<table>
<thead>
<tr>
<th>Title</th>
<th>Complex Copy Number Variation of AMY1 does not Associate with Obesity in two East Asian Cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Yong, Rita Y. Y.; Su'Aidah Binte Mustaffa; Wasan, Pavandip S.; Sheng, Liang; Marshall, Christian R.; Scherer, Stephen W.; Teo, Yik-Ying; Yap, Eric Peng Huat</td>
</tr>
<tr>
<td>Citation</td>
<td>Yong, R. Y. Y., Su'Aidah Binte Mustaffa, Wasan, P. S., Sheng, L., Marshall, C. R., Scherer, S. W., et al. (2016). Complex Copy Number Variation of AMY1 does not Associate with Obesity in two East Asian Cohorts. Human Mutation, 37(7), 669-678.</td>
</tr>
<tr>
<td>Date</td>
<td>2016</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/41899">http://hdl.handle.net/10220/41899</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2016 Wiley Periodicals Inc. This is the author created version of a work that has been peer reviewed and accepted for publication by Human Mutation, Wiley Periodicals Inc. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1002/humu.22996">http://dx.doi.org/10.1002/humu.22996</a>].</td>
</tr>
</tbody>
</table>
COMPLEX COPY NUMBER VARIATION OF AMY1 DOES NOT ASSOCIATE WITH OBESITY IN TWO EAST ASIAN COHORTS

Rita Y.Y. Yong1,2, Su’Aidah B. Mustaffa1,3, Pavandip S. Wasan1,2, Liang Sheng5, Christian R. Marshall6, Stephen W. Scherer6,7, Yik-Ying Teo2,4, Eric P.H. Yap1,2,3,*

1 Defence Medical and Environmental Research Institute, DSO National Laboratories, Singapore,
2 Saw Swee Hock School of Public Health, National University of Singapore, Singapore,
3 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore,
4 Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore.
5 Unit of Biostatistics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore,
6 The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada.
7 Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON, Canada.

* Corresponding Author
Abstract

The human amylase gene locus at chromosome 1p21.1 is structurally complex. This region contains two pancreatic amylase genes, AMY2B, AMY2A, and a salivary gene AMY1. The AMY1 gene harbours extensive copy number variation (CNV), and recent studies have implicated this variation in adaptation to starch-rich diets and in association to obesity for European and Asian populations. In this study, we showed that by combining quantitative PCR and digital PCR, coupled with careful experimental design and calibration, we can improve the resolution of genotyping CNV with high copy numbers (CN). In two East Asian populations of Chinese and Malay ethnicitiy studied, we observed a unique non-normal distribution of AMY1 diploid CN genotypes with even:odd CN ratio of 4.5 (3.3-4.7), and an association between the common AMY2A CN=2 genotype and odd CN of AMY1, that could be explained by the underlying haplotypic structure. In two further case-control cohorts (n = 932 & 145, for Chinese and Malays, respectively), we did not observe the previously reported association between AMY1 and obesity or body mass index (BMI). Improved methods for accurately genotyping multiallelic CNV loci and understanding the haplotype complexity at the AMY1 locus are necessary for population genetics and association studies.

KEYWORDS

Amylase 1 gene, Amylase 2A gene, Copy Number Variation, multiallelic, Chinese and Malay populations, obesity, quantitative PCR, digital PCR.
Introduction

Human salivary amylase is encoded by the \textit{AMY1A}, \textit{AMY1B} and \textit{AMY1C} genes (MIM*104700, *104701, *104702), while pancreatic amylase is encoded by two genes, \textit{AMY2A} (MIM*104650) and \textit{AMY2B} (MIM*104660). The \textit{AMY1} and \textit{AMY2} gene families share homology of 92-94\%, both in coding sequences and exon-intron structure. The three copies of \textit{AMY1} genes share > 99.9\% sequence identity, and the similarity extends over a region ~27 kb around each copy, suggesting a relatively recent origin of \textit{AMY1} gene duplication [Carpenter, et al., 2015]. In addition, two foreign elements have been identified in the promoter regions of the amylase genes, one of which is responsible for the tissue-specific expression of the salivary amylase [Meisler and Ting, 1993; Perry, et al., 2007]. These genes and a pseudogene \textit{AMYP1} are located as a cluster on chromosome 1p21.1, in a complex genomic locus characterised by inversions, deletions, and tandem duplications (Suppl. Fig. S1a). Though the chromosomal organisation of the amylase gene cluster was first reported 25 years ago through analysis of restriction mapping in pedigrees [Groot, et al., 1989], its implications for Copy Number Variation (CNV) allele structures have not been fully appreciated.

The human amylase genes cluster region is known to be copy number variable [Iafrate, et al., 2004; Sudmant, et al., 2010] and one of the most variable CNV among humans [Sudmant, et al., 2010]. Our initial work in three East Asian populations using standard CNV microarray genotyping methods correctly identified CNVs in the amylase locus but grossly underestimated the extent of CN variation (Supp. Fig. S1b). Studies using qPCR have reported \textit{AMY1} diploid copy number ranging from 2 to 20, with an approximately normal distribution pattern of copy number frequency in several populations [Perry, et al., 2007;
Santos, et al., 2012; Falchi, et al., 2014. AMY2A/2B genes also exhibit CNV, but to a lesser extent than AMY1 [Groot, et al., 1991; Sudmant, et al., 2010; Falchi, et al., 2014].

Recent studies suggest that these CNVs may have functional significance. AMY1 has undergone gene number expansion in the human lineage as chimpanzee harbours only 2 copies in its genome. It was argued that this may be related to low dietary starch intake by chimpanzee [Perry, et al., 2007]. In addition to the cross-species differences, the copy number of AMY1 is highly variable among humans, and this variation correlates positively with the expression of transcript and protein [Perry, et al., 2007; Mandel, et al., 2010; Falchi, et al., 2014]. Significantly higher AMY1 CN were found in populations with high-starch diets (eg. Japanese & Hadza) compared to those with traditional low-starch diets (eg. Yakut & Biaka), irrespective of their geographical or ethnic origin [Perry, et al., 2007]. These studies suggest adaptive change in the AMY1 CNV in response to adoption of agriculture and high-starch diet as a selective pressure.

AMY1 CNV has also been linked to human disease with a reduced copy number of AMY1 associated with increased BMI and predisposing to obesity [Falchi, et al., 2014]. The association was initially identified as gene dosage effect attributed to CNV at amylase genes locus by using microarray combining transcriptomics and GWAS analysis in Swedish sibling-pairs discordant for obesity. The association of AMY1 copy number to BMI, but not AMY2A copy number, was validated by qPCR in the same population, and was replicated in several European populations and one Chinese cohort from Singapore. Furthermore, the same group reported an association with similar effect size of AMY1 copy number and obesity risk in Mexican children [Mejia-Benitez, et al., 2015]. However, the association reported by Mejia-Benitez et al. [2015] was driven by AMY1 high CN outliers among the controls, while
the association in Falchi et al [2014] was driven by the bulk of samples producing a shift of the overall CN distribution. This suggests two different association findings rather than a replication. More recently, a negative association finding was reported in two European cohorts [Usher et al. 2015]. In addition, Usher et al. reported a differential even:odd AMY1 copy number frequency distribution, in contrast to the approximately normal distribution reported in Falchi et al. This differential even:odd pattern was corroborated by Carpenter et al. [2015], and by using different experimental approaches, both groups reported a AMY1 CN frequency distribution that closely resembled each other for HapMap European samples [Carpenter, et al., 2015; Usher, et al., 2015].

