Published in final edited form as:

Ann Hum Genet. 2011 January ; 75(1): 10–19. doi:10.1111/j.1469-1809.2010.00629.x.

Genome-Wide Association Scan Allowing for Epistasis in Type 2 Diabetes

Jordana T. Bell^{1,2,*}, Nicholas J. Timpson³, N. William Rayner^{1,4}, Eleftheria Zeggini^{1,5}, Timothy M. Frayling⁶, Andrew T. Hattersley⁶, Andrew P. Morris¹, and Mark I. McCarthy^{1,4}

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

²Department of Twin Research, King's College London, UK

³MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, Bristol University, Bristol

⁴Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

⁵Wellcome Trust Sanger Institute, Hinxton, UK

⁶Peninsula Medical School, Exeter, UK

Summary

In the presence of epistasis multilocus association tests of human complex traits can provide powerful methods to detect susceptibility variants. We undertook multilocus analyses in 1924 type 2 diabetes cases and 2938 controls from the Wellcome Trust Case Control Consortium (WTCCC). We performed a two-dimensional genome-wide association (GWA) scan using joint two-locus tests of association including main and epistatic effects in 70,236 markers tagging common variants. We found two-locus association at 79 SNP-pairs at a Bonferroni-corrected P -value = 0.05 (uncorrected P -value = 2.14×10^{-11}). The 79 pair-wise results always contained rs11196205 in *TCF7L2* paired with 79 variants including confirmed variants in *FTO*, *TSPAN8*, and *CDKAL1*, which are associated in the absence of epistasis. However, the majority (82%) of the 79 variants did not have compelling single-locus association signals (P -value = 5×10^{-4}). Analyses conditional on the single-locus effects at *TCF7L2* established that the joint two-locus results could be attributed to single-locus association at *TCF7L2* alone. Interaction analyses among the peak 80 regions and among 23 previously established diabetes candidate genes identified five SNP-pairs with case-control and case-only epistatic signals. Our results demonstrate the feasibility of systematic scans in GWA data, but confirm that single-locus association can underlie and obscure multilocus findings.

Keywords

Epistasis; simultaneous search; joint effects; genome-wide association

Introduction

Epistasis, or gene–gene interaction, has recently received much attention in gene-mapping studies in human genetics (Carlborg & Haley, 2004; Marchini et al., 2005; Evans et al.,

2006; Cordell, 2009). Conventional methods used to locate genes involved in human traits are primarily based on single-locus analyses and often ignore the joint effects of multiple genes on the phenotype. However, studies in model organisms have shown that epistatic interactions can have large effects, may occur frequently, and can involve more than two loci (see Mackay, 2001; Carlborg & Haley, 2004). For example, novel loci that act through epistatic pathways have been identified through multidimensional scans of complex traits in the mouse (Sen & Churchill, 2001; Shimomura et al., 2001), rat (Sugiyama et al., 2001; Ways et al., 2002), chicken (Carlborg et al., 2003; Carlborg et al., 2004), *Drosophila* (Montooth et al., 2003), and in plants (Holland et al., 1997; Li et al., 1997). These studies suggest that multiple interacting genes can influence complex phenotypes often in the absence of significant single-locus effects. Epistasis in this context typically refers to the interaction between genetic variants at multiple loci, or to the presence of non-additive genetic variance in the population. In the present study, we define epistasis as the statistical interaction between genetic variants at multiple loci, and the epistatic variance as the genetic variance not attributed to single-locus additive and dominance effects at the population level.

The findings of epistatic interactions across multiple species imply that epistasis should be taken into account when searching for susceptibility loci for human complex traits, particularly within the context of recent genome-wide association (GWA) analyses. Strategies to incorporate epistasis within GWA scans have been examined and results indicate that in the presence of epistasis, association studies of complex human traits can have more power to detect susceptibility variants under multilocus models (Marchini et al., 2005; Evans et al., 2006). However, if variants at two loci interact, then the estimated multilocus variance-components, and especially the observed epistatic variance, will depend on allele frequency (Lynch & Walsh, 1998; Hill et al., 2008). In particular, Hill et al., (2008) show under several population genetic models that if the allele frequency distribution is taken into account, a large proportion of the population genetic variance will be additive, even in the presence of epistasis. These results suggest that joint multilocus tests, including both main and epistatic effects, may prove advantageous over other multilocus approaches, such as epistasis-only or conditional analyses, for identifying novel susceptibility loci in GWA data even in the presence of interactions.

We considered joint multilocus effects on a genome-wide level with respect to type 2 diabetes (T2D), which is one of the most common heritable complex traits with rising prevalence throughout the world. To date over 30 T2D susceptibility loci have been identified (Saxena et al., 2007; Scott et al., 2007; Sladek et al., 2007; Steinthorsdottir et al., 2007; WTCCC, 2007; Zeggini et al., 2007; Unoki et al., 2008; Yasuda et al., 2008; Zeggini et al., 2008; Kong et al., 2009; Rung et al., 2009; Dupuis et al., 2010; Voight et al., 2010), but these variants explain only a small proportion of the heritability of T2D. We aimed to assess the contribution of genetic variants to the disease, incorporating the possibility of epistatic interactions within the GWA analyses by using the logistic regression framework. We performed a systematic two-locus GWA scan including epistasis in T2D cases and controls used in the WTCCC GWA scan.

