International Journal of Obesity (2013) **37**, 559–567

www.nature.com/ijo

REVIEW

Challenges in reproducibility of genetic association studies: lessons learned from the obesity field

A Li and D Meyre

A robust replication of initial genetic association findings has proved to be difficult in human complex diseases and more specifically in the obesity field. An obvious cause of non-replication in genetic association studies is the initial report of a false positive result, which can be explained by a non-heritable phenotype, insufficient sample size, improper correction for multiple testing, population stratification, technical biases, insufficient quality control or inappropriate statistical analyses. Replication may, however, be challenging even when the original study describes a true positive association. The reasons include underpowered replication samples, gene × gene, gene × environment interactions, genetic and phenotypic heterogeneity and subjective interpretation of data. In this review, we address classic pitfalls in genetic association studies and provide guidelines for proper discovery and replication genetic association studies with a specific focus on obesity.

International Journal of Obesity (2013) **37,** 559–567; doi:10.1038/ijo.2012.82; published online 15 May 2012

Keywords: genetic epidemiology; false-positive association; replication; guidelines

INTRODUCTION

The prevalence of obesity has reached epidemic proportion across the world.[1] Genetic influence has a substantial role in this disease,[2] and has led scientists to search for the specific genetic determinants of human obesity. Candidate gene and linkage approaches have given rise to the discovery of several genes associated with monogenic or syndromic obesity, but provided limited success to identifying genetic variants that are associated with polygenic obesity.[3] This situation was not exclusive to obesity, as most reported associations were not conclusive in human complex diseases.[4] These observations raised concerns about the reliability of such genetic association studies, and high-impact publications such as the *Journal of Clinical Investigation* even stated in their editorial policy that genetic linkage and association studies related to complex disorders were unlikely to be accepted (http://www.jci.org/kiosk/publish/policies). This situation has changed recently with new high-throughput single-nucleotide polymorphism (SNP) genotyping array technologies enabling genome-wide association studies (GWAS),[5] and with the establishment of international consortiums providing sample size large enough to identify gene variants with modest effect size (odds ratio (OR) < 2).[6] More than 1000 loci have now been convincingly associated with human complex traits using GWAS and close to 70 gene variants have been associated with obesity-related traits at a stringent level of significance ($P < 5 \times 10^{-8}$).[7] However, despite the remarkable progress made during the last decade in the genetic epidemiology field, several questions remain.[8] How can we explain that despite promising original association reports, some genetic associations are widely replicated whereas others are not? When can we be confident that a positive finding is truly positive? How can we differentiate good genetic association studies from inadequate ones? We searched literature using the keywords 'genetic AND obesity' (11 152) or 'genetic association studies' (47 368) in the PubMed NCBI database from January 1987 to February 2012. We selectively cited the more illustrative papers to address the classic pitfalls in genetic association studies and to provide guidelines for proper genetic association discovery and replication studies, specifically focused on obesity.

LACK OF REPLICATION DUE TO FALSE POSITIVE RESULT REPORTED IN ORIGINAL STUDY

Genetically complex diseases involve a large number of genetic, environmental factors and their interactions. In general, common risk alleles are less deterministic and more probabilistic than rare monogenic mutations, and only slightly increase the chance of disease. They are found in affected but also in unaffected people in populations.[4] Finding an association between such a variant and disease is statistic by essence and the risk of reporting an initial positive association by chance can never be totally excluded. However, the risk of reporting a false positive association is more likely to occur in certain circumstances as listed below.

The phenotype is not heritable

Heritability reflects the proportion of total phenotypic variability caused by genetic variance in a population. When genetic variation refers to additive genetic effect only, heritability is named narrow-sense heritability or just heritability (h^2); when genetic variation includes all additive, dominant and epistatic genetic effects, heritability is defined as broad-sense heritability (H^2).[9] Adoption and twin studies are the best study designs for heritability estimation because of their natural separation of genetic and environmental components, whereas family parent–offspring and sibling studies estimate less accurately the heritability owing to their partial similarity in genes and the same degree of environmental similarity.[9]

A phenotype with a substantial heritability is an important prerequisite to enable the identification of genetic determinants.[10]

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. Correspondence: Dr D Meyre, Department of Clinical Epidemiology and Biostatistics, McMaster University, Michael DeGroote Centre for Learning and Discovery, Room 3205, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada.
E-mail: meyred@mcmaster.ca
Received 19 December 2011; revised 13 April 2012; accepted 15 April 2012; published online 15 May 2012

Heritability studies have shown that obesity, body mass index (BMI) and BMI change are highly heritable traits (h^2 is 50–80%).[2,11,12] Intermediary endophenotypes related to obesity, such as percentage of body fat,[13] eating behavior[14] and energy expenditure,[15] have also a significant heritable component.

Insufficient sample size

A meta-analysis of 37 linkage studies for BMI did not detect significant linkage regions (LOD score > 3.6),[16] suggesting that multiple susceptibility variants with small effect sizes may affect obesity predisposition. Evidence from the variants consistently associated with obesity to date confirms this view.[7] To the rare exception of the *MC4R* I251L or *GPR120* R270H infrequent coding variants,[17,18] polymorphisms associated with obesity display modest OR values typically lower than 1.50.[7] This implies that large sample sizes are needed in case–control studies designed for gene discovery efforts (Table 1). Such large sample sizes have not been reported in obesity studies until recently,[19,20] suggesting that most of the association studies reported before were underpowered to identify the effect sizes expected for obesity predisposing common variants. As small discovery studies tend to overestimate the true effect of gene variants,[21] underpowered case–control studies are likely to prompt false positive associations and only few associations first reported from smaller studies prove to be reliable in larger replication data sets.[21] The *GAD2* gene story is illustrative of this. In 2003, an initial study reported an association between the *GAD2* -243A > G promoter variant and obesity in a case–control design of 575 morbidly obese and 646 control French subjects.[22] This sample was considered substantial at that time, but was far less than the sample requirements predicted by power calculation studies for an OR of 1.3 and a risk allele frequency of 0.17 in controls (Table 1). Logically, a replication effort in 680 class III obesity cases and 1186 lean controls of European ancestry did not replicate the original finding[23] and an adequately powered meta-analysis recapitulating the data of the two studies ($n = 3052$) did not find any association ($P = 0.28$).[23]