The association reported by Falchi et al. is the first genetic link between carbohydrate metabolism and obesity, an intriguing finding that potentially provides new insight into novel biological mechanisms underlying obesity. However, we also recognised the technical challenge of accurately genotyping multiallelic CNV in a large-scale well-powered association study, due to the large dynamic copy range of a multiallelic CNV such as AMY1. This technical difficulty could be the reason why the association of AMY1 with obesity was not detected in earlier GWAS studies, since studies carried out with array-CGH or SNP array have limitations in detecting copy number >5. Conversely, several multiallelic CNV association studies done with qPCR were fraught with reproducibility issues [Cantsilieris and White, 2013; Cantsilieris, et al., 2014]. Recently, digital PCR has been used for quantifying gene dosage by digital counts of single molecule amplifications and is considered more accurate because of its arithmetic linearity compared to the exponential nature of qPCR [Weaver, et al., 2010; Baker, 2012; Hindson, et al., 2013]. AMY1 CNV genotyping by digital PCR has been reported using emulsion droplets [Usher, et al., 2015] or by high density arrays.
of microwells as reported here; both methods require costly custom equipment and/or consumables.

In this study, we developed a novel calibration method combining quantitative real-time PCR (qPCR) and array-based digital-PCR (dPCR), and used this combined method in an attempt to validate the previously reported association of *AMY1* copy number with obesity in two East Asian populations. We also tested the association of obesity with *AMY2A* copy number. *AMY2B* was not studied because it was reported to be largely invariant in Asian populations [Sudmant, et al., 2010; Handsaker, et al., 2015]. We also report here the unexpected finding of an even-odd imbalance in CN genotypes in these populations, and study the underlying haplotype structure.

**Materials and Methods**

**Study sample set**

The study sample set was a legacy collection from a previous obesity project, which was granted ethics approval by the DSO IRB in March 2004 (ref no.: DMERI-20030260-R2). Samples were collected from volunteers with written consent. Non-obese control subjects had Body Mass Index (BMI) less than 23 kg/m², while obese subjects had BMI of at least 28 kg/m², and severely obese subjects had BMI ≥ 32 kg/m². All volunteers were male, of age ranged from 18 to 21, and of self-reported ethnicity either Chinese or Malay, with all four grandparents of the same race as the volunteer. This study sample set comprised of 519 Chinese controls, 413 Chinese obese, 30 Malay controls and 115 Malay obese. The demographic attributes of these samples are summarised in Table 1. Genomic DNA was
extracted from the buffy coat of whole blood using QIAamp DNA Mini Kit (Qiagen, Germany) and quantification by Nanodrop Spectrophotometer (Thermo Scientific, USA).

Sample Randomisation to avoid Differential Bias

All samples, both cases and controls, were subjected to identical work processes throughout this study. Samples were randomised in all laboratory processes, starting from sample collection, processing, DNA extraction, and qPCR, in order to avoid systematic bias, differential errors and batch effects.

Copy-number estimation for AMY1 and AMY2A by quantitative real-time PCR

The copy numbers of AMY1 and AMY2A genes were estimated by duplex quantitative real-time PCR (qPCR) on a QuantStudio™ 12K Flex Real-Time PCR System (Thermo Fisher Scientific Inc., USA) with 384-well plate. The analysis software was QuantStudio™ 12K Flex Software version 1.1.2. The duplex reactions consisted of two assays, each with two primers and a TaqMan® probe (Thermo Fisher Scientific Inc., USA); one specific for the target, AMY1 (Hs07226362_cn) or AMY2A (Hs04204136_cn), and one specific for the reference RNaseP (P/N4403328, Thermo Fisher Scientific, USA). These assays are similar as those used in Falchi et al. (2014) and hence facilitated direct comparison of results. The primers and probe of the Taqman qPCR assay employed for AMY1 specifically target a region within exon 1 of the AMY1 gene which is absent in the AMY1P pseudogene. This strategy, coupled with primer sequence specificity, ensured specificity of the qPCR assay for AMY1. Details of these Taqman assays are summarized in Supplementary Table S1. Detail protocols of qPCR and dPCR can be found in the Supplementary materials.
Each qPCR was carried out in replicates of four. For assay targeting \textit{AMY1}, a final volume of 10 µl per aliquot was used in order to achieve better precision to accommodate the high copy variation in \textit{AMY1}, while a final volume of 4 µl was used in assay for \textit{AMY2A}. Real-time PCR data was analysed with QuantStudio™ 12K Flex Software version 1.1.2, results were then exported to CopyCaller v2.0 software (Thermo Fisher Scientific Inc., USA) for copy number calculation. All analysis settings selected were as recommended in the user guides. Samples with standard deviation of ΔCt >0.25, or with z-score >1.75 were removed and repeated. Diploid copy numbers were estimated by ΔΔCt method using the calibrator DNA sample NA10851 (Coriell Cell Repositories, USA) which carries 6 copies of the \textit{AMY1} gene and 2 copies of \textit{AMY2A}.

The sample NA10851 was selected as a reference sample for several reasons. It is a common reference sample used in many CNV studies carried out globally using microarrays, and it was the calibrator used in our previous qPCR validation study for CNV discovered in-house. It has been compared and calibrated with several reference samples using both qPCR and digital-PCR, and has consistently shown good concordance and reproducibility. Prior to its selection, a few reference samples had been studied as a potential calibrators because their copy numbers had been independently measured and reported in several published studies using methods including fiber-FISH and qPCR. These included NA18956 which was reported to carry 6 copies of \textit{AMY1} and 2 copies of \textit{AMY2A} [Falchi, et al., 2014]; NA18972 with 14 copies of \textit{AMY1} [Perry, et al., 2007; Falchi, et al., 2014]; and NA10472 with 6 copies of \textit{AMY1} [Perry, et al., 2007]. However, the 14 copies of \textit{AMY1} for NA18972 was disputed by a recent study [Carpenter, et al., 2015] and also this study (Supp. Table S2, Supp. Figure S4). The samples NA18956 and NA10851 have consistently showed 6 copies for \textit{AMY1} in our qPCR results. But NA18956 has always displayed slightly lower copy number ratio than
the NA10851, possibly due to mosaicism in this cell line sample. NA18956 was subsequently typed as CN5 by digital PCR. On the other hand, NA10472 had displayed 6 copies of AMY1 in fiber-FISH [Perry, et al., 2007] and has been typed as CN6 by dPCR. Because NA10851 and NA10472 both consistently typed as CN6 in dPCR, we considered NA10851 calibrated against NA10472, and NA10851 was elected as the calibrator in this study.

The ΔΔCt method assumes equal amplification efficiency for the target and reference genes. PCR efficiency for the AMY1 target assay and its reference (RNase P) assay was 92.76% and 97.44%, respectively, and for AMY2A target assay and its reference assay was 96.56% and 98.59%, respectively (Supp. Figs. 2a and 2b). The efficiency between target and reference loci differ by less than 5% is considered sufficiently similar [Fernandez-Jimenez, et al., 2011].

Copy number estimation of AMY1 by Digital PCR

Digital PCR was performed on Quantstudio™ 3D Digital PCR System (Thermo Fisher Scientific) using Digital PCR 20K Chip which contains 20,000 reaction wells per chip. Thermal cycling was carried out in a Dual Flat Block GeneAmp® PCR System 9700 and analysis with Quantstudio™ 3D AnalysisSuite software. The human genomic DNA was first digested with an appropriate restriction enzyme to break up the tandemly repeated CNV copies. A final concentration of 2.5 ng/µl genomic DNA input into the dPCR chip is equivalent to approximately 750 copies/µl, which is within the recommended target sequence concentrations of between 200 to 2,000 copies/µl for accurate dPCR analysis (Thermo Fisher Scientific Inc., USA). As the AMY1 copy number increases in some samples, the input DNA was reduced accordingly so to avoid reaction chamber saturation, and to get both target and reference molecules concentration fall around 200 to 2,000 copies/µl in order to achieve
precise dPCR analysis. Each sample was genotyped in duplicate at least, and the number of replicates were increased accordingly for samples with higher copy numbers or showed higher variation between chips.

