

ARTICLE

A novel method, the Variant Impact On Linkage Effect Test (VIOLET), leads to improved identification of causal variants in linkage regions

Lisa J Martin^{*,1,2,3}, Lili Ding^{2,3}, Xue Zhang¹, Ahmed H Kissebah^{4,5,9}, Michael Olivier^{4,6,7}
and D Woodrow Benson^{3,8}

The Human Genome Project was expected to individualize medicine by rapidly advancing knowledge of common complex disease through discovery of disease-causing genetic variants. However, this has proved challenging. Although linkage analysis has identified replicated chromosomal regions, subsequent detection of causal variants for complex traits has been limited. One explanation for this difficulty is that utilization of association to follow up linkage is problematic given that linkage and association are not required to co-occur. Indeed, co-occurrence is likely to occur only in special circumstances, such as Mendelian inheritance, but cannot be universally expected. To overcome this problem, we propose a novel method, the Variant Impact On Linkage Effect Test (VIOLET), which differs from other quantitative methods in that it is designed to follow up linkage by identifying variants that influence the variance explained by a quantitative trait locus. VIOLET's performance was compared with measured genotype and combined linkage association in two data sets with quantitative traits. Using simulated data, VIOLET had high power to detect the causal variant and reduced false positives compared with standard methods. Using real data, VIOLET identified a single variant, which explained 24% of linkage; this variant exhibited only nominal association ($P=0.04$) using measured genotype and was not identified by combined linkage association. These results demonstrate that VIOLET is highly specific while retaining low false-negative results. In summary, VIOLET overcomes a barrier to gene discovery and thus may be broadly applicable to identify underlying genetic etiology for traits exhibiting linkage.

European Journal of Human Genetics (2014) 22, 243–247; doi:10.1038/ejhg.2013.120; published online 5 June 2013

Keywords: quantitative trait locus; statistical; variant discovery; quantitative traits; complex traits

INTRODUCTION

The Human Genome Project was expected to advance knowledge of the genetic basis of common complex disease. Unfortunately, identification of disease-causing genetic variants in complex traits has been challenging,^{1,2} a difficulty which may be related, in part, to analytical strategies for gene discovery. Linkage and association are analytic methods used in complex trait analysis.³ Although many replicated linkage and association signals exist, except for a few examples,^{4–6} peaks do not overlap. This overlap failure may be fundamental to analytic differences in that linkage requires familial segregation whereas association tests for co-occurrence.⁷ Further, association pinpoints common variants with small effects, whereas linkage identifies large chromosomal regions of moderate or large effect.³ As such, overlap of linkage and association may occur only in special circumstances, such as Mendelian inheritance, but cannot be expected universally.

Various methods have been proposed to test whether a genetic variant can account for an observed linkage signal.^{8–13} These methods model linkage and association jointly and thus may fail to identify

variants when linkage and association are not co-occurring. Thus, there is a need for methods that test for the effects of variants on the linkage signal regardless of association.

To overcome this barrier, we propose a novel method, the Variant Impact On Linkage Effect Test (VIOLET). VIOLET is unique because it identifies genetic variants that impact the quantitative trait locus' variance without any assumptions about association. Using simulated and real data, we demonstrate that VIOLET has reduced false positives but without corresponding increases in the false negatives when compared with measured genotype and combined linkage association. Thus, VIOLET may fill a gap in variant discovery.

METHODS

Data sets

To compare VIOLET with standard methods, simulated (Genetics Analysis Workshop: GAW17) and real (Metabolic Risk Complications of Obesity Genes Study: MRC-OB) data sets were used. GAW17 had regions with linkage and association evidence.^{14,15} Using MRC-OB, linkage on 7q36 for triglycerides was identified.¹⁶ After dense single-nucleotide polymorphism (SNP) genotyping,

¹Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ²Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ³Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH, USA; ⁴Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI, USA; ⁵Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA; ⁶Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, WI, USA; ⁷Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA; ⁸Herma Heart Center, Children's Hospital of Wisconsin, Milwaukee, WI, USA

*Correspondence: Dr LJ Martin, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 4006, Cincinnati, OH 45229, USA. Tel: +1 513-636-1244; Fax: +1 513 636 7297; E-mail: Lisa.martin@cchmc.org

⁹Deceased.