Materials and Methods

Genotype Data

The study consisted of 1924 T2D cases and 2938 controls used in the WTCCC GWA scan and genotypes assayed on the Affymetrix 500K platform. We considered genotype data for a subset of 393,143 common (minor allele frequency (MAF) ≥ 0.05) SNPs. These SNPs passed stringent quality control criteria as previously described (WTCCC, 2007; Zeggini et

al., 2008; 2007), including checks for Hardy-Weinberg equilibrium in the controls, and checks for differential missing rates in cases and controls.

Analyses were performed on a subset of SNPs in the data (tag-SNPs) that tagged genetic variation at all remaining genotyped variants in the T2D cases and WTCCC controls. We chose to analyse tag-SNPs as a compromise between reducing computational load and maintaining power to detect multilocus effects, although selecting tag-SNPs alone will reduce the power to detect both joint two-locus effects and non-additive genetic variance components (Hill et al., 2008). Our primary goal was to perform a genome-wide scan to test for evidence for joint two-locus association, rather than focus on specific pairs of regions using all available genetic variation. We defined tag-SNPs based on a pairwise r^2 of 0.2 in the WTCCC data. This resulted in a final set of 70,236 tag-SNPs for the two-locus GWA analyses.

Two-Locus GWA Scan Strategy

The goal of the study was to identify novel T2D loci with evidence for joint multilocus effects. We first fitted a two-locus joint-effects test including main and epistatic components for all pairs of non-syntenic tag-SNPs across the genome. We did not perform an epistatic-only GWA scan because our aim was to identify T2D susceptibility variants with joint multi-locus effects rather than identify evidence for epistasis alone. We applied a Bonferroni correction threshold ($P = 2.14 \times 10^{-11}$) to obtain genome-wide significant pairs of candidate loci with joint-effects. We obtained 79 genome-wide significant marker-pairs, for which we then fitted a series of nested models: single-locus, epistatic, and two-locus models conditional on the main effects at *TCF7L2*. We used different multiple testing thresholds to assess results from the nested models. In the single-locus analyses we used previously published thresholds (WTCCC, 2007) to qualify genome-wide significance ($P = 5.00 \times 10^{-7}$) and a relaxed threshold to establish suggestive evidence for association ($P = 5.00 \times 10^{-4}$). In the two-locus test conditional on the main effects at *TCF7L2* we used a Bonferroni corrected threshold ($P = 7.51 \times 10^{-7}$) to establish significance. In the epistasis analyses we used Bonferroni corrected thresholds for tests focusing on the peak 79 pairs of markers from the joint-effects analyses ($P = 6.30 \times 10^{-4}$). Lastly, we also performed epistatic-analyses specific to the 23 confirmed T2D regions available in our data (see *Epistasis-only analyses* Methods section) and applied a Bonferroni correction for multiple testing in these analyses ($P = 1.97 \times 10^{-4}$). Our two-locus scan strategy required only moderate computational demands, because the joint two-locus test is an exact test that has a closed form solution to the maximum likelihood, and the majority of the two-locus analyses over the genome were performed using this test. However, the epistasis-only tests were computationally intensive, therefore we only tested for pair-wise epistasis among the peak joint two-locus results and 23 confirmed T2D regions.

Two-Locus Case-Control GWA Analyses

We performed case-control two-locus analyses within the logistic regression framework. Our principal two-locus GWA analyses took into account the main effects and all possible interaction terms between two loci in their contribution to the phenotype. The contribution of two loci to the linear predictor of T2D status in individual i can be described as

$$\pi_i \sim \beta_0 + \beta_1 X_{A1} + \beta_2 X_{D1} + \beta_3 X_{A2} + \beta_4 X_{D2} + \beta_5 X_{A1} X_{A2} + \beta_6 X_{A1} X_{D2} + \beta_7 X_{D1} X_{A2} + \beta_8 X_{D1} X_{D2}, \quad 1$$

where β_0 is the intercept, β_1 and β_3 are the main-effects additive effects at loci 1 and 2, β_2 and β_4 are the main-effects dominance effects at loci 1 and 2, and β_5 – β_8 are the additive-by-additive (β_5), additive-by-dominance (β_6), dominance-by-additive (β_7), and dominance-by-

dominance (β_8) epistatic effects. We scaled the genotypes at each of the two biallelic loci on additive (X_A) and dominance (X_D) scales. For example, for biallelic locus 1 with genotypes 11, 12, and 22, on the additive scale we counted the number of copies of the minor (2) allele ($11:X_{A1} = 0$, $12:X_{A1} = 1$, $22:X_{A1} = 2$) and on the dominance scale we compared heterozygote to homozygote genotypes ($11:X_{A1} = 0$, $12:X_{A1} = 1$, $22:X_{A1} = 0$). In the principal two-locus GWA analyses we compared the likelihood of the full two-locus model to the null model with no genetic effects (including just the β_0 term) to assess the evidence for joint two-locus effects on T2D. Genome-wide significance thresholds were obtained by Bonferroni correction for the overall number of tests performed (2.34×10^9), where a nominal $P = 2.14 \times 10^{-11}$ corresponded to a corrected $P = 0.05$.