**Calibration Curve of qPCR with dPCR and Allele Binning**

A calibration curve was constructed to calibrate qPCR data to those of dPCR. A total of 31 samples (3 reference samples from Corielle Repository and 28 in-house study samples) were genotyped in both qPCR and dPCR. These 28 in-house samples used to construct the calibration curve have qPCR results covering the full range from CN2 to 62, and with representative sample(s) from respective clusters. The number of dPCR replicates genotyped increased accordingly with the *AMY1* CN of the sample (Supp. Table S2). Instead of using calculated copy numbers, the raw qPCR signal intensity delta-delta Ct (ΔΔCt) was plotted against the logarithm of copy numbers obtained from dPCR. By taking ΔCt of calibrator into consideration, this allowed the normalization of ΔCt across inter-plate experimental runs.

Subsequently, these calibrated *AMY1* CN values were binned into copy number classes on the basis of pre-defined thresholds at 0.5, and binned to the nearest integer.

**Results**

**Genotyping of *AMY2A* and *AMY1* copy number by qPCR and dPCR**

The technical challenge of genotyping multiallelic CNV increased with the dynamic range of copy numbers. This is demonstrated by the CN genotyping of *AMY2A* and *AMY1* initially carried out by the qPCR method alone. The distribution of diploid CN in this combined population of 1077 individuals was 1 to 4 copies for *AMY2A*, and 2 to 24 copies for *AMY1,*
respectively. The histogram of raw \( AMY2A \) CN computed from delta \( C_T \) showed four CN genotypes clustered discretely (Fig. 1a, Supp. Fig. S3). CN variation at \( AMY2A \) is low with most individuals having 2 diploid copies, and about 6.5% of Chinese and 7.6% of Malays with \( AMY2A \) CN that were not 2.

The large range of \( AMY1 \) CN presented a challenge to calling genotypes. Although the raw CN computed from qPCR delta Ct showed clustering, bins were unclear for CN\(>4 \), the frequency peaks were not integers, and apparent inflation at higher CN (with outliers of 35 and 62 far beyond what had been previously reported) were noted (Fig. 1b). These were similarly observed in the lower CN of \( AMY2A \) (Fig. 1a), suggesting a methodological rather than locus-specific artefact. In order to obtain a more accurate CN assignment of these clusters and to assess the reliability of the qPCR results, three reference samples and a set of 28 samples representative of the CN range of \( AMY1 \) were genotyped with digital PCR (dPCR) and their results compared directly with the qPCR data (Supp. Table S2).

While there was a high correlation between CN deduced from dPCR and qPCR (Spearman \( r=0.988 \)), the relationship was non-linear, with qPCR giving a higher CN estimate compared to dPCR (Fig. 1c). Assuming dPCR gives more accurate CN quantitation, the polynomial trend line from this dPCR calibration set of 31 samples was used to adjust and calibrate the qPCR CN calls for all samples in the study. The histogram of the calibrated qPCR CN for the 1077 samples showed a range of distribution from 2 to 24 diploid copies (Fig. 1d).

While this calibration did not discretise bins above CN4, the peaks fell at integer values, and allowed genotype calling using constant bin sizes and rounding qPCR calibrated CN to the nearest integer (Fig. 1d, 1e). While rounding off CN is routinely used in CNV association
studies [Gonzalez, et al., 2005; Aitman, et al., 2006; Fellermann, et al., 2006; Barnes, et al., 2008], in our work, we demonstrate downstream association analysis using 3 results - integer CN, calibrated CN without binning as well as the non-adjusted continuous qPCR signal intensity (delta-delta C<sub>T</sub>).

**AMY1 and AMY2A copy number and haplotype frequency distribution in East Asian populations**

Interestingly, it was noted that the major clusters of AMY1 were of even number with diploid CN 4, 6, 8, 10, and 12, and odd CN clusters formed smaller peaks in between (Fig. 1d). The ratio of even CN to odd CN was 4.7, 4.5, 3.3, 4.0 and 4.5 for Chinese control, Chinese obese, Malay control, Malay obese and combined samples, respectively. Repeated dPCR genotyping on selected samples per cluster produced consistent results (Supp. Table S2, Supp. Fig. S4).

This novel observation of differential even:odd diploid CN (Fig. 2) was further investigated by analysis of haplotype structures at the AMY1 locus. As both qPCR and dPCR provide only the diploid CN counts, frequencies of each haplotype were estimated computationally using the expectation-maximization algorithm (CoNVEM, http://apps.biocompute.org.uk/convem/) which assumes Hardy-Weinberg equilibrium [Gaunt, et al., 2010]. The distribution of haplotype frequencies for AMY1 and AMY2A CNVs are similar between the Chinese and Malay populations (Figs. 3a & 3b). There are 2 common haplotypes for AMY1, CN3 and CN5, accounting for ~45% and ~32%, respectively (Fig. 3a) and odd haploid CN occur more frequently than even ones. In AMY2A one haplotype containing a single copy of the gene accounts for ~95% in each population (Fig. 3b).
Linkage dysequilibrium (LD) between *AMY2A* and *AMY1* was studied, by correlating diploid CN at these two adjacent loci. Individuals with the major *AMY2A* CN allele of 2, almost always have an even *AMY1* CN, while the non-2 *AMY2A* CN are associated with odd *AMY1* CN in both Chinese (Supp. Table S5) and Malay samples (data not shown). The same pattern of LD was also seen when CN were analysed as integer calls, where it reached statistical significance in all 3 groups, except the Malay control because of small sample n=30 (Supp. Table S3). Similar correlation was also reported by Usher et al. and Carpenter et al. [Carpenter, et al., 2015; Usher, et al., 2015].

**Association testing of Obesity with *AMY1* and *AMY2A* copy number**

In both populations, the modal copy number appeared lower in obese than in lean (Chinese: CN6 vs CN8, Malay CN8 vs CN10) (Fig. 4a). In this study of *AMY1* CN-obesity association testing in two populations, we are unable to replicate the association, using multiple analyses and with sufficient statistical power to detect half the published effect size. The CN differences did not reach statistical significance in univariate analysis carried out for the four pair-wise comparisons; controls versus obese, and controls versus extreme obese, in Chinese and Malay populations separately. The mean difference in CN between each case-control set was small, as was the 95% confidence interval of mean difference (Table 2a). Both T-test and Mann-Whitney tests showed no significant association in all 4 groups of comparison (Table 2a). Logistic regression similarly indicated no significant association in all four case-control sets (Table 2a). *AMY1* copy numbers ≥13 (all CN classes with frequency <3%) were collapsed together into one single category but no significance difference could be detected either. Linear regression was done to test if *AMY1* copy number predicts BMI changes, and results were also negative (results not shown). Results remained not significant when analysis was done using calibrated copy number without rounding to integer. Statistical analysis also
was carried out on direct raw qPCR signal intensity ($\Delta\Delta C_T$) as a continuous measurement, but no association could be detected (Mann-Whitney test, p>0.05) (Figs. 4b & 4c). The average BMI at various AMY1 copy numbers did not show any significant trend, neither in the obese nor the controls of both Chinese and Malays (Suppl. Figs. S5a & S5b).

We calculated if this study had enough statistical power to detect association as in Falchi et al. (2014). It was estimated that the CN mean difference between case-control samples in one of their study cohorts (DESIR cohort, Supplementary table 9 in Falchi et al. 2014) was 0.95 with standard deviation of ~2.0. With these parameters, a sample size of 95 per arm would have 90% power. Assuming a smaller mean difference at 0.5 and standard deviation at 2.0, a sample size of 338 per arm would have 90% power. Hence the Chinese sample set in this study should have 90% power to detect even half the AMY1 genetic effect size seen in the previous study.

CNV-association testing was similarly carried out for AMY2A. The distribution of AMY2A copy number in controls and obese samples in Chinese and Malay cohorts were summarized in Fig. 4d. No association between AMY2A copy number and obesity was detected by T-test or logistic regression (Table 2b). Mann-Whitney test using $\Delta\Delta C_T$ as a continuous measurement also showed no association in all 4 case-control sample sets (Figs. 4e & 4f).

