Curr Diab Rep (2016) 16: 57
DOI 10.1007/s11892-016-0758-y

GENETICS (AP MORRIS, SECTION EDITOR)

Putting the Genome in Context: Gene-Environment Interactions in Type 2 Diabetes

Paul W. Franks[1,2,3] · **Guillaume Paré**[4,5,6,7]

Published online: 7 May 2016
© Springer Science+Business Media New York 2016

Abstract The genome is often the conduit through which environmental exposures convey their effects on health and disease. Whilst not all diseases act by directly perturbing the genome, the phenotypic responses are often genetically determined. Hence, whilst diseases are often defined has having differing degrees of genetic determination, genetic and environmental factors are, with few exceptions, inseparable features of most diseases, not least type 2 diabetes. It follows that to optimize diabetes, prevention and treatment will require that the etiological roles of genetic and environmental risk factors be jointly considered. As we discuss here, studies focused on quantifying gene-environment and gene-treatment interactions are gathering momentum and may eventually yield data that helps guide health-related choices and medical interventions for type 2 diabetes and other complex diseases.

Keywords Gene-lifestyle interaction · Cardiometabolic · Obesity · Genomic · Systems biology · Variance heterogeneity · Genotype-based recall · Recall by genotype

Introduction

Although malignant melanoma can be caused by sunburn, fatty liver disease by excessive alcohol consumption, lung cancer by cigarette smoking, and type 2 diabetes (T2D) by obesity, disease is not an inevitable consequence of excessive exposures to these risk factors; by contrast, in Mendelian diseases like familial partial lipodystrophy [1] or phenylketonuria [2], there is much higher certainty that exposure to refined carbohydrates and phenylalanine respectively can cause severe health detriments. The high sensitivity to certain environmental exposures and pharmacotherapies that some people experience and others do not may be governed by genetic

This article is part of the Topical Collection on *Genetics*

Electronic supplementary material The online version of this article (doi:10.1007/s11892-016-0758-y) contains supplementary material, which is available to authorized users.

✉ Paul W. Franks
paul.franks@med.lu.se

[1] Genetic and Molecular Epidemiology Unit, Lund University Diabetes Center, Department of Clinical Sciences, Clinical Research Center, Skåne University Hospital Malmö, Lund University, Building 91, Level 10, Jan Waldenströms gata 35, 205 02 Malmö, Sweden

[2] Department of Public Health and Clinical Medicine, Umeå University, 90188 Umeå, Sweden

[3] Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA

[4] Population Health Research Institute, McMaster University, Hamilton General Hospital Campus, DB-CVSRI, 237 Barton Street East, Room C3103, Hamilton, ON L8L 2X2, Canada

[5] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada

[6] Department of Clinical Epidemiology and Biostatistics, Population Genomics Program, McMaster University, Hamilton, ON, Canada

[7] Thrombosis and Atherosclerosis Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, ON, Canada

factors that interact with these exposures to determine risk. In some instances, genetic variation may be beneficial, rendering the bearers of these mutations especially sensitive to the health-enhancing effects of specific drugs, foods or types of exercise, for example, and in other people genetic factors may augment the detrimental effects of lifestyle and predispose those taking certain medicines to adverse events. Discovering, replicating, validating, and translating information about interactions between genetic variation and environmental exposures and medical therapies has important implications for the prediction, targeted prevention, and stratified treatment of T2D and many other diseases.

The literature on gene-environment interactions in diabetes-related traits is extensive, but few studies are accompanied by adequate replication data or compelling mechanistic explanations. Moreover, most studies are cross-sectional, from which temporal patterns and causal effects cannot be confidently ascertained. This has undermined confidence in many published reports of gene-environment interactions across many diseases; although interaction studies in psychiatry have been especially heavily criticized [3], many of the points made in that area relate to other diseases, not least to T2D, where the diagnostic phenotype (elevated blood glucose or HbA1c) is a consequence of underlying and usually unmeasured physiological defects (e.g., at the level of the pancreatic beta-cell, peripheral tissue, liver, and gut), and the major environmental risk factors are difficult to measure well. Nevertheless, several promising examples of gene-environment interactions relating to cardiometabolic disease exist, as discussed below and described in Table 1, and interaction studies with deep genomic coverage in large cohorts are now conceivable; the hope is that these studies will highlight novel disease mechanisms and biological pathways that will fuel subsequent functional and clinical translation studies. This is important, because diabetes medicine may rely increasingly on genomic stratification of patient populations and disease phenotype, for which gene-environment interaction studies might prove highly informative.

How Are Gene-Environment Interactions Defined?

The term *gene-environment interaction* has different meanings to different biomedical researchers (see Supplement 1for glossary of terms used). However, here, we focus on the concept of *effect modification*, where the genetic and environmental exposures convey synergistic effects, or, in other words, where the joint effects are more or less than additive and the estimated genetic effect on a trait differs in magnitude (and sometimes direction) across the spectrum of an environmental exposure. Figure 1 shows three types of interaction effects, and also illustrates why modeling interactions is challenged by *scale dependency* (i.e., where interaction effects are influenced by the scale on which the dependent variable is modeled). In clinical trials, *gene-treatment interactions* are usually considered to occur when the direction and/or magnitude of the treatment effects are conditional on the participant's genotype.

The Rationale for Studying Gene-Environment Interactions

It is often said that T2D is the consequence of gene-environment interactions [17]. Indeed, both the environment and the genome are involved in diabetes etiology, and there are many genetic and environmental risk factors for which very robust evidence of association exists. But when epidemiologists and statisticians discuss gene-environment interactions, they are usually referring to the synergistic relationship between the two exposures, and there is limited empirical evidence for such effects in the etiology of cardiometabolic disease. Indeed, in non-monogenic human obesity, a condition widely believed to result from a genetic predisposition triggered by exposure to adverse lifestyle factors, of the >200 human gene-lifestyle interaction studies reported since 1995, only a few examples of gene-environment interactions have been adequately replicated [18], and because these results are derived primarily from cross-sectional studies with little or no experimental validation, even those that have been robustly replicated may not represent causal interaction effects. The evidence base for T2D is thinner still. Nevertheless, other data support the existence of gene-environment interactions in complex disease, thus motivating the search for empirically defined interactions in T2D.

Some of the earliest empirical examples of gene-environment interactions come from studies in *Drosophila* that show that eye facet number varies both by genotype and temperature [19–21]; similar examples exist for other morphological features of the fly's eyes and head [22]. In agricultural genetics, the need to maintain or improve food security in the face of global population growth, climate change, and land challenges has demanded the cultivation of genetically engineered plants to maximize crop yields conditional on environmental characteristics (e.g., soil quality, precipitation, altitude, or temperature) [23]. Studies of gene-environment interactions in durum wheat, for example, illustrate that in low crop yield regions, the D3415 cultivar performs well, whereas other cultivars (Karel, W4267, M104 and Messapia) produce much higher yields than D3415 in high-yield regions [24]. Such studies emphasize how pairing a plant's genes with its environment can optimize selected phenotypes; similarly, matching appropriate environments and medical interventions to genotype is likely to be necessary for the optimization of health phenotypes in humans.