Single-Locus Analyses and Two-Locus Analyses Conditional on *TCF7L2*

Single locus analyses were performed within the logistic regression framework by assessing the fit of the main additive and dominant effects at each locus. In particular, at each locus we compared the likelihood of the following model:

$$\pi_i \sim \beta_0 + \beta_1 X_{A1} + \beta_2 X_{D1}, \quad 2$$

where β_1 and β_2 were the main additive and dominance effects at that locus, to the likelihood under a null model with no genetic effects.

For the 80 peak regions of interest we wished to assess whether the two-locus signals were primarily due to the main effects at *TCF7L2*. We performed two-locus analyses conditional on *TCF7L2* main effects, by comparing the fit of model (1) to the fit of model (2), where β_1 and β_2 in model (2) were the main-effect additive and dominance effects specifically at *TCF7L2*. To correct for multiple testing in these conditional analyses, we applied a Bonferroni correction for the number of pair-wise tests genome-wide that involve the tag-SNP rs11196205 in *TCF7L2* (66,451 tests, excluding tests within chromosome 10), resulting in a nominal $P = 7.51 \times 10^{-7}$ corresponding to corrected $P = 0.05$.

The significant findings from the joint two-locus analyses included marker-pairs that were always in combination with tag-SNP rs11196205 in *TCF7L2*, which was the strongest single-locus association signal for T2D in these data. Previous studies show that when single-locus effects are large relative to interaction effects and particularly if they are sufficient to identify at least one of the loci in the pair, interaction-based approaches are not more powerful than single-locus approaches in identifying susceptibility loci (Marchini et al., 2005). To explore this further in our dataset, we performed genome-wide permutations of the data by keeping the genotypes at rs11196205 in *TCF7L2* constant and permuting the genotypes at the secondary locus in the pair, but preserving LD across the genome, for 1000 replicates. For each replicate we obtained the joint two-locus and single-locus test statistics. Using the 1000 replicates we then obtained the null distribution of the joint two-locus test statistic under a single-locus null model. The resulting permutation-based multiple testing threshold for the joint-two locus test under a null single-locus model was $P = 1 \times 10^{-14}$ and corresponded to a corrected $P = 0.05$.

Epistasis-Only Analyses

In the top 79 marker-pairs from the joint two-locus test, we evaluated the nominal significance of the interaction components alone to estimate the evidence of epistasis. We assessed the significance of the additive-by-additive epistasis variance term for case-control analyses as implemented within PLINK v1.06 (Purcell et al., 2007). PLINK interaction analyses in the case-control setting were also performed for variants in 23 of 28 established T2D candidate loci at the time of these analyses (Saxena et al., 2007; Scott et al., 2007;

Sladek et al., 2007; Steinthorsdottir et al., 2007; WTCCC, 2007; Zeggini et al., 2007; Unoki et al., 2008; Yasuda et al., 2008; Zeggini et al., 2008; Rung et al., 2009; Kong et al., 2009; Dupuis et al., 2010). We selected 23 T2D loci by selecting regions that were well represented by tagSNPs on the Affymetrix 500K array (by using HapMap CEU as a reference panel) and that passed quality control criteria in our study. We chose one tag-SNP per T2D candidate region (Table S1) and performed all pair-wise comparisons resulting in 253 tests between 23 candidate regions.

We then proceeded to test for interaction in the T2D cases only, because case-only analyses may provide an advantage over case-control analyses in terms of detecting interactions that contribute to disease-susceptibility (Yang et al., 1999). Case-only analyses were first performed for the 79 pair-wise regions of interest in PLINK. For both case-control and case-only analyses we applied a multiple-testing correction for the number of SNP pairs (79), setting the corrected epistasis P -value threshold to $P = 0.00063$. Case-only analyses were then also performed for the 253 pair-wise tests between the 23 T2D candidate regions.

For SNP-pairs of interest we also estimated genotype relative risks (GRRs) by using the odds ratios corresponding to the two-locus genotypes. We fit the full two-locus logistic regression model with respect to the two-locus genotype to obtain the odds ratios of T2D for individuals from a particular two-locus genotype relative to the baseline genotype, which was the rare double homozygote. Overall, the standard errors associated with the odds ratios were consistent among the eight remaining genotype classes. The odds ratios were used as estimates for the two-locus GRRs. We also estimated power to detect joint two-locus effects under one of the top five epistatic two-locus models in our dataset shown in Figure 2. Assuming a population prevalence of 2.4% for T2D (from Barroso, 2005) and the two-locus GRRs and allele frequencies from Figure 2, we simulated 1924 cases and 2938 controls for 1000 replicates. Simulated genotypes were analysed using the joint-effects two-locus test and the two-locus test conditional on *TCF7L2*.