Discussion

The technical challenge in genotyping multiallelic CNV is exemplified by the AMY1 and AMY2A locus. While genotyping low copy number CNV such as AMY2A (diploid CN 1-4) is straightforward, genotyping high copy number CNV like AMY1 (diploid CN 2-24) is
challenging, as the relative difference between alleles becomes smaller as CN increases and higher measurement precision is needed to quantify high CN. In our AMY1 genotyping data, CN classes were discrete till CN4 and \( C_T \)-based CN overlapped above CN5 (Figs. 1d & 1e). To increase precision at higher copy number, one approach is to increase the number of technical replicates for samples that harbour high copy numbers. However, this does not overcome the non-linear relationship of \( C_T \) with CN.

The large number of replicates and linear arithmetic nature of dPCR quantitation provides a solution to these issues. Digital PCR achieves absolute quantification by counting end-point results of single-molecule amplification across a large number of PCR replicates, and its merits for CNV quantitation have been considered elsewhere [Baker M, 2012; Weaver, et al., 2010; Whale, et al., 2012]. Two main factors influence the reliability of dPCR measurements: the number of replicates analysed and the number of template molecules in the assay [Weaver, et al., 2010; Pinheiro, et al., 2012; Whale, et al., 2012]. If the reference template (2 copies per genome) is loaded at an optimal 0.6 molecules per reaction well \( (\lambda_r=0.6) \), the number of reaction wells required to achieve 1.1 fold discrimination (ie CN10 vs 11) is approximately 5500 to 8000 [Weaver, et al., 2010; Whale, et al., 2012]. This number is easily achieved within the 20,000 microwell array dPCR platform used in this study. Its limitation is cost and sample throughput, with a maximum 24 samples per run. Hence we used dPCR to validate and calibrate a subset of samples representing the CN clusters identified by qPCR. While the correlation between qPCR and dPCR was high, the non-linear calibration curve allowed the adjustment of high CN qPCR results and assignment of bins (Fig. 1c).
This study has shown that despite the higher variability of qPCR measurement at high copy numbers, it still could retain the relative pattern of the overall results, correctly identifying the even:odd frequency distribution pattern in AMY1. In addition, qPCR offers an attractive option for large-scale CNV genotyping owing to its high throughput capability, widespread availability of instrumentation and possibly lower costs. In this study, care has been taken to minimize systematic bias and batch effects. All samples have been subjected to identical laboratory processes, sample randomization was done during genotyping, and four technical replicates were included per sample in qPCR. Samples with high variability between the 4 replicates were removed and re-genotyped. Also, appropriate reference samples with known and validated copy-numbers were used as internal controls. The 2 controls used in Falchi et al. (2014), NA18596 and NA18972, were replaced in this study. NA18596 was found to contain mosaicism and typed by dPCR to be AMY1 CN5, while NA18972 has been found to contain CN16 by our dPCR assay rather than the reported CN14 or CN18 in other studies (Supp. Table S2, Supp. Fig. S4.) [Perry, et al., 2007; Falchi, et al., 2014; Carpenter, et al., 2015]. Instead, we used another known reference, NA10472 with AMY1 CN6 [Perry, et al., 2007], to calibrate a common CNV microarray standard NA10851 and both consistently showed AMY1 CN6 in dPCR (Supp. Fig. S4, Supp. Table S2). NA10851 containing AMY1 CN6 was corroborated with Usher et al. (Supplementary table 3 of Usher et al. 2015).

In our study of South East Asian populations comprising ethnic Chinese and Malay we found a novel pattern of differential even:odd AMY1 frequency distribution (Fig. 2), which could be explained by the underlying haplotype structure. However the evidence of specific AMY haplotype can be traced to studies conducted a quarter century ago [[Groot, et al., 1989; Groot, et al., 1990]. Using segregation analysis of Southern blot hybridization patterns in pedigrees, Groot defined three common AMY haplotypes [Groot, et al., 1990; Groot, et al.,
1991], and proposed that a general designation 2B-2A-(1A-1B-P1)n-1C could describe the majority of the AMY haplotypes (Fig. 3c). The repeat sequence of about 100kb in size and encompassing the genes AMY1A-AMY1B-AMYP1 are flanked by segmental duplications (designated as SD1 in Suppl. Fig. S1a), could mediate chromosomal rearrangement via Non-allelic Homologous Recombination (NAHR). This haplotype structure corroborates a recent analysis of sequencing read depths which reported that the salivary genes AMY1A, AMY1B and the pseudogene AMYP1 were much more variable in copy number than the pancreatic genes AMY2A and AMY2B, while AMY1C appeared to be largely invariant [Sudmant, et al., 2010].

Two recent studies independently expanded the range of AMY haplotypes and reported similar even:odd distribution pattern for AMY1 in three HapMap population groups, Chinese cum Japanese, European and Yoruba [Carpenter, et al., 2015; Usher, et al., 2015]. Carpenter et al. used a combination of four methods including Paralogous Ratio Test (PRT), Microsatellite analysis, sequence read depth and fiber-FISH in CNV genotyping. On the other hand, Usher et al adopted whole-genome-sequencing read depth analysis, droplet digital PCR (ddPCR) and genome mapping. The frequency distribution of AMY1 diploid copy number reported for HapMap CEU closely parallel each other [Carpenter, et al., 2015; Usher, et al., 2015]. While the frequency distribution reported for HapMap CHB+JPT [Usher, et al., 2015] closely resembled the results of this study, with similar modes at CN6 & CN8 (Fig. 2). The similar even:odd pattern was found in all populations in these 3 studies, and the even:odd differential was more pronounced in East Asian populations (Chinese & Malays in this study, and HapMap CHB+JPT) compared to CEU.
Furthermore, the common AMY haplotypes deciphered from both studies corroborated with Groot et al. [Groot, et al., 1989; Groot, et al., 1990; Carpenter, et al., 2015; Usher, et al., 2015]. The common haplotypes; AMY*H0, AMY*H1, and AMY*H2 defined by Groot et al. (Fig. 3c), or designated as AH1, AH3, and AH5 in Usher et al. (2015), should be equivalent to AMY1 haploid copy number 1, 3 and 5 in this study (Fig. 3a). The haplotype frequencies determined experimentally by Usher et al. for the 4 most common AMY haplotypes (with AMY1 haploid CNs 1, 3, 5 & 7) in Europeans closely resembled those determined for Singapore populations by CoNVEM, with AMY*H1 (AMY1 haploid CN3) being the most common with frequency >40%, and AMY*H2 (AMY1 haploid CN5) as second with frequency >30% (Figs. 3a & 3c). Importantly, the four most common AMY1 haplotypes, all with odd copy number of AMY1 (1, 3, 5 and 7 copies), constituted a total of 90% of all AMY haplotypes in European population [Usher, et al., 2015]. Similarly, the 4 odd copy number AMY1 haplotypes constituted a total of >88% in Chinese or Malays as determined by CoNVEM (Fig. 3a). Therefore, this would explain why majority of the samples have an even AMY1 diploid copy number in both this study and the two recent European studies [Carpenter, et al., 2015; Usher, et al., 2015].

The fact that this study has independently discovered the unique characteristics of differential frequency distribution between even and odd diploid copy numbers of AMY1, and correlation between AMY2A and AMY1 CN genotypes, has lent support to the reliability of this data set. The unusual even:odd distribution of AMY1 copy number is a novel finding which is markedly different from many previous publications [Perry, et al., 2007; Santos, et al., 2012; Falchi, et al., 2014]. Previous studies of various sample sizes ranging from hundreds to thousands (~6,200 samples in Falchi et al. 2014) showed histograms with an approximately normal distribution. This finding has now been corroborated by two recent studies which
have shown that this non-normal frequency distribution could be explained by the underlying structural haplotypes [Carpenter, et al., 2015; Usher, et al., 2015]. Furthermore, the AMY1 diploid copy number distribution in Chinese and Malays in this study (Fig. 2) is similar to that of the North Asian CHB+JPT populations [Usher, et al., 2015]. In addition, the calculated AMY1 haploid allele frequency in Chinese and Malays (~0.4 for CN3 and ~0.3 for CN5) (Fig. 3a) were also similar to those determined in Europeans (Fig. 3c). It is reassuring to see that different genotyping approaches and population samples are converging on a similar understanding of variation at this locus of AMY1.

