The importance of gene–environment interactions in human obesity

Hudson Reddon, Jean-Louis Guéant, David Meyre

To cite this version:

inserm-01373635
The importance of gene-environment interactions in human obesity

Hudson Reddon¹, Jean-Louis, Guéant², David Meyre¹²³

¹Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada
²Inserm U-954, Faculté de Médecine, University of Lorraine and University Hospital Center of Nancy, Nancy, France
³Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

Abstract

The worldwide obesity epidemic has been mainly attributed to lifestyle changes. However, who becomes obese in an obesity-prone environment is largely determined by genetic factors. In the last twenty years, important progress has been made in the elucidation of the genetic architecture of obesity. In parallel with successful gene identifications, the number of gene-environment interaction (GEI) studies has grown rapidly. This paper reviews the growing body of evidence supporting gene-environment interactions in the field of obesity. Heritability, monogenic and polygenic obesity studies provide converging evidence that obesity-predisposing genes interact with a variety of environmental, lifestyle and treatment exposures. However, some skepticism remains regarding the validity of these studies based on several issues, which include statistical modelling, confounding, low replication rate, underpowered analyzes, biological assumptions and measurement precision. What follows in this review includes (1) an introduction to the study of GEI, (2) the evidence of GEI in the field of obesity, (3) an outline of the biological mechanisms that may explain these interaction effects, (4) methodological challenges associated with GEI studies and potential solutions, and (5) future directions of GEI research. Thus far, this growing body of evidence has provided a deeper understanding of GEI influencing obesity and may have tremendous applications in the emerging field of personalized medicine and individualized lifestyle recommendations.

Summary statement

Current research has identified several environmental exposures that can moderate the impact of genetic risk factors on obesity. This paper reviews these studies (gene-environment interactions) in the obesity field and outlines the methodological challenges of these investigations.

Short title: A review of gene-environment interaction studies in obesity
Keywords: obesity; gene-environment interactions; heritability; monogenic; polygenic; methodology.
Abbreviations: EWAS (epigenome-wide association studies), genome-wide analysis of epigenetic modifications associated with disease phenotypes; SNP (single nucleotide polymorphism), a single nucleotide variation that is common in >1% of the population; eQTL (expression quantitative trait locus), a locus that influences the expression levels of mRNAs or proteins; GEWIS (genome-wide interaction studies), simultaneous testing of genetic variants theoretically covering the whole genome in interaction with an environmental factor; meQTL (methylation quantitative trait locus), a locus linked to different patterns of DNA methylation; pQTL (protein quantitative trait locus), a locus associated with variations in protein levels; QTL (quantitative trait locus), region of DNA containing or linked to genes that underlie a quantitative trait.

Correspondence: Dr. David Meyre (meyred@mcmaster.ca)
Introduction

Over the past three decades, the prevalence of obesity has reached epidemic proportions throughout the world (1). This recent epidemic cannot be explained by sudden changes in the human population gene pool and has been mainly attributed to lifestyle modifications (2). Over-nutrition and decline in physical activity are the two “usual suspects”, but additional factors (reduced gut microflora diversity, sleep debt, endocrine disruptors, reduction in variability of ambient temperatures) have emerged as significant contributors to the escalating prevalence of obesity (3). If obesity is a multifactorial disorder that requires environmental influences to manifest, some individuals are more susceptible than others to weight gain in an obesity-prone environment, and who becomes obese at the individual level is largely determined by genetic factors (4). Technological and methodological breakthroughs in the last twenty years have led to important progress in the elucidation of the genetic architecture of obesity (5). The first two genes (LEP and MKKS) associated with a Mendelian non-syndromic or syndromic form of obesity were identified in 1997 and 2000 (6, 7). Seven years later, the first common variant (located in the intron 1 of the FTO gene) reproducibly associated with polygenic obesity was identified (8, 9). At the time we are writing, over 40 monogenic obesity loci (with or without syndromic features) and 130 polygenic obesity loci have been described, and this list is destined to grow over the coming years (5). In parallel with successful gene identification efforts, the number of studies on gene-environment interactions has grown rapidly (10). In the first segment of this review, we summarize the findings supporting gene-environment interaction in obesity from heritability, monogenic and polygenic studies and provide a biological hypothesis to explain these statistical interactions. The final section will outline methodological challenges associated with GEI studies, provide potential solutions to these issues based on existing evidence and highlight future directions of GEI research.

Definitions

The genetic etiology of obesity can be classified into two categories. First, Mendelian (or monogenic) obesity describes individuals who carry a rare gene variant with a dramatic impact on adiposity (11). These variants are associated with a high lifetime risk of disease and exhibit a near one-to-one relationship between genotype and phenotype (12-14). Monogenic obesity can be classified as syndromic or non-syndromic. Syndromic obesity refers to Mendelian obesity that co-occurs with a distinct set of clinical phenotypes, such as mental retardation, dysmorphic features and organ-specific developmental abnormalities (15). Syndromic forms of obesity result from chromosomal abnormalities or point mutations, which can be autosomal or X-linked disorders (16). Non-syndromic forms are caused by pathogenic mutations or structural variations in genes involved in the leptin/melanocortin pathway, and are mainly characterized by hyperphagic obesity (17). Homozygous/compound heterozygous carriers of pathogenic mutations in genes from the leptin/melanocortin pathway are exceedingly rare and lead to a fully penetrant form of early-onset extreme obesity. In contrast, heterozygous incomplete penetrant mutations in the same pathway account for a greater proportion of oligogenic obesity cases in population (5). Heterozygous loss-of-function mutation
carriers in \(MC4R\) display mild obese phenotypes, which can be moderated by environmental factors that increase the risk of obesity (18, 19). Second, other cases of obesity can be attributed to the concerted presence of DNA variation in several genes (each with a relatively small effect), known as polygenic obesity (12). With respect to body weight regulation, recent simulations estimated that hundreds of variants with small to modest effect may account for the genetic architecture of complex traits such as obesity (20).

The concept of gene-environment interaction in the context of human diseases is not recent and has been discussed since proposed by J.B. Haldane in 1946 (21). The statistical definition of an interaction between two or more risk factors is simply the coefficient of the product term of the risk factors, also known as effect modification or effect modulation. Interaction is thus measured in terms of departure from a multiplicative or an additive model (22, 23). Alternatively, biological interaction between two factors is defined as their co-participation in the same causal mechanism to disease development (24). For statistical evidence of gene-environment interaction to be convincing, it is typically necessary to replicate the findings in additional samples and/or support the evidence with plausible underlying biological mechanisms (22).

**The importance of gene-environment interactions in obesity: evidence from heritability, monogenic and polygenic studies**

**Heritability estimates are influenced by the environment**

Early indications of the shared influence of genetics and the environment in shaping obesity originated from heritability studies involving environmental exposures among twins. Heritability is the proportion of total phenotypic variability caused by genetic variance in a population. Large pedigree, twin and adoption studies allow the calculation of heritability and they all evidence a strong genetic component in human obesity (25-27). Before the first obesity gene identification reports, scientists considered the possibility that heritability, a global estimate of genetic predisposition to obesity, may be modulated by specific environments (28). Specific environmental exposures known to mediate heritability estimates include biological, socio-economic factors and lifestyle factors.

*In utero* factors have been proposed to modulate offspring’s future risk for obesity (29). The higher estimates of heritability for BMI observed in mother-offspring pairs in comparison with father-offspring pairs suggest a possible modification effect of maternal *in utero* environment on the offspring’s genetic predisposition to obesity (30). Maternal weight gain during pregnancy may interact with genetic factors to render the offspring more susceptible to develop obesity in young adulthood (31).

Genetic influence on BMI may also interact with sex and age. Sex-specific genetic effects on BMI have been observed in adolescents as well as in adults (32, 33). Heritability of obesity also varies with age. A previous study of over 12,000 twin pairs reported a heritability of 4-9% for BMI at birth, which increased to more than 50% at 5 months of age (34). Heritability estimates increase from infancy to childhood (35), from childhood to pre-adolescence (36), from preadolescence to adolescence (37), and reach a plateau during adolescence and adulthood, and then slightly decrease in late adulthood.
Longitudinal BMI change from adolescence to young adulthood and from young adulthood to adulthood is a heritable trait, but genetic variants for change in BMI partially overlap with those affecting the level of BMI (39, 40). Moreover, heritability estimates of obesity increase with the severity of obesity status (41).