Animal studies of obesity and diabetes also provide useful examples of interactions, where phenotypic differences

Table 1 Summary of key literature on gene-environment interactions in obesity and type 2 diabetes

Topic and citation	Design/Population	Key findings	Replication?
(a) Lifestyle × GWAS variants in diabetes [4, 5]	Prospective cohort study of Swedish adults (n = 16,003, of whom 2063 developed type 2 diabetes during follow-up). Analyses focused on the interaction of physical activity and 17 established type 2 diabetes-associated loci, with diabetes as the outcome.	An experiment-wide significant interaction rs4430796 at *HNF1B* and baseline physical activity was observed, such that association of physical inactivity (vs. physically active) with diabetes incidence was strongest in the G/G genotype group, intermediate in the G/A group, and lowest in the A/A group.	No
	Nested case-cohort meta-analysis of 12,403 incident cases of type 2 diabetes and a representative sub-cohort of 16,154 non-disease participants from eight European countries. Analyses focused on 17 established type 2 diabetes-associated loci. Analyses focused on interactions of multiple lifestyle and demographic factors with a genetic risk score comprised of 49 established type 2 diabetes loci, with diabetes as the outcome.	The genetic score conveyed a significantly stronger effect on diabetes in younger and leaner adults than in older and fatter adults, respectively. The interaction of genetic risk score and waist circumference on diabetes risk was genome significant.	No
(b) Lifestyle × GWAS variants in obesity [6–12]	Cross-sectional cohort study of healthy female nurses (n = 6934) and male health professionals (n = 4423), as well as a replication cohort of 21,740 women, all from the USA. Analyses focused on the interaction of a genetic risk score comprised of 32 BMI-associated loci and sugar-sweetened beverage consumption [8].	The genetic score was more strongly associated with obesity in persons reporting high consumption of sugar-sweetened beverages.	Yes
	Cross-sectional cohort study of healthy female nurses (n = 9623) and male health professionals (n = 6379), as well as a replication cohort of 21,421 women, all from the USA. Analyses focused on the interaction of a genetic risk score comprised of 32 BMI-associated loci and fatty food consumption [9].	The genetic score was more strongly associated with obesity in persons reporting high consumption of fatty foods.	Yes
	Cross-sectional cohort study of adults from the UK (n = 20,430). Analyses focused on the interaction of physical activity with a genetic risk score comprised of 12 established obesity loci, with BMI as the outcome [11].	The genetic score was more strongly associated with obesity in persons reporting low levels of physical activity.	Yes (in [6])
	Analyses focus on replicating the findings reported in [11]. Cross-sectional cohort meta-analysis of 111,421 adults from European ancestry populations in Europe and North America [6].	The genetic score was more strongly associated with obesity in persons reporting low levels of physical activity, although the magnitude of the interaction effect was substantially less than reported in [11].	Yes (of [11])
	Cross-sectional cohort study of 5000 adults from Denmark. Analyses focused on testing the interaction between physical activity and the rs9960939 variant at the *FTO* locus [12].	The *FTO* variant was more strongly associated with obesity in persons reporting low levels of physical activity. This was the original report of the now widely replicated interaction at this locus. A clinical trial [13] and US cross-sectional study [14] reported compatible findings around the same time.	Yes (in [7])

Table 1 (continued)

Topic and citation	Design/Population	Key findings	Replication?
	Analyses focus on replicating the findings reported in [12].Cross-sectional cohort meta-analysis of cohorts from Europe and North America (45 studies of adults (n = 218,166 adults from 45 cohorts and n = 19,268 children from nine cohorts) [7].	The *FTO* variant was more strongly associated with obesity in persons reporting low levels of physical activity, although the interaction effect was substantially weaker than in the original report [12].	Yes (of [12])
	Cross-sectional cohort study of healthy female nurses (n = 7740) and male health professionals (n = 4564). Analyses focused on the interaction of a genetic risk score comprised of 32 BMI-associated loci and television viewing [10].	The genetic score was more strongly associated with obesity in persons reporting a large degree of television viewing.	No
(c) Genome-wide interaction analysis: Waist/hip ratio × SNP in total cholesterol [15]	Cross-sectional study (max. n = 32,225), with a replication cohort collection (N = 17,102) of European ancestry adults. Analyses focused on genome-wide interaction analyses with waist/hip ratio (WHR).	Genome-wide significant interactions with WHR on total cholesterol localizing to rs6448771 (proximal to *PCDH7-CCKAR*), with corresponding differential expression levels for in adipose tissue.	Yes
(d) Genome-wide interaction analysis: BMI × SNP in fasting insulin and glucose [16].	Cross-sectional meta-analysis of 96,496 non-diabetic individuals from 52 cohorts of European ancestry. Analyses focused on using the *joint meta-analysis* (JMA) approach to modeling joint effects of the margin effects of genotypes and gene-BMI interactions.	Six previously unreported loci were associated (at a genome-wide level of statistical significance) with fasting insulin concentrations were reported, although no genome-wide significant interactions were identified.	N/A

SNP single nucleotide polymorphism, *BMI* body mass index, *OR* odds ratio

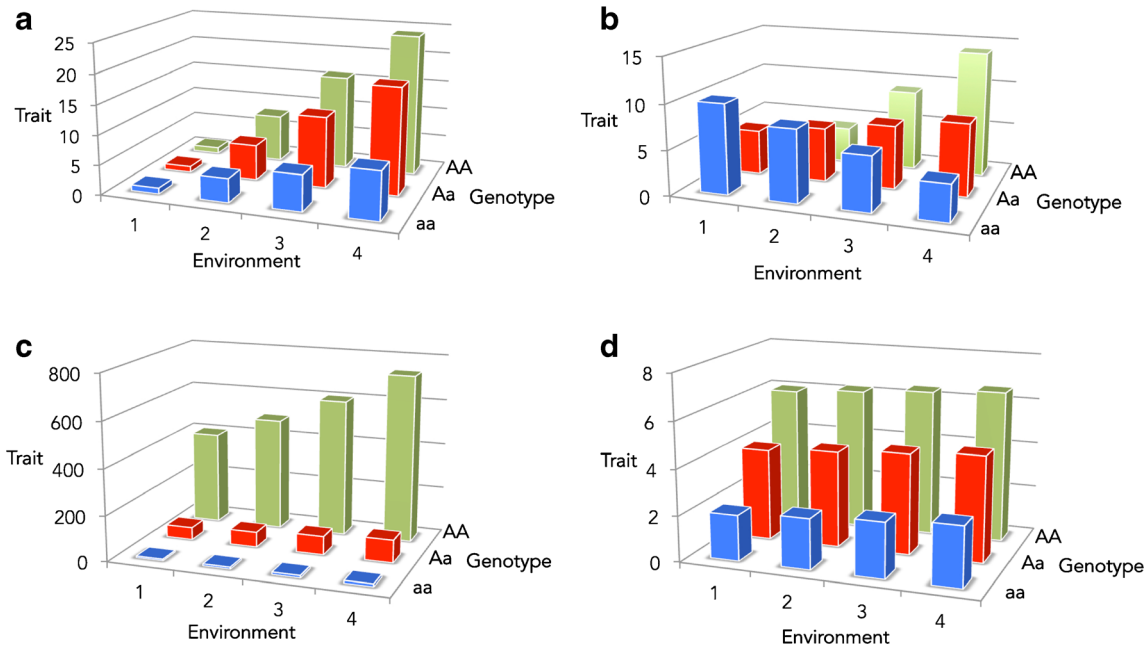

Fig. 1 Types of gene-environment interactions. **a** A non-removable, "pure" interaction. **b** A non-removable "cross-over" interaction. **c** A removable interaction with the trait expressed on the linear scale. **d** Exactly the same data used in (**c**), but the interaction is removed by expressing the trait on the natural log scale