Received 4 May 2012; revised 28 March 2013; accepted 19 April 2013; published online 5 June 2013

associations were found, but only a modest portion of the linkage was explained.^{17,18}

Simulated data set – GAW17-simulated data set. GAW17 used the 1000 Genomes¹⁹ exome sequence data^{14,15} to generate a data set with 697 individuals in 8 pedigrees. Fully informative markers were used to compute identical-by-descent (IBD) allele sharing.¹⁵ A quantitative phenotype (Q1) influenced by 39 SNPs in 9 genes was simulated. GAW17 data providers have given permission for data use.

To identify regions exhibiting consistent evidence of linkage with Q1, all 200 simulations were evaluated. VEGFA, on chromosome 6p21.1, showed linkage (median LOD = 3.1). Across chromosome 6, 856 SNPs were polymorphic, including the causal variant, C6S2981.

Real data set – MCR-OB. MRC-OB was established in 1994, when families were recruited from the Take Off Pounds Sensibly Inc. membership.^{16,20} Fasting plasma triglycerides were determined spectrophotometrically in triplicate. A total of 2209 individuals from 507 families of Northern European descent formed the cohort.¹⁶ All protocols were approved by the Institutional Review Board of the Medical College of Wisconsin.

A genome-wide linkage scan identified a quantitative trait locus (QTL) on chromosome 7q36 linked to triglycerides (LOD = 3.7).¹⁶ From the initial cohort, 1235 individuals from 258 families contributing to the linkage were selected for dense genotyping (Table 1) of 1048 tag SNPs using an Affymetrix MegAllele custom-designed array (Affymetrix, Santa Clara, CA, USA).^{17,18,21} Additionally, 354 SNPs from chromosome 14 were available for analysis. Chromosome 14 SNPs were used to determine VIOLET's specificity. SNPs exhibiting Mendelian inconsistencies were blanked.

Statistical methods

Data preparation. Q1 and triglycerides were examined for normality. Q1 exhibited a normal distribution, so no transformation was applied. Triglyceride levels exhibited right skewing, so the data were natural log (*ln*) transformed. Data were re-examined and observations exceeding 4 SD units were removed.¹⁶

Measured Genotype Association (MGA). To test a SNP's phenotypic effect, we used MGA.²² Briefly, genotypes were assigned as 0, 1, and 2 according to the number of minor alleles.²³ To account for phenotypic correlation between family members, variance component analysis in SOLAR was used (Texas Biomedical Research Institute, San Antonio, TX, USA).²⁴ Mixed effects models are applied where fixed effects are covariates. Random effects are defined by

Table 1 Descriptive statistics for the take off pounds sensibly (MCR-OB) cohort

Trait	Original cohort	Follow-up cohort – linking to 7q36
<i>N</i>	2209	1235
No. of Families	507	258
Family size (mean (range))	5.0 (2–29)	4.8 (2–27)
Age (years)	48.12 ± 15.10	47.92 ± 14.80
Triglycerides	128.73 ± 81.19	129.57 ± 83.70
Ln triglycerides	4.71 ± 0.53	4.71 ± 0.54
Sex (% male)	23.5	23.8
Asthma (% Y)	8.5	9.5
Menopausal (% Post)	40.0	40.5
Estrogen use (% Y)	17.7	18.5
Ever smoke (% Y)	22.0	16.0
Cholesterol/LDL medications	3.5	2.9
Type 2 diabetes (% Y)	15.9	16.9
Diabetes medications (% Y)	5.2	5.7

genetic and environmental deviations:

$$y = \mu + \beta(\text{SNP}) + g + e$$

$$\Omega = 2\Phi\sigma_g^2 + I\sigma_e^2 \quad (\text{MGA})$$

where μ is the grand mean, β is the SNP effect, and g and e are the genetic and environmental deviations, respectively. Assuming g and e are uncorrelated random normal variables with expectation 0, the phenotypic covariance of relative pairs (Ω) can be partitioned into additive genetic and environmental components, where Φ is the kinship matrix, I is the identity matrix, and σ_g^2 and σ_e^2 are the variance due to additive genetic (g) and residual (e) effects, respectively. To test a SNP effect, log likelihood of the model estimating the SNP effect is compared with the log likelihood of the model in which the SNP effect is constrained to zero. Assuming that trait values follow a multivariate normal distribution, twice the difference in the log likelihoods of these two models is asymptotically distributed as χ_1^2 .