The investigation of socio-economic factors and lifestyle behaviours has revealed many additional conditions that impact heritability estimates. One may presume that the emergence of a society characterized by food abundance and physical inactivity may increase the impact of environment (and therefore decrease the impact of genes) in the determination of the obese phenotypes. Counter intuitively, the proportion of variability in BMI attributable to genetic variation is increased among people born after the establishment of a modern ‘obesogenic’ environment (42-44). These results are congruent with the seminal work by Claude Bouchard and colleagues showing that the BMI response to long-term overfeeding in young adult male twins is mainly influenced by genetic factors (28). Twin studies have shown that a high level of physical activity can substantially reduce the influence of genetic factors on BMI in both young and older adults (45, 46). PT. Williams studied the parental contribution to offspring’s BMI in 47,691 adult runners and showed that vigorous physical activity (running distance ≥ 9 km/day) decreased the parental contribution to BMI, by 48-58 %, in comparison with runners with moderate physical activity (running distance < 3 km/day)(47). Socio-economic research indicates that higher educational status is associated with decreased risk of obesity (48), but heritability estimates for BMI in late childhood/adolescence are positively correlated with the level of education of parents (49). Sleep duration is negatively associated with obesity (50). In a twin study, the heritability of BMI (h² = 70%) in short-sleepers (< 7 hours/day) was more than twice the heritability of BMI (h² = 32%) when sleep duration was longer (≥ 9 h/day) (51). Weight gain is a well-known adverse effect of antipsychotic medication (52), but a considerable degree of inter-individual variability has been described in literature (53). Two pilot twin/sibs comparison studies have reported heritability estimates of 60-80% for body weight gain in response to antipsychotics in adolescents and adults (54, 55). Weight loss in response to vigorous exercise, diet restriction or bariatric surgery is also highly variable which suggests a heritable component (56-58).

A recent analysis of the Framingham Heart Study analysed how the heritability of BMI was influenced by historical period, life course and physical activity (59). These authors reported that: 1) the heritability estimates of BMI were considerably larger after the mid 1980’s compared to the 3 preceding decades; 2) the genetic influence on BMI appears to decrease across the lifespan, with the greatest genetic influence observed during reproductive ages across historical period and 3) the heritability of BMI was considerably smaller among physically active individuals aged 21-50 years, but not among those >50 years old (59).

**Obesity predisposing gene variants interact with the environment**

Although heritability studies provided early evidence for the genetic contribution to obesity, recent efforts have focused on the identification of specific gene variants that impact obesity risk. Our knowledge about the genetic architecture of Mendelian (syndromic and non-syndromic) and polygenic forms of obesity has greatly expanded in
the last 20 years (17). It is noteworthy that even some forms of Mendelian (syndromic and non-syndromic) obesity can display a somewhat variable phenotype (60-63). This can be attributed not only to genetic heterogeneity, gene-gene interactions and inheritance model (64, 65), but interactions with environmental factors should be considered as one of the causes for the variability in obese phenotypes (23). Since the rapid increase in obesity prevalence over the last few decades indicates a strong environmental influence on BMI (e.g. physical activity, diet, educational status, age, sex) (3), many researchers have worked on the identification of specific environmental factors that interact with monogenic and polygenic obesity predisposing genes. The existing evidence regarding the study of obesity indicates that lifestyle factors can significantly modify the impact of obesity predisposing gene variants.

**Obesity predisposing gene variants interact with non-modifiable biological factors**

*Obesity predisposing gene variants interact with pregnancy and in utero factors*

Pre-pregnancy maternal obesity and excessive weight gain during pregnancy are both associated with increased birth weight, higher rate of macrosomia in the offspring (66, 67) and higher risk of adiposity in offspring during childhood, adolescence and adulthood (29, 68, 69). Recently, a morbidly obese female patient with a rare homozygous LEPR mutation was reported to gain 110 lbs during pregnancy, far beyond the 11-40 lbs gestational weight gain range recommended by the Institute of Medicine, and gave birth to a baby with macrosomia (70). These data suggest that a Mendelian predisposition for obesity increases gestational weight gain and offspring’s birth weight.

However, no such effect on gestational weight gain was observed for a polygenic gene score composed of four common obesity-predisposing common variants in or near FTO, MC4R, TMEM18 and GNPDA2 (71). Studies with gene scores including more SNPs are needed to further investigate this hypothesis.

Prenatal exposure to maternal cigarette smoking was found to interact with genetic variation in OPRM1 to modulate fat intake in offspring (72). Among 956 adolescents, the T allele in OPRM1 was associated with lower fat intake but only in those without prenatal exposure to cigarette smoke (72). DNA methylation was significantly reduced within several CpGs across OPRM1 among adolescents exposed to prenatal maternal cigarette smoking compared to those not exposed (72).

*Obesity predisposing gene variants interact with sex*

Females are generally more likely to develop morbid obesity than males (73) and these discrepancies may be explained in part by sex-specific genetic effects. In line with this hypothesis, pathogenic monogenic mutations in MC4R have an effect on BMI about twice as strong in females as in males (18, 74).

Seven out of 14 polygenic loci convincingly associated with waist-to-hip ratio displayed sexual dimorphism, all with a stronger effect on the phenotype in women than in men (75, 76). A recent genome-wide interaction meta-analysis did not report any sex-specific for variants associated with BMI, but found 44 loci with sex-specific effects on waist-to-hip ratio adjusted for BMI (28 of the 44 loci displayed larger effects in women than men) (77).
Obesity predisposing gene variants interact with age

The syndrome of Prader-Willi has two distinct phenotypic stages. In infancy, it is characterized by poor suck, feeding problems and failure to thrive, followed by hyperphagia in later childhood that leads to excessive weight gain (78). Rare deletions in the region p11.2 of the chromosome 16 have been associated with a highly penetrant mendelian form of obesity with additional developmental features (79). These individuals generally have early feeding and growth difficulties, and start to gain excessive weight around 5-6 years of age. As a result, an incomplete penetrance for childhood obesity but a complete penetrance for adult obesity has been observed for the carriers of the chromosome 16p11.2 deletion (79, 80). The longitudinal study of adult MC4R mutation carriers shows an increasing age-dependent penetrance (18).

The life-course analysis of the intronic FTO gene variant and BMI in longitudinal studies indicates that this polygenic obesity-predisposing variant is negatively associated with BMI during infancy (age: 0-2.5 years) but positively associated with BMI from the age of 4 years, with an age-dependent increase during childhood, adolescence and young adulthood (36, 81-84). Most of the effect of the FTO intron 1 SNP on BMI gain occurs during this period, and no appreciable effect of FTO on BMI increase is observed during adulthood and agedness (83, 85-88). Studying the association of an obesity gene score from multiple markers in longitudinal cohorts provided similar results: the genetic predisposition score displayed a moderately positive association with birth weight, and more strongly associated with BMI gain during early infancy and childhood, but no association with BMI change during adulthood was observed (89-91). A negative genotype x age interaction between the PCSK1 rs6232 SNP and obesity traits was observed in two independent studies, and a recent meta-analysis of up to 331,175 individuals confirmed this result and identified a similar interaction between age and the PCSK1 rs6235 SNP on obesity (92-94). A genome-wide interaction meta-analysis also identified 15 BMI loci with age specific effects, 11 of which showed greater effect sizes in younger (<50 years) compared to older (≥ 50 years) adults (77).

Obesity predisposing gene variants interact with lifestyle factors

Obesity predisposing gene variants interact with an obesity-prone environment

The promotion and globalization of societal changes leading to an imbalance between calorie intake and calorie expenditure partly explain the current obesity epidemic, but interactions between genes and this obesity-prone environment also contribute to the development of obesity. Dudley et al. reported a significant cohort effect on the prevalence of obesity in Prader-Willi syndrome (62). Prevalence of obesity was higher in patients born after 1990 than before (62). A generation-dependent penetrance of MC4R pathogenic monogenic mutations on obesity was also found in multigenerational pedigrees, with the effect of mutations on the obesity phenotype being amplified by the emergence of an "obesogenic" environment (18, 95).