between genetically engineered animals are augmented with interventions that perturb the molecular pathways upon which the gene(s) of interest reside. For example, high-fat feeding is a common intervention used to accentuate phenotypic differences between genetically distinct animals; in a study of glucose and lipid metabolism, the effects of 8-week high and low fat feeding regimes on metabolic phenotypes of five inbred mouse strains (C57BL/6J, 129X1/SvJ, BALB/c, DBA/2, FVB/N) were compared; the study showed that metabolic sensitivity to dietary fat varied considerably by genotype. Elsewhere, the NOD mouse strain has provided a longstanding murine model for autoimmune type 1 diabetes owing to its predisposition to early-onset disease [25]; the NOD mouse is especially susceptible when reared in a germ-free environment, but much less so when reared in standard "dirty" cages [26]. This phenomenon, which is not observed in wild-type mice, is thought to reflect immune adaptations in the NOD mouse that require exposure to foreign microbes early in life [26].

Complex metabolic diseases such as non-autoimmune diabetes are often uncommon in indigenous populations living traditional substance farming or hunter-gatherer lifestyles, yet phylogenetically similar people living industrialized lifestyles are often disproportionally afflicted [3]; these observations are consistent with the presence of susceptibility loci whose effects are triggered by environmental exposures. This phenomenon is most apparent in ethnic groups whose recent evolution is characterized by migration and frequent exposure to famine, cold, and other metabolic stressors. This process, which is described in detail elsewhere [27], might have led to enrichment of alleles that predispose to metabolic efficiency, particularly after meals. Other intriguing examples are those from certain populations that cope unusually well living at high altitudes [17], in nutrient deficient settings [18], or in cold climates [28]. Whilst these ecological observations are especially prone to confounding, bias, and reverse causality, they provide tentative support for gene-environment interactions in human disease.

Heritability studies conducted in intervention settings also provide suggestive evidence of gene-treatment interactions. Studies of overfeeding, underfeeding, and aerobic exercise training in twins and nuclear families indicate that changes in body composition are more highly correlated between members of the same kinship than between those of different kinships. For example, Bouchard et al. implemented a long-term overfeeding protocol (structured diet containing 1000 kcal/day above the baseline energy requirement) in 12 pairs of monozygotic (MZ) twins [29]; the intraclass correlation (ICC) for change in body weight in MZ pairs was $r = 0.55$. The ICC in non-twin pairs was not reported, but the ratio of the trait variance explained between pairs to that within pairs (F ratio) was 3.43, suggesting that body weight adaptation to long-term overfeeding is heritable. Elsewhere, adaptation of maximal oxidative capacity (a measure of aerobic fitness that is a strong predictor of diabetes) following a 20-week standardized exercise intervention protocol was examined in 720 individuals from 450 nuclear families [30]. As with the overfeeding study, aerobic adaptation was strongly correlated in biologically related participants, and much less so in those who were unrelated (F ratio = 2.50). Importantly though,

defining heritability in this way incorporates both genetic and shared non-genetic (e.g., shared familial environment) sources of trait variance; moreover, the heritable basis of baseline body weight and aerobic fitness is substantial and because these short-term studies did not partition out these factors, it is difficult to determining the extent to which phenotypic adaptation is under genetic control.

Discovery Strategies

Numerous approaches, varying by study design, data type and analytical method, have been used to discover gene-environment interactions; some approaches address similar objectives, whilst others are complementary and can be applied in sequence. Below we describe several of these approaches, and refer the reader to another excellent review of gene-environment interaction methods [31].

(a) Established statistical approaches

Until 2008, almost all studies of gene-environment interactions focused on testing hypotheses based on existing biological evidence, typically focusing on a small number of genetic variants. Linkage studies were the first generation of genome-wide interaction studies (GWIS) [32] but were generally unsuccessful and are seldom used in contemporary studies of complex traits. With few exceptions (see Table 1), neither approach led to convincing evidence of gene-environment interactions.

The advent of genome-wide association studies (GWAS) in 2005 facilitated a new era of genetic association studies and the rapid discovery of thousands of loci for many complex traits; GWAS triggered a quantum leap in population genetics, largely because it is agnostic to prior biological knowledge, which directly contrasts most previous gene discovery approaches. By 2008, researchers were exploring if environmental risk factors modified the effects of GWAS loci, an approach that now predominates in gene-environment interaction research. There is appeal to this approach because few statistical tests are performed, which helps preserve statistical power, and it is analytically simple. Indeed, several of the few adequately replicated examples of gene-environment interactions have been discovered in this way (Table 1). There are, however, good arguments for why loci derived from GWAS may not, on average, be good candidates for interactions [33]. For instance, heterogeneous SNP association signals are generally filtered out in standard GWAS meta-analyses, yet as we discuss below, variance across genotypes is a characteristic of interactions. Indeed most, perhaps all, comprehensive studies focused on determining whether established GWAS-derived loci interact with environmental risk factors or clinical interventions have yielded predominantly negative results [4, 5, 34–36].

With GWAS came the possibility to conduct GWIS at a much higher variant density, and in samples of unrelated individuals, not only in family pedigrees as with earlier linkage studies. The simplest approach involves testing all SNPs for interaction with one or more environmental variables. Whilst computationally feasible [37], conventional GWIS for complex traits require sample sizes that are often unachievable to be adequately powered. To help preserve power, restricting the number of variants tested to those with nominally significant marginal associations (e.g., $P = 0.10$) may help [38]. Other statistical tricks to minimize multiple testing involve the joint estimation of SNP and SNP × environment regression coefficients (2 *df* tests), which are relatively powerful, especially when an interacting locus also conveys a detectable marginal effect [39]. This approach has also been adapted for meta-analysis [40], and in some empirical situations has been shown to be more powerful than testing for marginal or interaction effects separately [16], although no novel loci have yet been confirmed using this approach for T2D.

(b) Data reduction approaches

A number of data reduction strategies for the analysis of gene-environment interactions have been proposed for use in observational studies. A common feature of these approaches is reduction of multiple hypothesis testing through selection of a subset of variants (step 1) for explicit interaction testing (step 2). One such approach is the "case-only" design, whereby the association between SNPs and an interacting variable is first tested only in disease cases and associated SNPs are then tested for interaction in the full cohort of cases and controls. Statistical power is preserved because the first screening step only involves association tests, which generally yield higher power than interaction tests when all else is equal. Although somewhat counterintuitive, in the presence of gene-environment interactions, SNPs are associated with the interacting environmental exposure only in cases, providing an opportunity to shortlist candidate SNPs for subsequent pairwise interaction tests in the full cohort using an interaction effect test [41]. A caveat to this approach is that when the genetic and environmental variables are correlated in controls, variants will be inappropriately prioritized for interaction testing, thereby reducing the power of the test; this problem may be enhanced when using GWAS, owing to the large number of variants tested.