Combined Linkage Association (CLA). To test the impact of variants on the QTL effect, we adapted a variance components CLA (implemented in SOLAR).²⁴ Briefly, the standard linkage model is defined by:²⁴

$$y = \mu + a + g + e$$

$$\Omega = \hat{\Pi}\sigma_a^2 + 2\Phi\sigma_g^2 + I\sigma_e^2 \quad (\text{Linkage})$$

Where $\hat{\Pi}$ provides the predicted proportion of alleles that related individuals share IBD at locus A and σ_a^2 is the variance due to locus A. The significance of linkage is estimated through a LOD score, which is calculated by comparing models with and without a . In this model, the grand mean account for the trait mean but not for the SNP effects. CLA is a linkage model conditional on a SNP fixed effect, such that:

$$y = \mu + \beta(\text{SNP}) + a + g + e$$

$$\Omega = \hat{\Pi}\sigma_a^2 + 2\Phi\sigma_g^2 + I\sigma_e^2 \quad (\text{CLA})$$

if a SNP accounts for all of the linkage, evidence of linkage should disappear; pragmatically when the LOD score drops <0.5 indicating that the linkage is fully explained.²⁵

As both the simulated and the real data exhibited differences in the base LOD score (simulation due to differences in replicates; real data due to some missing genotype data), examination of the LOD score from CLA did not provide a complete picture on the magnitude of change. To account for these differences, percentage of LOD drop ((LOD_{no SNP} – LOD_{SNP})/LOD_{no SNP}) was examined. However, it is important to note that the percentage of LOD drop is simply used to provide an assessment in the change in the LOD score while accounting for the baseline LOD.

VIOLET. To test the significance of the impact of a variant on the QTL, VIOLET builds upon CLA. However, VIOLET explicitly tests whether the variance explained by the QTL changes with SNP inclusion. This is operationalized by comparing the CLA model with a model that is identical to the CLA model except that the σ_a^2 is constrained to be equal to the variance due to the locus when the SNP effect is constrained to zero (σ_{a0}^2). Thus

$$y = \mu + \beta(\text{SNP}) + a + g + e$$

$$\Omega = \hat{\Pi}\sigma_{a0}^2 + 2\Phi\sigma_g^2 + I\sigma_e^2 \quad (\text{VIOET})$$

To test for significance, twice the difference in the log likelihoods of model CLA and VIOLET are evaluated, this test statistic is named V. This statistic differs from CLA because in CLA the major comparison is between a freely estimated σ_{a0}^2 to one constrained to zero. Given that the likelihood function of VIOLET's model is a function of σ_{a0}^2 , which itself is a maximum likelihood solution, Wilks' Theorem on the asymptotic approximations of test statistic distributions under the null hypothesis (that there is no difference in the goodness-of-fit) may not hold. As such, V's distribution under the null was examined empirically to determine appropriate thresholds. For the real data, we evaluated V derived when genotypes were randomly assigned across 1000

permutations; however, microsatellite data for linkage retained in their original structure.

RESULTS

Simulated data from GAW17

To evaluate the performance of VIOLET, measured genotype, and combined linkage association, two thresholds were utilized, 99% power and multiple testing corrected type I error ($P=0.0000584$). The power threshold was set to the 1% quantile from causal SNPs ($V=4.17$ for VIOLET, $\chi^2=33.16$ for measured genotype, and percentage of LOD drop = 90.98 for combined linkage association). The type I error rate threshold was set to the 99.99416% quantile of non-causal variants ($V=4.08$, $\chi^2=42.15$, and percentage of LOD drop = 97.61) (Table 2).