Lack of proper correction for multiple testing

To account for multiple testing, researchers typically apply Bonferroni correction in which a threshold P of 0.05/n ($n =$ number of statistical tests performed in the study) is set. For instance, in a GWAS using 1×10^6 genetic markers, a threshold P of 5×10^{-8} is needed to declare an association at the corrected threshold of type I error of 5%. Testing multiple hypotheses in one study (multiple genetic markers, phenotypes and subgroup analyses) without proper adjustment for multiple comparisons increases the risk of a false positive report. In the original report showing an association between rs7566605 in *INSIG2* gene and BMI in 694 participants from the Framingham Heart Study ($P = 0.0026$), the lack of correction for multiple testing (a Bonferroni adjusted P-value threshold of 5.7×10^{-7} was needed after performing 86604 statistical tests)[24] in combination with the insufficient sample size of the discovery cohort may explain the lack of replication of this association in three well-powered follow-up studies,[25–27] in a meta-analysis of 74 345 subjects[28] and more recently in the large-scale GIANT GWAS initiative,[29] even if others have pointed out a context dependent association of the rs7566605 with obesity as an alternative explanation.[28,30]

Table 1. Sample sizes needed in a case–control design to detect significant association with a power of 90% and a two-sided P of 0.001 by odds ratio and allele frequency for risk allele

MAF in control	*0.01*	*0.05*	*0.1*	*0.2*	*0.3*	*0.4*
Allelic OR						
1.1	443 854	92 868	49 252	27 974	21 518	19 010
1.2	116 354	24 434	13 018	7460	5792	5162
1.3	54 110	11 404	6102	3526	2760	2480
1.5	21 208	4498	2426	1424	1132	1032
2.0	6386	1374	754	458	376	354

Abbreviations: MAF, minor allele frequency: OR, odds ratio. Calculations assume multiplicative effect on disease risk. Sample sizes presented are total number of cases and controls needed, assuming an equal number of cases and controls. The estimated sample sizes are derived assuming power = 90% and two-sided $P = 0.001$.

Population stratification and confounding

In a case–control study, controls should ideally be selected from the same population as cases. Allele frequency of genetic variants can vary depending on the ethnic background or even the geographical location. When cases and controls are drawn from multiple ethnic or geographic groups whose phenotype prevalence and allele frequencies differ for the gene variants tested, confounding unintentionally occurs, leading to false positive associations.[4,8,31] This is likely the case for obesity predisposing variants, as several of them show highly varying levels of inter-ethnic or inter-geographic allele frequency variation, possibly due to positive diversifying selection (for example, *ENPP1* rs1044498, *FTO* rs9939609 or *LCT* rs4988235).[32–34] In addition, prevalence of obesity strongly varies according to the ethnic background in a specific country.[35] Even if such confounding seems to have modest consequences in case–control studies, it may generate spurious association signal if not properly taken into account, especially in large-scale association studies.[36] GWAS for obesity-related traits have reported modest but real evidence of population stratification.[6,19]

Genotype misclassification and reliability in quality control

Genotyping errors are frequent in genetic association studies,[37] leading to non-differential genotype misclassifications (same probability of being misclassified for all study subjects) or differential genotype misclassifications (varying probability of being misclassified according to the study groups). A 1% increase in genotyping errors will add up sample size by 2–8% to keep the same type I and II errors.[38] Genotyping errors from batch to batch, laboratory to laboratory or preferential rejection of particular genotypes (usually heterozygotes) can result in differential genotype misclassifications and significant differences between case and control groups, leading to false positive association. A recent large-scale family-based study using TaqMan technology excluded a role of VNTR *INS* polymorphism in childhood obesity[39] despite previous positive association using PCR-based restriction fragment length polymorphism (RFLP).[40] As the reproducibility of RFLP genotyping data has been questioned, this method being highly subjective,[41] the authors suggested that the lack of replication may be a result of previous genotype misclassification from the RFLP method.[39]

Inappropriate statistical analyses

Numerous studies use the χ^2 or Fisher exact test to estimate the difference of genotype distribution in cases and controls. Linear/logistic regression is therefore a more relevant approach to make comparisons between groups as it allows adjustment for confounding or mediating factors, and introduction of gene × gene or gene × environment interaction terms.[42] The *FTO* identification study by Frayling *et al.*[43] is illustrative of the importance of properly adjusted association studies for confounding or mediating factors. The rs9939609 in *FTO* gene was found initially by GWAS for type 2 diabetes. This strong

association disappeared after further adjustment in the linear regression model for BMI, suggesting that the *FTO* intron 1 variant predisposition for diabetes was mediated through an effect on BMI.[43]

In some association studies, when no expected main effect is observed, researchers tend to perform post-hoc subgroup analyses to dredge some apparent positive association.[31] Positive association drawn from such analyses should be viewed with caution.

REPLICATION MAY BE CHALLENGING EVEN WHEN THE ORIGINAL RESULT IS A TRUE POSITIVE ASSOCIATION

There is a consensus that a single study, even when well-designed and powered, is not sufficient to establish a valid genotype–phenotype association and further replication is needed. The main interest of replication is not only to increase the statistical significance of the result but also to prove that the initial association is true and not the result of methodological inadequacies.[44] Given that many published associations could not be replicated, several high-impact journals such as *Nature Genetics* stipulate in their publication guidelines that replication of an original association in a second independent cohort is requested prior to consideration for publication.[45] The *FTO* gene identification story is illustrative of the interest of independent replications to be confident about a genetic discovery. In 2007, four different teams independently discovered that variation in the intron 1 of fat mass- and obesity-associated (*FTO*) gene is the major contributor to polygenic obesity in populations of European ancestry.[43,46–48] However, replication does not always follow this ideal route, and the lack of association in follow-up studies does not automatically imply that the initial study did not pick-up a true association signal. Below we describe the main reasons why replication of a true association may be challenging.

Sample size

A well-powered replication design is crucial to avoid the risk of both false negative and false positive association reports.[49] Power estimation of the replication samples has become an issue with the emergence of large-scale discovery studies through international consortiums. The current largest GIANT meta-analysis of 249 796 individuals in obesity association study has confirmed 10 previously identified loci associated with BMI and 4 loci associated with waist circumference, and revealed 18 new loci associated with BMI.[50] Among these 32 loci, rs1558902 (*FTO*) explains 0.34% of variance in BMI, and the remaining loci harbor smaller effect size. A recent study by Zhao *et al.*[51] examined all 32 loci identified from GIANT study in 1097 childhood obesity cases and 2760 controls, but the associations of only 9 SNPs out of 32 were confirmed. Phenotypic heterogeneity (adult BMI versus childhood obesity) may explain why some genetic variants are not replicated, but beyond this limitation the Zhao *et al.*[51] study illustrates the difficulty to replicate association signals with small effect size issued from large GWAS meta-analyses. With the sample size used in this study, the statistical power was 99% to detect an OR of 1.5, 73% to detect an OR of 1.15, but only 19% to detect an OR of 1.05 (minor allele frequency = 0.2, one-sided $\alpha = 0.05$), whereas the range of the effect size of obesity predisposing variants is typically between 1.05 and 1.5.