Results

Joint Two-Locus GWA Analyses

We performed a two-dimensional genome-wide scan of T2D using a joint two-locus test of association including main and epistatic effects in a subset of 70,236 markers tagging common SNP variation at $r^2 > 0.2$ and $MAF \geq 0.05$. We examined 2.34×10^9 comparisons involving pairs of inter-chromosomal SNPs, and found that 79 pairs showed evidence for two-locus association at a Bonferroni-corrected P -value = 0.05, or uncorrected P -value = 2.14×10^{-11} (Fig. S1). The 79 pairs consisted of 80 markers (Fig. 1), where all of the 79 pairs involved loci that showed evidence for joint effects on diabetes in cooperation with rs11196205 in *TCF7L2*, which had the strongest single-locus effects (Table 1, Table S2). In CEPH individuals from the HapMap Project (Frazer et al., 2007), marker rs11196205 is in linkage disequilibrium with previously established T2D variants in *TCF7L2*, rs7903146 ($D' = 1$, $r^2 = 0.47$) and rs7901695 ($D' = 1$, $r^2 = 0.497$). Overall, we observed an enrichment of two-locus signals over the expectation under the null model that neither locus contributes to T2D, as shown in the quantile–quantile plot of these two-locus GWA results (Fig. S1).

The peak joint pair-wise results included confirmed variants in *FTO*, *TSPAN8*, and *CDKALI*, which exhibited significant main-effects and no evidence for epistasis with rs11196205 in *TCF7L2*. However, the majority (82%) of the 79 tag-SNPs did not have compelling single-locus evidence for association (P -value = 5×10^{-4}), suggesting the possibility that epistatic effects may contribute to T2D susceptibility.

TCF7L2 Conditional Analyses

Single-locus analyses for rs11196205 in *TCF7L2* resulted in the strongest evidence for single-locus association ($P = 6.99 \times 10^{-11}$). Because *TCF7L2* was the strongest single-locus association signal in the data, we next assessed whether the 79 peak two-locus signals were primarily due to the main effects at *TCF7L2* using a *TCF7L2* conditional test. None of the 79 pairs of loci surpassed genome-wide significance for the conditional test (Fig. S1), indicating that the majority of the genome-wide significant joint two-locus signals can be attributed to the main effects at *TCF7L2* alone. In addition, we also performed genome-wide permutations of the data under a null model where we kept the single-locus genotypes at rs11196205 in *TCF7L2*, but randomized the effect at the secondary locus across the genome. The resulting permutation-based joint two-locus threshold under this single-locus null model ($P = 1 \times 10^{-14}$) was more extreme than the observed joint two-locus findings, confirming the conditional test findings that the joint two-locus effects in our data are predominantly due to *TCF7L2* alone (Fig. S1).

However, the two-locus analyses highlighted pairs of loci where the secondary locus did not have single-locus evidence for association, but where the joint two-locus results were genome-wide significant and the *TCF7L2* conditional results for the pair were strongly suggestive of association. For example, in SNP-pairs rs1122637 (*RPP40*) – rs11196205 (*TCF7L2*), rs12447958 (*A2BPI*) – rs11196205 (*TCF7L2*), and rs2945770 (*ST3GAL1*) – rs11196205 (*TCF7L2*), the main effects at the secondary locus were not significant, but there was moderate evidence for association in the conditional analyses ($P = 5 \times 10^{-4}$). These findings imply that epistatic interactions may contribute to the joint two-locus effects and suggest further analyses assessing the significance of the interaction terms alone in the peak 79 SNP-pairs.

Epistasis-Only Analyses

We performed analyses assessing the evidence for gene–gene interactions alone, both in a case-control design and within the type 2 diabetes cases only, for the peak 79 pairs of loci (Table S2), testing specifically the additive-by-additive interaction component. In the case-control analyses 27 pairs of SNPs surpassed nominal significance ($P = 0.05$) for additive-by-additive epistasis and in the case-only analyses, there were 21 pairs, of which 17 overlapped with the case-control epistatic findings. When we took into account multiple testing, five of the 79 pairs of SNPs showed evidence for additive-by-additive epistasis both in the case/control analyses and in analyses restricted to the T2D cases alone (Table 2). In these five SNP-pairs both the case-only and the case-control epistasis findings were nominally significant, and at least one surpassed the multiple threshold correction. The five loci pairs included rs11196205 (*TCF7L2*) in combination with tag-SNPs located near *SLC9A9* (rs17635531), and regions on chromosomes 4 (rs1450140), 6 (rs1122637; rs1935683) and 11 (rs12803308). We examined pair rs1935683 (*RFPL4B*) – rs11196205 (*TCF7L2*) in more detail and obtained the odds ratios corresponding to the two-locus genotypes at the two SNPs. We used the odds ratios as estimators of the GRRs to obtain an insight into the underlying two-locus genetic model for this pair of regions (Fig. 2). Using the observed GRRs at these two loci, we then simulated case-control data to estimate power to detect these loci in our study. Power to detect multilocus joint effects at these two loci was estimated at 94.1%. However, when we tested power to detect rs1935683 (*RFPL4B*), conditional on *TCF7L2* main effects, only 28.6% of the replicates surpassed the corrected threshold ($P = 7.51 \times 10^{-7}$) for the conditional test. These results indicate that the conditional test has low power to detect epistatic loci of modest single-locus effects that interact with *TCF7L2*, which has strong single-locus effects in T2D.

