No evidence for association of body mass index with salivary amylase gene copy number in the UK 1958 Birth Cohort

Nzar A. A. Shwan[1] and John A. L. Armour*

School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK

[1] current address: Erbil Technical Health College, Erbil Polytechnic University, Hadi Chawshli Street, Kirkuk Road, Kurdistan Region, Iraq

*corresponding author

Keywords: Obesity; BMI; genetics; amylase

Running title: *AMY1* copy number and BMI

Contact information for corresponding author:

john.armour@nottingham.ac.uk, ORCiD 0000-0002-8558-8506
School of Life Sciences,
University of Nottingham,
Queen's Medical Centre,
Nottingham NG7 2UH, UK
Tel +44-115-8230308

Word count (main text): 3955

Disclosure: The authors declared no conflict of interest.

Author Contributions: J.A.L.A. conceived and designed the study, and coordinated access to the 1958 Birth Cohort data and DNA samples. N.A.A.S. used PRT to measure amylase copy number in genomic DNA samples. Both authors analysed CGH data, PRT data and association with BMI. Both authors contributed to all stages of manuscript preparation.

Study Importance Questions

What is already known about this subject?

- Gene copy number of salivary amylase (*AMY1*) has been reported to be associated with BMI
- Attempts to replicate the finding have led to inconsistent outcomes
- The strength of *AMY1*/BMI association may be age-dependent

What does your study add?

Our study adds
- Application to association tests and validation of new PRT methods to measure amylase copy numbers
- A well-powered specific test of the hypothesis of age-dependent association between *AMY1* copy number and BMI
- An additional independent and well-powered test of the general hypothesis of association (i.e., at all age points)

Abstract

Objective

In a 2014 publication, evidence was presented supporting association of body mass index (BMI) with copy number of the salivary amylase (*AMY1*) gene, with an unprecedented effect size of −0.15kg/m^2 (SE 0.02) per copy of *AMY1*. Most well-powered attempts to reproduce these findings have not been successful. However, because of different study designs, a significant association may still apply under restricted conditions, such as in particular age groups. We specifically tested BMI-*AMY1* association at different age points in the same individuals, using longitudinal BMI information from participants in the UK 1958 Birth Cohort study.

Methods

We measured *AMY1* copy number by paralogue ratio tests (PRT) in genomic DNA and using array-CGH data. BMI data from 1958 Birth Cohort subjects were available from eight different age points between 7 and 50 years old.

Results

We found no evidence, even at nominal significance, for association of amylase gene copy number with BMI at any age point in approximately 1400 members of the 1958 Birth Cohort, nor in 2835 people from two disease cohorts from the Wellcome Trust Case Control Consortium.

Conclusions

Our results do not support association between BMI and *AMY1* copy number at any age point.

Introduction

Obesity remains a critical public health problem in many populations, and numerous individual factors have been defined that contribute to the heritable component of variation in adiposity (most commonly represented by measurement of body mass index, BMI). In addition to relatively strong genetic factors, such as the well-characterised variation in the gene *FTO* (1), GWAS analyses of very large cohort size have defined increasing numbers of single nucleotide polymorphism (SNP) loci associated with variation in BMI (2).

By contrast with SNP diversity, gene copy number variation (CNV) poses practical difficulties for accurate measurement (3, 4), but may have strong functional consequences through gene dosage effects. Furthermore, while variation at most biallelic CNVs can be effectively captured in SNP data, SNPs often have very little predictive power for more complex multiallelic CNVs (4), so that direct typing of copy number is necessary for effective analysis of association. This extra dimension of variation was explored in an analysis of association between BMI and copy number of the *AMY1* (salivary amylase) gene; array-CGH and real-time PCR methods were used to analyse *AMY1* copy number in more than 6,000 adult subjects, demonstrating a significant association of low *AMY1* copy number with increased BMI (5). The study used adults, and drew representative population samples from the Twins UK (N = 1479, median age 53y) and DESIR (N = 2137, median age 52y) cohorts. Taken at face value, this effect (β = −0.15kg/m^2 per copy of *AMY1*) would easily be the most substantial heritable factor yet discovered in obesity.

In nearly all studies with best power to replicate Falchi *et al.*'s core finding of population-wide association between BMI and *AMY1* copy number (5), no significant

associations have been observed. However, because different studies use cohorts with different properties, and in some cases also test different hypotheses, it is unclear whether the proposed association may still apply in some contexts. A very well-powered study in three cohorts of European adults (6), totalling over 4,000 individuals and including people selected for being at the extremes of the BMI distribution, strongly excluded a major effect in the general population.