Gene × gene, gene × environment interaction

Gene × gene or gene × environment interactions specific to the discovery sample may challenge the ability to replicate the initial discovery in independent study designs.[52] If the other genetic and relevant environmental factors are not balanced in populations between initial and subsequent studies, the variance in effect size from gene × gene and gene × environment interactions may in part explain the inability in replicating the association. Statistical power issues are critical in gene × gene or gene × environment interaction studies and meta-analytic approaches are recommended to reach sufficient power.[53,54] Different study designs enable the study of gene × gene or gene × environment interactions, but power calculations have shown that population-based or nested case–control studies have a higher capacity in detecting interactions than retrospective case–control studies.[55–57] Case–control studies may be relevant in detecting gene × gene, gene × environment interactions in specific situations, as several parameters (for example, allele frequency, magnitude of interactions, genetic model and linkage disequilibrium with the causal variant) impact the statistical power.[55,58] Convincing evidence of gene × gene or gene × environment starts to emerge in the field of obesity. Interaction between the *FTO* intron 1 variant, the level of physical activity and BMI or obesity has been described in 14 independent studies and confirmed in a meta-analysis of 218 166 adults where physical activity attenuated the odds of obesity conferred by the variant by 27%.[59] A recent study conducted in 62 245 east Asian subjects identified a significant gene-gene interaction ($P = 2.0 \times 10^{-8}$) between two SNPs in the *KLF9* and *MSTN* loci.[60] These data support the existence of gene × gene and gene × environment interactions that may interfere with the replication of an initial association.

Heterogeneity

Genetic/ allelic heterogeneity exists where multiple alleles at the same locus are associated with the same disease, so that different studies may find different alleles associated with the same disease. Linkage disequilibrium varies with the ethnic background and most of the disease-associated SNPs issued from GWAS are proxy (a genetic marker of the disease in a specific population) but not causal, which results in misclassification of the functional risk alleles. Some SNPs are causal but may be associated in an ethnic-specific manner.[61] If a negative replication is obtained in a different population, this might lead to reject a true association in that specific population as a consequence of ignoring allelic heterogeneity. Another consequence of genetic heterogeneity is that different polymorphisms can be associated with obesity-related traits by independent teams, and it can puzzle researchers when they select genetic markers for replication studies. In 2007, four SNPs in high linkage disequilibrium (rs9939609, rs1421085, rs9930506, rs1121980) were described as lead signals for the association between *FTO* and obesity by different teams.[43,46–48] As no consensus has been made to select a unique lead SNP for replication, follow-up studies using any of the four SNPs have been published, puzzling subsequent meta-analysis initiatives.[62]

The lack of replication in follow-up studies may sometimes be explained by phenotypic heterogeneity (different phenotypes used in the discovery and the follow-up studies). For instance, BMI can be studied as a quantitative or a categorical trait. Several thresholds can be used to define the status (overweight, class I, II or III obesity). Although highly correlated, these different phenotypes cannot be considered as identical. Following the initial report of association between the *GAD2* -243 A > G variant and morbid adult obesity among the French,[22] Groves *et al.*[63] assessed its role in 573 UK pedigrees ascertained for type 2 diabetes, and found a nominal but directionally inconsistent association between *GAD2* -243A > G and BMI. The authors concluded that their data 'did not replicate the previous associations' while acknowledging the phenotypic heterogeneity in the two studies.[63]

Depending on the study design and recruitment strategy, ascertainment heterogeneity can partly explain the lack of consistency between initial and replication association studies. Even if the same phenotype is used to classify subjects as cases

(for example, BMI $\geqslant 30\,kg\,m^{-2}$), different subpopulations of patients may be collected according to the recruitment procedure (population-based, pedigrees selected for a family history of obesity, hospital, type 2 diabetes cases, incident cases). The age window of the collected sample is also crucial. Obesity in childhood, adolescence, adulthood and agedness are likely to harbor a complex pattern of inheritance, as recently shown for *FTO* intron 1 gene variant and BMI variation.[64]

Inheritance model

Information on the inheritance model is useful in replication studies or biological experiments.[65] When a formal procedure to attribute the best-fitting inheritance model for obesity was applied to the *ENPP1* K121Q gene variant, a significant departure from the additive to the recessive model was observed,[66] and later confirmed for the association between K121Q SNP and type 2 diabetes.[67] Association under a different genetic model from that reported in the original study cannot be considered as a formal replication.

Subjective interpretation of data

Subjective overstatement or understatement in the interpretation of replication data can be observed.[68] Researchers may over-interpret positive findings in the follow-up studies when attempting to replicate their own original findings, or negatively interpret follow-up studies when attempting to refute their competitors' original claim of association, especially if the original claim was published in a prestigious journal (the Proteus phenomenon).[69]

GUIDELINES FOR PROPER DISCOVERY AND REPLICATION ASSOCIATION STUDY DESIGNS

Discovery association study designs

Study designs. Four designs are classically used in genetic association studies: prospective cohort, case–control, family-based and quantitative trait association study. The prospective cohort study is often considered as the gold standard in epidemiology, but the rates of collecting disease cases (other than common ones) and their follow-up are slow. Consequently, the statistical power is usually weak.[49]

A nested case–control can be designed for a prospective study. Disease cases collected during the follow-up are matched to non-disease controls selected from a portion of the entire cohort subjects. This design minimizes the recall bias, selection bias and inadequate/unreliable records of the environmental exposure from a retrospective case–control study, particularly when gene $\times$ environment hypotheses are being tested.

Family-based design is optimal in specific situations, such as the identification of disease-associated variants subjected to parental imprinting, or in haplotype studies (the reconstruction of the haplotype phase is improved by availability of parental genotypes). It is robust to population stratification bias,[70] but its main limitation is the lack of power, especially if the effect sizes are small.[71] If the power is 90%, two-sided $P = 0.001$ and control allele frequency is 20%, 1765 trios will detect an OR of 1.30, and for an OR of 1.20, 3731 trios will be necessary, representing 50% more participants than for a case–control study.