The initial impetus of this study was to investigate if the reported association of AMY1 copy number and obesity [Falchi, et al., 2014] could be replicated in an independent Singapore Chinese cohort. The reported association has been tested on 6,200 subjects, comprising four European cohorts and one Singapore Chinese cohort. The obesity risk (Odds Ratio or OR) per copy reduction of AMY1 was reported to be 1.19 in European (95% confidence interval (CI) = 1.13 - 1.26), and 1.17 in Singapore Chinese cohort (95% CI = 1.05 – 1.29). On the other hand, AMY2A copy number was reported to have no association with BMI or fat mass [Falchi, et al., 2014]. However, various statistical analyses in this study failed to replicate any association between AMY1 copy number and obesity in both Chinese and Malay sample sets. Like Falchi et al. (2014), no association could be detected between AMY2A copy number and obesity in the same populations.

The AMY1 results are discordant between these two studies, with Falchi reporting near normal CN distribution in all three of its cohorts. In our study, significant effort was placed in minimizing systematic bias and batch effects. Both case and control samples were collected
at a single centre by the same research team, samples were randomized into batches for processing, and a novel calibration method for qPCR genotyping.

The same research group also reported an inverse association between *AMY1* copy number with obesity risk in Mexican children, with similar Odd Ratio at 1.19 but a different allelic architecture from the original study for European adults [Mejia-Benitez, et al., 2015]. In the study of Mejia-Benitez et al. [2015], *AMY1* genotyping was carried out in dPCR using Fluidigm chip with each sample typed in 4x 770 reaction chambers but no pre-dPCR enzyme digestion was indicated. As the *AMY1* gene tandem repeat is about 27 kb, adjacent repeats could co-segregate in high molecular weight preparations of genomic DNA, and cause under-estimation of CN by dPCR. In our hands, dPCR quantification without enzyme digestion resulted in a lower CN call, particularly for samples with high copy numbers (Supp. Fig. S6). These technical differences in methodology could result in differences in results.

The *AMY2A* result in this study corroborated with Falchi et al. (2014), who suggested that the positive association result for *AMY1* identified in Falchi et al. could not be due to linkage with the neighbouring *AMY2A* gene. We found that *AMY2A* copy number is not associated to obesity, consistent with earlier findings in European cohort [Falchi et al. 2014]. Taking all analyses into consideration, this study concludes that *AMY1* and *AMY2A* diploid copy numbers are not associated with obesity in the Chinese sample (while the Malay sample was too small to be statistically powered). A similar conclusion was reached by Usher et al., whose study involved 3 European cohorts, totalling >3,500 subjects and with 99% power to detect the reported effect size in Falchi et al (2014), but found no association of *AMY1* copy number with obesity or BMI [Usher, et al., 2015].
Acknowledgements

We thank Mahesh Uttamchandani for his comments on this manuscript, all volunteers for their participation and provision of samples, and the Singapore Ministry of Defence for funding support. We also thank the anonymous reviewers for the insightful comments and constructive suggestions. We declare no conflict of interest.

References


Figure Legends

Figure 1. Distribution of the diploid copy numbers for AMY2A and AMY1 CNVs. Calibration and Binning of AMY1 CNV. Total sample size 1077 comprising controls and obese from both Chinese and Malays populations.

A: Histogram of AMY2A diploid copy number, direct from qPCR not rounded to integer. Discrete delineation of AMY2A CN into 4 copy classes is visible.

B: Pre-calibrated AMY1 diploid copy number, direct from qPCR not rounded to integer.

C: Calibration curve correlating qPCR ΔΔCT to dPCR Log2(CN) for AMY1 CNV.

D: Post Calibration AMY1 CN. Distribution of the calibrated diploid copy numbers for AMY1 in all 1077 samples. Histogram shows main clusters at even diploid copy numbers and small clusters at odd copy numbers. AMY1 diploid copy number ranges from CN2 to 24.

E: AMY1 calibrated CN sorted in order and binned with pre-defined threshold to nearest integer. Two samples with CN19 and CN24 not shown.
Figure 2. *AMY1* calibrated diploid copy number distribution in 519 Chinese controls and 115 Malay obese. Distribution patterns appeared similar in the 413 Chinese obese and combining all 1077 samples across 2 populations. All displayed the unique even:odd pattern.

*Frequency distribution for HapMap samples combining Chinese from Beijing and Japanese from Tokyo (Total n=197). Obtained from Supplementary table T2, Usher et al. 2015.

§Frequency distribution for HapMap samples combining European and Asians (Total n=209). Obtained from Supplementary table file 1, Carpenter et al. 2015.

Figure 3. Haploid copy number allele frequencies in the 2 Singapore populations. Allele frequencies were estimated through the EM algorithm implemented in CoNVEM using 519 Chinese control and 115 Malay obese samples. A: *AMY1* CNV, B: *AMY2A* CNV.

C. Structure of the human amylase haplotypes; AMY*H1, AMY*H1, and AMY*H2 defined by Groot et al. (Modified from Groot et al. 1989). AMY*H1, AMY*H1, and AMY*H2 are equivalent to AH1, AH3, and AH5, as designated in Usher et al. 2015.

Abbreviation; 2B, *AMY2B*; 2A, *AMY2A*; 1A, *AMY1A*; 1B, *AMY1B*; 1C, *AMY1C*; P1, *AMYP1* pseudogene. The arrow indicates a 100kb homologous sequence which is flanked by segmental duplication listed as SD1 in Supplementary Fig. S1A. (Modified from Groot et al. 1990). aHaplotype frequency in HapMap European samples in Usher et al. 2015. b*AMY1* Haploid CN estimated by CoNEVM in this study as in Fig. 3A.

D. Association of *AMY2A*-AMY1 diploid copy number (CN) in the Chinese sample set. Even diploid copy numbers of *AMY1* tends to associated with *AMY2A* with 2 diploid CN, while *AMY1* odd copy numbers associated with *AMY2A* with CN non-2. Copy numbers of *AMY1* were calibrated by the qPCR-dPCR calibration curve and without rounding to integer. Grey - 519 Chinese controls. Black - 413 Chinese obese samples.
Figure 4. Association testing of AMY1 and AMY2A diploid CN with obesity.

A: AMY1 diploid CN frequency distribution in two Singapore populations. 519 Chinese controls versus 413 obese, and 30 Malay controls versus 115 obese. AMY1 CN is calibrated CN rounded to integer.

B & C: Mann-Whitney test of ∆∆CT in controls, obese, and extreme obese samples. Low ∆∆CT values correspond to high AMY1 copy numbers. B: Chinese samples. C: Malay samples.

D: AMY2A diploid CN frequency distribution in two Singapore populations. 519 Chinese controls versus 413 obese, and 30 Malay controls versus 115 obese.

E & F: Mann-Whitney test of ∆∆CT in controls, obese, and extreme obese samples. Low ∆∆CT values correspond to high AMY2A copy numbers. E: Chinese samples, and F: Malay samples.

Table Legends

Table 1. Summary information of the two study cohorts.

Table 2. Association testing of Amylase genes diploid copy numbers with obesity.

Results of T-test and logistic regression are shown. P-values from Mann-Whitney tests are similarly not significant and results not shown. A: AMY1 CN and obesity. Results were calculated using AMY1 copy numbers calibrated from qPCR-dPCR calibration curve and rounding to integer. B: AMY2A CN and obesity. Discrete AMY2A copy numbers deduced directly from qPCR.
Figure 1. Distribution of the diploid copy numbers for AMY2A and AMY1 CNVs. Calibration and Binning of AMY1 CNV. Total sample size 1077 comprising controls and obese from both Chinese and Malays populations.

A: Histogram of AMY2A diploid copy number, direct from qPCR not rounded to integer. Discrete delineation of AMY2A CN into 4 copy classes is visible.
Fig B: Pre-calibrated *AMY1* diploid copy number, direct from qPCR not rounded to integer.
Fig 1C: Calibration curve correlating qPCR ΔΔCT to dPCR Log2(CN) for AMY1 CNV.