VIOLET. The V null distribution for non-causal variants is highly skewed with 99.5% of observations falling below 0.13 (Figure 1). Controlling for type I error ($V \geq 4.08$), the causal variant, C6S2981 (MAF = 0.033), was identified (median $V = 8.73$, range 2.50–16.10) in 198 out of 200 simulations (Figure 2a). In all but two simulations, C6S2981 exhibited the highest V . Controlling for power ($V \geq 4.17$), we detected a very low false-positive rate ($9/171\,000 = 0.005\%$) (Table 2). These results demonstrate that VIOLET has a high degree of specificity, with little overlap in the distribution of V between non-causal and causal variants.

Comparison of VIOLET with MGA and CLA. Like VIOLET, MGA and CLA identified C6S2981 (median P -value = 2.1×10^{-14} , LOD = 0) (Figure 2). All methods had high power to detect C6S2981 (Table 2) when controlling for type I error. MGA and CLA exhibited 93% and 88% power to detect C6S2981 as compared with 99% power for VIOLET. Further, when evaluating the percentage of LOD drop, there was 93% power.

When controlling for power, VIOLET exhibited fewer false positives than the other methods. Out of 200 replicates, VIOLET identified 9 false positives, whereas MGA identified 58 and CLA identified 106 using the LOD and 37 using LOD drop. Importantly, two non-causal SNPs were identified as associated using MGA after Bonferroni correction in over half of the simulations (median P -values = 2.1×10^{-5} and 3.0×10^{-7} ; Figure 2b).

Real data from MCR-OB – analysis of linkage for serum triglycerides

Dense genotyping. Dense genotyping was performed on 1235 individuals. There were 1023 and 352 polymorphic SNPs on chromosomes 7 and 14, respectively. There were no major phenotypic

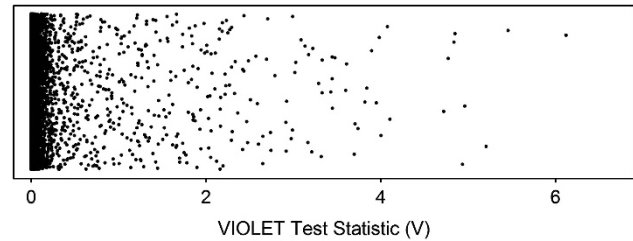


Figure 1 Distribution of the VIOLET test statistic (V) in non-causal variants. Results were from 200 simulations, each with 855 non causal SNPs.

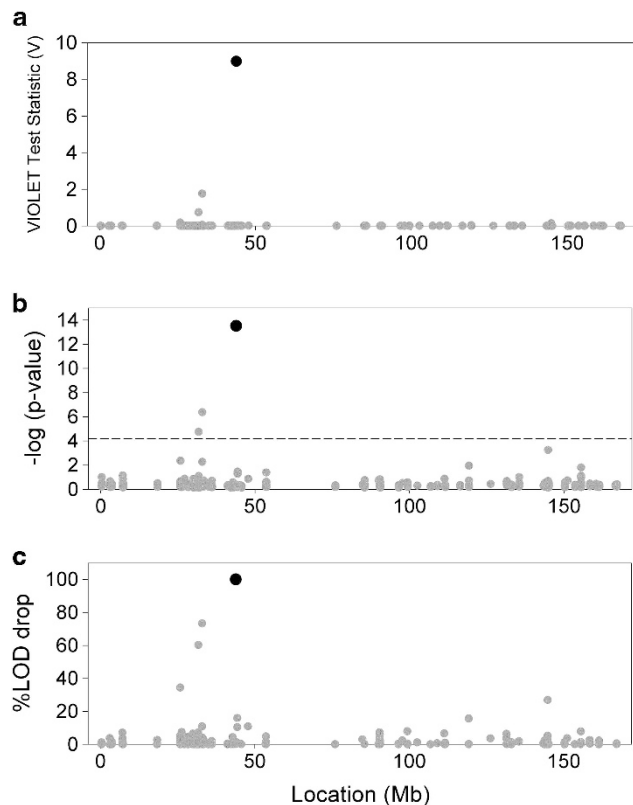


Figure 2 Comparison of results using the Genetics Analysis Workshop 17 (GAW17) simulated data set. Panel **a** presents VIOLET, panel **b** presents MGA, and panel **c** presents CLA. All data are presented as the median results from 200 simulated replicates. The black dot identifies C6S2981, the causal variant.

