Title

Colocalization of GWAS and eQTL signals at loci with multiple signals identifies additional candidate genes for body fat distribution.

Journal

Human Molecular Genetics, 28(24)

Authors

Wu, Ying
Broadaway, K
Raulerson, Chelsea
et al.

Publication Date

2019-12-15

DOI

10.1093/hmg/ddz263

Peer reviewed

Human Molecular Genetics, 2019, Vol. 28, No. 24 4162–4173

doi: 10.1093/hmg/ddz263
Advance Access Publication Date: 6 November 2019
General Article One

GENERAL ARTICLE ONE

Colocalization of GWAS and eQTL signals at loci with multiple signals identifies additional candidate genes for body fat distribution

Ying Wu[1,†‡], K. Alaine Broadaway[1,†], Chelsea K. Raulerson[1], Laura J. Scott[2], Calvin Pan[3], Arthur Ko[3], Aiqing He[4], Charles Tilford[4], Christian Fuchsberger[2,5], Adam E. Locke[2,6], Heather M. Stringham[2], Anne U. Jackson[2], Narisu Narisu[7], Johanna Kuusisto[8], Päivi Pajukanta[3], Francis S. Collins[7], Michael Boehnke[2], Markku Laakso[8], Aldons J. Lusis[3], Mete Civelek[3,9,†] and Karen L. Mohlke[1,†*]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA, [2]Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA, [3]Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA, [4]Bristol–Myers Squibb, Pennington, NJ 08534, USA, [5]Center for Biomedicine, European Academy of Bolzano/Bozen, University of Lübeck, Bolzano/Bozen, Italy, [6]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA, [7]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA, [8]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio 70210, Finland, and [9]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22904, USA

*To whom correspondence should be addressed at: Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. Tel: +1 919 966 2913; Fax: +1 919 966 0401; Email: mohlke@med.unc.edu

Abstract

Integration of genome-wide association study (GWAS) signals with expression quantitative trait loci (eQTL) studies enables identification of candidate genes. However, evaluating whether nearby signals may share causal variants, termed colocalization, is affected by the presence of allelic heterogeneity, different variants at the same locus impacting the same phenotype. We previously identified eQTL in subcutaneous adipose tissue from 770 participants in the Metabolic Syndrome in Men (METSIM) study and detected 15 eQTL signals that colocalized with GWAS signals for waist–hip ratio adjusted for body mass index (WHRadjBMI) from the Genetic Investigation of Anthropometric Traits consortium. Here, we reevaluated evidence of colocalization using two approaches, conditional analysis and the Bayesian test COLOC, and show that providing COLOC with approximate conditional summary statistics at multi-signal GWAS loci can reconcile disagreements in colocalization classification between the two tests. Next, we performed conditional analysis on the METSIM

[†]These authors contributed equally to this work.
[‡]Present address: Pfizer Worldwide Research, Development, and Medical, Cambridge, MA 02139, USA.

Received: June 14, 2019. **Revised:** October 23, 2019. **Accepted:** October 30, 2019

subcutaneous adipose tissue data to identify conditionally distinct or secondary eQTL signals. We used the two approaches to test for colocalization with WHRadjBMI GWAS signals and evaluated the differences in colocalization classification between the two tests. Through these analyses, we identified four GWAS signals colocalized with secondary eQTL signals for *FAM13A*, *SSR3*, *GRB14* and *FMO1*. Thus, at loci with multiple eQTL and/or GWAS signals, analyzing each signal independently enabled additional candidate genes to be identified.

Introduction

Genome-wide association studies (GWAS) have discovered thousands of loci associated with hundreds of complex diseases and related traits (www.ebi.ac.uk/gwas), yet the underlying genes that influence disease susceptibility often remain unknown. One approach to identify causal genes is to map expression quantitative trait loci (eQTL) that contribute to variation in gene expression level (1–3), and then assess evidence of colocalization between overlapping GWAS and eQTL signals—that is, we test whether the same variant(s) is(are) likely responsible for the signals in both studies. Several studies have interrogated available databases/resources for eQTL, and identifying GWAS-colocalized eQTL has enabled identification and interpretation of likely functional genes and potential biological pathways underlying the disease/trait associations (4–10). Previously, we analyzed subcutaneous adipose tissue gene expression using microarrays in 770 participants in the Metabolic Syndrome in Men (METSIM) study and identified novel eQTL colocalized with 109 GWAS loci for cardiometabolic diseases and traits, suggesting new candidate genes mediating the variant associations with cardiometabolic disorders (11).