242x168mm (300 x 300 DPI)
Fig 1D: Post Calibration AMY1 CN. Distribution of the calibrated diploid copy numbers for AMY1 in all 1077 samples. Histogram shows main clusters at even diploid copy numbers and small clusters at odd copy numbers. AMY1 diploid copy number ranges from CN2 to 24.

257x80mm (300 x 300 DPI)
Fig 1E: AMY1 calibrated CN sorted in order and binned with pre-defined threshold to nearest integer. Two samples with CN19 and CN24 not shown.

255x191mm (300 x 300 DPI)
Figure 2. *AMY1* calibrated diploid copy number distribution in 519 Chinese controls and 115 Malay obese. Distribution patterns appeared similar in the 413 Chinese obese and combining all 1077 samples across 2 populations. All displayed the unique even:odd pattern. *Frequency distribution for HapMap samples combining Chinese from Beijing and Japanese from Tokyo (Total n=197). Obtained from Supplementary table T2, Usher et al. 2015. §Frequency distribution for HapMap samples combining European and Asians (Total n=209). Obtained from Supplementary table file 1, Carpenter et al. 2015.

![Graph showing frequency distribution for *AMY1* calibrated diploid copy number.](image-url)
Figure 3. Haploid copy number allele frequencies in the 2 Singapore populations. Allele frequencies were estimated through the EM algorithm implemented in CoNVEM using 519 Chinese control and 115 Malay obese samples.

A: *AMY1* CNV,

249x114mm (300 x 300 DPI)
Figure 3. Haploid copy number allele frequencies in the 2 Singapore populations. Allele frequencies were estimated through the EM algorithm implemented in CoNVEM using 519 Chinese control and 115 Malay obese samples.

B: AMY2A CNV.

245x152mm (300 x 300 DPI)
For Peer Review

Fig 3C. Structure of the human amylase haplotypes; AMY*H1, AMY*H1, and AMY*H2 defined by Groot et al. (Modified from Groot et al. 1989). AMY*H1, AMY*H1, and AMY*H2 are equivalent to AH1, AH3, and AH5, as designated in Usher et al. 2015. Abbreviation: 2B, AMY2B; 2A, AMY2A; 1A, AMY1A; 1B, AMY1B; 1C, AMY1C; P1, AMYP1 pseudogene. The arrow indicates a 100kb homologous sequence which is flanked by segmental duplication listed as SD1 in Supplementary Fig. S1A. (Modified from Groot et al. 1990). *Haplotype frequency in HapMap European samples in Usher et al. 2015. †AMY1 Haploid CN estimated by CoNEVM in this study as in Fig. 3A.

250x76mm (300 x 300 DPI)
Fig 3D. Association of AMY2A-AMY1 diploid copy number (CN) in the Chinese sample set. Even diploid copy numbers of AMY1 tends to associated with AMY2A with 2 diploid CN, while AMY1 odd copy numbers associated with AMY2A with CN non-2. Copy numbers of AMY1 were calibrated by the qPCR-dPCR calibration curve and without rounding to integer. Grey - 519 Chinese controls. Black - 413 Chinese obese samples.
Figure 4. Association testing of AMY1 and AMY2A diploid CN with obesity.
A: AMY1 diploid CN frequency distribution in two Singapore populations. 519 Chinese controls versus 413 obese, and 30 Malay controls versus 115 obese. AMY1 CN is calibrated CN rounded to integer.

254x159mm (300 x 300 DPI)
Fig 4B: Mann-Whitney test of ΔΔCT in controls, obese, and extreme obese samples. Low ΔΔCT values correspond to high AMY1 copy numbers. B: Chinese samples
Fig 4C: Mann-Whitney test of ∆ΔCT in controls, obese, and extreme obese samples. Low ∆ΔCT values correspond to high AMY1 copy numbers. C: Malay samples.
Fig D: *AMY2A* diploid CN frequency distribution in two Singapore populations. 519 Chinese controls versus 413 obese, and 30 Malay controls versus 115 obese.

272x178mm (300 x 300 DPI)
Fig 4E: Mann-Whitney test of ΔΔCT in controls, obese, and extreme obese samples. Low ΔΔCT values correspond to high AMY2A copy numbers. E: Chinese samples
Fig 4F: Mann-Whitney test of $\Delta\Delta$CT in controls, obese, and extreme obese samples. Low $\Delta\Delta$CT values correspond to high AMY2A copy numbers. F: Malay samples.

97x108mm (300 x 300 DPI)
### Table 1. Summary information of the two study cohorts.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Chinese</th>
<th></th>
<th>Malay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Extremely Obese</td>
<td>Control</td>
</tr>
<tr>
<td>Number</td>
<td>519</td>
<td>413</td>
<td>173</td>
</tr>
<tr>
<td>Average Age (SD)</td>
<td>19.4 (1.1)</td>
<td>19.2 (1.2)</td>
<td>19.3 (1.1)</td>
</tr>
<tr>
<td>Average BMI in kg/m² (SD)</td>
<td>20.8 (2.0)</td>
<td>34.2 (3.6)</td>
<td>37.5 (2.7)</td>
</tr>
<tr>
<td>Min to Max BMI</td>
<td>15.8 - 25.2</td>
<td>28.1 - 49.3</td>
<td>32.5-49.3</td>
</tr>
<tr>
<td>Median BMI (1st-3rd quartiles)</td>
<td>21.0 (19.3-22.3)</td>
<td>33.5 (31.4-36.7)</td>
<td>37.2 (35.6-38.6)</td>
</tr>
</tbody>
</table>

For Peer Review

Chinese Malay
Sample Control Obese Extreme Obese Control Obese Extreme Obese
Number 519 413 173 30 115 64
Average Age (SD) 19.4 (1.1) 19.2 (1.2) 19.3 (1.1) 19.7 (0.9) 19.4 (1.1) 19.3 (1.1)
Average BMI in kg/m² (SD) 20.8 (2.0) 34.2 (3.6) 37.5 (2.7) 20.8 (2.2) 36.2 (4.8) 39.3 (4.2)
Min to Max BMI 15.8 - 25.2 28.1 - 49.3 32.5-49.3 17.0-24.7 28.7-51.5 33.8-51.5
Median BMI (1st-3rd quartiles) 21.0 (19.3-22.3) 33.5 (31.4-36.7) 37.2 (35.6-38.6) 20.5 (19.1-22.2) 35.1 (32.6-39.1) 38.8 (35.8-41.1)
Table 2. Association testing of Amylase genes diploid copy numbers with obesity. Results of T-test and logistic regression are shown. P-values from Mann-Whitney tests are similarly not significant and results not shown. A: AMY1 CN and obesity. Results were calculated using AMY1 copy numbers calibrated from qPCR-dPCR calibration curve and rounding to integer. B: AMY2A CN and obesity. Discrete AMY2A copy numbers deduced directly from qPCR.
COMPLEX COPY NUMBER VARIATION OF *AMY1* DOES NOT ASSOCIATE WITH OBESITY IN TWO EAST ASIAN COHORTS

Rita YY Yong¹,², Su’Aidah B. Mustaffa¹,³, Pavandip Singh Wasan¹,², L Sheng⁵, C Marshall⁶, S Scherer⁶, YY Teo²,⁴, Eric PH Yap²,³,*

¹ Defence Medical and Environmental Research Institute, DSO National Laboratories, Singapore,
² Saw Swee Hock School of Public Health, National University of Singapore, Singapore,
³ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore,
⁴ Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore,
⁵ Unit of Biostatistics, Yong Loo Lin School of Medicine, National University of Singapore,
⁶ The Centre of Applied Genomics, Hospital of the Sick Children, Toronto, Canada

* Corresponding Author
Supplementary Notes for Methods

gPCR & dPCR Protocols

Each qPCR reaction is carried out in a final volume of 10 or 4 µl. Each 10 µl aliquot would contain 10 ng genomic DNA, 0.5 µl each of the Taqman target assay (20x) and the reference assay RNRaseP (20x), 5 µl of Taqman genotyping master mix (2x), and 3 µl of sterile water. Cycling condition consisted of one cycle of 95ºC 10 min, followed by 40 cycles of 95ºC 15 sec, 60ºC 1 min.