Table 2 False-positive and false-negative rates for MGA, CLA, and VIOLET using GAW 17-simulated data

	Statistic	VIOLET V	MGA χ^2	CLA LOD	% LOD drop
Constant power (99%)	Threshold	4.17	33.16	0.53	90.98
	#	9	58	106	37
	Type I error (%)	0.005	0.034	0.062	0.022
Constant type I error (5.8×10^{-5})	Threshold	4.08	42.45	0.04	97.61
	#	2	14	24	14
	Power	99%	93%	88%	93%

Results were from 200 simulations, each with 856 SNPs, one of which was causal. Constant power thresholds derived from the 1% quantile of the causal SNPs. Constant type I error thresholds derived from the 99.99416% quantile of non-causal variants.

differences between the full cohort and the dense genotyping (Table 1). The chromosome 7 LOD score was 8.2, which is higher than the full cohort as families were selected because they positively contributed to the linkage.

VIOLET. When using VIOLET, one SNP (rs39179) exhibited an increased test statistic ($V = 9.0$; Figure 3a); beyond rs39179 there is no evidence of a variant contributing to the linkage (mean $V_{\text{excluding rs39179}} = 0.00003 \pm 0.000081$; Figure 3a). When evaluating 1000 permutations of rs39179, mean $V = 0.0007 \pm 0.0019$, with no value exceeding 0.02 (empirical P -value < 0.001). When 352 SNPs from chromosome 14 were examined, there is no evidence that these variants account for the chromosome 7 linkage ($V < 0.02$), suggesting a high degree of specificity for V .

Comparison of VIOLET with MGA and CLA. Neither MGA nor CLA identified any significant variant (Bonferroni corrected P -value ($P < 0.000048$) and adjusted LOD < 0.5 , respectively). Using MGA, 109 chromosome 7 variants exhibited nominal evidence of association ($P \leq 0.05$; Figure 3b); minimum P -value = 0.00007. Using CLA, the mean percentage of LOD drop (\pm SD) was 0.008 ± 0.014 with a range of 0–0.24 (Figure 3c). Interestingly, the SNP identified by VIOLET (rs39179) exhibits the largest percentage of LOD drop; the nominal P -value from the measured genotype approach was 0.04. When MGA P -values are ranked, 88 other SNPs showed stronger association than rs39179. When examining an unlinked chromosome 14 region, both VIOLET and CLA exhibited little evidence of an effect (CLA mean LOD drop = 0.001 ± 0.01). However, using MGA, 23 SNPs exhibited

nominal association; no SNPs reached Bonferroni correction ($P < 0.00014$, minimum P -value = 0.0015).

DISCUSSION

Identification of causal variants accounting for linkage has been difficult.^{26–32} This is a problem because failure to identify causal variants within linkage regions may impede gene discovery. Part of the difficulty may be related to the analytical strategy using association-based methods to follow up linkage. Using association-based methods to follow up linkage signals may miss variants with little evidence of association but substantial effects on the linkage signal. Thus, we propose a novel method, VIOLET, to examine the impact of a specific variant on linkage without any assumptions about association. VIOLET has considerable advantage over MGA because only variants contributing to the linkage are identified. Additionally, VIOLET offers an advantage over CLA, as it provides a formal test statistic to evaluate the significance of variants that do not completely explain a linkage peak. This is accomplished by comparing two models whose only difference is the proportion of variation explained at a locus. We demonstrate that VIOLET identifies variants underlying linkage in a highly specific manner. As such, VIOLET may expedite causal variant discovery.