A study of *AMY1* copy number in 4800 nondiabetic adult subjects did not replicate the simple association found by Falchi *et al.* between *AMY1* CN and BMI (P=0.8) (7), but provided evidence for dependence of the relationship between *AMY1* copy number and BMI on dietary starch intake. Plasma enzyme activity was significantly associated both with *AMY1* gene copy number and with BMI in the well-characterised DESIR cohort, but association between BMI and *AMY1* copy number was only nominally significant (P=0.023) (8). In the same study, analysis of two further case-control cohorts for obesity, comparing 1179 affected adults with 2220 controls, and 785 affected children and adolescents with 712 controls, found a significant association of obesity with *AMY1* copy number (P = 6.8×10^{-3} overall), mostly driven by association (P = 7.1×10^{-3}) among the children and adolescents (8).

Studies specifically analysing obesity in children and young adults have yielded results apparently supporting the association of BMI with amylase copy number. A study comparing 293 Mexican children with obesity against 304 normal weight controls (mean age 9.5 years) found a highly significant association (9). Significant associations of *AMY1* with BMI were reported in Finnish females (N = 29, P = 0.045) but not males among young adults (mean ages 19-20 years) affected by early-onset obesity (10) and in boys (P = 0.033) but not in girls among 744 Italian children with mean age 8.4 years (11). In contrast, no evidence for association was observed in

over 1000 Chinese and Malay males aged 18-21, in which *AMY1* copy number was convincingly determined using droplet digital PCR (12).

Taking all these studies together, the general population-wide association claimed by Falchi *et al.* is not supported by independent replication, but neither is an age-restricted effect decisively excluded by all studies. Any genuine *AMY1* association operating in specific, restricted contexts could still have great potential power in both understanding aetiology and in guiding specific interventions in clinical management. An effect of *AMY1* variation on metabolism might be strongest at earlier ages, because of changes in carbohydrate intake and turnover during an individual's lifetime. We particularly wanted to clarify the possible age-dependence of *AMY1*-BMI associations, and for this purpose we exploited the potential of the British 1958 Birth Cohort (National Child Development Study, (13)), determining *AMY1* copy number both by direct measurement from genomic DNA and by custom analysis of existing array-CGH data. Samples from the 1958 Birth Cohort made up some of the unaffected control samples used in the Wellcome Trust Case-Control Consortium (WTCCC) studies (3, 14), from which CGH data were available. The 1958 Birth Cohort is a representative (unselected) population sample, and each 1958 Birth Cohort participant has multiple BMI measurements at intervals from birth to age 50, allowing longitudinal analysis of association between *AMY1* copy number and BMI.

Methods

Subjects and DNA Samples

The main association analyses described in this study used the subset of the UK 1958 Birth Cohort included as control subjects in the WTCCC CNV analysis (3). The UK 1958 Birth Cohort study initially sampled 17416 children born in the UK in the same week in 1958, and data sweeps including interviews and biometric measurements were conducted at ages 7, 11, 16, 23, 33, 42, 44 and 50 (13). Loss to follow-up meant that by age 50 there were data from 9790 participants. DNA was prepared from immortalised cell lines derived from blood taken at age 44, and DNA and array CGH data were available for our study from 1448 people. The use of this cohort was therefore intended as a representative population sample, with no enrichment of people with extreme values. We also included two disease cohort samples from the WTCCC CNV study for which BMI, age and sex data were available as well as raw array-CGH data; these were 1027 samples (425 females) from the type II diabetes (T2D) cohort and 1808 samples (366 females) from the Coronary Artery Disease (CAD) cohort. All samples were from UK individuals; in the WTCCC CNV study, 26 individuals had been removed on the basis of Principal Component Analysis of genotypes, or of self-reported non-European ancestry (3). Recruitment, measurement and blood sampling of these cohorts were carried out with local ethical approval in the original studies.

DNA samples from 1958 Birth Cohort participants were made available from the ALSPAC Laboratory, University of Bristol (under the management of Dr Sue Ring) after application to the METADAC Secretariat (at Bristol UK, subsequently

Newcastle, see https://www.metadac.ac.uk/1958bc/) and approval from the Access Committee for CLS Cohorts (ACCC, see http://www.cls.ioe.ac.uk/).

BMI data

Longitudinal BMI data for 1958 Birth Cohort participants were made available after approval from METADAC (see above); BMI was calculated as the mass in kg divided by the square of the height in metres, for which measurements had been recorded at age 7, 11, 16, 23, 33, 42, 44 and 50. Participants had a mean of 7.00 BMI measurements, and 90% of participants had at least 6 out of 8 possible BMI measurements; all age points had BMI data for at least 1100 people. BMI values (with corresponding age and sex information) for the WTCCC T2D and CAD cohorts were made available through the relevant WTCCC study leads (Andrew Hattersley [Exeter], Mark McCarthy [Oxford] and Will Rayner [Oxford] for T2D; Nilesh Samani [Leicester] and Alistair Hall [Leeds] for CAD).

Array CGH data from WTCCC CNV study (including 1958BC, T2D and CAD cohorts)

Raw array-CGH data generated as part of the WTCCC CNV study (3, 15) were made available by the European Genome-phenome Archive after application to the WTCCC Data Access Committee (cdac@sanger.ac.uk, see www.wtccc.org.uk/). Including controls and duplicates, genome-wide array data from 21058 DNA samples were available for analysis, including 1448 samples from the 1958 Birth Cohort.