This trend is supported by a recent analysis of the Framingham Heart Study (FHS), which demonstrated that risk allele carriage in FTO rs9939609 was associated with a greater increase in BMI among individuals born after 1942 compared to those who were born before 1942 (96). The FTO intron 1 variant is weakly associated with BMI in South Asian Indian populations, but its effect on weight is stronger in urban compared to
rural dwellers (97, 98). A lack of association of FTO with obesity-related traits was also observed in a Gambian rural population (99). The authors speculate that the impact of genetic variance in FTO rs9939609 on BMI may be marginal in lean populations where excess food is scarce, compared to populations where food is abundant (99). Lastly, the growing influence of obesity predisposing genes in ‘obesogenic’ environments has also been supported by the positive interaction between birth year and the impact of 32 obesity predisposing genes (100). Together, these data suggest a stronger influence of genetic factors on obesity in obesity-prone environments.

**Obesity predisposing gene variants interact with physical activity**

Recent data indicate that genetic predisposition to obesity can be blunted in part through physical activity. Over twenty independent studies reported an interaction between the FTO obesity risk genotype and physical activity on BMI variation or obesity in children, adolescents and adults (47, 101-122). An interaction between FTO intron 1 variant and the level of physical activity on obesity was recently confirmed in a meta-analysis of 218,166 adults where physical activity attenuated the odds of obesity by 27% conferred by the variant (123). No such interaction was found in 19 268 children and adolescents (123). We recently studied more accurate surrogates of physical activity and adiposity in a multi-ethnic study of 17 423 participants recruited in 17 low-, middle- and high-income countries and we observed that the effect of FTO rs1421085 on the variation of body adiposity index was reduced by 56% in the higher versus lower metabolic equivalent score tertiles (121). Similar results were obtained for a genetic predisposition score combining the information of 12 obesity-associated SNPs, and a high level of physical activity was associated with a 40% reduction in the genetic predisposition to obesity in adults (N=20 430), as well as for BMI level and BMI change across time (110). Physical activity was also found to attenuate the effect of a 28 SNP obesity gene score on BMI among a sample of East Asians and Europeans (124). We did not evidence any significant interaction between the quantitative level of physical activity and a 14 SNP obesity gene score on BMI or body adiposity index in an international multi-ethnic study of 17 423 participants (121). Our data, consistent with the conclusions of a recent meta-analysis in participants of European ancestry living in North America and Europe, suggest that the benefits of being physically active may be optimal in genetically predisposed people living in the more sedentary countries (121, 125).

A number of recent studies have analysed the interaction between sedentary behaviours and genetic risk for BMI, independent of physical activity level (111, 126, 127). The initial report by Qi et al demonstrated that prolonged television watching accentuated the impact of a 32 SNP genetic risk score on BMI (111), and a second study of an adolescent sample reported that screen time increased the impact of two SNPs (FLJ35779, GNPDA2), although these interactions were ethnic specific and of nominal significance (126). The most recent study of this interaction analyzed how total sitting time impacted the association between FTO rs9939609 and BMI among the Framingham Heart Study (FHS) and the Women’s Health Initiative Study (WHI), but the results were not significant (127).

**Obesity predisposing gene variants interact with diet**
Rouskas et al. reported that the penetrance of *MC4R* loss-of-function heterozygous mutations on obesity is exceptionally low (6.3 %) in the Greek population, in comparison with those observed in other European countries (60-100%) (128). A possible explanation of this ‘Greek paradox’ may be a protective effect of the Mediterranean diet against MC4R deficiency-induced obesity (128).

Several studies have characterized the impact of diet patterns on genetic predisposition to obesity. Independent cross-sectional and longitudinal samples of Caucasians and Latin Americans, suggest that a high daily energy intake, high fat intake or high saturated fat intake can amplify the effect of the *FTO* genotype on obesity risk in children, adolescents and adults (105, 114, 129-132). Higher intake of fried foods has also been shown to increase the impact of a 32 SNP gene score (and an *FTO* variant individually) on BMI over follow-up (133). These interactions were replicated in an independent cohort of 21,421 women (133). A recent 25 year follow-up study in Australia reported an interaction between rs9939609 in *FTO* and diet on BMI change (134). The prudent/healthy diet was associated with a greater BMI change among AA compared to TT genotypes. This interaction was observed at 17 years of follow-up, but was restricted to females (134). Increased intake of sugar-sweetened beverages has also been shown to increase the impact of a 32 SNP genetic risk score on BMI (135). This interaction effect was also observed for incident obesity cases and replicated in an independent cohort (N=21,740) (135).

Despite the many studies demonstrating that diet patterns can moderate the genetic risk for obesity, two recent meta-analyses did not detect significant interactions between diet patterns and obesity-associated gene variants (136, 137). Data from 177,330 adults (87% Whites, 10% Asian, 3% African American) did not indicate any significant interactions between the *FTO* variant and dietary intake of total energy, fat, protein or carbohydrate on BMI (137). A second meta-analysis of 68,317 Europeans did not detect an interaction between a 32 SNP genetic risk score and a multifactorial diet score on BMI (136). The diet score moderated the impact of two SNPs on BMI (*LRRN6C* and *MTIF3*), although these effects were nominal and the impact of these risk variants appeared to be greater among those consuming healthier diets (136). The authors speculate that the broad diet assessment in their analysis may have masked interactions of varying directions and magnitudes that were identified in previous studies (136).

The Apolipoprotein A-II (*APOA2* -265 T>C promoter functional polymorphism appears to interact with high-saturated fat to increase BMI and obesity risk in several independent populations (Mediterranean, Asian, Caucasian, Hispanic and Caribbean) (138, 139). High saturated fat intake was associated with significant increases in the genetic risk for obesity across populations (139). Specifically, the C allele homozygotes with high saturated fat intake displayed a 1.84 (95% CI, 1.38-2.47) odds of obesity compared to a 0.81 (95% CI, 0.59-1.11) odds in those with low saturated fat intake (139). A separate analysis of 1,225 obese adults demonstrated the C allele homozygotes with a high saturated fat intake (≥ 20.7 g/day) had higher waist circumference values than individuals with any other genotype in the high saturated fat intake group (140). While the underlying biological mechanism explaining this association is not fully understood, the *APOA2* -265 T>C SNP has been associated with obesity risk eating behaviours such as meal skipping, and dietary modulation of plasma ghrelin (140).
The Apolipoprotein A5 protein influences plasma triglyceride concentrations in humans and regulators of the APOA5 gene (peroxisome proliferator-activated receptors, insulin, thyroid hormone) have been implicated in obesity risk (141, 142). In a weight loss study of 606 men with hyperlipidemia, C allele carriers of the -1131 T>C variant in the APOA5 gene displayed significantly greater BMI reduction while on a fat restriction diet (143). Additional evidence from the Framingham risk study suggests that carriers of the mutant C allele may have a lower risk of obesity compared to T allele homozygotes when consuming a diet high in monounsaturated fats (144). This interaction was also tested in a Mediterranean sample and greater fat intake was associated with obesity among T allele homozygotes while no association was observed among carriers of the mutant C allele (145). These studies suggest that the C allele in APOA5 may have a protective effect against obesity among individuals consuming a high fat diet.

Several studies have examined the interaction between diet patterns and PPARG Pro12Ala polymorphism with regards to obesity (146), (147). The risk allele 12Ala has been linked to increased obesity risk through meta-analysis, although some heterogeneous effects of this mutation have also been observed in interaction studies (148-150). Lamri et al analysed the interaction between PPARG Pro12Ala polymorphism and fat intake, and observed that AlaAla individuals displayed greater BMI values than Pro carriers among high fat consumers (151). In contrast, a study of 720 French Canadians found that higher amounts of saturated or total fat consumption were associated with greater waist circumference in Pro allele homozygotes but not in 12Ala carriers (147). Similar results were observed from studies analysing BMI. An investigation of the Health Nurses’ Study demonstrated that high total fat intake was associated with greater BMI among participants homozygous for the Pro allele but not among 12Ala carriers (152). This study also reported that monounsaturated fat intake was associated with decreased BMI among 12Ala allele carriers and this interaction was replicated in an independent weight loss study (153). Additional interaction studies of PPARG related to body composition have shown sex-specific effects (154, 155) and weight change analyses have reported inconsistent results (153, 156-159).