Analytical strategies have also emerged that focus on modeling genetic effects for quantitative signatures of gene-environment interactions. These approaches pivot on the notion that interaction effects are characterized by heteroscedastic phenotypic variances that are conditional upon genotype (termed *variance heterogeneity*) (see Fig. 2);

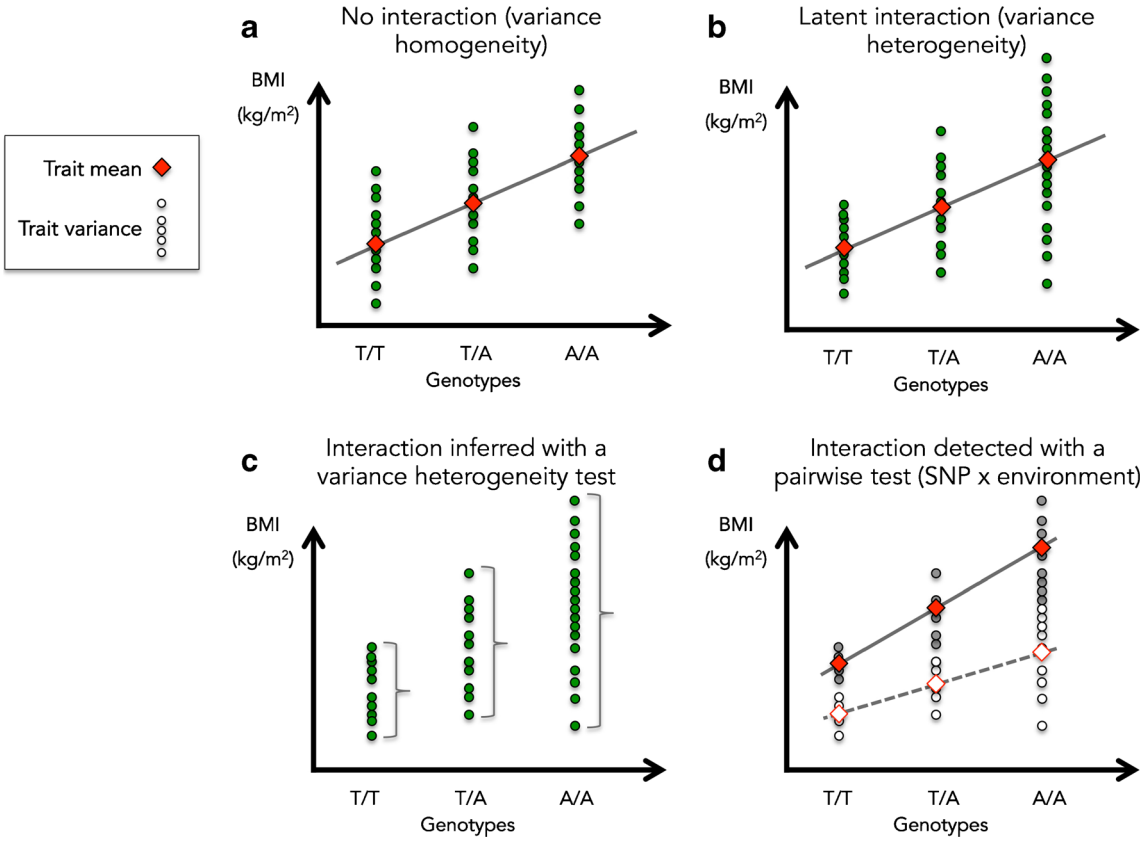

Fig. 2 Outline of how variance heterogeneity tests can be used to discover gene-environment interactions. **a** Conventional linear regression analysis in the presence of variance homogeneity. **b** Conventional linear regression analyses in the presence of variance heterogeneity. **c** Linear regression analyses intended to model variance heterogeneity. **d** Linear regression analyses intended to unmask the underlying gene-environment interaction

various methods have been proposed that exploit this characteristic, approaches that have proven somewhat successful for discovering gene-environment interactions in cardiometabolic traits [42•, 43•]. Thus, identifying differences in variance conditional upon genotype allows for the shortlisting of SNPs for explicit interaction testing. In the seminal description of this approach [42•], SNPs with genome-wide significant ($P < 5 \times 10^{-8}$) heterogeneity of variance estimates were identified for plasma C-reactive protein and soluble ICAM1, which were subsequently shown to interact with BMI and smoking ($P < 5 \times 10^{-8}$). Although in this example, the interaction would have been detectable in a conventional GWIS analysis, in other examples, where the explicit interaction test (stage 2) is not genome-wide significant, a less conservative significance threshold might be sufficient, owing to the orthogonal nature of the two sets of evidence.

An important advantage of variance heterogeneity tests is that the environmental exposure does not need to be explicitly characterized, as heterogeneity of variance will be present even when the interacting environmental factor is unmeasured or unknown. Indeed, many large datasets exist with genetic and phenotypic data that lack good environmental exposure data, and even where environmental exposure data are available, standardizing measurements across cohorts can results in a substantial loss of power in meta-analyses [6]. A caveat of this approach, as with most tests of gene-environment interaction, is that it is prone to confounding by linkage disequilibrium (synthetic associations and rare variant effects), scale dependency, and population stratification.

(c) Causal inference models

Causality is often uncertain in epidemiology when an association between an exposure and outcome is observed. Genetics is well suited to causal inference, because genetic variants are randomly assorted at meiosis and are usually not correlated with factors that can confound non-genetic associations in epidemiology. Using an approach termed *Mendelian randomization*, genotypes can be used as instrumental variables in experiments that resemble randomized controlled trials (RCT) [44, 45]. Because there are now many established associations between gene variants and diabetes-related exposures (e.g., smoking [46], coffee consumption [47], macronutrient intake [48]), it is possible to undertake a special type of *Mendelian randomization* experiment that focuses on modeling gene-environment interactions using genotypes as proxies

for environmental exposure, although interaction studies of this kind are yet to be reported. A limitation of this approach is that suitable instruments (genetic variants that are strongly correlated with the exposures of interest) for the environmental exposures in gene-environment interaction tests are often unavailable.

Causal interactions between genetic and environmental factors can also be modeled using types of Bayesian Network Analysis, such as the Bayesian Epistasis Association Mapping tool [49] and hierarchical modeling [50]. Approaches like these utilize multiple layers of data to estimate directional relationships between variables and hence permit some degree of causal inference. Bayesian Network Analysis in general works well when accurate and precise data are included and where gene ontologies are well defined, and much less so when these conditions do not hold. One of the major appeals of Bayesian Network Analysis is its capacity to integrate data across multiple biologic systems gathered within the same participant, which is likely to be particularly relevant for the functional elucidation of gene-environment interaction effects.

Translation of Gene-Environment Interaction Effects

Research on the genetics of complex disease has two principal objectives: (i) to elucidate understanding of pathobiology and (ii) to aid the prevention or treatment of disease. The major advances in human genetics during the past 15 years, made possible primarily through huge developments in high-throughput genomic technologies combined with a greater willingness of scientists to collaborate, have facilitated discovery of thousands of disease-associated loci that with appropriate follow-up will substantially further our understanding of disease biology. The second objective, however, is yet to be realized to any meaningful degree.

(a) Theoretical considerations

Two common characteristics of established complex disease-associated variants discovered using hypothesis-free high-throughput approaches is that the magnitude of effect is relatively small and homogeneous across a range of environmental settings and treatment arms of clinical trials [4, 34–36]. Whilst the discovery of these loci helps define novel aspects of human biology, this information has proven relatively ineffective for the stratification of medical interventions, probably in part because of the way in which the variants were discovered. To identify gene variants that are of use for stratified medicine will likely require explicit strategies that seek to discover loci that predict a person's susceptibility to disease given specific environmental exposures or that predict treatment response. The strategies needed to detect such interactions will be distinct from those used to detect genetic associations per se.