Quantitative trait study (for example, BMI) in large-scale population-based samples has proved to be an efficient method to identify novel susceptibility loci.[29] However, it requires larger sample size than case–control studies to reach the same statistical power,[72] and this limitation can become critical if expensive technologies are used (for example, genome-wide DNA arrays).

Thus, a case–control study represents the most powerful and cost-efficient method to perform genetic association studies,[73] if the two study groups are selected properly. Population-based obesity cases can be recruited, but the power of a case–control study can be increased potentially by applying an enrichment sampling strategy.[74] Obesity cases coming from hospital clinics, having a strong family background of the disease, an early age of onset or a more severe phenotype are likely to be enriched for genetic susceptibility in comparison with obese subjects randomly selected in the general population. Though enrichment sampling strategy is useful to improve power in genetic association studies, it inflates the relative risk and population attributable risk of the associated gene variants, and population-based follow-up cohort studies will be needed to obtain a reliable estimation of these parameters.[49]

To decrease potential sources of heterogeneity between the two study groups in case–control studies, controls and cases need to be selected from the same population. Potential confounders need to be ascertained in both and adjusted. Ideally, subjects in two study groups should be matched for ethnic/geographical origin, age, sex and additional variables for obesity (for example, level of physical activity) to adjust the regression model for confounding factors and to reduce residual error. 'Super control' subjects can also be selected (normal-weight subjects with no familial background of obesity[75] or extremely lean phenotypes).[76]

Phenotype. An accurate and specific phenotype of interest is critical in the design of genetic association studies.[8] A phenotype with a substantial heritability ($h^2 > 50\%$) should be favored and may allow the finding of etiological genetic variants in realistically achievable sample sizes. The ideal phenotype should be clinically and biologically relevant, not too rare, inexpensive and easy to identify in different places, thus allowing large-scale and feasible discovery and replication studies. It should be well defined so that measurement error, misclassification and heterogeneity can be minimized. BMI as a quantitative or qualitative trait is the more commonly used phenotype in genetic association studies for obesity. It is a highly heritable trait, easy and inexpensive to measure. As BMI is available in large data sets, genetic association meta-analyses have been performed in up to 250 000 subjects.[29] However, it also presents several limitations. Subjects in the higher BMI range tend to under-report their weight, resulting in inaccurate BMI measurement and misclassification in case–control studies. Thus, self-reported BMI data by the participants need to be considered with caution, and BMI data measured by medical staff is preferred.[77] The value of BMI to estimate the degree of adiposity may be questioned. Although it is strongly correlated with fat mass in obese subjects, in normal weight and underweight subjects the correlation is weaker.[78,79] At a given BMI, substantial variation in fat mass was reported.[78] Because phenotypic precision is critical for genetic investigations, the use of 'deep phenotyping' information (for example, body fat content and percentage of fat mass estimated by dual-emission X-ray absorptiometry or bioimpedance, behavioral food intake measured by *ad libitum* meal test[80] and energy expenditure estimated by room respiration calorimetry[80]) in large sample sizes may dissect the genetic influences more accurately on obesity.[81] Consistent with this hypothesis, a recent GWAS for body fat percentage in 32 626 individuals, with a replication of the best association signals in 39 576 individuals, identified three loci associated at a genome-wide level of significance: *FTO*, *IRS1* and *SPRY2*. Two out of three loci were not identified by previous large-scale GWAS meta-analysis for BMI.[82]

Recent GWAS for obesity-related traits have collected phenotype information in individuals living in widely heterogeneous environments. Although successful, this approach may have missed gene variants associated with BMI in specific environmental exposures. As the genetic basis of body weight regulation is unlikely to be fully discernible in individuals who are at stable body weight,[83] it may be of particular interest to perform genetic association studies for adiposity change in response to a major

environment modification (antipsychotic drug use, smoking cessation, rapid change in the level of physical activity or diet habits, caloric restriction, obesity surgery). Such phenotypes may provide a more comprehensive molecular basis for genetic predisposition to obesity.

Gene identification strategies. In the field of genetic epidemiology, we usually distinguish hypothesis-driven candidate gene and hypothesis-free genome-wide studies.[84] Candidate genes are selected based on the prior evidence of its role in the disease or phenotype of interest. Arguments to select a specific candidate gene may come from biology, functional genomics, pharmacology, animal models or genetics. The success rate of candidate gene approaches has been poor in the obesity field.[3] We consider the candidate gene approach as a relevant strategy as far as a careful gene selection process is applied. For instance, candidate gene studies focusing on *MC4R* and *PCSK1* genes led to the identification of common variants reproducibly associated with obesity risk/protection.[17,85,86] However, the candidate gene selection was based on three strong prior arguments: (1) an obesity phenotype in genetic mouse models; (2) the involvement in monogenic human obesity; (3) a likely functional role of the selected polymorphisms.

Hypothesis-free GWAS have been exceptionally successful and are to date the most efficient way to identify common variants (minor allele frequency $>1\%$) associated with obesity-related traits. The main limitation of genome-wide approaches in comparison with the candidate gene strategy is the high level of significance needed to adjust for the multiple tests performed and the situation may be even worth in the context of whole-exome or whole-genome studies. Testing the association of 30 000 genes with obesity rather than only one is certainly more exhaustive, but as a result a higher level of statistical evidence for genes with little supporting biological information is needed before significance is attributed. As pinpointed by John Ioannidis, 'the greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true'.[68]

A hybrid approach, the hypothesis-driven genome-wide association analysis, may lead to an overall decreased number of statistical tests and to more relaxed significance thresholds by the prior statement of specific hypotheses to be tested. Genome-wide analyses restricted to a specific biological pathway,[87] to a linkage region,[88] to genes specifically expressed in one important tissue,[89] to genes showing different expression patterns related to the disease,[90] to prioritized candidate genes,[19] to potentially damaging SNPs, or to SNPs harboring an evolutional signature, may increase success of picking up association signals that have been missed by conventional GWAS analyses.