A more recent analysis found an interaction between an eight SNP obesity gene score and mono and polyunsaturated fatty acid intake (160). Among 2 346 children with low unsaturated fat intake, the gene score was associated with increased body fat mass index yet no association was present among unexposed children (160).

Obesity predisposing gene variants interact with psychosocial stress

A recent genome-wide interaction analysis identified a significant interaction between psychosocial stress and five SNPs within the Early B-cell Factor 1 (EBF1) gene and hip circumference (161). The interaction reached genome-wide significance among the subset of 2460 Whites in the Multi-Ethnic Study of Atherosclerosis (MESA) but was not significant among Chinese Americans, Blacks or Hispanics. This study reported that the impact of risk allele carriage in EBF1 on hip circumference was greater among participants with a greater chronic stress burden (161). The authors also replicated the interaction between psychosocial stress and three of the original five SNPs (rs17056278 C>G, rs17056298 C>G, and rs17056318 T>C) in EBF1 in the Framingham Offspring cohort(161). A subsequent analysis by the same research group replicated the EBF1 x psychosocial stress interaction on obesity (waist circumference or BMI) in the Family
Heart Study Whites and at trend level in the Duke Caregiver study (162). The direction of the interaction effect was consistent across each of the studies: chronic psychosocial stress amplified the effect of EBF1 variation on BMI (162).

**Obesity predisposing gene variants interact with educational status**

Epidemiological studies have shown an association between a low level of education and higher risk of overweight and obesity (48). A significant negative association between BMI and educational status was found in non-carriers of MC4R mutations but not in MC4R loss-of-function mutation carriers issued from the same pedigrees (18). These results show that a high level of education has no protective effect on obesity risk in presence of MC4R pathogenic mutations.

On the contrary, a significant gene x education interaction has been found in the intron 1 variant in FTO, the significant effect of the SNP on BMI and obesity risk restricted to subjects with no university education (163). This finding is supported by a recent study of European children (N=16 228) indicating that favourable socioeconomic status is protective against obesity, yet this effect was only observed in participants with the low risk genotype TT in FTO rs9939609 (164).

**Obesity predisposing gene variants interact with smoking status**

A meta-analysis of nine European study samples (N=24 198) demonstrated that smoking status moderated the association between genetic variation at the CHRNA5-CHRNA3-CHRNB4 locus (rs1051730) and BMI (165). While there was no evidence of association between variation at rs1051730 and BMI in never smokers, each additional risk allele (T) was associated with a BMI decrease of 0.16 and 0.33 kg/m² among former and current smokers, respectively (165). A separate study of 14 131 Pakistani adults reported another gene x smoking interaction: the minor allele (T) in FLJ33534 was associated with lower BMI in current smokers and positively associated with BMI among adults who had never smoked (166). A number of gene x smoking interactions were identified when African Americans and Caucasians were analysed separately, but no significant interactions were observed in the overall sample from the Southern Community Cohort Study (167). Four nominally significant gene x smoking interactions were reported in a recent study of 16 157 Pakistani adults, with current smoking status amplifying the effect of PTBP2 rs11165643, HIP1 rs1167827 and GRID1 rs7899106 SNPs, and decreasing the effect of C6orf106 rs205262 SNP (120).

**Obesity predisposing gene variants interact with alcohol consumption**

Among a sample of 3 522 East Asians, increased alcohol consumption was associated with an increase in the effect of a 29 SNP GRS on BMI (122). Increased alcohol intake was also reported to increase the impact of PPARGC1A rs4619879 among African Americans, but this interaction was not significant in Caucasians or in the combined sample (167).

**Obesity predisposing gene variants interact with disease status/response to treatment**

**Obesity predisposing gene variants interact with specific health conditions**
Beyerlein et al suggest that pre-existent overweight may double the effect of an obesity genetic predisposition score on body fat mass in 4,837 European children (168). This association is supported by an independent study of 7,225 children of European ancestry which found that previously identified obesity predisposing loci had a greater impact on BMI among obese children compared to their non-obese counterparts (169). Similar results were observed in 1,930 adults of European descent (170). If true, it signifies that obesity predisposing genes may have an even more detrimental effect on weight gain once overweight/obesity is established. Depression predicts subsequent development of obesity (171) and depression status has been shown to amplify the effect of FTO SNPs on BMI (172). Moreover, obesity has an important role in the etiology of polycystic ovary syndrome (173) and FTO intronic SNP has larger effects on BMI in patients with polycystic ovary syndrome than in subject from the general population (174, 175).

**Obesity predisposing gene variants interact with lifestyle modifications**

A strict, fat-reduced, and carbohydrate-modified diet leads to a long-term marked weight reduction in adolescents with Prader-Willi syndrome who are already overweight (176). Importantly, if diagnosis is made at an early age and intensive diet management starts early, reasonable weight control is achieved in non-obese patients with Prader-Willi syndrome (177, 178). Regular exercise training has beneficial effects on body composition and weight loss in Prader-Willi syndrome patients (179, 180), especially as they tend to be less physically active than obese non-syndromic individuals (181). MC4R or POMC monogenic patients respond as well as non-monogenic obese patients to hypocaloric dietary or multidisciplinary (exercise, behavior, nutrition therapy) interventions (182, 183) but MC4R monogenic patients fail to maintain weight loss after intervention (183).

The obesity risk variant rs9939609, an allele in FTO, does not modify the weight loss response to lifestyle intervention (184-186) or caloric restriction (187, 188), but is associated with lower additional weight loss and higher risk of weight regain during the weight maintenance phase that follows the caloric restriction program (188). Carriers of the FTO intron 1 obesity risk variant experience a higher rate of dropout when they are submitted to a high-fat/low carbohydrate (in comparison with a low-fat / high carbohydrate) hypocaloric diet (189), but they achieve better weight maintenance than wild-type individuals during a 3-year intervention with a Mediterranean diet (190). FTO variation has also been shown to interact with protein intake to moderate the response to weight-loss interventions (191, 192). A two-year randomized control trial (RCT) found that higher protein intake was associated with improved weight loss and body composition among carriers of the FTO rs1558902 risk allele compared to non-carriers (191). Another analysis of the same trial (N >700) showed that individuals with the FTO rs9939609 risk (A) allele achieved more favourable changes in food cravings and appetite when consuming a higher-protein weight-loss diet (192). Carriers of the FTO obesity risk alleles also lose less weight in response to exercise training (193, 194). Among five childhood obesity susceptibility loci identified in a French-German genome-wide association studies (GWAS) meta-analysis (195), only one (SDCCAG8) was associated with differential weight loss after lifestyle intervention in 401 children and adolescents (196). Eight out of 15 obesity predisposing gene variants recently identified by GWAS
showed trends of association with weight loss or weight regain during lifestyle intervention in 3356 adults of the Diabetes Prevention Program (197).

The *PLIN* gene has received increasing support for its role in obesity risk and insulin resistance, and genetic variation at the perilipin locus has been shown to interact with diet behaviours (198-200). Two separate studies have shown that A carriers of the 11482 G>A SNP at the perilipin locus lost less weight in response to weight loss interventions compared to non-carriers (199, 201).

**Obesity predisposing gene variants interact with therapeutic treatment**

As most obese persons are resistant to the weight-reducing effects of leptin, administration of recombinant leptin to obese subjects does not generally result in significant weight loss (202). However, patients with congenital leptin deficiency markedly reverse obesity and associated phenotypic abnormalities when they are treated with daily injections of recombinant human leptin (203, 204). Leptin administration reduces energy intake, fat mass, hyperinsulinemia, and hyperlipidemia, restores normal pubertal development, endocrine and immune function and improves neurocognitive performances (205). Although patients with complete leptin deficiency are extremely rare, leptin supplementation may eventually help a far greater number of obese patients with partial leptin deficiency (heterozygous for a loss-of-function mutation in the *LEP* gene) based on the observation that leptin therapy induces more significant weight loss in subjects with low leptin levels (206, 207).

