbers at *AMY1* and BMI was assessed for the subgroups of subjects also genotyped at each tested SNP, and *AMY1* estimated copy-numbers were dichotomized into low (<6) or high (≥6) categories.

The most significant association in the *FTO* region in TwinsUK was observed at rs3751812 (β [*se*]= 0.45[0.18] kg/m²; $P = 1.44 \times 10^{-2}$) while *AMY1* estimated copy-number association for the same subset of subjects showed a level of significance two orders higher and a stronger effect (β [*se*] = -0.86 [0.25] kg/m²; $P = 5.52 \times 10^{-4}$; **Supplementary Table 14**). In DESIR the most significant association observed at the *FTO* locus was at the SNP rs3751813 (β [*se*] = -0.26[0.10] kg/m²; $P = 1.12 \times 10^{-2}$), 248bp from the most significant association observed in TwinsUK (rs3751812), while the *AMY1* association was $P = 5.38 \times 10^{-6}$ (β [*se*] = -0.64[0.14]) **(Supplementary Table 15**). The *FTO* SNP rs1558902 showed the most significant association with BMI variation in the largest GIANT meta-analysis[26]. In Twins UK and DESIR, SNP rs1558902 was significantly associated with increased BMI (β [*se*]= 0.39 [0.19] kg/m² and $P = 3.47 \times 10^{-2}$; and β [*se*]= 0.25 [0.10] kg/m², $P = 1.56 \times 10^{-2}$; respectively) while *AMY1* estimated copy-number association showed a much stronger level of significance and effect (β [*se*] = -0.86 [0.25] kg/m²; $P = 5.52 \times 10^{-4}$; and β [*se*] = -0.64 [0.14] kg/m²; $P = 5.38 \times 10^{-6}$; respectively; **Supplementary Tables 14-15**).

3.13. qPCR-estimated copy-number association analysis in AOB

Association analysis was carried out between cases and controls using a generalized mixed effects model including age and sex as covariates and genotyping plate as a random effect. Furthermore, association between carriers of low and high copy-numbers was also carried out using a binary variable to categorize subjects below and above the first (≤3 copy-numbers) and last (≥10 copy-numbers) decile of the estimated copy-number distribution, respectively.

To assess if any spurious association was resulting from rounding of the calculated copy-number to the closest integer value, we carried out associations with both the residuals (integer copy-number value – calculated copy-number value) and its absolute value against obesity. Association was not significant (P = 0.31; $P_{absolute}$ = 0.62).

3.14. Meta-analyses

Meta-analyses for both BMI and obesity were performed using the software program METAL[1] (www.sph.umich.edu/csg/abecasis/metal) which performs an inverse-variance method of meta-analysis with fixed effects, by combining effect sizes and weighting them by their standard errors (**Table 1** and **Supplementary Tables 10-11**).

3.15. Estimation of the proportion of genetic variance for BMI and obesity explained by copy-number variation in *AMY1*

We assessed the proportion of genetic variance for BMI explained by inferred copy-numbers at *AMY1*. We calculated correlation coefficients *r* between inferred copy-numbers and BMI, adjusted for age, sex (DESIR), qPCR plate, and family (TwinsUK) and transformed into Fisher *z-scores*:

$$\zeta = (0.5)\ln\left|\frac{1+r}{1-r}\right|$$

with derivation of their standard error as:

$$SE_{\zeta} = \sqrt{\frac{1}{n-3}}$$

where *n* is the number of samples in each study.

z-scores were combined using the inverse variance method, with the function *metagen* of the package *meta* in R. Combined *z-score* and its 95% confidence interval were then back-transformed to *r* values :

$$r = \frac{e^{2\xi} - 1}{e^{2\xi} + 1}$$

and squared to assess the BMI variance explained by variation in inferred copy-numbers at the *AMY1* gene.

Correlations in DESIR (n=2,137) and TwinsUK (n=1,479) were r=0.085 and r=0.111, respectively. Combined correlation was 0.1001 [95% CI = 0.0682; 0.1327; $P < 1.00 \times 10^{-4}$; Heterogeneity $P = 0.45$].

Assuming heritability for BMI between 40-70% the variance explained by inferred copy-numbers at *AMY1* lies between 0.66% and 4.40%.

The proportion of genetic variance for obesity explained (Vg) by inferred *AMY1* copy-numbers was estimated according to So *et al*[27] using applied by post-hoc binning of our case-control samples into low, intermediate and high categories of *AMY1* estimated copy-number. The population frequencies of copy-numbers were based on a weighted average of the frequencies in cases and controls, the weight being the prevalence of obese and non-obese in the population. Amylase effect on obesity between carriers of high estimated copy-numbers (greater than 7; 25% of each sample) and either carriers of the most common estimated copy-numbers of (between 5 and 7 representing 50% of each case control samples), and carriers of low estimated copy-numbers (less than 5; 25% of each sample) are reported in **Supplementary Table 11**. Prevalence of obesity was taken to be 16.85%[28]. The 95% confidence interval was estimated by a simulation approach, assuming a normal distribution of *ln*(OR) with *SE* as obtained from the generalized linear mixed model. Random normal variables were generated and the corresponding Vg calculated.