Genotyping methodology and quality control procedures. Quality control to prevent, detect and minimize biases and errors during genotyping should be one of the top priorities in genetic association studies.[91] From earlier genotyping method of RFLP to PCR-based high-throughput genotyping, the genotyping error rate has significantly decreased, but errors and biases still exist. Sample tracking methods are encouraged to ensure a perfect correspondence in DNAs, genotypes and phenotypes, and an accurate analysis. Differences in extraction methods and storage of DNA samples collected from cases and controls or in different study centers should be mentioned. DNA quality and concentration need to be first documented to exclude low-quality DNA samples from further genotyping. Highly reliable and objective genotyping methods are recommended and genotyping protocols need to be described. Internal controls, duplicated samples and blank controls should be distributed in plates to ensure correct orientation and absence of DNA contamination. Genotyping quality control procedures include call rate calculation and deviations from Hardy–Weinberg equilibrium (HWE) separately in cases and controls. Call rate higher than 95% and HWE with P higher than 0.005 in the controls must be targeted, as an association may induce a modest deviation from HWE in affected subjects.[92] The assay reproducibility should be estimated by the calculation of a concordance rate in a sufficiently representative sample size (usually $>10\%$ of the whole genotyped sample). A concordance rate higher than 99% is needed to allow further analyses of the genetic markers. If a family-based design is used, the consistency of genotypes with mendelian expectations should be checked. Additional quality control procedures for the most critical results include the comparison of marker allele frequency in the experiment to public human genome databases and to individuals of comparable ethnicity, careful examination of the genotyping cluster plots to avoid a technical artifact, validation by an independent genotyping method or genotyping of additional genetic markers in strong linkage disequilibrium to confirm the association with the phenotype of interest.[46] A flow chart scheme summarizing the quality control procedures should be provided at least as supplementary material.

Statistical analyses. Prior statistical analyses should be stated and a table with the estimated power of the study design to detect a wide range of effect and minor allele frequency should be provided. As only few associations first reported from small studies are confirmed in larger replication data sets,[21] the scientific community must be warned against underpowered discovery studies and promote the generalization of large-sized and well-powered discovery study. A limited number of additional analyses of subgroups such as age, sex, or categories of obesity should be specified in a prior hypothesis in addition to tests in the whole sample, and interpretation of results from subgroup analyses should be done cautiously and accompanied by interaction tests to prove significant between-group heterogeneity. We recommend using statistical tests that enable the adjustment for confounding variables. In multiple SNP haplotype analyses, the investigators should detail the methodology used. Although the importance of adjusting for multiple comparisons in genetic association studies is generally accepted, there is no universal criterion on how it can be achieved. False discovery rate, Bayesian procedures and adjusted Bonferroni correction are acceptable.[31,93,94] Application of the Bonferroni correction tends to be over-conservative because the tests performed in a genetic association study are unlikely to be fully independent (lack of independence of the three inheritance models, correlation between different phenotypes and linkage disequilibrium between SNPs). In GWAS, we recommend a more relaxed P threshold of 5×10^{-6} to select a limited number of promising genetic markers for further large-scale replication.[29] In that situation, a Bonferroni corrected P of 5×10^{-8} should be applied to the discovery plus replication (stage 1 + stage 2) joint analysis to confer significance at the genome-wide level.[95] In the recent GIANT GWAS meta-analysis for BMI in 123 865 individuals, the application of a more relaxed threshold of $P<5 \times 10^{-6}$ for SNP follow-up rather than a threshold of P of 5×10^{-8} induced a modest burden of false positive association signals (24%) but substantially increased the harvest of disease-associated validated loci (from 68%).[29] For most original associations of common gene variants with common diseases, the genetic model of inheritance is not properly addressed.[65] A formal procedure to attribute the best-fitting inheritance model for the phenotype studied is recommended from the discovery study and should be used for replication efforts. If several SNPs are associated at the same locus, conditional regression analyses should be applied to determine if these SNPs are independent and to select the lead SNP signal(s) to follow up in replication efforts.

Population stratification. The first step to lower population stratification is to collect cases and controls in the same geographic area,

Table 2. Guidelines in designing and implementing discovery and replication studies

Discovery association study	
Study designs	
Prospective cohort study	Gold standard
	Slow process
	Low statistical power
	Expensive to follow-up
	Possible high withdrawal rate
Family-based design	Useful in the context of parental imprinting
	Useful in the context of haplotype studies
	Robust to population stratification bias
	Low power when effect size is small
	Expensive design
Quantitative trait study	Large sample size needed
	Unable to detect genetic effects at the extremes of the trait distribution
Case–control study	Most powerful and cost-efficient
	Cautions needed in selecting cases and controls to lower heterogeneity
Phenotype selection	
BMI	High heritability
	Clinically and biologically relevant
	Common
	Non-negligible measurement error/ risk of misclassification if self-reported
	Not accurate to measure adiposity body fat content and percentage of fat mass
'deep phenotyping'	Food intake/ behavior
	Energy expenditure
	Adiposity response to environmental changes
Gene identification strategies	
Candidate genes	Hypothesis-driven
	Prior evidence from other disciplines such as biology, functional genomics, pharmacology, animal models or genetics
Genome-wide studies	Hypothesis-free
	Efficient in identifying novel common variants without previous candidacy
	High level of significance needed
Hypothesis-driven genome-wide association	Lower number of statistical tests
	Useful to pick up the association signals missed by conventional GWAS
	Focus on biological pathways, linkage regions, transcriptomics data
Genotyping methodology and quality control procedures	
DNA sample handling	
DNA quality	
Description of genotyping protocol	
Genotyping quality control: DNA contamination, call rate, Hardy–Weinberg equilibrium, concordance rate, external control from public human genome databases, genotyping cluster plots, mendelian inconsistencies, validation by another technology and analysis of neighboring SNPs in strong linkage disequilibrium	
Statistical analyses	
Analytic methodology selected according to the study design	
Prior statement of hypotheses to be tested	
Multiple testing correction	
Best-fitting inheritance model	
Selection of lead SNP for replication efforts	
Population stratification	
Cases and controls selected from the same population and geographic area	
Adjustment for population stratification	
Replication studies	
Systematic replication of promising original reports	
Systematic follow-up of promising associations	
Objective interpretation of data	
Statistical power	
Sample size calculation corrected by the 'Winner's curse effect'	
Homogeneity	
Same ethnicity and similar age, sex, geographic origin	
Same phenotype, measurement method	
Same inheritance model	
Same study design, ascertainment strategy	
Same genetic marker, statistical analysis, adjustment for confounding variables	
Meta-analysis	
Powerful to confirm genetic associations at a stringent level of significance	
Estimation of publication bias and between-study heterogeneity	
Additional studies	
Extension of confirmed variants to other populations (different ethnic backgrounds, age, environmental risks)	
Fine-mapping	
Gene × environment, gene × gene interaction studies	
Functional assay to confirm the biological role of gene variants	
Additional biological evidence	