The guanine nucleotide binding protein beta polypeptide 3 (*GNB3*) C825T functional gene variant predicts that obese individuals will benefit more from the anti-obesity drug sibutramine treatment. Sibutramine is a serotonin and norepinephrine reuptake inhibitor and given that *GNB3* variance is associated with an altered response to G protein subunit activation (α 2-adrenergic activation) (208), there is biological evidence to support this interaction. Two independent studies showed that the carriers of the 825 T allele lose more weight in response to sibutramine administration than C allele homozygotes (209, 210). Lastly, obesity predisposing gene variants in *FTO* and *MC4R* are associated with more weight gain in response to antipsychotic treatments (211-215).

**Obesity predisposing gene variants interact with bariatric surgery**

Bariatric surgery is the most effective long-term treatment for severe obesity (216). Laparoscopic adjustable gastric banding did not result in a long-term weight reduction in a 18-years old patient with complete MC4R deficiency (217), and was associated with a high risk of conversion to bypass operations in individuals with partial MC4R deficiency (218). On the contrary, three studies confirmed that Roux-en-Y gastric bypass surgery was an efficient strategy to lose weight in *MC4R* mutation carriers (219-221). These results suggest that diversionary operations, which are more invasive but efficiently improve the neuro-hormonal control of satiety than gastric banding procedures, be recommended in the context of non-syndromic monogenic forms of hyperphagic obesity.

*FTO* risk allele carriers lose less weight than common allele homozygous individuals after banding surgery (222, 223), but experience a similar level or more weight reduction after gastric bypass surgery (222, 224).
Biological processes underlying statistical gene-environment interactions

Epigenetic changes are believed to be one of the primary mechanisms explaining interactions between environmental exposures and genetic variation (225, 226). Other potential mechanisms propose that the transcription changes induced by environmental exposures may vary across genotypes (225). Epigenetics is defined as heritable changes in gene function that cannot be explained by changes in the deoxyribonucleic acid (DNA) sequence, and the three main types of epigenetic modification in mammals include DNA methylation, histone modification and non-coding RNA (226, 227). Of these mechanisms, DNA methylation has received the most attention in human studies, and in mammals this process mainly occurs at CpG dinucleotides (228). Specifically, covalent bonding of a methyl group to the cytosine base creates a barrier that inhibits transcription factors from binding to the DNA helix (228, 229). CpG DNA methylation at gene promoters is typically associated with gene silencing, whereas CpG methylation in gene bodies is linked to gene activity (228). Given that epigenetic differences have been linked to obesity status, as well as genetic variation and a variety of pre and postnatal environmental factors, these processes likely represent a plausible mechanism of gene-environment interactions.

The emergence of new approaches to study epigenetic variation, such as epigenome-wide association studies (EWAS), has led to the identification of methylation patterns associated with obesity (230). Increased BMI among adults was found to be positively associated with increased methylation at the hypoxia-inducible factor 3 alpha (HIF-3α) locus in blood cells and adipose tissue (231). This finding was confirmed in two replication cohorts in the initial analysis (231), as well as two additional independent studies (232, 233). A separate EWAS of an African American sample identified an association between methylation at 37 CpG sites and BMI, which were replicated in two cohorts of European ancestry (232). Analyzing whole-genome DNA methylation and expression data in human adipose tissue from 96 males and 94 females revealed that DNA methylation and expression of 2825 genes was correlated with BMI (234).

Existing evidence also supports the association between environmental exposures on DNA methylation patterns. Monozygotic twins, who are epigenetically indistinguishable at birth, exhibited drastically different overall content and genomic distribution of DNA methylation and histone acetylation in later life (235). Moreover, methylation and expression of 1050 genes have been found to vary with age (234, 236). The epigenetic divergence that occurs with aging likely reflects the accumulation of environmental exposures that influence methylation patterns. Prenatal factors including maternal BMI and variations in maternal methyl donor intake during pregnancy have been linked to methylation changes in the offspring (237), and multiple studies have shown that maternal vitamin B12, folate and cobalamin levels during pregnancy are associated with offspring adiposity (238, 239). Folate, vitamin B12 and choline are methyl donors and involved in the synthesis of methionine, the precursor of the universal donor of methyl groups needed for DNA methylation (S-adenosylmethionine). As a result, disregulation in any of these components can alter the epigenomic regulation of gene expression (240). With respect to postnatal determinants of DNA methylation, exercise interventions have been shown to alter the DNA methylation of 2817 genes in skeletal muscle and 7663 genes in adipose tissue (18 of which were obesity candidate
genes) (241). The effect of exercise on DNA methylation appeared to be tissue specific, with the majority of genes in skeletal muscle displaying decreased DNA methylation (242), and the majority of genes in adipose tissue showed increased DNA methylation (243). These changes mirrored the patterns observed for gene expression: most of the genes showing concurrent changes in DNA methylation and expression displayed increased expression in skeletal muscle and decreased expression in adipose tissue (241). A recent review of 25 studies (16 observational 9 interventional) found that both acute and chronic exercise significantly influenced DNA methylation, and these changes occurred in a tissue- and gene-specific manner (244). DNA methylation changes have also been observed in response to high-fat intake (245-248), and after weight loss interventions the methylation profiles of adipose and muscle tissue among those formerly obese became more similar to lean individuals (249-252). These methylation changes involved a number of known obesity-associated loci, including LEPR, STAB1, ZNF608, HMGA1, MSRA, TUB, NRXN3, FTO, MC4R and BDNF (250, 251). Both exercise and diet have long been recognized as central determinants of body weight regulation in epidemiological studies (253, 254). The epigenetic changes induced by these exposures likely explain a portion of their association with BMI as well as their interaction with obesity predisposing gene variants (102, 125).

Although environmental factors have the potential to influence the epigenetic environment, it is estimated that approximately 20-40% of epigenetic variation can be attributed to genetic differences (226, 255, 256). Early evidence demonstrated that the risk allele of FTO promotes increased methylation of sites within intron 1 of the FTO gene, as well as greater methylation of additional genes (257). Other evidence identified 28 obesity-associated SNPs that were associated with differential methylation at 107 proximal CpG sites (258). A recent study of Trim28 haploinsufficiency used findings from mice and humans to demonstrate the variation in obesity phenotypes that can be induced through epigenetic changes (259). These authors also reported that FTO expression was decreased among Trim28_Low obese children compared to Trim28_Low lean individuals (259).

These epigenetic findings support an emerging mechanistic model to explain gene-environment interactions in obesity. Existing evidence indicates that obesity, pre and postnatal environmental factors and multiple obesity associated gene variants are linked with epigenetic patterns (Figure 1). Given that gene variants, such as those in FTO, and environmental factors both play a role in the methylation of obesity genes, the balance between these two effects likely impact the manifestation of genetic obesity. This biological model provides support for many of the statistical interactions reported to date, and further integration of genomic, epigenomic and transcriptomic data with gene-environment interaction studies will aid in uncovering these biological mechanisms.

Methodological considerations in gene-environment interaction studies of obesity: recent developments and future options

Despite the rapid growth of GEI studies in the field of obesity, there has been increasing scrutiny regarding the validity of these studies based on several issues including statistical modelling (260, 261), confounding (260, 262), a low replication rate (263-265), underpowered analyzes (266), lack of biological assumptions (267, 268) and
measurement precision (269) (Table 1). The relevance of testing interactions between individual genetic variants and specific environmental exposures has also been questioned (260). Based on these concerns, some leaders of opinion have suggested that a large proportion of significant gene x environment interaction (GEI) findings are in fact false positives (23, 260). This scepticism has been adopted by multiple journals, which have implemented stringent criteria for candidate gene and interaction studies considered for review (270, 271).

This portion of the review will focus on issues in GEI studies related to (1) statistical modelling of interaction terms, (2) modelling of confounding variables, (3) timing of environmental exposure across the life span and (4) measurement of predictor and outcome variables. The final section will provide suggestions to these issues based on existing evidence and will outline future directions of GEI research.