3.16. Analyses between qPCR-estimated copy-number, serum protein levels and BMI in ABOS

The relationship between copy-number at *AMY1* and *AMY2* inferred by qPCR and the salivary and pancreatic amylase levels measured in serum for 468 patients was assessed using multiple regression models with qPCR-typed copy-number at both *AMY1* and *AMY2* genes as predictors. The relationship between salivary and pancreatic amylase levels and BMI was assessed using multiple regression models with the protein levels as predictors. Age and sex were used as covariates. No significance gender differences between males and females were observed in either salivary (P = 0.13) or pancreatic amylase (P = 0.39) mean levels, nor estimated between copy-numbers at *AMY1* (P = 0.94) or *AMY2* (P = 0.54).

AMY1 copy-number distribution in obese subjects of the ABOS cohort was identical to the one observed in the AOB obese cases (Wilcoxon P = 0.98), however we did not include the ABOS sample in the case-control association as copy-number estimates in this sample were generated in a separate experiment.

3.17. *AMY1* expression levels in adipose tissue

Despite several historical reports of amylase enzyme activity in a number of different tissue types other than the salivary glands, including adipose tissue, its transcriptional levels in such tissues have not been well documented. In our Swedish family sample, amylase transcript levels (probeset 208498_s_at) in subcutaneous adipose tissue were ranked among the top 11% most-expressed transcripts.

To validate these results, we examined an independent, publicly-available gene expression dataset derived from subcutaneous adipose tissue of 856 subjects, generated using a different microarray platform and deposited in ArrayExpress

(experiment E-TABM-1140). Data were available for 48,638 probes measured on the IlluminaHT12 v3.0 array, including three probesets with high specificity for *AMY1* (ILMN_2294762, ILMN_1726327 and ILMN_1663313). These data confirmed strong expression of *AMY1* in adipose tissue, with the three probes ranked in the top 23%, 27% and 3% most-expressed genes (average across all three probesets ranked within top 13%).

3.18. Copy number measurement outlier analyses

In our study, we took particular care to design our experiments so as to ensure even BMI distribution across plates for copy number estimation, in order to avoid false positive associations that may arise through batch effects. However, we did observe some differences across plates in the *AMY1* copy number distribution, which are likely to be a consequence of the limitations in the available technologies for copy number estimation at such multi-allelic loci in large experiments.

Therefore, we conducted a multiple comparison test after Kruskal-Wallis using the function *kruskal* (R package *Agricolae*) in order to identify potential "outlier" plates in DESIR and TwinsUK, to exclude the possibility of a false-positive association driven by potential batch effects. By removing 13 plates (n = 587 subjects) from DESIR (Kruskal-Wallis $P < 0.05$) the association with BMI remained significant with β [se] -0.17[0.03]; $P = 5.14\times10^{-7}$. Analogously, by removing 12 plates (n = 444 subjects) from TwinsUK (Kruskal-Wallis $P < 0.05$) the association with BMI remained significant with β [se] -0.17[0.06]; $P = 5.85\times10^{-3}$.

References

1. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
2. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336-7.
3. Walley AJ, Jacobson P, Falchi M, et al. Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. Int J Obes (Lond) 2012;36:137-47.
4. Spector TD, Williams FM. The UK Adult Twin Registry (TwinsUK). Twin Res Hum Genet 2006;9:899-906.
5. Hysi PG, Young TL, Mackey DA, et al. A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. Nat Genet 2010;42:902-5.
6. Guo W, Jiang L, Bhasin S, Khan SM, Swerdlow RH. DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination. Mitochondrion 2009;9:261-5.
7. Fode P, Jespersgaard C, Hardwick RJ, et al. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PLoS ONE 2011;6:e16768.
8. Hollox EJ. Beta-defensins and Crohn's disease: confusion from counting copies. The American journal of gastroenterology 2010;105:360-2.
9. 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 & metabolism 1997;23:428-34.
10. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255-68.
11. Lin LI. A note on the concordance correlation coefficient. Biometrics 2000;56:324-5.
12. Meyre D, Delplanque J, Chevre JC, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet 2009;41:157-9.
13. Wen W, Cho YS, Zheng W, et al. Meta-analysis identifies common variants associated with body mass index in east Asians. Nat Genet 2012;44:307-11.
14. Perry GH, Dominy NJ, Claw KG, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet 2007;39:1256-60.
15. Sudmant PH, Kitzman JO, Antonacci F, et al. Diversity of human copy number variation and multicopy genes. Science 2010;330:641-6.
16. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. In.
17. Eleftherohorinou H, Andersson-Assarsson JC, Walters RG, et al. famCNV: copy number variant association for quantitative traits in families. Bioinformatics 2011;27:1873-5.
18. Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B (Methodological) 2002;64 part 3:479-98.
19. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. Nature 2010;464:704-12.
20. Brown MB. A Method for Combining Non-Independent, One-Sided Tests of Significance. Biometrics 1975;31:987-92.
21. Atkinson B, Therneau TM. Kinship: Mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees (Mayo Foundation for Medical Education and Research, Rochester, MN). 2008.
22. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-73.

23. Alkan C, Kidd JM, Marques-Bonet T, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 2009;41:1061-7.
24. Hach F, Hormozdiari F, Alkan C, Birol I, Eichler EE, Sahinalp SC. mrsFAST: a cache-oblivious algorithm for short-read mapping. Nat Methods 2010;7:576-7.
25. Field SF, Howson JM, Maier LM, et al. Experimental aspects of copy number variant assays at CCL3L1. Nat Med 2009;15:1115-7.
26. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937-48.
27. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 2011;35:310-7.
28. International Obesity Taskforce (2011). http://www.iaso.org/iotf/. In.