Finally, we tested for epistasis across pair-wise combinations of tag-SNPs tagging variation in 23 regions that were type 2 diabetes susceptibility loci. We did not observe evidence for additive-by-additive interaction when correcting for the number of candidate-region tests performed (nominal $P = 1.97 \times 10^{-4}$ corresponding to corrected $P = 0.05$). The peak signals in these analyses were obtained for *IRS1* (rs2943634) with *IGF2BP2* (rs4402960), with suggestive evidence for epistasis from both case-control ($P = 2.67 \times 10^{-4}$) and case-only ($P = 0.01$) tests, and for *NOTCH2* (rs10923931) with *PPARG* (rs1801282) with modest evidence for epistasis from case-control analyses ($P = 5.97 \times 10^{-3}$).

Discussion

We performed a two-locus GWA scan including epistasis in T2D by selecting variants tagging common variation across the human genome. We obtained genome-wide evidence for joint two-locus effects at 79 pairs of markers, all of which included a single variant (rs11196205) in *TCF7L2*. Our conditional test results and permutation results indicate that the genome-wide significant joint two-locus effects are due to the strong main-effects at *TCF7L2* and not to genome-wide significant epistatic effects. We obtained interesting signals for epistasis at five pairs of regions and more detailed analyses of the SNP-pair with the strongest evidence for epistasis (rs1935683 in *RFPL4B* and rs11196205 in *TCF7L2*) gave an indication into potential underlying multilocus models.

Overall, we found weak evidence that epistatic effects contribute to the multilocus joint effects obtained in this study. These results are likely in part influenced by the strong single-locus effects of rs11196205 in *TCF7L2* in our dataset. Previous studies have shown that although multilocus approaches have on average more power than single-locus approaches to detect susceptibility loci in the presence of epistasis, that is not necessarily the case when one of the loci in the pair has strong single-locus effects that are large relative to the interaction effects and are sufficient to detect this locus in a single-locus GWAS (Marchini et al., 2005). Our findings confirm that in multilocus association testing very strong single locus effects can obscure the two-locus positive results. For example, for one of the observed suggestive models of epistasis in Figure 2, we estimated over 94% power to detect susceptibility variants under the joint two-locus test, however, power to detect the effects of the secondary locus (rs1935683 in *RFPL4B*) in the pair under the conditional test was only 28.6%. These results suggest that although interactions may contribute to T2D, there is currently insufficient power to detect them even in the single largest GWA study for T2D to date (WTCCC, 2007). Power to detect loci in the presence of epistasis depends on many factors, including sample-size, underlying model of epistasis, and disease allele frequencies at the two loci. Several approaches to detect genetic interactions exist (see Cordell, 2009) and the majority of methods used in our study focus on detecting susceptibility variants in the presence of epistasis, rather than targeting interactions alone. Tests for epistasis in GWA data may prove challenging, both computationally and due to the effect of allele frequency on the corresponding epistatic variance components. Computationally tractable methods to detect epistasis alone have previously been developed (Ritchie et al., 2001; Purcell et al., 2007; Steffens et al., 2010; Wan et al., 2010), and one of these has been applied to the T2D data used in our study (Wan et al., 2010) also reporting negative evidence for epistasis among genome-wide SNPs at least 1Mb away from each other. These findings confirm our results for the conditional and epistasis-only tests, when interpreted in a genome-wide context. In our study, single-locus tests together with the conditional test indicate the relative magnitude of epistatic effects. The joint two-locus analyses may provide a sense of the proportion of two locus effects in the genome that are epistatic, for example, the joint two-locus test results and associated quantile-quantile plot show that the majority of the deviation from the null model of no two-locus association is due to *TCF7L2* alone, although there is some minimal enrichment of two-locus signals for tests excluding *TCF7L2* (Fig.

S1). Based on these findings, it is unlikely that interactions between common variants will be detected at genome-wide significant thresholds in present-day single GWA scans, even if they contribute appreciably to the missing heritability in T2D. Replication of our negative findings is important, but the replication sets for our study will be smaller than the original sample set used, therefore obscuring the distinction between lack of power to detect signals and negative results. Larger sample sizes and meta-analysis approaches may help detect epistatic effects, but do not address the problems of low power to detect epistasis under certain underlying genetic models, and interpreting the epistatic effects in the context of the allele frequency spectrum. Furthermore, the best strategy to perform meta-analyses on results from multiple two-locus GWA scans remains unclear.