Abbreviations: BMI, body mass index; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

first using self-reported ethnicity to equilibrate cases and controls in multi-ethnic studies, or restricting the analysis to one ethnic background if a particular ethnic group predominates. As relying solely on self-reported ethnicity is not sufficient to exclude population stratification[96] and subtle population stratification is observed within ethnically homogeneous populations,[97] several methods (genomic control, EIGENSTRAT, PCA-L and LAPSTRUCT) have been developed to correct for population stratification[98] and are now routinely used.[6,19] These methods are based on using a set of ancestry informative markers,[99] which are easy to derive from GWAS data sets,[100] but require further genotyping in candidate gene studies. These markers may also be useful to account for cryptic relatedness, another potential source of spurious association in case–control designs.[101] Alternatively, family-based association tests can be used to confirm an association derived from a case–control study, as they are insensitive to population stratification biases.[102]

Replication study designs

Systematic replication of promising original reports. Several teams working in the field of obesity should be commended for systematically replicating promising association reports in their own cohorts. However, other replication efforts may not have been as systematically conducted even if the original report is well designed and statistically sound. To move on the right direction, researchers, but also editors of influential journals, need to consider well-designed initial and follow-up (negative or positive) reports of association as equally important. Replication studies need to be conducted more systematically and less selectively, and an objective interpretation of data is crucial.

Statistical power. When the original association report describes a true association signal, the effect size is often overestimated due to the Winner's curse effect.[103] Subsequent follow-up studies often fail to replicate the association owing to underestimation of the required sample size. Hence, when calculating the statistical power of the replication design, the investigators may correct the OR from initial study by taking into account the Winner's curse effect.[103]

Heterogeneity. If both original and replication studies are well designed and well executed, the inability to replicate an original finding indicates a genuine difference that may be due to heterogeneity from many sources.[104] Heterogeneity among studies represents a major problem in replicating a true original signal. Moonesinghe *et al.*[52] have shown that in the presence of large between-study heterogeneity, true associations may not be possible to replicate with consistency, no matter how large the study is. When designing a study in the initial rounds of replication, investigators need to avoid all the potential sources of heterogeneity and focus on one main question to increase the chance to reproduce the original association finding. To reduce the genetic and environmental heterogeneity, populations in the initial and in the follow-up studies should share the same ethnicity and ideally come from the same country. Correction for population stratification should be applied as well. Same age window and sex ratios should be targeted. The study design, genetic marker, statistical analysis and genetic model should be the same as those in the initial study. To lower phenotypic heterogeneity, the same phenotype, measurement method, recruitment strategy and adjustment for confounding variables should be applied.

Meta-analysis. There are always limitations in individual association study. Common genetic variants harboring modest OR or rare variants with higher OR rarely reach a conclusive threshold of association in a single study. Meta-analysis of data sets from comparable studies improves the power to confirm genetic associations at a stringent level of significance. Unfortunately, the heterogeneity in the design of follow-up studies (genetic marker, phenotype and genetic model) and insufficient data description may reduce the number of eligible studies and the statistical power in subsequent meta-analyses. Hence, Gallo *et al.*[105] recently provided detailed guidelines to accurately report the findings of epidemiological studies involving biomarkers through the STrengthening the Reporting of OBservational studies in Epidemiology - Molecular Epidemiology (STROBE-ME) statement. Estimation of the degree of between-study heterogeneity and of potential publication bias are important for the interpretation of meta-analyses.[106]

Additional studies. Once the risk of false positive association has been ruled out by initial replication studies, the focus of the association can be extended to different age windows, different study designs and ascertainment criteria for cases and controls, additional phenotypes related to obesity, or to the study of gene $\times$ gene/ gene $\times$ environment interactions. The association can be studied in different ethnic backgrounds to estimate the worldwide contribution of a given locus, and to fine-map the association signal. Using a dense SNP map in the associated interval is strongly recommended, as the initial SNP associated with obesity is probably in linkage disequilibrium with the causal locus, while linkage disequilibrium patterns vary according to ethnicity. Once the putative causal susceptibility variant(s) has been identified, functional assays should be used to examine whether a SNP variant alters protein expression or function. As an association does not always imply causality, biological insights are essential in increasing the credibility of the observed genetic association,[107] as recently illustrated by the *FTO* gene. In 2007, a link between *FTO* and obesity was established in Europeans through association studies when *FTO* was 'a gene of unknown function in an unknown pathway'.[43,46] Subsequently, more than 400 articles have been published on *FTO*, providing strong biological evidence that *FTO* is a causal gene underlying obesity.[108]

CONCLUSION

We have reviewed the reasons for the lack of reproducibility in obesity genetic association studies and offered guidelines for proper discovery and replication association designs (see Table 2). 'Learn from yesterday, live for today, hope for tomorrow': a quote attributed to the physicist Albert Einstein perfectly depicts our position in the young field of genetic epidemiology of obesity. Genetic association studies may first appear simple but actually face numerous challenges. After some initial disappointments, methodological, technological and scientific developments have recently led to more robust associations and renewed attention to genetic epidemiology. We expect a prolific period of discovery in human obesity genetics ahead and hope that this review will contribute to the advancement of these noble goals.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Dipika Desai and Emeritus Professor Patricia Chang for reviewing and editing of the manuscript, and the reviewers for their helpful comments.

REFERENCES

1 Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ *et al.* National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet* 2011; **377**: 557–567.