**Statistical modelling issues in gene-environment interaction research**

Using a multiple linear regression model with the inclusion of a cross-product term signifying the product of environmental (E) and genetic (G) variables is the most common method to assess interactions (107, 114, 117, 125). Coding genetic polymorphisms is either performed to create a binary variable (under a recessive or dominant model) or a three-category variable based on an additive model, with the latter often used when the true functional model of a given marker is not known (261). A recent analysis demonstrated that modelling gene-environment interactions with a simple cross-product term (G x E) often produces misleading results when assuming an additive model (261). First, the simple cross-product model always forces the regression lines to be ordered (0, 1, 2 and never 0, 2, 1). While this assumption may be intuitive from a biological perspective, this approach will always predict an ordered effect of genotypic differences even when the data do not reflect this assumption (265). Second, the differences in slopes between the adjacent regression lines are always assumed to be the same (261). There is no rationale for this assumption and in practice, sampling error alone would be expected to create uneven differences between regression slopes (272). Alternatively, non-linear gene-environment interaction effects may be present, which could not be estimated accurately with only a cross-product interaction term (261, 273). Third, this model constrains all three regression lines to cross at the same point when interaction effects are present (261). This implies that there is a certain level of environmental exposure that confers the same level of risk/disease for all three different genotypes. There is no statistical or biological evidence to justify this assumption, especially since the specific genetic model is often not established for the genetic marker of interest (260). These simulations indicate that the models using only the cross-product term are more vulnerable to Type 1 and Type 2 errors (261). In all cases, including two additional coefficients, one to model non-linear genetic effects (βG²) and another to account for non-linear interaction effects (βGE x E), represented the interaction (or lack of interaction) more accurately. Many authors recommend this model for genetic variants following an additive or unknown genetic model, and emphasize that failure of an interaction to match a plausible biological interaction likely indicates a false positive result (261, 274). In summary, re-conceptualizing interaction models to account for non-
linear effects removes the constraints of traditional regression techniques and provides a more accurate representations of gene-environment interaction effects (261).

Other authors contend that traditional GEI analyses neglect to test the \textit{a priori} hypotheses that form the basis of these studies (275). The implicit framework adopted in traditional GEI analyses is the diathesis-stress model of environmental action (276), which specifies that certain individuals are more vulnerable to the adverse consequences of some exposures than others (277). As an extension of this limitation, exploratory approaches also fail to test or compare the competing predictions from alternative theoretical frameworks such as the more recent differential-susceptibility framework (278). This theory posits that some individuals are more susceptible to not only negative exposures, but to positive environmental influences as well (279). Based on this characterization some authors have proposed that gene variants classically referred to as ‘vulnerability genes’ be reclassified as ‘plasticity genes’ to correspond with the differential susceptibility framework (280). This theory has been proposed by several authors (278, 281), and has been applied to the study of GEI (280, 282, 283). In response to these competing frameworks, many statistical criteria were developed to distinguish differential-susceptibility interactions from those representing diathesis-stress theory (279, 284), and Widaman’s confirmatory method appears to be the most efficient (285). This technique directly evaluates alternative theoretical frameworks by aligning the analyses with each hypothesis (285). Specifically, this method systematically adjusts the parameters included in the regression equation to compare different theoretical frameworks, and specifies where the regression lines (representing each genetic subgroup) will cross relative to the value of the environmental exposure (285). With respect to the frameworks discussed above, the diathesis-stress theory models an ordinal interaction whereby the predicted outcome value for the genetically vulnerable subgroup is always less than that of the genetically low-risk group. The differential-susceptibility framework predicts that the risk of the outcome for the genetically malleable group can be higher or lower than the genetically non-malleable group depending on the level of the environmental exposure (275). It is important to note that Widaman’s confirmatory approach can be used for dominant/recessive or additive genetic models, and can be applied to other forms of statistical interaction involving competing hypotheses about the nature of the interaction (275, 286). Integration of this technique into interaction studies where the theoretical framework is uncertain may help to improve the accuracy and replication rates of interaction studies. This model is supported by biological evidence from a recent study, which demonstrated that TRIM28 knockout mice are alternatively lean or obese depending on subtle environmental changes (259). Other consequences of this framework have important implications for variants identified in GWAS. If a ‘plasticity gene’ displays opposite effects in two exposure groups, the main effect of this variant may not be identified in GWAS and this interaction effect could be missed. Another concern for GWAS is the possibility that interaction effects may only exist among subgroups with very rare exposures. Analyzing both ‘vulnerability’ and ‘plasticity’ gene frameworks may be a method to prevent these issues and identify additional gene-environment interactions in future studies.

Another consideration when analyzing interaction effects is the selection of either an additive or multiplicative interaction scale (287-289). This decision has important implications given that different scales can lead to different conclusions and
consequently, different public health recommendations (290). An additive interaction exists when the combined genetic and environmental risk is significantly greater than would be expected if their effects were additive, whereas a multiplicative interaction describes a joint genetic and environmental risk that is greater than expected from multiplying their effects (23, 291). Some authors argue that the selection of measurement scale is less crucial when the underlying biological processes are not known, and both scales can be appropriate in certain situations (292-294). If the pathophysiology consists of a multistage process, such as cancer initiation and promotion stages, two factors that act at the same stage will generally fit an additive model and those acting at different stages will typically fit a multiplicative model (295, 296). It has also been suggested that if the main objective is to study public health impact, an additive scale is better suited to identify heterogeneous effects across subgroups, while the multiplicative scale is more appropriate for studying disease etiology (295).

Confounding issues in gene-environment interaction research

Several modelling strategies have been proposed to address the impact of confounding in gene-environment interaction studies (260, 262). Variables with the potential to offer alternative explanations of an interaction are typically entered into the regression equation as covariates to control for their potential confounding effects (265). While this method controls the influence of confounding on the main effect of the genotype and environment, it does not adjust for potential confounding of the interaction term (260). An alternative method has been proposed whereby all covariate x gene and covariate x environment interaction terms are included in the model that tests the gene x environment interaction of interest (260). If significant covariate interactions are observed, the validity of any gene x environment interactions may be compromised by the covariate and warrant additional analysis. Although there are potential objections to this modelling technique, the justifications of this approach are outlined by Keller (260). First, even though over-fitting the model may prevent accurate estimations of the many covariate interactions, the purpose of including covariate interactions is to control for their effects on the gene x environment interaction rather than producing accurate parameter estimates. Second, multicolinearity between the many interaction terms may diminish the strength of the main gene x environment interaction. This however is the purpose of this procedure and if inclusion of the covariates weakens the main interaction, then the covariates may be significantly influencing the interaction. Lastly, it is reassuring to recognize that the gene x environment interaction term is only marginally affected if there is no ‘true’ relationship between the covariate and the gene x environment interaction (260). One caveat to this approach is that shared heritability between the covariates and the outcome can introduce bias and increase the risk of false-positive results (297). Therefore, including heritable covariates in the model should be avoided if they are associated with the gene variant being tested (297).

Confounding issues can be further complicated if the interacting genetic variant and environmental exposure of interest are correlated (297, 298). Under these circumstances, simulations have demonstrated that uncontrolled confounding will bias the estimates of the main genetic effect and the gene-environment interaction even if the genetic and environmental factor are not associated with the outcome (299).
variant and environmental factors are independent, this is no longer an issue as long as unmeasured environmental confounders are not related to genetic factor. The issue of gene-environment dependence has been highlighted in extreme cases where the genetic variants are associated with both the environmental factor and the outcome. For example, variants on 15q25 have been linked to both smoking behaviours and lung cancer (300-302). As a result, some authors suggest directly analyzing the relationship between the interacting genetic variant and environmental exposure (299).

**Considering time of exposure in gene-environment interaction research**

Given that gene expression and silencing varies significantly throughout development, it may be important to consider time of exposure when modelling exposures that can have differential effects throughout the life cycle (303). Evidence from toxicology research indicates that many environmental exposures display distinct dose response curves that vary based on the developmental stage at which exposure occurs (304, 305). The identification of these developmental windows suggests a need to include time of exposure as a third interacting factor when analysing gene-environment interactions (268). However, the inclusion of a three-way interaction term dramatically increases the necessary sample size (260, 306) and this information is rarely available. Simulation studies indicate that the sample size required to detect three-way interactions is four-fold the sample size necessary to detect a two-way interaction of the same magnitude (307). Another statistical method to address this issue involves considering environmental exposure as a time-varying factor to analyze the lag effects of gene x time-varying environment interactions (268). Yet, the repeated measurements needed to measure lag effects are often not feasible due to the cost of repeated measurement in large studies. This constraint explains the high prevalence of cross-sectional case-control designs to study gene-environment interactions (268). The challenge of measuring variations in the impact of environmental exposures is compounded by changes in the heritability of the outcome across time. A meta-regression of heritability studies of BMI found that the genetic contribution to BMI varies with age: heritability was positively associated with age among child studies and negatively associated with age among studies of adults (308). A recent genome-wide interaction meta-analysis identified 15 BMI loci that interacted with age, 11 of which had a greater effect impact in younger (<50 years) compared to older (≥ 50 years) adults (77). Failure to address the time-varying effects of environmental exposures and heritability may account for some of the challenges with replicating gene-environment interactions (23, 260, 309).