Copy number estimates for *AMY2B* were derived from raw array-CGH data by combining ratios (of quantile-normalised red:quantile normalised green) for the 10 probes used in (15) to define CNVR266.13, followed by probe variance scaling (3) and calibration to the known range of integer copy numbers. *AMY2A* copy number

was estimated from the ratio (of quantile-normalised red:quantile normalised green) for six probes in the region unique to *AMY2A*, followed by probe variance scaling, correction for cross-hybridization from *AMY1*, and integer calibration. Finally, copy number determination for *AMY1* used the mean quantile-normalised ratio values for eight probes from the copy-variable *AMY1* region; this initial estimate was further refined by a correction to remove the effect of cross-hybridization to other similar sequences (represented by seven probes) in the wider amylase gene region that did not bear a simple relationship to *AMY1* copy number. Estimates of *AMY1*, *AMY2A* and *AMY2B* copy number derived by these methods showed the known properties of the variation, including predominance of even numbers for *AMY1*, and the parity for odd or even copy number between *AMY1* and *AMY2A* (6, 16) (Figure 1a).

PRT measurement of copy number in genomic DNA

Measurement of copy number of *AMY1*, *AMY2A* and *AMY2B* in 1958 Birth Cohort genomic DNA samples followed the paralogue ratio test (PRT) methods described by Shwan *et al.* (17); in summary, *AMY1* gene copy number was estimated after measurement of two *AMY1AB*:*AMY1*C ratios, and copy number estimates for *AMY2A* and *AMY2B* were deduced after measurement of the *AMY2A*/*2B* duplication sequence combined with measurement of the *AMY2A*:*2B* ratio as described (16, 17). Fluorescently labelled PCR products from these assays were combined and separated by electrophoresis in a single ABI3130xl capillary for each sample. Out of the 1448 samples tested, copy number measures were successfully obtained for *AMY1* from 1308 samples, for *AMY2A* from 1300 samples and for *AMY2B* from 1309 samples (Table S1). Amylase copy numbers derived from PRT analysis of genomic DNA from 1958 Birth Cohort samples have been submitted (in PLINK generic variant file format) to the central 1958 Birth Cohort database.

Association analysis and power calculations

Analysis of association between BMI and amylase copy number was carried out using linear regression in the context of the PLINK toolset (v1.07) for genome-wide association studies (18). Amylase copy number was encoded in the generic variant (gvar) file format for multiallelic CNV; sex was included as a covariate, as was age, except in analysis of single age-points for 1958 Birth Cohort data. For analysis in children (ages 7, 11 and 16), we used BMI Z-scores relative to the mean and standard deviation of the cohort to examine association, but at older age points (age 23+) BMI was used without transformation. The power of different cohorts was simulated using Monte-Carlo simulations (each with a million iterations) applying the effect size observed by Falchi *et al.* to the numbers of samples investigated in each cohort in this work. For the 1958 Birth Cohort samples, different numbers of individuals had both *AMY1* copy number information and BMI data at different age points, but the minimum sample size (at age 16, N = 1107) has 58% power to demonstrate association significant at $P = 0.01$ and 78% power for $P = 0.05$; the mean number of informative 1958 Birth Cohort samples across the different age points (N = 1266) had 60% power to demonstrate association at $P = 0.01$, and 83% power for $P = 0.05$. For the WTCCC disease cohorts, 1808 CAD samples had 92% power, and 1027 T2D samples 76% power, to demonstrate association at nominal ($P = 0.05$) levels of significance. Finally, a combined cohort of 4237 adult samples had 99.1% power to detect association significant at $P = 0.01$, and 99.9% power to detect association at nominal ($P = 0.05$) significance.

Results

Falchi *et al.*'s initial demonstration of association between BMI and *AMY1* copy number mostly relied on real-time PCR typing of amylase copy number (5), and subsequent studies highlighted the apparently low accuracy of copy number measurement (6, 16). In seeking to replicate the association we therefore paid particular attention to accurate measurement of gene copy number for *AMY1*. We started with analysis of existing array CGH data for 20158 samples from the WTCCC CNV study, which included 1448 samples from the 1958 Birth Cohort (3). Variation in *AMY1* gene number is complex at the sequence level, and presumably for this reason *AMY1* variation was not among the copy-variable regions that could be called with high confidence in the WTCCC CNV study (3, 15). In this work, customised analysis of raw array-CGH data, informed by knowledge of the structural basis of the variation (6, 16), allowed us to extract measurements of *AMY1* representation that clustered around integer values and matched the known properties of the variation (see Methods). The placement of clusters in Figure 1a demonstrates the known relationship between *AMY1* and *AMY2A* copy number parity – when *AMY2A* has an odd number of copies, so does *AMY1*, and when *AMY2A* has an even number, so does *AMY1* (6, 16).