Using simulated data, VIOLET had higher power and lower type I error as compared with MGA and CLA. A major challenge with the simulated data was that there was a single causal SNP contributing to the linkage for a quantitative phenotype, as such all methods performed well. Further, it is important to note that for both the simulated and the real data set, the variant identified had MAF $< 5\%$. Variants of lower frequency are of concern in association (including MGA) studies due to possible stratification. Thus, future studies should examine VIOLET's performance in scenarios when multiple variants contribute to the linkage, when variation is non-additive,³³ when the outcome is dichotomous,³⁴ and when causal variants differ in frequency.

Using MCR-OB data, VIOLET was applied to a linkage peak on chromosome 7.¹⁶ Although this region has been densely genotyped, MGA yielded associations that explained little of the observed linkage.^{17,18} Using VIOLET, a single SNP (rs39179) accounting for 24% of the LOD score was identified (but it did not reach the traditional CLA threshold). This SNP was only nominally associated in MGA ($P = 0.04$) and was ranked eighty-ninth in the P -value ranking. However, causal variants do not always result in the highest ranking P -values.³⁵ The problem with this scenario is that based on MGA results there are too many promising candidates to be experimentally validated; thus only the top ranking associations are likely to be examined for biological plausibility. Indeed, our research team had not considered rs39179 (minor allele frequency 2.6% in our cohort; present in 25 of the 258 families and explained 0.7% of the variation in triglycerides) a promising candidate and rather focused on other SNPs.^{17,18} Our results suggest that either rs39179 or SNPs in strong linkage disequilibrium (LD) with rs39179 may be causal. Using SNAP,³⁶ a single ungenotyped SNP (rs10276884) in strong LD with rs39179 was identified. This variant is in the promoter of *DPP6* and predicted to change a SF2/ASF motif. Clearly, additional studies are required.

It may seem counterintuitive that linkage would not require association as there are examples of association and linkage overlapping.^{4,6,37,38} However, for complex traits, linkage and association overlap may be the exception. Mouse strain-dependent variability supports such lack of overlap.^{39–43} Indeed, fibronectin defects cause cardiovascular malformations,⁴⁴ but there is substantial phenotypic

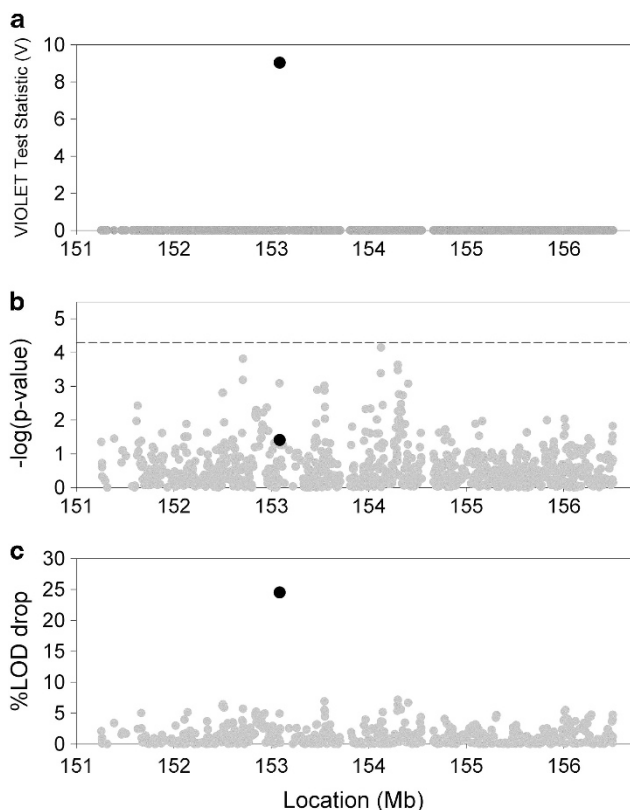


Figure 3 Comparison of results using data from the Metabolic Risk Complications of Obesity Genes Study (MCR-OB) Study. Panel **a** presents VIOLET, panel **b** presents MGA, and panel **c** presents CLA. The black dot represents the variant, rs39179, accounting for a 24% LOD drop.

heterogeneity by strain.^{42,45} Thus, even for severe genetic changes such as gene deletion, other loci may contribute to the phenotype. As complex traits are expected to be the combination of multiple genetic factors, lack of strong association is not unexpected. Indeed, most genome-wide association studies exhibit small effects.^{1,2} However, VIOLET tests the impact of a variant on linkage in a highly specific manner and thus is optimally positioned to identify variants that contribute to the linkage regardless of association evidence.