**Measurement issues in gene-environment interactions research**

The measurement of the exposure and outcome also represent important considerations for gene-environment interaction research. Major determinants of power include allele frequency, genetic effect size, the magnitude of interaction effect, risk allele frequency, degree of genetic misclassification and measurement error associated with the exposure and outcome (269, 306, 310). Although the trade-off between precision and feasibility is common to most study designs, the large samples required to study interaction effects make this balance particularly important. Currently, the most notable
gene-environment interactions in obesity have measured diet patterns or physical activity as environmental exposures (102, 107, 311, 312). The gold standard criterion measure for these exposures are a 7-day weighed diary and doubly labelled water, respectively. Unfortunately, the large number of participants required for these studies have restricted the measurement of these exposures to less precise instruments. The error associated with exposure measurement generally attenuates the estimate of the true effect size (313, 314). Similar problems occur when the outcome used is an indirect measure for the true outcome of interest. In gene-environment analyses of obesity, BMI is commonly used as the outcome (102, 125, 315, 316), which further contributes to this error given that BMI fails to distinguish between fat and fat-free mass (317).

Simulations have characterized how varying different determinants of power can impact the required sample size of gene-environment interaction studies (269, 306). As an example of these analyses (269), genetic misclassification was fixed at 2.5% to be consistent with prior empirical studies (318, 319) and the magnitude of effect for the common allele was also constant. With a correlation between the true and observed exposure and outcome of 0.6 and 0.7, respectively, a sample size of just over 9500 is needed to detect an interaction at a significance of $10^{-4}$ with 95% power (17). However, the correlations of 0.6 and 0.7 between the true and observed exposure and outcome are unusually high for gene-environment interactions in obesity due to the cost of precise measurement tools (125, 310, 320). With more typical correlations of 0.3 and 0.4, the required sample size can increase to over 100 000 participants with all other variables held fixed (269). If precise instruments are not available to mitigate this error, performing repeated measurements is a useful strategy on condition that the error in repeated measures is not correlated (313). As an example, performing two independent repeated measures using a tool with a validity coefficient of 0.6 increases the overall validity coefficient to almost 0.8. With all other variables being fixed, this reduces the necessary sample size by more than a factor of six (269).

The value of improving measurement accuracy as opposed to increasing sample size can be reinforced with the example of physical activity measurement, a common exposure analyzed in gene-environment interactions of obesity (117, 125, 269). Physical activity is usually assessed by questionnaire, and even comprehensive instruments that address occupational and leisure activity rarely correlate with objective measures of energy expenditure above 0.3 (321). The physical activity assessment used in the EPIC-Norfolk study displayed an overall correlation of 0.44 with objective measures, although this fell to 0.28 after adjustment for age and sex (322). The error associated with measuring this exposure is compounded by the moderate correlation (0.5) of BMI with body fat percentage as measured by dual-energy x-ray absorptiometry (323). Using the EPIC-Norfolk questionnaire with BMI as an outcome would require almost 90 000 participants to detect an interaction that doubled the effect of a genetic variant, when the variant is present in 20% of the population (269). Since a doubling of genetic risk from an environmental exposure is at the upper limit of interaction effect estimates reported for common variants and exposures (324-326), some authors speculate that the majority of published interaction studies are underpowered and report “lucky” true-positives or false-positive results (23). A recent study by our group provides an empirical example of how measurement precision can influence statistical power. We analyzed physical activity x FTO interactions on BMI using two measures of physical activity: a three-level
categorical variable and a quantitative estimate of energy expenditure. The categorical data was available in 99% of the sample while the quantitative energy expenditure data was only available in 63% of the sample. Despite this disparity, similar interactions were detected using both measures, which may suggest that the added precision of the energy expenditure data compensated for the decrease in sample size (121).

Given the sample size requirements imposed by this type of data, more direct measurement techniques have been proposed. Objective measures such as heart rate monitors carry increased precision while maintaining feasibility in moderately sized epidemiological studies (327). Heart rate monitor data have demonstrated a correlation with the gold standard of energy expenditure methods (doubly labelled water) of 0.73 (328). Two repeated measures can increase this correlation to over 0.88. Substituting this method of exposure measurement for questionnaire methods would decrease the necessary sample size to 9453, a decrease by a factor of 10 (269). Therefore, the gain in precision associated with more accurate measurements of exposure may be less resource intensive than accruing large sample sizes. The power implications of using precise measurement techniques suggest that smaller studies with more accurate measures of exposure and outcome may be better suited to detect gene-environment interactions than large sample sizes with imprecise measurement (269). The issue of measurement imprecision has long been debated in the nutrition field and ‘deep phenotyping’ strategies (measuring metabolic markers such as circulating plasma lipids as a surrogate of a high-fat diet) may be worthwhile alternatives to traditional self-report measures (329-331). Other assessments that may mitigate the concerns associated with traditional diet measures include ad libitum energy intake tests or analyzing the dietary information of food consumed in regulated settings such as cafeterias or restaurants (332).

The issue of direct and indirect measurement of genetic variants also has important implications for statistical power. In many current GWAS and GEI studies, the true susceptibility loci involved in the disease etiology is not known (or unavailable) and the linkage disequilibrium (LD) between marker alleles and the actual disease loci is used to study associations between gene variants and the phenotype under study (333, 334). Since this is an indirect approach, the effect estimate will be underestimated if the LD between the two variants is incomplete ($r^2 < 1$) (335). Previous studies have demonstrated that the sample size requirements of GEI studies can be strongly influenced by the marker allele frequency, disease allele frequency, the LD between these loci, as well as the main genetic and environmental effect, the prevalence/impact of the environmental exposure and the magnitude of the interaction (334, 335).

Future directions for gene-environment interactions and obesity

Given that specific environments can greatly impact the magnitude of genetic predisposition to obesity, the systematic study of gene-environment interactions constitutes an important field of investigation in order to inform public health strategies to prevent and manage obesity and other complex diseases. Gene-environment interaction studies in the context of various forms of obesity (monogenic, polygenic) and in diverse experimental designs (observational, interventional) (22) may lead to a better understanding of the protective or detrimental environmental exposures that modify the impact of certain genetic variants. Existing interactions need to be studied in additional
obesity-prone (e.g. response to smoking cessation, response to insulin therapy in diabetic subjects) or obesity-protective (e.g. lifestyle intervention, response to the anti-obesity drug orlistat administration or to bariatric surgery) conditions. Gene-environment interaction studies are complementary to observational epidemiology, interventional study or clinical trials, and will certainly help to elaborate efficient strategies to reverse the obesity epidemic.

Currently, GWAS for obesity-related traits have focused on the marginal gene effect ignoring gene-environment interaction entirely (336). Gene-environment interactions are nevertheless frequent in obesity, and statistical models that do not properly account for gene-environment interactions may attenuate the marginal effect size and reduce the power to detect true associations (23, 337). Applying a joint test for a main genotype effect and gene-environment interaction may increase the power to identify an individual SNP associated with a disease outcome (338-340). As many completed GWAS for obesity have been conducted on samples with large amounts of existing environmental data, performing gene-environment-wide interaction studies (GEWIS) in these existing datasets is a cost-effective strategy to find additional obesity-associated gene variants that interact with specific environments but have been missed by conventional GWAS (341). Since large sample sizes and meta-analytical approaches are required to reliably detect SNPs with subtle gene-environment interaction patterns (342), GEWIS for obesity have been initiated in the context of large international obesity consortiums like GIANT (343). Although these methods show promise, recent simulations indicate that this technique is more appropriate for analyzing interactions between genetic risk scores rather than individual SNPs, due to the reduced power when analyzing the small effect sizes of individual SNPs (344). As a potential solution, Marigorta and Gibson suggest selecting participants who are at a high-risk for obesity based on environmental exposure (344). This strategy has the potential to identify environmental exposures that can modulate the impact of specific variants associated with obesity (344). However, challenges associated with GEWIS include identifying adequately sized cohorts with appropriate genetic and phenotypic data, as well as issues with statistical power. As a novel alternative to these techniques, variance prioritization was developed as a method to model genetic associations with genetic variance, without requiring knowledge of the interacting variables (345). Since the main effects of gene variants involved in interactions are typically associated with a large degree of variance, this strategy exploits this pattern to rank and prioritize variance estimates to identify gene variants associated with a large degree of variance in a quantitative trait (345).