The extent to which genetic information enhances the accuracy of established disease prediction models or improves the degree to which disease occurrence is correctly predicted in prospective analyses is likely to vary considerably across diseases. Importantly though, because germline DNA variants are salient biomarkers, their predictive accuracy relative to non-genetic biomarkers can improve as the time between the baseline assessment and disease incidence lengthens [51]. Thus, genotypes provide a rare example of disease biomarkers that could be measured very early in life to predict diseases occurring several decades later. Whilst many studies have reported on the discriminative or predictive accuracy of models including genetic and environmental data, most do not consider their joint, synergistic, effects and generally treat these two types of exposure as independent factors. However, Aschard et al. [52•] examined the discriminative value and reclassification potential of simulation models including two-way gene-gene and gene-environment interaction effects in relation to breast cancer, rheumatoid arthritis (RA), and T2D. The authors found that the inclusion of up to ten interaction effects of fairly modest magnitude improved discriminative accuracy (ROC AUC) for breast cancer by approximately 4 %, RA by approximately 2 %, and T2D by approximately 1 %. The net improvement in case–control classification for the model including all 10 interaction effects was approximately 30 % for each of these traits compared with the null model. Increasing the number (up to 20) and magnitude (risk ratio = 10) of the simulated interaction effects included in the model substantially increased both its discriminative accuracy and net reclassification.

Aschard et al.'s analyses focused on discriminating between people with and without prevalent disease, which is unlikely to be directly comparable with analyses focused on predicting incident events; although few discovery genetic association studies have been performed using longitudinal data, some prospective studies have estimated the predictive value of established prevalent disease-associated gene variants for change in quantitative biomarkers [53] or disease events [54]. Those studies suggest that genetic variants that are strongly associated with cross-sectional traits do not always predict change in the trait, and vice versa. Moreover, the primary metric used in these analyses was the C-statistic, a measure of discriminative accuracy whereby a value of 50 % reflects accuracy equivalent to tossing a coin and a value of 100 % reflects perfect discrimination; importantly, this particular approach to quantifying discriminative accuracy is sensitive to the frequency of the disease and its risk factors, with models focused on rarer diseases and exposures generally yielding lower values than those focused on common diseases and risk factors. Nevertheless, Aschard et al.'s study provides valuable information that may help quantify assumptions

about the extent to which data on gene-environment interactions can help classify and predict disease events.

(b) Mechanisms of action

The mechanisms underlying observations of gene-environment interactions in T2D are rarely discussed, probably because few functional studies have been performed around explicit interaction effects. However, more than half a century ago Jacob and Monold [55] outlined the mechanisms underlying the synthesis of enzymes in bacteria, which they described as requiring genetic *repressors* that can be activated or inactivated by specific metabolites present in the cellular environment [55]. In pharmacogenetics, mechanisms are often eloquently described; take for example, activating mutations in *KCNJ11*, the gene encoding the Kir6.2 subunit that controls gating of the ATP-sensitive K^+ channels (K_{ATP}) in the pancreatic beta cells. Here, carriers of the mutations can produce but not secrete insulin in response to glucose; however, treatment with sulfonylureas, which binds to the SUR1 subunit of the sulfonylurea receptor/potassium channel complex on the beta-cell membrane, depolarizes the K^+ channels, leading to the activation of voltage-gated Ca^{2+}channels thus increasing the secretion of insulin [56].

Most gene-environment interactions are likely to include one of four mechanisms: (i) *ligand binding* interactions (mutations that disrupt the binding of ligands to the cell membrane receptor(s) or the nuclear receptor(s)); (ii) *epigenetic* interactions (mutations that in the presence of certain environmental exposures cause epigenetic changes that differentially affect gene transcription); (iii) *double hit* interactions (where environmental exposures cause somatic mutations that interact with existing germline variants); and (iv) *gating* interactions (where mutations in regulatory elements pathogenically modulate the activity of biologic processes, such that, for example, without exercise, diet modification or pharmacotherapy, disease occurs).

Whilst understanding mechanisms of action may not be necessary for translating knowledge of interactions into the clinical context, defining mechanisms is necessary to identify therapeutic targets. Thus, emphasis should be placed on elucidating the functional processes underlying any valid observation of gene-environment interaction.

(c) Genotype-based recall (GBR)

Specially designed intervention studies, where large sample frames are used to identify two equally sized subgroups that are highly distinct in their genetic predisposition to disease (e.g., minor vs. major allele homozygotes at a given rare variant) and who are subsequently enrolled into a randomized controlled trial, represent a powerful test-bed through which gene-environment interaction effects can be validated (Fig. 3; Supplement 2). The earliest example of a genotype-based recall study focused on in vivo effects of the *PPARG* Pro12Ala genotypes on adipose tissue free fatty acid metabolism [57]. A second recent intervention study focused on administering 0, 10, or 20 mg of yohimbine in people selected for genotypes at the α(2A)-adrenergic receptor locus (*ADRAD2A*) [58]. The main outcome was early insulin response (30 min) insulin concentrations following a 75-g oral glucose load. The study was one of the first GBR trials to be reported and showed that treatment response is conditional on *ADRAD2A* genotype.

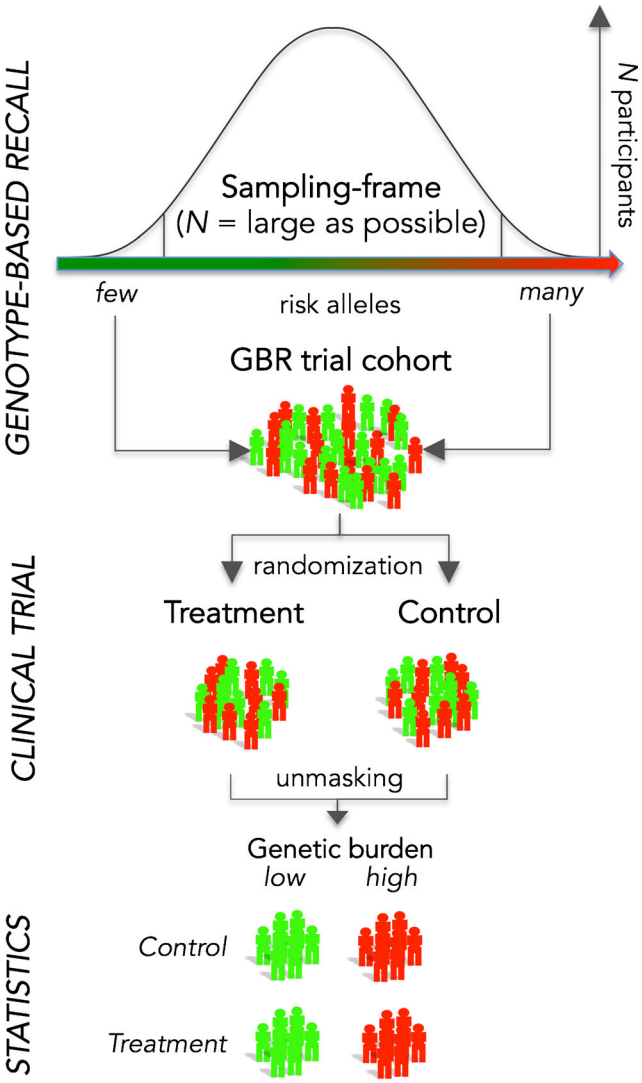

Fig. 3 Figure describing the genotype-based recall trial paradigm. Participants are selected from low- and high-burden genetic risk groups within a large sampling frame and are subsequently randomized to treatment (e.g., intensive lifestyle modification) or control arms of a clinical trial. Treatment allocation and genotype ideally remain masked until the trial has ended (although this is often difficult or impossible with lifestyle interventions), at which time gene-treatment interactions can be quantified