A comparative Ct method was used in estimating diploid copy number. The ∆∆Ct method assumes equal amplification efficiency for the target and the reference genes. PCR efficiency was evaluated for both the target assays (AMY1 & AMY2A) and the reference RNRaseP assay using a standard curve of serial dilutions of a DNA sample of known concentration. Log10 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:

\[ E = 10^{(-\text{slope})} \]

PCR Efficiency (%) = (E – 1) x 100

For dPCR protocol, the human genomic DNA was first digested with an appropriate restriction enzyme to break up the tandemly repeated CNV copies. 4 µl of genomic DNA (100ng/µl) was digested with 1µl RsaI (10U/µl) (New England Biolabs, USA), with 1 µl of 10x enzyme buffer in a final volume of 10 µl. Digestion was carried out at 37ºC for 1h followed by enzyme inactivation at 65ºC for 20mins. Digested DNA was diluted 4x to a final volume of 40µl. 4ul of the diluted DNA (10ng/µl) was used in each digital-PCR reaction, which would contain 8µl of the 2x digital-PCR master mix, 0.8µl of 20x target Taqman assay, 0.8µl of 20x reference RNaseP assay, and 2.4µl sterile water to constitute a final volume of 16µl. About 14.5µl of the dPCR reaction product was loaded per
chip. Thermal cycling condition was 1 cycle of 96°C 10 min, followed by 39 cycles of 60°C 2 min
98°C 30sec, and a last cycle of 60°C 2 min.

Leave-one-out Analysis to test sensitivity of the qPCR-dPCR calibration curve

A leave-one-out analysis was carried out to test the sensitivity of the calibration curve if the number
of sample used in constructing the calibration curve is reduced. One sample from the total 31 samples
was removed at a time, the calibration curve was re-calculated, and the newly derived calibration
curve was used to translate the qPCR CN calls to the calibrated CN calls. Of the total 31 leave-one-
out experiments, 6 (19%) showed a reduction of accuracy. In each of these 6 leave-one-out
experiments, discordance between dPCR CN and calibrated CN increased by 1.

In addition, an early analysis had been carried out using a calibration curve established from 18
samples. The table below contrasts the AMY1 CN frequency distribution of the 1077 samples using
the calibration curves established with 18 samples or 31 samples. There were slight difference in
frequency but the difference is not significant (Fisher Exact test two-sided p-value = 0.103).

Association testing using the calibrated CN calls derived from the 18 samples calibration curve
similarly produced a negative association result.

<table>
<thead>
<tr>
<th>AMY1 CN</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>18</th>
<th>19</th>
<th>24</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 samples</td>
<td>Calibration curve</td>
<td>0.3%</td>
<td>0.4%</td>
<td>6.5%</td>
<td>2.5%</td>
<td>24.7%</td>
<td>6.6%</td>
<td>25.3%</td>
<td>6.3%</td>
<td>15.5%</td>
<td>4.5%</td>
<td>4.4%</td>
<td>1.5%</td>
<td>1.2%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td>31 samples</td>
<td>Calibration curve</td>
<td>0.3%</td>
<td>0.4%</td>
<td>6.5%</td>
<td>1.9%</td>
<td>24.8%</td>
<td>3.8%</td>
<td>26.6%</td>
<td>3.1%</td>
<td>16.4%</td>
<td>4.3%</td>
<td>6.1%</td>
<td>1.5%</td>
<td>0.8%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Statistical analysis

Student’s t-test, Mann-Whitney test and Correlation test were carried out in GraphPad Prism 6.
Fisher’s exact test, Cochran-Armitage trend test and power calculation were done in Excel XLSTAT.
Logistic and linear regressions were carried out using SPSS v15.0.
Supplementary Figures and Tables

Supp. Figure S1A. Schematic representation of α-amylase gene cluster on chromosome 1.2p1.1. Genome co-ordinates according to assembly NCBI37/hg19. (modified from Santos et al. 2012)

A γ-actin processed pseudogene was inserted in 5’ position of all AMY genes except the pseudogene AMYP1. The amount of insertions of the retroviral element is marked with +. Approximate locations of two segmental duplications are indicated, which have implication in generating the more common haplotype structures at this AMY locus. The approximate location of the 2 qPCRs for AMY1 and AMY2A are indicated.
**Supp. Figure S1B.** Human amylase gene cluster on chromosome 1.21p.1 as displayed in UCSC Genome Browser. Genome assembly is NCBI36/hg18. Positions of AMY2B, AMY2A, AMY1A, AMY1B, AMYP1 and AMY1C are shown. Two groups of CNVs are identified by SNP array in DMERI database, overlapping AMY2A and AMY1, of sizes 50 kb and 100 kb, respectively. CN2 and CN6 were non-CNV for AMY2A and AMY1, respectively. Microarray results underestimated the extent of CN variation for both CNVs, especially for AMY1. Chinese in red, Malays in green, and Indians in blue lines representing CNV locations, with each line representing one sample. Segmental duplications (SDs) are shown in lowest section. Orange SD, > 99% sequence similarity, the long orange SD corresponds to SD1 of Figure 1 in main paper. Light to dark yellow, 98 - 99% similarity, corresponding to SD2. Light to dark gray, 90 - 98% similarity.
**Supp. Figure S2.** Real-time PCR standard curve representing PCR efficiency. A: Taqman assay hs07226362_cn targeting AMY1 and the RNaseP reference assay. B: taqman assay hs04204136_cn targeting AMY2A and the RNaseP reference assay.

**Supp. Figure S3.** Distribution of the diploid copy numbers for AMY2A CNV. Total sample size 1077 comprising controls and obese from both Chinese and Malays populations. AMY2A CN sorted in order to show the discrete delineation of the 4 copy classes. Correspond to Figure 2A in main text.
Supp. Figure S4. AMY1 diploid copy number determined by dPCR for 32 samples, four of which were cell line reference samples from Coriell Repository. Error bar represents standard deviation. Number of replicates per sample as in Supplementary Table 2.
Supp. Figure S5. Average BMI at different AMY1 diploid copy numbers. Error bar represents standard deviation. Some points with no error bar because they are singletons. A: 519 Chinese controls versus 413 obese. B: 30 Malay controls versus 115 obese.
Supp. Figure S6. Comparison of dPCR performance with or without restriction enzyme Rsa I digestion. Each dot represents one dPCR run. Grey - With digestion. Black - No digestion.
Supp. Table S1. Characteristics of Taqman® copy number assays targeting AMY1 and AMY2A genes. Genome co-ordinate according to NCBI37/hg19.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Assay ID</th>
<th>Cytogenetic Band</th>
<th>Chromosome Location on NCBI Genome Assembly</th>
<th>Location on Transcript or Gene</th>
<th>Reporter Dye</th>
<th>Quencher</th>
<th>Amplicon size (bp)</th>
<th>Context Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMY1A, AMY1B, AMY1C</td>
<td>Hs07226362_cn</td>
<td>1p21.1</td>
<td>1 104,158,366, 104,238,841, 104,292,504</td>
<td>Within Exon 1</td>
<td>FAM</td>
<td>NFQ</td>
<td>101</td>
<td>ATGTTGTCAGGGCAGTTGTTCTGAG</td>
</tr>
<tr>
<td>AMY2A</td>
<td>Hs04204136_cn</td>
<td>1p21.1</td>
<td>1 104,161,065</td>
<td>Within Intron 2</td>
<td>FAM</td>
<td>NFQ</td>
<td>111</td>
<td>TTAGGGTCACTTGCTCTCATCCTGT</td>
</tr>
<tr>
<td>RPPH1</td>
<td></td>
<td>14q11.2</td>
<td>14 20,811,565</td>
<td>within single exon</td>
<td>VICTAMRA</td>
<td>87</td>
<td></td>
<td>GAGCTTCCCTCGGCGCTGGAAA</td>
</tr>
</tbody>
</table>