Bayesian methods have also been developed to integrate variations in multiple SNPs within a given gene/region, and examine how an environmental exposure moderates the risk of these genetic profiles (346). This method was applied to the Environment and Genetics of Lung Cancer Etiology (EAGLE) study and detected a smoking x genetic profile interaction that was not detected by conventional interaction tests (346). Artificial neural networks have been applied to interaction analyses and simulations suggest that this technique may be particularly valuable for detecting non-linear penetrance and interaction effects (347). Other analytical approaches have been developed to test interactions while addressing the common concern of statistical power. These techniques, termed “cocktail methods,” involve a three-step approach to testing genome-wide gene-environment interactions while preserving the type 1 error rate and
increasing power by 30-40% under certain circumstances (348). These three steps include screening, multiple testing and GxE testing, and current simulations of this technique have been applied to binary environmental variables, although this approach is applicable to continuous environmental data (348). While early analysis of these novel techniques has been positive, further real data application of these methods will reveal the generalizability of these approaches.

Recent GWAS for obesity have collected phenotypic information in individuals living in a broad range of environments. While successful, this approach may fail to identify potential gene variants associated with obesity-related traits in a context dependent manner. Gene identification efforts may therefore be targeted in populations that display homogeneous environment and lifestyle factors across time and across the community, as observed in the Plain people (349). Performing genetic association studies for adiposity change in response to a standard major environment modification (antipsychotic drug use, smoking cessation, intensive caloric restriction, anti-obesity drug therapy, obesity surgery) is another valuable way to control the environmental exposure, lower sources of heterogeneity and provide a more comprehensive molecular basis for genetic predisposition to obesity.

In order to refine the search for interaction variants, statistical GEI tests could be combined with methylation quantitative trait loci (meQTL), expression quantitative trait loci (eQTL), and protein quantitative trait loci (pQTL) to focus on SNPs with a plausible biological mechanism for interaction (258). Specifically, a joint test could be applied to (1) identify genetic variants that statistically interact with a given environmental exposure (e.g. physical activity level) to modulate an outcome (e.g. BMI), (2) test if the same genetic variants are also eQTL, meQTL and/or pQTL for a given locus and (3) determine if the methylation, expression and protein level of the same locus is modulated by the same environmental factor (243). A similar test could be applied to analyse the interaction between an individual SNP and multiple environmental factors. Since methylation is influenced by several environmental exposures (physical activity (243), diet (248), sleep (350)) identifying SNPs that redundantly interact with multiple exposures may be a method to exploit this pattern. The ‘Identifying REdundant Gene-environment InteractionS’ (REGIS) method may increase the probability of detecting ‘true’ and replicable gene-environment interactions. Another avenue for future research is to study gene-environment interactions jointly in mouse and human studies (351). The development of the clustered regularly interspaced short palindromic repeat (CRISPR) system for gene targeting and editing creates a new opportunity to study ‘humanized’ genetically modified mice carrying human mutations (352, 353). Combining this biological data from animal studies with statistical evidence of interaction form human epidemiological studies is also likely to improve the validity of gene-environment interaction studies (23).

Conclusion

Heritability, syndromic, monogenic and polygenic obesity studies provide converging evidence that obesity predisposing genetic factors strongly interact with environment, from birth to agedness and in a wide range of situations. A prolific period of discovery is foreseen in this fast-moving field, especially with the many
methodological innovations that attempt to address the ‘missing heritability’ of obesity. To effectively tackle this knowledge gap, prospective studies need to incorporate current methodological evidence to optimize the validity of emerging evidence. Emerging epigenetic studies have demonstrated that obesity, genetic variants and environmental exposures can influence DNA methylation, which provides a mechanistic model to support the statistical interactions from genetic epidemiology. A comprehensive understanding of gene-environment interactions in obesity may lead to tremendous applications in the emerging field of personalized medicine and individualized lifestyle recommendations. Evidence from interaction studies suggests that specific subgroups of individuals may have an increased risk to develop obesity in specific environments but may also benefit more from lifestyle interventions, a treatment or a surgical procedures (354). This information will help determine if population-wide or personalized subgroup interventions are the best suited to fight the worldwide obesity epidemic (355, 356).

ACKNOWLEDGMENTS
We would like to thank Michelle Turcotte for her technical assistance with figure design and creation.

DECLARATION OF INTEREST
The authors declare no financial interests.

FUNDING
D. M. holds a Canada Research Chair in Genetics of Obesity.
References


34. Dubois L, Ohm Kyvik K, Girard M, Tatone-Tokuda F, Perusse D, Hjelmhorg J, et al. Genetic and environmental contributions to weight, height, and BMI from birth to 19


32


186. Lappalainen TJ, Tolppanen AM, Kolehmainen M, Schwab U, Lindstrom J, Tuomilehto J, et al. The common variant in the FTO gene did not modify the effect of


273. Van Hulle CA, Lahey BB, Rathouz PJ. Operating characteristics of alternative statistical methods for detecting gene-by-measured environment interaction in the


Table 1. Summary of methodological issues and solutions for gene-environment interaction studies in obesity.

<table>
<thead>
<tr>
<th>Methodological Issue</th>
<th>Suggested Solution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modelling the G x E cross-product terms</td>
<td>Include an additional coefficient to model non-linear genetic effects ($\beta G^2$), and a second to account for non-linear interaction effects ($\beta G^2 x E$)</td>
<td>Aliev, <em>Bavav Genet</em>, 2014</td>
</tr>
<tr>
<td>Comparing biological frameworks (e.g. diathesis-stress model vs. differential susceptibility framework)</td>
<td>Adjust the parameters in the regression equation to compare alternate theoretical frameworks</td>
<td>Belsky, <em>Psychol Bull</em>, 2009 Widamen, <em>Psychol Methods</em>, 2012</td>
</tr>
<tr>
<td>Selection of interaction scale (e.g. additive vs. multiplicative)</td>
<td>Consider the application of the interaction test <em>a priori</em>. Additive scales have been recommended for identifying heterogeneous effects across subgroups in public health settings, while multiplicative scales are suggested for studying disease etiology</td>
<td>Ottman, <em>Prev Med</em>, 1996</td>
</tr>
<tr>
<td>Confounding of the G x E interaction term</td>
<td>Include all covariate x gene and covariate x environment interaction terms</td>
<td>Keller, <em>Biol Psychiatry</em>, 2014</td>
</tr>
<tr>
<td>Shared heritability between the outcome and covariates</td>
<td>Avoid the inclusion of heritable covariates that are associated with the gene variant being tested</td>
<td>Aschard, <em>Am J Hum Genet</em>, 2015</td>
</tr>
<tr>
<td>Correlation between the gene variant under study and the interacting environmental factor</td>
<td>Directly analyze the relationship between the interacting gene variant and environmental exposure to ensure that they are not correlated</td>
<td>VanderWeele, <em>Am J Epidemiol</em>, 2013</td>
</tr>
<tr>
<td>Variations in gene expression/silencing, and changing the heritability of BMI throughout development</td>
<td>Use a repeated measures design or include a G x E x Time term if the sample size is sufficient</td>
<td>Liu, <em>Environ Health</em>, 2012</td>
</tr>
<tr>
<td>Measurement error associated with the environmental exposure and outcome</td>
<td>Consider more accurate measurement tools or repeated measures in favour of large sample sizes with less accurate measures</td>
<td>Wong, <em>Int J Epidemiol</em>, 2003</td>
</tr>
</tbody>
</table>
Figure 1. Biological model to explain gene-environment interactions in obesity.