In parallel, we obtained independent duplicate copy number measurements by direct analysis of genomic DNA in 1448 DNA samples from the 1958 Birth Cohort, using ratio-based (Paralogue Ratio Test, PRT) methods (17), again demonstrating concordance with previous copy number distributions (Table S1) and the odd-even parity between *AMY1* and *AMY2A* (Figure 1b). The mean copy numbers found in 1958 Birth Cohort samples were 6.71 (range 2-15) for *AMY1*, 2.0 (range 0-5) for *AMY2A* and 2.15 (range 2-5) for *AMY2B*. Measurements of *AMY1* copy number by

array-CGH and by PRT showed good agreement, with an r^2 value of 0.872, and 90% of samples concordant for the inferred integer *AMY1* copy number (Figure 1c). Similarly, both array-CGH and DNA typing methods were used to derive copy number estimates of the pancreatic amylase genes *AMY2A* and *AMY2B* (Figures 2a and 2b). Overall, the appearance of integer clusters, with a predominance of even copy numbers for *AMY1* (6, 16), and demonstration of the known relationships between *AMY1* and *AMY2A* gene copy numbers, provided validation of the accuracy of copy number estimates.

We used linear regression with sex as a covariate to examine association between BMI (or BMI Z-score for ages 7, 11 and 16) and amylase gene copy number in the 1958 Birth Cohort samples, at each of eight age points between ages 7 and 50 (Table 1). Even at nominal levels of significance, no association was observed with copy number of *AMY1*, *AMY2A* or *AMY2B*, using either copy numbers measured by direct typing (Table 1) or from array-CGH (data not shown). To reduce the effects of error in copy number measurement, which might be most substantial at high copy numbers of *AMY1*, we also analysed association of BMI with *AMY1* gene copy number after rounding to integer values conditional on *AMY2A* copy number: if *AMY2A* copy number was even, the measured *AMY1* copy number was rounded to the nearest even number, and likewise for odd numbers. Again, no association of BMI with *AMY1* was observed at any age point, even at nominal significance (data not shown).

No significant association was observed ($P > 0.05$) when specifically testing the childhood data (from ages 7, 11 and 16) for association between *AMY1* copy number and obesity as a categorical state (defining obesity as a BMI above the 95th centile, with controls below the 85th centile (9)). We also examined the trajectory of BMI with

age by testing the interaction between *AMY1* copy number, BMI and age, detecting no significant interaction term (P = 0.27).

We tested for associations specific to WTCCC disease cohorts, specifically hypothesising that type 2 diabetes (T2D) might be a state in which handling of dietary starch is more important for overall metabolism. BMI data were available for members of two of the WTCCC disease cohorts (type 2 diabetes [T2D] and Coronary Artery Disease [CAD]), and amylase gene copy numbers were estimated for these samples using the array-CGH analyses described in this work. Using sex and age as covariates, no significant associations were observed between BMI and amylase gene number in subjects with T2D or CAD (Table S2). Although use of the WTCCC disease cohorts may lead to bias and introduce heterogeneity into the populations studied, we combined data from different cohorts to get a single test of association with maximum power from the samples we had analysed. Combining T2D and CAD samples with the 1958 Birth Cohort samples gave a total of 4237 individuals to test the simple hypothesis of association of BMI with amylase gene copy number. This analysis used sex and age as covariates, and a single age point (age 44) was chosen for the 1958 Birth Cohort. No significant association was observed between BMI and gene copy number for *AMY1*, *AMY2A* or *AMY2B* (P >0.05 in all cases, data not shown).

Discussion

In attempting to replicate the observations of Falchi *et al.* (5) our data provide no
support, even at nominal levels of significance, for association between BMI and
amylase gene copy number at any age point in subjects from the 1958 Birth Cohort,
or in the WTCCC T2D and CAD cohorts. In this respect our results are concordant
with those of Usher *et al.* (6) and Yong *et al.* (12). Although we evaluated copy
number directly by typing genomic DNA for only about 1300 samples from the 1958
Birth Cohort (Table S1), all participants had array-CGH data available from the
WTCCC CNV study, so that in practice power was limited by the availability of BMI
data. In the sample from the UK 1958 Birth Cohort, different age points have
different numbers of individuals with BMI data (from N=1107 at age 16 to N=1402 at
age 44), and hence variable power to detect association at any one age point. At the
most conservative extreme, if we assume that association of BMI with amylase gene
copy number is only operating at a single age point in the 1958 Birth Cohort, then
our study has between 78% and 86% power to detect association at nominal
significance, assuming an effect of the magnitude claimed by Falchi et al 2014 (beta
$= -0.15$ kg/m^2 per copy of *AMY1*). The results of Monte Carlo simulations of power at
different individual age points are given in Table S3. These values offer conservative
estimates of power, assuming that each individual age point is the only opportunity to
detect association. Because in reality BMI at different age points for the same
individual will not be independent, the overall power to test association in children
might be conservatively estimated using the total number of unique individuals
sampled at ages 7, 11 or 16 (N = 1406), which has more than 85% power to detect
the full effect claimed by Falchi *et al.* at nominal significance (Table S3). If we could
assume that the association applies equally and independently at all age points, then

14

effectively the longitudinal series of eight age points would create a very powerful resampling, increasing the power to nearly 100% to detect association at nominal significance *in at least one* of the age points. Furthermore, if the effect is age-independent and unaffected by disease status, then our combined analysis of 4237 independent samples also has close to 100% power to detect an effect of the magnitude reported by Falchi *et al.* (Table S3).