Supp. Table S2. Comparison of dPCR and qPCR results for 28 samples and 3 reference specimens.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Replicate</th>
<th>CN Mean</th>
<th>CN SD</th>
<th>Log2 (CN)</th>
<th>Replicate</th>
<th>ΔΔCt</th>
<th>ΔΔCt SD</th>
<th>CN Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>OB0273</td>
<td>3</td>
<td>1.95</td>
<td>0.03</td>
<td>0.96</td>
<td>4</td>
<td>1.84</td>
<td>-0.11</td>
<td>1.68</td>
</tr>
<tr>
<td>OB1189</td>
<td>5</td>
<td>2.97</td>
<td>0.10</td>
<td>1.57</td>
<td>4</td>
<td>1.14</td>
<td>0.03</td>
<td>2.71</td>
</tr>
<tr>
<td>OB0662</td>
<td>5</td>
<td>4.08</td>
<td>0.11</td>
<td>2.03</td>
<td>4</td>
<td>0.61</td>
<td>0.01</td>
<td>3.93</td>
</tr>
<tr>
<td>OB0562</td>
<td>2</td>
<td>4.95</td>
<td>0.03</td>
<td>2.31</td>
<td>4</td>
<td>0.26</td>
<td>0.01</td>
<td>5.01</td>
</tr>
<tr>
<td>NA18956</td>
<td>12</td>
<td>5.31</td>
<td>0.13</td>
<td>2.41</td>
<td>48</td>
<td>0.13</td>
<td>0.01</td>
<td>5.50</td>
</tr>
<tr>
<td>NA10851</td>
<td>11</td>
<td>5.85</td>
<td>0.14</td>
<td>2.50</td>
<td>48</td>
<td>0.00</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td>OB0984</td>
<td>4</td>
<td>6.16</td>
<td>0.12</td>
<td>2.62</td>
<td>4</td>
<td>-0.07</td>
<td>0.00</td>
<td>6.29</td>
</tr>
<tr>
<td>OB0191</td>
<td>2</td>
<td>6.21</td>
<td>0.10</td>
<td>2.63</td>
<td>4</td>
<td>-0.01</td>
<td>0.00</td>
<td>6.05</td>
</tr>
<tr>
<td>OB0129</td>
<td>2</td>
<td>6.78</td>
<td>0.19</td>
<td>2.76</td>
<td>4</td>
<td>-0.33</td>
<td>0.00</td>
<td>7.52</td>
</tr>
<tr>
<td>271</td>
<td>2</td>
<td>6.96</td>
<td>0.22</td>
<td>2.80</td>
<td>48</td>
<td>-0.38</td>
<td>0.01</td>
<td>7.82</td>
</tr>
<tr>
<td>OB0978</td>
<td>2</td>
<td>7.32</td>
<td>0.09</td>
<td>2.87</td>
<td>4</td>
<td>-0.59</td>
<td>0.01</td>
<td>9.01</td>
</tr>
<tr>
<td>OB1096</td>
<td>2</td>
<td>7.83</td>
<td>0.03</td>
<td>2.97</td>
<td>4</td>
<td>-0.74</td>
<td>0.02</td>
<td>9.99</td>
</tr>
<tr>
<td>OB1125</td>
<td>2</td>
<td>7.93</td>
<td>0.15</td>
<td>2.99</td>
<td>4</td>
<td>-0.58</td>
<td>0.02</td>
<td>8.97</td>
</tr>
<tr>
<td>OB0561</td>
<td>2</td>
<td>8.06</td>
<td>0.10</td>
<td>3.01</td>
<td>4</td>
<td>-0.66</td>
<td>0.02</td>
<td>9.50</td>
</tr>
<tr>
<td>OB0076</td>
<td>2</td>
<td>8.18</td>
<td>0.17</td>
<td>3.03</td>
<td>4</td>
<td>-0.50</td>
<td>0.00</td>
<td>8.50</td>
</tr>
<tr>
<td>OB0395</td>
<td>3</td>
<td>9.69</td>
<td>0.30</td>
<td>3.28</td>
<td>4</td>
<td>-0.84</td>
<td>0.06</td>
<td>10.76</td>
</tr>
<tr>
<td>OB1216</td>
<td>3</td>
<td>9.87</td>
<td>0.19</td>
<td>3.30</td>
<td>4</td>
<td>-0.75</td>
<td>0.01</td>
<td>10.12</td>
</tr>
<tr>
<td>OB0768</td>
<td>2</td>
<td>10.03</td>
<td>0.03</td>
<td>3.33</td>
<td>4</td>
<td>-1.13</td>
<td>0.05</td>
<td>13.16</td>
</tr>
<tr>
<td>OB0130</td>
<td>2</td>
<td>10.11</td>
<td>0.07</td>
<td>3.34</td>
<td>4</td>
<td>-0.94</td>
<td>0.05</td>
<td>11.48</td>
</tr>
<tr>
<td>OB0246</td>
<td>4</td>
<td>10.19</td>
<td>0.31</td>
<td>3.35</td>
<td>4</td>
<td>-0.99</td>
<td>0.03</td>
<td>11.88</td>
</tr>
<tr>
<td>OB0357</td>
<td>4</td>
<td>11.81</td>
<td>0.58</td>
<td>3.56</td>
<td>4</td>
<td>-1.37</td>
<td>0.05</td>
<td>15.50</td>
</tr>
<tr>
<td>OB1054</td>
<td>4</td>
<td>12.04</td>
<td>0.31</td>
<td>3.59</td>
<td>4</td>
<td>-1.30</td>
<td>0.01</td>
<td>14.81</td>
</tr>
<tr>
<td>OB0513</td>
<td>5</td>
<td>12.36</td>
<td>0.21</td>
<td>3.63</td>
<td>4</td>
<td>-1.55</td>
<td>0.02</td>
<td>17.60</td>
</tr>
<tr>
<td>OB0413</td>
<td>4</td>
<td>13.38</td>
<td>0.37</td>
<td>3.74</td>
<td>4</td>
<td>-1.94</td>
<td>0.06</td>
<td>23.07</td>
</tr>
<tr>
<td>OB0379</td>
<td>4</td>
<td>14.15</td>
<td>0.55</td>
<td>3.82</td>
<td>4</td>
<td>-1.70</td>
<td>0.02</td>
<td>19.51</td>
</tr>
<tr>
<td>OB0752</td>
<td>9</td>
<td>14.64</td>
<td>0.42</td>
<td>3.87</td>
<td>4</td>
<td>-2.05</td>
<td>0.04</td>
<td>24.81</td>
</tr>
<tr>
<td>OB1170</td>
<td>9</td>
<td>15.39</td>
<td>0.47</td>
<td>3.94</td>
<td>4</td>
<td>-1.98</td>
<td>0.01</td>
<td>23.64</td>
</tr>
<tr>
<td>OB0479</td>
<td>5</td>
<td>15.58</td>
<td>0.82</td>
<td>3.96</td>
<td>4</td>
<td>-1.91</td>
<td>0.04</td>
<td>22.48</td>
</tr>
<tr>
<td>NA18972</td>
<td>21</td>
<td>15.96</td>
<td>0.63</td>
<td>4.00</td>
<td>48</td>
<td>-2.25</td>
<td>0.01</td>
<td>28.63</td>
</tr>
<tr>
<td>OB0433</td>
<td>10</td>
<td>18.73</td>
<td>0.44</td>
<td>4.23</td>
<td>4</td>
<td>-2.55</td>
<td>0.02</td>
<td>35.13</td>
</tr>
<tr>
<td>OB0350</td>
<td>17</td>
<td>24.61</td>
<td>1.13</td>
<td>4.62</td>
<td>4</td>
<td>-3.37</td>
<td>0.06</td>
<td>62.17</td>
</tr>
</tbody>
</table>

John Wiley & Sons, Inc.
**Supp. Table S3.** Association testing of *AMY2A* and *AMY1* diploid copy numbers. a) 519 Chinese controls. b) 413 Chinese obese samples. c) 30 Malay controls. d) 115 Malay obese samples.