Although we did not find evidence for epistasis genome-wide, our few interaction-only analysis findings point towards loci for which there is weak or no evidence for single-locus association. Previous studies have shown that association analyses can detect large epistatic effects in the complete absence of main effects (Culverhouse et al., 2002), but another factor which can influence the power of multilocus GWA search strategy is the presence of LD between marker and disease locus (Marchini et al., 2005). We applied a tagging approach to select a subset of markers for the joint two-locus GWA tests, as this approach significantly reduced the computational load of the analyses. However, tagging may both reduce power to detect multilocus association, and may yield misleading inferences from interaction follow-up results due to the effect of linkage disequilibrium on the underlying two-locus models (Lynch & Walsh, 1998; Hill et al., 2008). Future studies need to improve our understanding of the effect of linkage disequilibrium on the epistatic components of the test statistic under various underlying models of epistasis.

The extent to which findings of statistical interactions relate to biological epistasis remains unclear. As pointed out by Lynch and Walsh (1998), unless the allele frequencies are known, the estimated variance components provide limited insight into the biological gene action. Some approaches incorporate additional biological data to prioritize, improve, or interpret statistical interaction analyses (Aylor & Zeng, 2008; Emily et al., 2009). However, there are many genetic models of epistasis for which the biological mechanisms are difficult to envision. The optimal approach would be to test for statistical epistasis in a situation where biological epistasis exists, unfortunately, examples of such studies are rare. The underlying genetic models for most complex traits are likely models of higher-dimensional biological complexity. However, an assessment of pair-wise joint effects is useful as a first step towards obtaining further insight into the genetic etiology of common complex traits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Broxholme for computational assistance and J. Randall for helping create Figure 1. JTB is supported by a Sir Henry Wellcome Trust Fellowship. NJT is supported by the MRC as part of the MRC CAiTE Centre (G0600705). EZ and APM are supported by the Wellcome Trust (Grant nos. WT088885/Z/09/Z to EZ and WT081682/Z/06/Z to APM). This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

References

Aylor D, Zeng Z. From classical genetics to quantitative genetics to systems biology: Modeling epistasis. *PLoS Genet.* 2008; 4:e1000029. [PubMed: 18369448]

- Barroso I. Complex disease: Pleiotropic gene effects in obesity and type 2 diabetes. *Eur J Hum Genet.* 2005; 13:1243–1244. [PubMed: 16251891]
- Carlborg O, Haley CS. Epistasis: Too often neglected in complex trait studies? *Nat Rev Genet.* 2004; 5:618–625. [PubMed: 15266344]
- Carlborg O, Hocking PM, Burt DW, Haley CS. Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genet Res.* 2004; 83:197–209. [PubMed: 15462413]
- Carlborg O, Kerje S, Schutz K, Jacobsson L, Jensen P, Andersson L. A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res.* 2003; 13:413–421. [PubMed: 12618372]
- Cordell H. Detecting gene–gene interactions that underlie human diseases. *Nat Rev Genet.* 2009; 10:392–404. [PubMed: 19434077]
- Culverhouse R, Suarez BK, Lin J, Reich T. A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet.* 2002; 70:461–471. [PubMed: 11791213]
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparso T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proenca C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'connell J, Luan J, Elliott A, Mccarroll SA, Payne F, Roccasacca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, De Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllenstein U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC, Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jorgensen T, Julia A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoecur C, Li Y, Lyssenko V, Mahley R, Mangino M, Manning AK, Martinez-Larrad MT, Mcateer JB, Mcculloch LJ, Mcpherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orru M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurdsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Thorand B, Tichet J, Tonjes A, Tuomi T, Uitterlinden AG, Van Dijk KW, Van Hoek M, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Rios M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP, Wichmann HE, Illig T, Rudan I, Wright AF, Stumvoll M, Campbell H, Wilson JF, Bergman RN, Buchanan TA, Collins FS, Mohlke KL, Tuomilehto J, Valle TT, Altshuler D, Rotter JJ, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Frayling TM, Ferrucci L, Kong A, Thorsteinsdottir U, Stefansson K, Van Duijn CM, Aulchenko YS, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Abecasis GR, Wareham NJ, Sladek R, Froguel P, Watanabe RM, Meigs JB, Groop L, Boehnke M, McCarthy MI, Florez JC, Barroso I. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* 2010; 42:105–116. [PubMed: 20081858]