Barriers and Limitations

Epidemiology has yielded most of the evidence garnered during the past 20 years on gene-environment interactions in T2D and related traits, much from small cross-sectional studies. However, several large prospective cohort studies exist with good measures of environmental and genetic exposures, repeated measures of quantitative outcomes, and long-term follow-up for incident disease [59–61], rendering them excellent resources for generating hypotheses about gene-environment interactions. However, epidemiological studies are prone to various forms of chance, bias, and confounding as well as reverse causality, which make the determination of causal effects especially challenging [62]. Owing to the salient nature of germline DNA variants, genetic association studies are robust to reverse causality, but there are other sources of bias and confounding, such as population stratification, synthetic association, and survival bias that may provide alternative explanations for an apparent effect of a genotype on a disease trait [63]. Epidemiological studies of gene-environment interactions are prone to the limitations of both genetic and non-genetic epidemiology, as well as other limitations that are idiosyncratic to this type of research. *Scale dependency* is one such limitation, which occurs when data conversions drive the presence or absence of statistically significant interactions (see Fig. 1).

The term *error* relates to the imprecision of an estimate and the term *bias* describes the extent to which error is disproportionate between two or more groups under investigation. The large size of many epidemiological studies necessitates that environmental exposures are usually assessed with fairly imprecise methods such as questionnaires and outcomes with proxy variables. This can cause underestimation of the true magnitude of the marginal and interaction effects and diminishes power to detect interactions [64]. Under a set of reasonable assumptions about interaction studies, Wong et al. [65] described sample size requirements to detect interactions with low type 1 and type 2 error rates; when the exposure and outcome are good proxies for the true (latent) exposure ($\rho_{Tx}=0.8$) and outcome ($\rho_{Ty}=0.8$) ~2410 participants are required to detect a reasonably sized interaction, but when exposure and outcome are poorly assessed ($\rho_{Tx}=0.4$; $\rho_{Ty}=0.4$), the required sample size booms (N ~84,787).

Recognizing that many existing interaction studies may have been underpowered, studies of interaction are now often performed by combining results from large cohort collections using meta-analysis. Palla et al. illustrated why *retrospective* meta-analysis of published interaction studies may yield meaningless results [66•], owing largely to bias and confounding, and difficulty standardizing results. Thus, most gene-environment interaction studies involving multiple cohorts focus on *prospective* meta-analyses, where each participating cohort performs new analyses according to a standardized analyses plan, and their summary results are subsequently pooled.

Meta-analyses of data from multiple cohorts have obvious appeal, as sample sizes that far exceed most individual study of gene-environment interaction can be collated. A caveat to the approach though is that the assessments of exposures and outcomes in these cohorts often differ on multiple levels (e.g., type and validity of measures, data structure, reference timeframe, data processing approaches). Methodological differences demand that environmental exposure variables are standardized before analysis, which typically involves collapsing exposure data to a parsimonious level, which can substantially reduce statistical power. We recently conducted a large meta-analysis examining the interaction between an *FTO* variant and physical activity in obesity [7]; the study involved meta-analyzing summary statistics from 45 adult and nine pediatric cohorts. Although some cohorts had very detailed physical activity data (e.g., objective continuous assessments of physical activity), others had very crude (binary) subjective physical activity data; thus, all cohorts were asked to reduce their physical activity exposure variables to a simple binary variable where approximately 80 % of participants were defined as physically active and the remaining 20 % were defined as inactive.

This approach, whilst pragmatic, diminishes statistical power in at least two key ways: first, stratification of continuous data often results in loss of power [67]; second, where interaction effects are approximately linear, asymmetrical stratification of exposure data also diminishes power. We provide several relevant examples elsewhere [6]; for example, a study with ~15,000 participants and the environmental exposure variable stratified at the median of its distribution would be adequately powered (80 %) to detect the interaction effect. But if the exposure variable is stratified at the 80th centile of its distribution, with all else equal, a sample size of ~24,000 would be equivalently powered to detect the same interaction effect. Power is lost primarily owing to increased variance in the exposure variable. Pooling multiple heterogeneous cohorts causes an increase in the dependent variable's variance, which also leads to a substantial loss of statistical power. Hence, meta-analyses of gene-environment interactions composed of data from multiple diverse cohorts may not be as powerful a strategy for replication as many hope, and focusing on a handful of large, well-characterized and comparable cohorts for replication is likely to be a considerably more efficient strategy. The recent availability of large cohorts with data that are suitable for gene-environment interaction modeling such as UK Biobank [68], seem set to change the research community's dependency on meta-analysis to conduct large genetic studies, as many of the caveats to the latter (described in detail in [6]) are likely offset in this single large study.

Interaction effects are also prone to a specific form of confounding that can occur when the outcome variable is a proxy

for the phenotype of interest, as is often true in epidemiological studies. Consider the example of gene-lifestyle interactions in obesity, where anthropometric measures such as height and weight are used to derive BMI, a proxy for total adiposity. Because BMI is not a perfect correlate of adipose mass, there are relatively lean people within any population who are muscular and heavy, with a high BMI [69]. Those persons may plausibly exercise and avoid other unhealthful lifestyle behaviors (e.g., consuming fatty foods or sugar-sweetened beverages) more than those with a high BMI and high-fat percentage; thus the magnitude of the effect of a genetic risk score on BMI will likely be stronger in inactive than active people (causing a statistical interaction) purely because the outcome measure in the inactive group is more valid. This problem emphasizes the need to validate epidemiological observations of interaction in other studies that have the ability to elucidate the target phenotypes.

Conclusions/Perspective

The major recent breakthroughs in complex trait genetics have boosted confidence that similar successes might be achievable in the field of gene-environment interaction research. The derivation of massive amounts of genetic and phenotypic data, along with an understanding that those data should be used and reused, has encouraged investigators to dig deep into their databases to explore whether genetic association signals are modulated by non-genetic factors. Thus, the once esoteric topic of gene-environment interaction is now becoming mainstream and appealing to investigators across diverse disciplines; this has propelled major methodological innovations for the discovery, replication, validation and translation of gene-environment interactions. The exponentiation of data resources for these purposes has demanded analytical solutions that address data dimensionality reduction. Although not yet extensively implemented, systems-medicine approaches for interaction modeling in complex human disease, which might build on the eQTL-based methods developed in yeast [70, 71] and human dendritic cells [72], and other system-based approaches [73], are growing in popularity and will accelerate gene-environment interaction research as large systems genetics-focused studies come online [74] (Fig. 4).