2 Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr* 2008; **87**: 398–404.
3 Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G, Walts B *et al.* The human obesity gene map: the 2005 update. *Obesity* 2006; **14**: 529–644.
4 Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002; **4**: 45–61.
5 Cordell HJ, Clayton DG. Genetic association studies. *Lancet* 2005; **366**: 1121–1131.
6 Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**: 25–34.
7 Choquet H, Meyre D. Molecular basis of obesity: current status and future prospects. *Curr Genomics* 2011; **12**: 154–168.
8 Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G *et al.* Replicating genotype-phenotype associations. *Nature* 2007; **447**: 655–660.
9 Visscher PM, Hill WG, Wray NR. Heritability in the genomics era-concepts and misconceptions. *Nat Rev Genet* 2008; **9**: 255–266.
10 Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet* 2005; **366**: 941–951.
11 Haworth CM, Plomin R, Carnell S, Wardle J. Childhood obesity: genetic and environmental overlap with normal-range BMI. *Obesity* 2008; **16**: 1585–1590.
12 North KE, Graff M, Adair LS, Lange EM, Lange LA, Guo G *et al.* Genetic epidemiology of BMI and body mass change from adolescence to young adulthood. *Obesity* 2010; **18**: 1474–1476.
13 Malis C, Rasmussen EL, Poulsen P, Petersen I, Christensen K, Beck-Nielsen H *et al.* Total and regional fat distribution is strongly influenced by genetic factors in young and elderly twins. *Obes Res* 2005; **13**: 2139–2145.
14 Carnell S, Haworth CM, Plomin R, Wardle J. Genetic influence on appetite in children. *Int J Obes* 2008; **32**: 1468–1473.
15 Bouchard C, Tremblay A, Nadeau A, Despres JP, Theriault G, Boulay MR *et al.* Genetic effect in resting and exercise metabolic rates. *Metabolism* 1989; **38**: 364–370.
16 Saunders CL, Chiodini BD, Sham P, Lewis CM, Abkevich V, Adeyemo AA *et al.* Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity* 2007; **15**: 2263–2275.
17 Stutzmann F, Vatin V, Cauchi S, Morandi A, Jouret B, Landt O *et al.* Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. *Hum Mol Genet* 2007; **16**: 1837–1844.
18 Ichimura A, Hirasawa A, Poulain-Godefroy O, Bonnefond A, Hara T, Yengo L *et al.* Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. *Nature* 2012; **483**: 350–354.
19 Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 2009; **41**: 157–159.
20 Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI *et al.* Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genet* 2010; **6**: e1000916.
21 Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003; **361**: 567–571.
22 Boutin P, Dina C, Vasseur F, Dubois S, Corset L, Seron K *et al.* GAD2 on chromosome 10p12 is a candidate gene for human obesity. *PLoS Biol* 2003; **1**: E68.
23 Swarbrick MM, Waldenmaier B, Pennacchio LA, Lind DL, Cavazos MM, Geller F *et al.* Lack of support for the association between GAD2 polymorphisms and severe human obesity. *PLoS Biol* 2005; **3**: e315.
24 Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T *et al.* A common genetic variant is associated with adult and childhood obesity. *Science* 2006; **312**: 279–283.
25 Dina C, Meyre D, Samson C, Tichet J, Marre M, Jouret B *et al.* Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187 author reply 187.
26 Loos RJ, Barroso I, O'Rahilly S, Wareham NJ. Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187 author reply 187.
27 Rosskopf D, Bornhorst A, Rimmbach C, Schwahn C, Kayser A, Kruger A *et al.* Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007; **315**: 187 author reply 187.
28 Heid IM, Huth C, Loos RJ, Kronenberg F, Adamkova V, Anand SS *et al.* Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? *PLoS Genet* 2009; **5**: e1000694.
29 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
30 Lyon HN, Emilsson V, Hinney A, Heid IM, Lasky-Su J, Zhu X *et al.* The association of a SNP upstream of INSIG2 with body mass index is reproduced in several but not all cohorts. *PLoS Genet* 2007; **3**: e61.
31 Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; **361**: 865–872.
32 Chandalia M, Grundy SM, Adams-Huet B, Abate N. Ethnic differences in the frequency of ENPP1/PC1 121Q genetic variant in the Dallas Heart Study cohort. *J Diabetes Complications* 2007; **21**: 143–148.
33 Klimentidis YC, Abrams M, Wang J, Fernandez JR, Allison DB. Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions. *Hum Genet* 2011; **129**: 407–418.
34 Kettunen J, Silander K, Saarela O, Amin N, Muller M, Timpson N *et al.* European lactase persistence genotype shows evidence of association with increase in body mass index. *Hum Mol Genet* 2010; **19**: 1129–1136.
35 Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA* 2006; **295**: 1549–1555.
36 Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.
37 Sen S, Burmeister M. Hardy-Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci. *Hum Genomics* 2008; **3**: 36–52.
38 Gordon D, Finch SJ, Nothnagel M, Ott J. Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered* 2002; **54**: 22–33.
39 Bouatia-Naji N, De Graeve F, Bronner G, Lecoeur C, Vatin V, Durand E *et al.* INS VNTR is not associated with childhood obesity in 1,023 families: a family-based study. *Obesity* 2008; **16**: 1471–1475.
40 Le Stunff C, Fallin D, Schork NJ, Bougneres P. The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. *Nat Genet* 2000; **26**: 444–446.
41 Peters DL, Barber RI, Flood EM, Garner HR, O'Keefe GE. Methodologic quality and genotyping reproducibility in studies of tumor necrosis factor -308G->A single nucleotide polymorphism and bacterial sepsis: implications for studies of complex traits. *Crit Care Med* 2003; **31**: 1691–1696.
42 Saito YA, Talley NJ, de Andrade M, Petersen GM. Case-control genetic association studies in gastrointestinal disease: review and recommendations. *Am J Gastroenterol* 2006; **101**: 1379–1389.
43 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.
44 Page GP, George V, Go RC, Page PZ, Allison DB. 'Are we there yet?': Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet* 2003; **73**: 711–719.
45 Framework for a fully powered risk engine. *Nat Genet* 2005; **37**: 1153.
46 Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 2007; **39**: 724–726.
47 Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J *et al.* Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 2007; **3**: e115.
48 Hinney A, Nguyen TT, Scherag A, Friedel S, Bronner G, Muller TD *et al.* Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS One* 2007; **2**: e1361.
49 Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet* 2005; **366**: 1315–1323.
50 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
51 Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT *et al.* Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans. *Obesity* 2011; **19**: 2436–2439.
52 Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and non-replicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci USA* 2008; **105**: 617–622.
53 Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J *et al.* Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009; **38**: 263–273.
54 Aschard H, Hancock DB, London SJ, Kraft P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered* 2010; **70**: 292–300.
55 Luan JA, Wong MY, Day NE, Wareham NJ. Sample size determination for studies of gene-environment interaction. *Int J Epidemiol* 2001; **30**: 1035–1040.