Given the major role played by pancreatic amylase enzymes in intestinal digestion of dietary starch, it would be natural to predict that variation in the copy-variable pancreatic amylase (*AMY2A/2B*) genes could have a more substantial influence on carbohydrate metabolism than the salivary amylase genes. As part of these same analyses we were also able to derive copy number measurements for the pancreatic amylase genes *AMY2A* and *AMY2B*; no association was observed between BMI and the copy number of *AMY2* genes, even at nominal levels of significance.

The 1448 samples we studied from the UK 1958 Birth Cohort were used as controls in the WTCCC studies (3, 14), and were selected from the 9,790 participants still undergoing follow-up at age 50 from an original birth cohort of 17,406. It is not clear whether any bias may have been introduced alongside these reductions in participant numbers, and if so what influence such bias might have on our conclusions. Using all age points, our longitudinal analysis allowed us to address the question of whether *AMY1* copy number might be associated with the pattern of BMI change with age, but again we were unable to demonstrate a significant association. Our study has the obvious limitation that BMI is the only phenotype analysed, with sex and (in some analyses) age as covariates; we have no information on potentially relevant environmental or lifestyle factors such as nutrition. An association conditional on lifestyle factors not recorded in our study, such as dietary starch

intake (7), could therefore escape detection if the association could not be detected without allowance for this covariate. Indeed, the study of 4800 nondiabetic subjects with mean age 57y by Rukh *et al.* demonstrated no significant association of BMI with *AMY1* copy number alone ($P = 0.8$), but found a significant interaction ($P = 0.007$) conditional upon dietary starch intake. That study used droplet digital PCR to determine *AMY1* copy number, and the excess of samples with even numbers of copies is reassuring that the accuracy of copy number determination is likely to be high. The use of the same individuals at different ages in our study provides a strong basis for evaluating age-dependent effects of *AMY1* copy number on BMI, but suffers from a corresponding lack of constancy for environmental conditions; at earlier time points the subjects were younger, but they were also subject to different environmental conditions relevant to BMI, which have changed considerably during the lifetime of these individuals.

Conclusions

The original association claimed by Falchi *et al.* (5), which remains the largest study showing support for simple association of BMI with amylase copy number, had no conditional element other than correcting for age and sex as covariates. As a direct replication of the original findings of Falchi *et al.*, the conclusions of our study remain unambiguously negative. In analysis of the 1958 Birth Cohort samples with age, we find no evidence to support an association between BMI and *AMY1* copy number that is age-dependent.

Acknowledgements

16

Data sharing

The amylase copy number data obtained in this study, and the BMI data obtained from 1958 Birth Cohort participant measurement, have been submitted and are available to *bona fide* researchers after application to the METADAC Secretariat (at Newcastle, see https://www.metadac.ac.uk/1958bc/) and approval from the Access Committee for CLS Cohorts (ACCC, see http://www.cls.ioe.ac.uk/).

Ethics and consent

This study did not undertake any new recruitment of subjects, but instead used data and DNA collected under existing studies; ethical approvals and informed consent arrangements were confirmed with the relevant original review boards for the original

WTCCC and 1958 Birth Cohort studies, and our access to the data and DNA samples were approved (a) for the 1958 Birth Cohort by the METADAC Secretariat (now at Newcastle, https://www.metadac.ac.uk/1958bc/) and the Access Committee for CLS Cohorts (ACCC, see http://www.cls.ioe.ac.uk/) and (b) for the WTCCC by the consortium DAC (cdac@sanger.ac.uk, see www.wtccc.org.uk/).

References

1.   Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM*, et al.* A Common Variant in the *FTO* Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* 2007;**316:** 889-894.

2.   Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR*, et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518:** 197-206.

3.   Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;**464:** 713-720.

4.   Usher CL, McCarroll SA. Complex and multi-allelic copy number variation in human disease. *Briefings in Functional Genomics* 2015;**14:** 329-338.

5.   Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC*, et al.* Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet* 2014;**46:** 492-497.

6.   Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR*, et al.* Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nat Genet* 2015;**47:** 921-925.

7. Rukh G, Ericson U, Andersson-Assarsson J, Orho-Melander M, Sonestedt E. Dietary starch intake modifies the relation between copy number variation in the salivary amylase gene and BMI. *The American Journal of Clinical Nutrition* 2017;**106:** 256-262.