The paucity of replicated gene-environment interaction effects may reflect an abundance of false-positive findings in the published literature [3], although other explanations for why true-positive interaction effects fail to replicate should not be dismissed [6]. Most if not all complex traits probably result from the accumulation of many small-magnitude gene-environment interactions, gene-gene interactions, and marginal effects. If so, most existing interaction studies will likely be underpowered to detect real effects owing to their small sample sizes. Accordingly, interaction meta-analyses are increasingly performed on data from multiple cohorts. Although successful for genetic association studies, meta-analysis may not work well in the context of gene-environment interaction owing to the diversity of measurements and data across cohorts, which degrades statistical power [6]. Thus, it seems logical to focus gene-environment interaction analyses on cohorts that are either very large

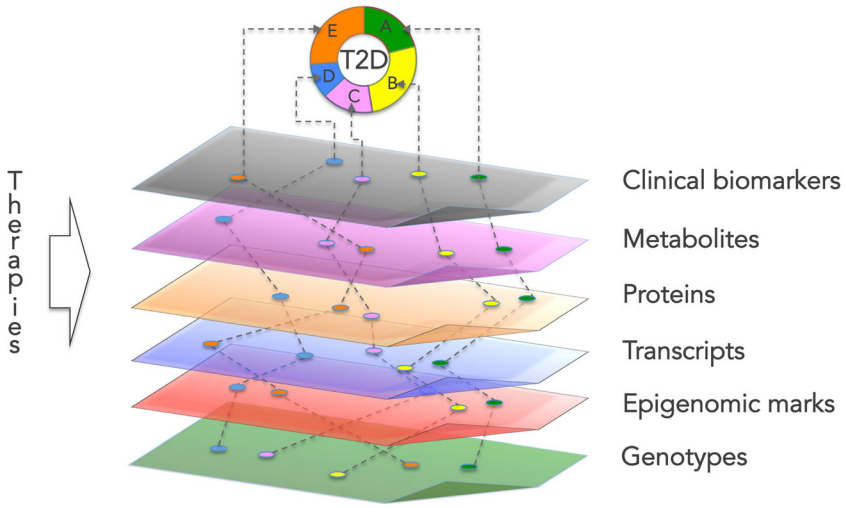

Fig. 4 We can now quantify the key molecular events that link germline DNA variation with disease processes in a range of settings, from cell lines to human populations, and major advances have been made in coupling these complex datasets with information about extrinsic environmental exposures including drug prescription in ways that allow the logical interrogation of gene-drug and gene-lifestyle interactions. Doing so may teach us about disease etiology and help stratify type 2 diabetes (T2D) into subclasses that can be treated more effectively, with fewer side effects and at lesser cost than before

and that include well-validated and standardized assessment methods, or those that are smaller in size but which include accurate and precise measures of exposures and outcomes [65]. By exception, variance prioritization meta-analyses are likely to be less prone to loss of power, because the environmental exposure is inferred by comparing phenotypic variances by genotype rather than through direct assessment. Although most published gene-environment interaction studies focus on cross-sectional data, longitudinal interaction studies are also needed, especially those that include repeated measures of exposures and outcomes, as this will facilitate temporal inference and help preserve statistical power [75].

Emphasis is frequently placed on *translation* when gene-environment interaction data are discussed. The logic is appealing, as identifying genetic markers that define patients who are at substantially greater or lesser risk of disease than the general population given exposure to modifiable risk factors, or who will respond much better of worse to treatment, could help optimize medical interventions. However, there are as yet no translatable examples of gene-environment interactions that are sufficiently convincing to guide medical interventions for T2D. Nevertheless, numerous examples from Mendelian disorders and pharmacogenetics fuel hope that genetic data may eventually help tailor prevention or treatment strategies for complex diseases focused on lifestyle modification. Specially designed intervention studies, such as genotype-based recall trials (Fig. 3), will also facilitate clinical translation of data on gene-environment interactions.

Acknowledgments The authors thank M-F Hivert (Boston, MA) and F Renström (Malmö, Sweden) for thoughtful comments on this manuscript. The ideas and perspectives described in the paper are those of the authors unless otherwise stated; however, these views have evolved through many previous and ongoing interactions with trainees and peers. PWF specifically thanks N Wareham (Cambridge, UK), P Kraft (Boston, MA), CA Franks (Vejbystrand, Sweden), R Hanson (Phoenix, AZ), and members of the Genetic and Molecular Epidemiology Unit (Malmö, Sweden) for many illuminating discussions around the topic of gene-environment interactions. The authors also thank the editors (JC Florez and AP Morris) for helpful feedback on this paper and D Shungin for input on Fig. 2. PWF was supported by grants from the Novo Nordisk Foundation, Swedish Research Council, Swedish Diabetes Association, Påhlssons Foundation, Swedish Heart-Lung Foundation, EXODIAB, Region Skåne, the Medical Faculty of Umeå University, the Innovative Medicines Initiative of the European Union (grant agreement no. 115317—DIRECT), and the European Research Council. GP is supported by the Canada Research Chair in Genetic and Molecular Epidemiology and the CISCO Professorship in Integrated Health Systems.

Compliance with Ethical Standards

Conflict of Interest Paul W. Franks has received consulting honoraria from Eli Lilly Inc and Sanofi Aventis in 2015.

Guillaume Paré declares that he has no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- Of importance

1. National Cholesterol Education Program Expert Panel on Detection, E, Treatment of High Blood Cholesterol in, A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106:3143–421.
2. Horner FA, Streamer CW. Effect of a phenylalanine-restricted diet on patients with phenylketonuria; clinical observations in three cases. J Am Med Assoc. 1956;161:1628–30.
3. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry. 2011;168:1041–9.
4. Brito EC et al. Previously associated type 2 diabetes variants may interact with physical activity to modify the risk of impaired glucose regulation and type 2 diabetes: a study of 16,003 Swedish adults. Diabetes. 2009;58:1411–8.
5. Langenberg C et al. Gene-lifestyle interaction and type 2 diabetes: a case-cohort study. PLoS Med. 2016.
6. Ahmad S et al. Gene × physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. PLoS Genet. 2013;9:e1003607.
7. Kilpelainen TO 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.
8. Qi Q et al. Sugar-sweetened beverages and genetic risk of obesity. N Engl J Med. 2012;367:1387–96.
9. Qi Q et al. Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. BMJ. 2014;348:g1610.
10. Qi Q et al. Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. Circulation. 2012;126:1821–7.
11. Li, S. et al. Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study. *PLoS Med* 7 (2010).
12. Andreasen CH et al. Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. Diabetes. 2008;57:95–101.
13. Franks PW et al. Assessing gene-treatment interactions at the FTO and INSIG2 loci on obesity-related traits in the Diabetes Prevention Program. Diabetologia. 2008;51:2214–23.
14. Rampersaud E et al. Physical activity and the association of common FTO gene variants with body mass index and obesity. Arch Intern Med. 2008;168:1791–7.
15. Surakka I et al. A genome-wide screen for interactions reveals a new locus on 4p15 modifying the effect of waist-to-hip ratio on total cholesterol. PLoS Genet. 2011;7:e1002333.
16. Manning AK et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44:659–69.