- Emily M, Mailund T, Hein J, Schauer L, Schierup M. Using biological networks to search for interacting loci in genome-wide association studies. *Eur J Hum Genet.* 2009; 17:1205–1206. [PubMed: 19277066]
- Evans D, Marchini J, Morris A, Cardon L. Two-stage two-locus models in genome-wide association. *PLoS Genet.* 2006; 2:e157. [PubMed: 17002500]
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altschuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, De Bakker PI, Barrett J, Chretien YR, Maller J, Mccarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, Mcvean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwdimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, Mclay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–861. [PubMed: 17943122]
- Hill W, Goddard M, Visscher P. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 2008; 4:e1000008. [PubMed: 18454194]
- Holland JB, Moser HS, O'donoghue LS, Lee M. QTLs and epistasis associated with vernalization responses in oat. *Crop Sci.* 1997; 37:1306–1316.
- Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Parental origin of sequence variants associated with complex diseases. *Nature.* 2009; 462:868–874. [PubMed: 20016592]
- Li Z, Pinson SR, Park WD, Paterson AH, Stansel JW. Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics.* 1997; 145:453–465. [PubMed: 9071598]
- Lynch, M.; Walsh, B. Genetics and analysis of quantitative traits. Sinauer Association, Ltd; Sunderland, MA: 1998.
- Mackay TF. The genetic architecture of quantitative traits. *Annu Rev Genet.* 2001; 35:303–339. [PubMed: 11700286]
- Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet.* 2005; 37:413–417. [PubMed: 15793588]
- Montooth KL, Marden JH, Clark AG. Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics.* 2003; 165:623–635. [PubMed: 14573475]

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
- Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet.* 2001; 69:138–147. [PubMed: 11404819]
- Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, Bacot F, Balkau B, Belisle A, Borch-Johnsen K, Charpentier G, Dina C, Durand E, Elliott P, Hadjadj S, Jarvelin MR, Laitinen J, Lauritzen T, Marre M, Mazur A, Meyre D, Montpetit A, Pisinger C, Posner B, Poulsen P, Pouta A, Prentki M, Ribel-Madsen R, Ruokonen A, Sandbaek A, Serre D, Tichet J, Vaxillaire M, Wojtaszewski JF, Vaag A, Hansen T, Polychronakos C, Pedersen O, Froguel P, Sladek R. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet.* 2009; 41:1110–1115. [PubMed: 19734900]
- Saxena R, Voight BF, Lyssenko V, Burt NP, De Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Riche D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* 2007; 316:1331–1336. [PubMed: 17463246]
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science.* 2007; 316:1341–1345. [PubMed: 17463248]
- Sen S, Churchill GA. A statistical framework for quantitative trait mapping. *Genetics.* 2001; 159:371–387. [PubMed: 11560912]
- Shimomura K, Low-Zeddies SS, King DP, Steeves TD, Whiteley A, Kushla J, Zemenides PD, Lin A, Vitaterna MH, Churchill GA, Takahashi JS. Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice. *Genome Res.* 2001; 11:959–980. [PubMed: 11381025]
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007; 445:881–885. [PubMed: 17293876]
- Steffens M, Becker T, Sander T, Fimmers R, Herold C, Holler DA, Leu C, Herms S, Cichon S, Bohn B, Gerstner T, Griebel M, Nothen MM, Wienker TF, Baur MP. Feasible and successful: Genome-wide interaction analysis involving all 1.9×10^{11} pair-wise interaction tests. *Hum Hered.* 2010; 69:268–284. [PubMed: 20357478]
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorraddottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Dumaty A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, Van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet.* 2007; 39:770–775. [PubMed: 17460697]
- Sugiyama F, Churchill GA, Higgins DC, Johns C, Makaritsis KP, Gavras H, Paigen B. Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics.* 2001; 71:70–77. [PubMed: 11161799]

- Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP, Holmkvist J, Borch-Johnsen K, Jorgensen T, Sandbaek A, Lauritzen T, Hansen T, Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y, Kashiwagi A, Kaku K, Kawamori R, Tai ES, Pedersen O, Kamatani N, Kadowaki T, Kikkawa R, Nakamura Y, Maeda S. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet.* 2008; 40:1098–1102. [PubMed: 18711366]
- Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, Mcculloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, Mccarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, Van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieveve A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, Van Dam RM, Van Haeften TW, Van Herpt T, Van Vliet-Ostaptchouk JV, Bragi Walters G, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloy AL, Gyllenstein U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, Van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* 2010; 42:579–589. [PubMed: 20581827]
- Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NL, Yu W. BOOST: A fast approach to detecting gene–gene interactions in genome-wide case-control studies. *Am J Hum Genet.* 2010; 87:325–340. [PubMed: 20817139]
- Ways JA, Cicila GT, Garrett MR, Koch LG. A genome scan for Loci associated with aerobic running capacity in rats. *Genomics.* 2002; 80:13–20. [PubMed: 12079278]
- WTCCC. A genome-wide scan of 14,000 non-synonymous coding SNPs in 5,500 individuals: The Wellcome Trust Case Control Consortium. *Nat Genet.* 2007; 39:1329–1337. [PubMed: 17952073]
- Yang Q, Khoury MJ, Sun F, Flanders WD. Case-only design to measure gene–gene interaction. *Epidemiology.* 1999; 10:167–170. [PubMed: 10069253]
- Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY, Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet.* 2008; 40:1092–1097. [PubMed: 18711367]
- Zeggini E, Scott L, Saxena R, Voight B, Marchini J, Hu T, De Bakker P, Abecasis G, Almgren P, Andersen G. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008; 40:638–645. [PubMed: 18372903]
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007; 316:1336–1341. [PubMed: 17463249]

Figure 1. Single-locus GWA results. GWA P -values for type 2 diabetes are presented for all 393,143 common SNPs included in the study, from the 2df test statistic assessing the significance of the additive and dominance main effects at each SNP of interest. In red are shown the 80 markers (tag-SNPs) included in the peak 79 joint genome-wide significant two-locus results.