8. Bonnefond A, Yengo L, Dechaume A, Canouil M, Castelain M, Roger E*, et al.* Relationship between salivary/pancreatic amylase and body mass index: a systems biology approach. *BMC Medicine* 2017;**15:** 37.

9. Mejía-Benítez M, Bonnefond A, Yengo L, Huyvaert M, Dechaume A, Peralta-Romero J*, et al.* Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children. *Diabetologia* 2015;**58:** 290-294.

10. Viljakainen H, Andersson-Assarsson JC, Armenio M, Pekkinen M, Pettersson M, Valta H*, et al.* Low Copy Number of the AMY1 Locus Is Associated with Early-Onset Female Obesity in Finland. *PLOS ONE* 2015;**10:** e0131883.

11. Marcovecchio ML, Florio R, Verginelli F, De Lellis L, Capelli C, Verzilli D*, et al.* Low AMY1 Gene Copy Number Is Associated with Increased Body Mass Index in Prepubertal Boys. *PLOS ONE* 2016;**11:** e0154961.

12. Yong RYY, Mustaffa SB, Wasan PS, Sheng L, Marshall CR, Scherer SW, *et al.* Complex Copy Number Variation of AMY1 does not Associate with Obesity in two East Asian Cohorts. *Hum Mutat* 2016;**37:** 669-678.

13. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* 2006;**35:** 34-41.

14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447:** 661-678.

15. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 2010;**464:** 704-712.

16. Carpenter D, Dhar S, Mitchell L, Fu B, Tyson J, Shwan N, *et al.* Obesity, starch digestion and amylase: Association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. *Hum Mol Genet* 2015;**24:** 3472-3480.

17. Shwan NAA, Louzada S, Yang F, Armour JAL. Recurrent Rearrangements of Human Amylase Genes Create Multiple Independent CNV Series. *Hum Mutat* 2017;**38:** 532-539.

18.   Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira Manuel AR, Bender D*,
       et al.* PLINK: a tool set for whole-genome association and population-based
       linkage analyses. *Am J Hum Genet* 2007;**81:** 559-575.

Table 1 Association study between copy number estimates of AMY1, AMY2A or AMY2B measured directly by PRT methods, and BMI in the 1958 Birth Cohort samples, with sex as a covariate. *At ages 7, 11 and 16 association analysis used Z-scores, so that beta values are standardised (i.e., are measured in SDs per copy); at other ages BMI was used directly, so that the beta values shown have the unit kg/m^2 per copy.

| Age group in years	AMY1				AMY2A				AMY2B			
	Total	Median BMI (1st-3rd quartile) in kg/m^2	Beta*	p-value	Total	Median BMI (1st-3rd quartile) in kg/m^2	Beta*	p-value	Total	Median BMI (1st-3rd quartile) in kg/m^2	Beta*	p-value
7	1102	15.7 (14.9-16.7)	-0.006	0.63	1096	15.7 (14.9-16.7)	-0.004	0.94	1103	15.7 (14.9-16.7)	0.022	0.78
11	1119	16.9 (15.8-18.4)	0.0066	0.62	1114	16.9 (15.8-18.4)	-0.079	0.13	1120	16.9 (15.8-18.4)	-0.051	0.52
16	1035	20.1 (18.7-21.9)	-0.001	0.91	1028	20.1 (18.6-21.9)	0.005	0.93	1032	20.1 (18.6-21.9)	-0.043	0.612
23	1144	22.2 (20.6-24.1)	-0.023	0.57	1137	22.2 (20.6-24.1)	-0.154	0.33	1142	22.2 (20.6-24.1)	-0.208	0.38
33	1203	24.3 (22.1-26.8)	0.023	0.67	1194	24.3 (22.1-26.8)	-0.087	0.68	1201	24.3 (22.1-26.8)	-0.504	0.11
42	1257	25.3 (22.8-28.1)	0.135	0.49	1251	25.3 (22.9-28.1)	1.272	0.1	1258	25.3 (22.8-28.1)	1.461	0.21
44	1266	26.7 (24.1-30)	0.008	0.89	1261	26.7 (24.1-30)	0.358	0.14	1268	26.7 (24.1-30)	0.17	0.64
50	1131	26.6 (24.1-30.1)	0.028	0.69	1125	26.6 (24.1-30.2)	0.374	0.15	1131	26.6 (24.1-30.1)	0.406	0.31

Figure 1

(a)

(b)