CONCLUSION

In summary, we propose a novel method, VIOLET, to follow up linkage. This method differs from the MGA and CLA because VIOLET measures the change in the estimate of the linkage effect when the SNP is included. Using real and simulated data, VIOLET is shown to be highly specific and reduce false-negative findings when following up linkage.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the MRC-OB Study team, as well as TOPS Inc. and its members for participating in and supporting the MRC-OB study. Preparation of GAW 17 data was supported, in part, by NIH R01 MH059490 and used data from the 1000 Genomes Project (www.1000genomes.org). This work was supported by grants from the NIH HL069712 (DWB), HL074728 (LJM, DWB), HL74168 (MO, LJM) GM031575 (Genetic Analysis Workshop data), and MH059490 for SOLAR.

- 1 Perola M: Genetics of human stature: lessons from genome-wide association studies. *Horm Res Paediatr* 2011; **76**(Suppl 3): 10–11.
- 2 Need AC, Goldstein DB: Whole genome association studies in complex diseases: where do we stand? *Dialogues Clin Neurosci* 2010; **12**: 37–46.
- 3 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
- 4 Klein RJ, Zeiss C, Chew EY *et al*: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**: 385–389.
- 5 Sladek R, Rocheleau G, Rung J *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
- 6 Grant SF, Thorleifsson G, Reynisdottir I *et al*: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320–323.
- 7 Chung RH, Hauser ER, Martin ER: Interpretation of simultaneous linkage and family-based association tests in genome screens. *Genet Epidemiol* 2007; **31**: 134–142.
- 8 Almasy L, Williams JT, Dyer TD, Blangero J: Quantitative trait locus detection using combined linkage/disequilibrium analysis. *Genet Epidemiol* 1999; **17** (Suppl 1): S31–S36.
- 9 Li M, Boehnke M, Abecasis GR: Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. *Am J Hum Genet* 2005; **76**: 934–949.
- 10 Biernacka JM, Cordell HJ: A composite-likelihood approach for identifying polymorphisms that are potentially directly associated with disease. *Eur J Hum Genet* 2009; **17**: 644–650.
- 11 Biernacka JM, Cordell HJ: Exploring causality via identification of SNPs or haplotypes responsible for a linkage signal. *Genet Epidemiol* 2007; **31**: 727–740.
- 12 Cardon LR, Abecasis GR: Some properties of a variance components model for fine-mapping quantitative trait loci. *Behav Genet* 2000; **30**: 235–243.
- 13 Li M, Boehnke M, Abecasis GR: Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am J Hum Genet* 2006; **78**: 778–792.
- 14 Ziegler A, Ghosh S, Dyer TD, Blangero J, Maccluer J, Almasy L: Introduction to genetic analysis workshop 17 summaries. *Genet Epidemiol* 2011; **35**(Suppl 1): S1–S4.
- 15 Almasy L, Dyer TD, Peralta JM *et al*: Genetic Analysis Workshop 17 mini-exome simulation. *BMC Proc* 2011; **5**: S2.
- 16 Sonnenberg GE, Krakower GR, Martin LJ *et al*: Genetic determinants of obesity-related lipid traits. *J Lipid Res* 2004; **45**: 610–615.
- 17 Smith EM, Zhang Y, Baye TM *et al*: INSIG1 influences obesity-related hypertriglyceridemia in humans. *J Lipid Res* 2010; **51**: 701–708.
- 18 Zhang Y, Smith EM, Baye TM *et al*: Serotonin (5-HT) receptor 5A sequence variants affect human plasma triglyceride levels. *Physiol Genomics* 2010; **42**: 168–176.
- 19 Durbin RM, Abecasis GR, Altshuler DL *et al*: A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
- 20 Kissebah AH, Sonnenberg GE, Myklebust J *et al*: Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci USA* 2000; **97**: 14478–14483.