56 Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 2002; **155**: 478–484.
57 Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schafer H. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur J Hum Genet* 2008; **16**: 1164–1172.
58 Hein R, Beckmann L, Chang-Claude J. Sample size requirements for indirect association studies of gene-environment interactions (G x E). *Genet Epidemiol* 2008; **32**: 235–245.
59 Kilpelainen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E *et al.* Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med* 2011; **8**: e1001116.
60 Okada Y, Kubo M, Ohmiya H, Takahashi A, Kumasaka N, Hosono N *et al.* Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* 2012; **44**: 302–306.
61 Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. *Lancet* 2000; **355**: 308–311.
62 Peng S, Zhu Y, Xu F, Ren X, Li X, Lai M. FTO gene polymorphisms and obesity risk: a meta-analysis. *BMC Med* 2011; **9**: 71.
63 Groves CJ, Zeggini E, Walker M, Hitman GA, Levy JC, O'Rahilly S *et al.* Significant linkage of BMI to chromosome 10p in the U.K. population and evaluation of GAD2 as a positional candidate. *Diabetes* 2006; **55**: 1884–1889.
64 Sovio U, Mook-Kanamori DO, Warrington NM, Lawrence R, Briollais L, Palmer CN *et al.* Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet* 2011; **7**: e1001307.
65 Salanti G, Southam L, Altshuler D, Ardlie K, Barroso I, Boehnke M *et al.* Underlying genetic models of inheritance in established type 2 diabetes associations. *Am J Epidemiol* 2009; **170**: 537–545.
66 Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoeur C, Vatin V *et al.* Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet* 2005; **37**: 863–867.
67 McAteer JB, Prudente S, Bacci S, Lyon HN, Hirschhorn JN, Trischitta V *et al.* The ENPP1 K121Q polymorphism is associated with type 2 diabetes in European populations: evidence from an updated meta-analysis in 42,042 subjects. *Diabetes* 2008; **57**: 1125–1130.
68 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2**: e124.
69 Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; **58**: 543–549.
70 Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. *Nat Rev Genet* 2011; **12**: 465–474.
71 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
72 Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genet Epidemiol* 2010; **34**: 254–257.
73 Putter C, Pechlivanis S, Nothen MM, Jockel KH, Wichmann HE, Scherag A. Missing heritability in the tails of quantitative traits? A simulation study on the impact of slightly altered true genetic models. *Hum Hered* 2011; **72**: 172–180.
74 Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. *Nat Protoc* 2007; **2**: 2492–2501.
75 Rouskas K, Kouvatsi A, Paletas K, Papazoglou D, Tsapas A, Lobbens S *et al.* Common variants in FTO, MC4R, TMEM18, PRL, AIF1, and PCSK1 show evidence of association with adult obesity in the Greek population. *Obesity* 2012; **20**: 389–395.
76 Allison DB, Schork NJ. Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. *Behav Genet* 1997; **27**: 401–421.
77 Visscher TL, Viet AL, Kroesbergen IH, Seidell JC. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity* 2006; **14**: 2054–2063.
78 Muller MJ, Bosy-Westphal A, Krawczak M. Genetic studies of common types of obesity: a critique of the current use of phenotypes. *Obes Rev* 2010; **11**: 612–618.
79 Gradmark AM, Rydh A, Renstrom F, De Lucia-Rolfe E, Sleigh A, Nordstrom P *et al.* Computed tomography-based validation of abdominal adiposity measurements from ultrasonography, dual-energy X-ray absorptiometry and anthropometry. *Br J Nutr* 2010; **104**: 582–588.
80 Fisher JO, Cai G, Jaramillo SJ, Cole SA, Comuzzie AG, Butte NF. Heritability of hyperphagic eating behavior and appetite-related hormones among Hispanic children. *Obesity* 2007; **15**: 1484–1495.
81 Tracy RP. 'Deep phenotyping': characterizing populations in the era of genomics and systems biology. *Curr Opin Lipidol* 2008; **19**: 151–157.
82 Kilpelainen TO, Zillikens MC, Stancakova A, Finucane FM, Ried JS, Langenberg C *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* 2011; **43**: 753–760.
83 Muller MJ, Bosy-Westphal A, Later W, Haas V, Heller M. Functional body composition: insights into the regulation of energy metabolism and some clinical applications. *Eur J Clin Nutr* 2009; **63**: 1045–1056.
84 Hoh J, Ott J. Genetic dissection of diseases: design and methods. *Curr Opin Genet Dev* 2004; **14**: 229–232.
85 Geller F, Reichwald K, Dempfle A, Illig T, Vollmert C, Herpertz S *et al.* Melanocortin-4 receptor gene variant I103 is negatively associated with obesity. *Am J Hum Genet* 2004; **74**: 572–581.
86 Benzinou M, Creemers JW, Choquet H, Lobbens S, Dina C, Durand E *et al.* Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat Genet* 2008; **40**: 943–945.
87 Liu YJ, Guo YF, Zhang LS, Pei YF, Yu N, Yu P *et al.* Biological pathway-based genome-wide association analysis identified the vasoactive intestinal peptide (VIP) pathway important for obesity. *Obesity* 2010; **18**: 2339–2346.
88 Johansson A, Marroni F, Hayward C, Franklin CS, Kirichenko AV, Jonasson I *et al.* Linkage and genome-wide association analysis of obesity-related phenotypes: association of weight with the MGAT1 gene. *Obesity* 2010; **18**: 803–808.
89 Du H, Vimaleswaran KS, Angquist L, Hansen RD, van der AD, Holst C *et al.* Genetic polymorphisms in the hypothalamic pathway in relation to subsequent weight change-the DiOGenes study. *PLoS One* 2011; **6**: e17436.
90 Greenawalt DM, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J *et al.* A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res* 2011; **21**: 1008–1016.
91 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–369.
92 Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 1998; **63**: 1531–1540.
93 Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004; **96**: 434–442.
94 Lin DY. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* 2005; **21**: 781–787.
95 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**: 209–213.
96 Serre D, Montpetit A, Pare G, Engert JC, Yusuf S, Keavney B *et al.* Correction of population stratification in large multi-ethnic association studies. *PLoS One* 2008; **3**: e1382.
97 Price AL, Helgason A, Palsson S, Stefansson St H, Clair D, Andreassen OA *et al.* The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet* 2009; **5**: e1000505.
98 Wu C, DeWan A, Hoh J, Wang Z. A comparison of association methods correcting for population stratification in case-control studies. *Ann Hum Genet* 2011; **75**: 418–427.
99 Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG *et al.* Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003; **72**: 1492–1504.
100 Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 2008; **17**(R2): R143–R150.
101 Sillanpaa MJ. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity* 2011; **106**: 511–519.
102 Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
103 Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* 2007; **80**: 605–615.
104 Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* 2007; **2**: e841.
105 Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JP, Kirsch-Volders M *et al.* STrengthening the Reporting of OBservational studies in Epidemiology - Molecular Epidemiology (STROBE-ME): An Extension of the STROBE Statement. *PLoS Med* 2011; **8**: e1001117.
106 Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The quality of meta-analyses of genetic association studies: a review with recommendations. *Am J Epidemiol* 2009; **170**: 1333–1343.
107 Khoury MJ, Little J, Gwinn M, Ioannidis JP. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol* 2007; **36**: 439–445.
108 Choquet H, Meyre D. Genetics of obesity: what have we learned? *Curr Genomics* 2011; **12**: 169–179.