Figure 2. Epistasis between SNPs in *RFPL4B* and *TCF7L2* gene regions. For type 2 diabetes, two-locus GRRs estimated from odds ratios from the full two-locus logistic regression model of the two-locus genotype classes. GRRs are estimated relative to the baseline CC/CC double homozygote.

Table 1

Peak[†] twenty joint two-locus GWA results ranked in order of the joint effects P-value.

Chromosome	SNP	Joint P	Single-locus P	TCF7L2 conditional P [‡]	Nearest gene (50 kb)
16	rs9926289	2.00×10^{-13}	2.75×10^{-7}	7.87×10^{-6}	<i>FTO</i>
12	rs11057970	2.60×10^{-13}	8.77×10^{-3}	2.70×10^{-5}	<i>DHX37</i>
3	rs1122341	9.50×10^{-13}	5.47×10^{-4}	4.22×10^{-5}	–
3	rs1552489	9.70×10^{-13}	2.35×10^{-2}	6.44×10^{-5}	<i>FYCO1</i>
15	rs2903265	9.70×10^{-13}	4.44×10^{-5}	7.06×10^{-5}	<i>ZFAND6</i>
9	rs10512091	1.09×10^{-12}	1.57×10^{-2}	3.06×10^{-5}	<i>GNAQ</i>
4	rs6813642	1.18×10^{-12}	3.12×10^{-2}	6.55×10^{-5}	–
21	rs6517434	1.75×10^{-12}	3.32×10^{-5}	2.27×10^{-4}	<i>KCNJ6</i>
20	rs1998190	1.88×10^{-12}	3.86×10^{-2}	8.55×10^{-5}	<i>BMP7</i>
4	rs13141142	1.94×10^{-12}	4.95×10^{-4}	6.11×10^{-5}	–
2	rs6436555	2.64×10^{-12}	1.12×10^{-3}	1.17×10^{-4}	–
9	rs638541	2.88×10^{-12}	9.48×10^{-2}	1.20×10^{-4}	–
16	rs2099106	3.05×10^{-12}	3.45×10^{-5}	1.21×10^{-4}	–
6	rs1122637	3.30×10^{-12}	5.73×10^{-1}	1.35×10^{-4}	–
8	rs6997046	3.48×10^{-12}	3.55×10^{-3}	1.39×10^{-4}	<i>NAT2</i>
12	rs1495377	3.61×10^{-12}	7.19×10^{-6}	1.57×10^{-4}	<i>TSPAN8</i>
16	rs12447958	4.00×10^{-12}	2.22×10^{-1}	3.89×10^{-4}	<i>A2BPI</i>
8	rs2945770	4.06×10^{-12}	4.16×10^{-1}	2.87×10^{-4}	<i>ST3GAL1</i>
5	rs6451305	4.76×10^{-12}	7.64×10^{-2}	2.40×10^{-4}	<i>SLC1A3</i>
6	rs9348440	5.24×10^{-12}	1.01×10^{-5}	1.69×10^{-4}	<i>CDKALI</i>

[†]Table S2 lists all 79 genome-wide significant SNP-pairs.

[‡]The single-locus P-value for *TCF7L2* was 6.99×10^{-11} and ranged between 1.27×10^{-9} and 3.96×10^{-11} when estimating the *TCF7L2* main-effects within each pair of regions due to missing data, considering only individuals with available two-locus genotypes.

Table 2

Results for SNP pairs with evidence for epistasis in both case-control and case-only analyses.

Chromosome	SNP	Joint P	Single-locus P	Epistasis P [†]	Nearest gene
6	rs1935683	8.62×10^{-12}	3.38×10^{-1}	2.92×10^{-5}	<i>RFPL4B</i>
4	rs1450140	1.83×10^{-11}	7.34×10^{-1}	1.34×10^{-4}	<i>ODZ3</i>
3	rs17635531	7.83×10^{-12}	2.06×10^{-1}	2.02×10^{-4}	<i>SLC9A9</i>
11	rs12803308	2.09×10^{-11}	1.02×10^{-2}	2.68×10^{-4}	<i>IRKL</i>
6	rs1122637	3.30×10^{-12}	5.73×10^{-1}	6.27×10^{-4}	<i>PPP1R3G</i>

[†]The Epistasis *P*-value is the most significant nominal *P*-value from the case-control and case-only epistasis tests in PLINK. In each case both nominal interaction *P*-values were < 0.05 , and at least one was $< 6.30 \times 10^{-4}$.