- 21 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 22 Boerwinkle E, Chakraborty R, Sing CF: The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 1986; **50**: 181–194.
- 23 Falconer D: *Introduction to Quantitative Genetics*, 3rd edn. New York, NY, USA: Longman Scientific and Technical, 1989.
- 24 Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**: 1198–1211.
- 25 Almasy L, Blangero J: Exploring positional candidate genes: linkage conditional on measured genotype. *Behav Genet* 2004; **34**: 173–177.
- 26 Delplanque J, Barat-Houari M, Dina C *et al*: Linkage and association studies between the proopiomelanocortin (POMC) gene and obesity in caucasian families. *Diabetologia* 2000; **43**: 1554–1557.
- 27 Comuzzie AG, Hixson JE, Almasy L *et al*: A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* 1997; **15**: 273–276.
- 28 Hixson JE, Almasy L, Cole S *et al*: Normal variation in leptin levels is associated with polymorphisms in the proopiomelanocortin gene, POMC. *J Clin Endocrinol Metab* 1999; **84**: 3187–3191.
- 29 Abraham R, Myers A, Wavrant-DeVrieze F *et al*: Substantial linkage disequilibrium across the insulin-degrading enzyme locus but no association with late-onset Alzheimer's disease. *Hum Genet* 2001; **109**: 646–652.
- 30 Bertram L, Blacker D, Mullin K *et al*: Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science* 2000; **290**: 2302–2303.
- 31 Lee SH, Park JS, Park CS: The search for genetic variants and epigenetics related to asthma. *Allergy Asthma Immunol Res* 2011; **3**: 236–244.
- 32 Nothen MM, Nieratschker V, Cichon S, Rietschel M: New findings in the genetics of major psychoses. *Dialogues Clin Neurosci* 2010; **12**: 85–93.
- 33 Zuk O, Hechter E, Sunyaev SR, Lander ES: The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 2012; **109**: 1193–1198.
- 34 Duggirala R, Williams JT, Williams-Blangero S, Blangero J: A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genet Epidemiol* 1997; **14**: 987–992.
- 35 Zaykin DW, Zhivotovskiy LA: Ranks of genuine associations in whole-genome scans. *Genetics* 2005; **171**: 813–823.
- 36 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI: SNP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938–2939.
- 37 Goldberg YP, Telenius H, Hayden MR: The molecular genetics of Huntington's disease. *Curr Opin Neurol* 1994; **7**: 325–332.
- 38 Tirado I, Soria JM, Mateo J *et al*: Association after linkage analysis indicates that homozygosity for the 46C->T polymorphism in the F12 gene is a genetic risk factor for venous thrombosis. *Thromb Haemost* 2004; **91**: 899–904.
- 39 Doetschman T: Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol* 2009; **530**: 423–433.
- 40 Lu SY, Jin Y, Li X *et al*: Embryonic survival and severity of cardiac and craniofacial defects are affected by genetic background in fibroblast growth factor-16 null mice. *DNA Cell Biol* 2010; **29**: 407–415.
- 41 Winston JB, Erlich JM, Green CA *et al*: Heterogeneity of genetic modifiers ensures normal cardiac development. *Circulation* 2010; **121**: 1313–1321.
- 42 George EL, Georges-Labouesse EN, Patel-King RS, Rayburn H, Hynes RO: Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development* 1993; **119**: 1079–1091.
- 43 Threadgill DW, Dlugosz AA, Hansen LA *et al*: Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* 1995; **269**: 230–234.
- 44 Astrof S, Hynes RO: Fibronectins in vascular morphogenesis. *Angiogenesis* 2009; **12**: 165–175.
- 45 Astrof S, Kirby A, Lindblad-Toh K, Daly M, Hynes RO: Heart development in fibronectin-null mice is governed by a genetic modifier on chromosome four. *Mech Dev* 2007; **124**: 551–558.