17. Tuomilehto J et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med. 2001;344:1343–50.
18. Ahmad S, Varga TV, Franks PW. Gene x environment interactions in obesity: the state of the evidence. Hum Hered. 2013;75:106–15.
19. Krafka J. The effect of temperature upon facet number in the bar-eyed mutant of Drosophila: part I. J Gen Physiol. 1920;2:409–32.
20. Krafka J. The effect of temperature upon facet number in the bar-eyed mutant of Drosophila: part II. J Gen Physiol. 1920;2:433–44.
21. Krafka J. The effect of temperature upon facet number in the bar-eyed mutant of Drosophila: part III. J Gen Physiol. 1920;2:445–64.
22. Hansen AM, Gardner EJ. A new eye phenotype in Drosophila melanogaster expressed only at temperatures above 25 degrees C. Genetics. 1962;47:587–98.
23. FAO 2010. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture (Rome).
24. Annicchiarico P, Mariani G. Prediction of adaptability and yield stability of durum wheat genotypes from yield response in normal and artificially drought-stressed conditions. Field Crop Res. 1996;46:71–80.
25. Makino S et al. Breeding of a non-obese, diabetic strain of mice. Jikken Dobutsu. 1980;29:1–13.
26. Singh B, Rabinovitch A. Influence of microbial agents on the development and prevention of autoimmune diabetes. Autoimmunity. 1993;15:209–13.
27. Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am J Hum Genet. 1962;14:353–62.
28. Hancock AM, Clark VJ, Qian Y, Di Rienzo A. Population genetic analysis of the uncoupling proteins supports a role for UCP3 in human cold resistance. Mol Biol Evol. 2011;28:601–14.
29. Bouchard C et al. The response to long-term overfeeding in identical twins. N Engl J Med. 1990;322:1477–82.
30. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. Med Sci Sports Exerc. 2001;33:S446–51. **discussion S452-3**.
31. Cornelis MC et al. Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. Am J Epidemiol. 2012;175:191–202.
32. Schmidt S, Schmidt MA, Qin X, Martin ER, Hauser ER. Linkage analysis with gene-environment interaction: model illustration and performance of ordered subset analysis. Genet Epidemiol. 2006;30:409–22.
33. Franks PW. Gene x environment interactions in type 2 diabetes. Curr Diab Rep. 2011;11:552–61.
34. Hivert MF et al. Updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. Diabetes. 2011;60:1340–8.
35. Nettleton JA et al. Meta-analysis investigating associations between healthy diet and fasting glucose and insulin levels and modification by loci associated with glucose homeostasis in data from 15 cohorts. Am J Epidemiol. 2013;177:103–15.
36. Travis RC et al. Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. Lancet. 2010;375:2143–51.
37. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet. 2005;37:413–7.
38. Kooperberg C, Leblanc M. Increasing the power of identifying gene × gene interactions in genome-wide association studies. Genet Epidemiol. 2008;32:255–63.
39. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Hum Hered. 2007;63:111–9.
40. Manning AK et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. Genet Epidemiol. 2011;35:11–8.
41. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. Am J Epidemiol. 2009;169:219–26.
42.• Pare G, Cook NR, Ridker PM, Chasman DI. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. PLoS Genet. 2010;6:e1000981. **The paper describes the use of variance prioritization to discover gene-environment interactions on a genome-wide scale. The paper also reports genome-wide significant interaction effects for a number of loci and BMI in relation to blood biomarker levels**.
43.• Visscher PM, Posthuma D. Statistical power to detect genetic Loci affecting environmental sensitivity. Behav Genet. 2010;40:728–33. **This paper describes an alternative approach to reference 36 for discovering gene-environment interactions using variance prioritization**.
44. Davey Smith G, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330:1076–9.
45. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27:1133–63.
46. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010;42:441–7.
47. Sulem P et al. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. Hum Mol Genet. 2011;20:2071–7.
48. Tanaka T et al. Genome-wide meta-analysis of observational studies shows common genetic variants associated with macronutrient intake. Am J Clin Nutr. 2013;97:1395–402.
49. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case–control studies. Nat Genet. 2007;39:1167–73.
50. Conti DV, Cortessis V, Molitor J, Thomas DC. Bayesian modeling of complex metabolic pathways. Hum Hered. 2003;56:83–93.
51. Lyssenko V et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med. 2008;359:2220–32.
52.• Aschard H et al. Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. Am J Hum Genet. 2012;90:962–72. **The paper reports simulation analyses to determine the extent to which the inclusion of data on gene-environment interactions is likely to improve the ability to discriminate between diseases and non-diseased individuals. The authors conclude that the inclusion of up to 20 small magnitude interaction effects in discriminative models is unlikely to have a major impact on discriminative accuracy for type 2 diabetes, rheumatoid arthritis and prostate cancer**.
53. Renstrom F et al. Genetic predisposition to long-term nondiabetic deteriorations in glucose homeostasis: ten-year follow-up of the GLACIER study. Diabetes. 2011;60:345–54.
54. Meigs JB et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008;359:2208–19.
55. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3:318–56.
56. Pearson ER et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med. 2006;355:467–77.
57. Tan GD et al. The in vivo effects of the Pro12Ala PPARgamma2 polymorphism on adipose tissue NEFA metabolism: the first use of the Oxford Biobank. Diabetologia. 2006;49:158–68.

58. Tang Y et al. Genotype-based treatment of type 2 diabetes with an alpha2A-adrenergic receptor antagonist. Sci Transl Med. 2014;6: 257ra139.
59. Hindy G et al. Role of TCF7L2 risk variant and dietary fibre intake on incident type 2 diabetes. Diabetologia. 2012;55:2646–54.
60. InterAct C et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia. 2011;54:2272–82.
61. Qi Q, Workalemahu T, Zhang C, Hu FB, Qi L. Genetic variants, plasma lipoprotein(a) levels, and risk of cardiovascular morbidity and mortality among two prospective cohorts of type 2 diabetes. Eur Heart J. 2012;33:325–34.
62. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet. 2006;7:812–20.
63. Franks PW, Nettleton JA. Invited commentary: gene X lifestyle interactions and complex disease traits–inferring cause and effect from observational data, sine qua non. Am J Epidemiol. 2010;172: 992–7. discussion 998–9.
64. Wong MY, Day NE, Luan JA, Wareham NJ. Estimation of magnitude in gene-environment interactions in the presence of measurement error. Stat Med. 2004;23:987–98.
65. Wong MY, Day NE, Luan JA, Chan KP, Wareham NJ. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? Int J Epidemiol. 2003;32:51–7.
66.• Palla L, Higgins JP, Wareham NJ, Sharp SJ. Challenges in the use of literature-based meta-analysis to examine gene-environment interactions. Am J Epidemiol. 2010;171:1225–32. **The authors set forth structured arguments for why meta-analyses of retrospective (published) data on gene-environment interactions are likely to fail. They highlight in particular the problems with standardizing data that has been analyzed in different ways, and the extent to which inherited**.
67. Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. Epidemiology. 1992;3:434–40.
68. Collins R. What makes UK Biobank special? Lancet. 2012;379: 1173–4.
69. Prentice AM, Jebb SA. Beyond body mass index. Obes Rev. 2001;2:141–7.
70. Gagneur J et al. Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. PLoS Genet. 2013;9:e1003803.
71. Parts L, Stegle O, Winn J, Durbin R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. PLoS Genet. 2011;7.
72. Lee MN et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science. 2014;343: 1246980.
73. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014;15:34–48.
74. Koivula RW et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia. 2014.
75. Franks PW et al. Interaction between an 11betaHSD1 gene variant and birth era modifies the risk of hypertension in Pima Indians. Hypertension. 2004;44:681–8.