(c)
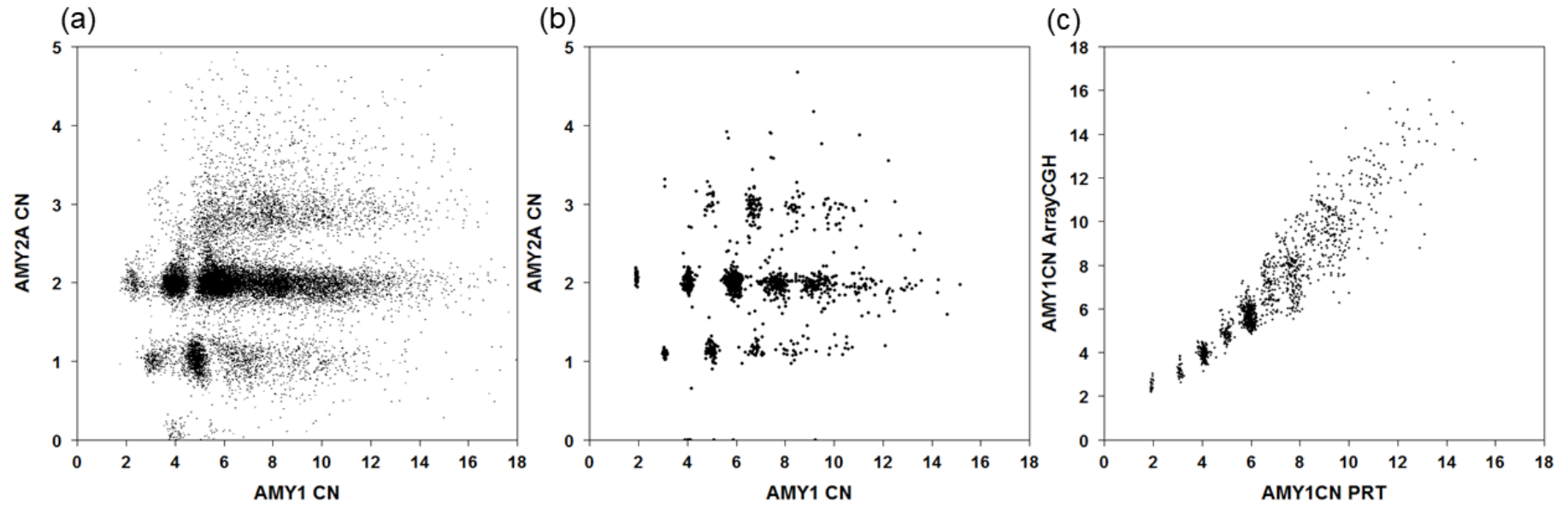
Figure 1

The relationship between measured *AMY1* and *AMY2A* copy number in (a) array

CGH data from 21058 samples from the WTCCC and (b) PRT data from 1287

samples from the 1958 birth cohort. The inclusion in (a) of many WTCCC subjects

not contributing to our BMI analysis illustrates the quality and uniformity of the

WTCCC array CGH data we used, even at this high volume of throughput, with most

individuals simply classifiable into discrete copy number classes for *AMY1* and

AMY2A. In (c), the concordance is shown for *AMY1* copy numbers for 1308 samples

from the 1958 birth cohort measured by both array CGH and by PRT.

Figure 2
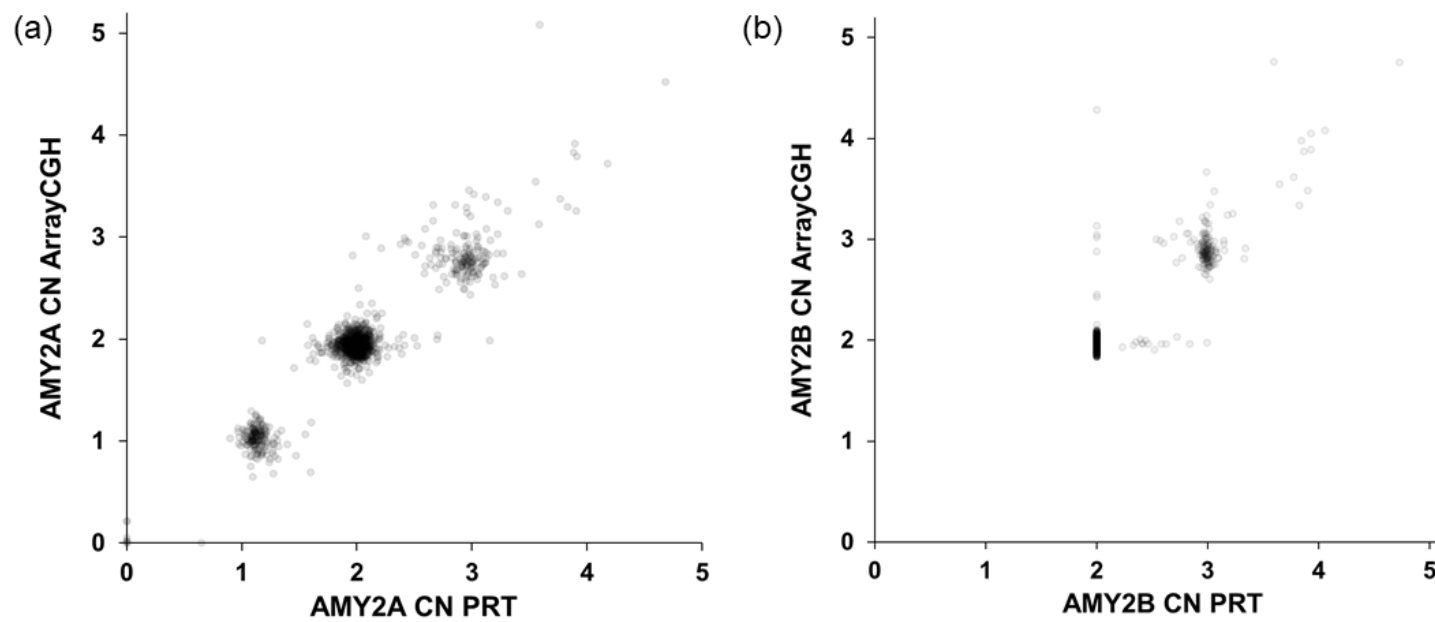
Figure 2

Concordance of copy number measurements derived from array CGH and PRT for the copy numbers of (a) *AMY2A* in 1300 samples and (b) *AMY2B* in 1309 samples from the 1958 birth cohort. In (b), the PRT analysis infers a copy number of exactly 2 from the absence of junction fragment products, so that over 1000 data points are superimposed around the coordinate (2,2).

Supporting Information (Tables S1-S3)

Table S1. Observed integer copy number distributions among 1308 (*AMY1*), 1300 (*AMY2A*) or 1309 (*AMY2B*) 1958 birth cohort participants, as measured from genomic DNA. In each entry the number of individuals observed in the integer copy number category is followed by the percentage population frequency in parentheses.

copy number	*AMY1*	*AMY2A*	*AMY2B*
0	0 (0%)	7 (0.54%)	0 (0%)
1	0 (0%)	173 (13.31%)	0 (0%)
2	22 (1.68%)	951 (73.15%)	1140 (87.09%)
3	31 (2.37%)	158 (12.15%)	158 (12.07%)
4	176 (13.46%)	10 (0.77%)	10 (0.76%)
5	103 (7.87%)	1 (0.08%)	1 (0.08%)
6	388 (29.66%)	0 (0%)	0 (0%)
7	157 (12.00%)	0 (0%)	0 (0%)
8	177 (13.53%)	0 (0%)	0 (0%)
9	106 (8.10%)	0 (0%)	0 (0%)
10	72 (5.50%)	0 (0%)	0 (0%)
11	39 (2.98%)	0 (0%)	0 (0%)
12	19 (1.45%)	0 (0%)	0 (0%)
13	11 (0.84%)	0 (0%)	0 (0%)
14	5 (0.38%)	0 (0%)	0 (0%)
15	2 (0.15%)	0 (0%)	0 (0%)
More	0 (0%)	0 (0%)	0 (0%)
(total)	1308 (100%)	1300 (100%)	1309 (100%)

Table S2. Results of association analysis between array-CGH extracted copy number estimates of *AMY1*, *AMY2A* or *AMY2B* and BMI in type 2 diabetes (T2D) and coronary artery disease (CAD) cohorts. The analysis used linear regression, with sex and age as covariates.

cohort	samples	trait	*AMY1*		*AMY2A*		*AMY2B*	
			Beta	p-value	Beta	p-value	Beta	p-value
T2D	1027	BMI	-0.06	0.44	-0.16	0.72	-0.16	0.72
CAD	1808	BMI	-0.003	0.95	-0.003	0.98	-0.15	0.44

Table S3. Representative results of power simulations, showing the percentage power predicted from Monte Carlo simulations to detect association between BMI and *AMY1* copy number at α = 0.01 or 0.05, with either the full effect size reported by Falchi et al. (β = −0.15kg/m^2 per copy) or half this effect size. For the 1958 birth cohort BMI data are available from different numbers of people at different age points, so the first three rows show results using the least (age 16), average and greatest (age 44) numbers available. Among the children (ages 7, 11, 16) we simulated power assuming the total number of unique individuals for whom BMI data were available (N= 1406). One million iterations were used for each simulation. T2D = type II diabetes; CAD = coronary artery disease.

N	Subset	Falchi *et al.* effect size		Half effect size	
		α = 0.01	α = 0.05	α = 0.01	α = 0.05
1107	1958 cohort age 16	53.8%	77.9%	13.6%	33.5%
1266	1958 cohort mean	60.4%	82.5%	15.3%	36.6%
1402	1958 cohort age 44	65.3%	85.7%	17%	39%
1406	1958 cohort unique children	65.5%	85.8%	17%	39.1%
1027	WTCCC T2D	50.7%	75.5%	12.7%	32.1%
1808	WTCCC CAD	77.5%	92.3%	21.8%	46.1%
4237	Combined 1958BC+WTCCC cohorts	99.1%	99.87%	50.9%	75.9%