Allelic heterogeneity, in which more than one genetic variant at the same locus influences the same phenotype, is a common characteristic of complex traits (12). Fine-mapping at GWAS loci routinely identifies many loci with multiple conditionally distinct association signals (defined as signals that remain or become significantly associated with the outcome after modeling the effect of other nearby signals) that increase the proportion of phenotypic variance explained by genetic variation at the locus (13–16). Fine-mapping at eQTL loci also has identified eQTL with multiple conditionally distinct signals (7); (17); (18). Identifying multi-signal loci is becoming more common as sample sizes of eQTL studies increase (14); (19), and testing for colocalization at each signal within a locus will help identify additional candidate genes that contribute to a trait. For example, two eQTL signals were identified in peripheral blood for the gene *FAM117B* at the total cholesterol locus *FAM117B*. After accounting for linkage disequilibrium (LD), only the secondary eQTL signal was colocalized with the total cholesterol GWAS signal (18), emphasizing the utility of conditional analysis.

Body fat distribution is a heritable trait related to cardiometabolic risk (20); (21). One GWAS by the Genetic Investigation of Anthropometric Traits (GIANT) consortium reported 68 conditionally distinct signals at 49 loci (49 primary signals and 19 additional signals after accounting for primary signals) associated with waist-to-hip ratio adjusted for body mass index (WHRadjBMI), a measure of body fat distribution (10). Based on available eQTL resources, the consortium reported that the lead GWAS variants at 21 of these 49 loci were in strong LD ($r^2 > 0.8$) with variants associated with transcript levels in subcutaneous adipose tissue, omental adipose tissue, liver and/or blood (10). Using our subcutaneous adipose eQTL data from METSIM, we reported 15 WHRadjBMI-eQTL colocalized signals, including seven GWAS-colocalized eQTL at six loci that had not been detected by the GIANT consortium (11). However, both analyses were limited to primary eQTL signals.

Here, we extended our analysis of initial, primary subcutaneous adipose tissue eQTL in the METSIM study to identify secondary eQTL signals. We evaluate colocalization of primary and secondary eQTL signals with primary and secondary GWAS signals for WHRadjBMI (10). We identify colocalization by pairwise LD and conditional analysis on the lead GWAS and lead eQTL variants (10); (11), and compare our findings to those obtained using COLOC, a Bayesian test of colocalization (22). The results demonstrate the value of separating signals at eQTL and GWAS loci to identify additional candidate *cis*-regulated genes that may influence disease etiology.

Results

In the METSIM microarray study of subcutaneous adipose tissue, we previously identified primary *cis*-eQTL that showed association between a genetic variant and expression level of at least one gene (FDR < 1%, equivalent to $P < 2.4 \times 10^{-4}$ in genome-scale eQTL mapping) (11). Here, we focused on the lead variants for each of 68 conditionally distinct GWAS signals identified previously at 49 WHRadjBMI loci (10). Of these 68 variants, 40 were associated with expression level of at least one gene within 1 Mb (FDR < 1%), while 28 were not associated with expression level of any gene. Some variants were associated with expression level of more than one gene; the 40 variants were associated with expression level of a total of 71 genes. For each of the 71 genes, we also identified the variant that exhibited the strongest association with expression level, which we denote as the lead eSNP. We further define a 'signal pair' as a lead GWAS variant that is associated with a gene's expression level, and the lead eSNP for that gene. Details on the 71 signal pairs are in Supplementary Material, Table S1.

We then used two strategies to determine whether the GWAS signals were colocalized with the primary eQTL signals for those genes. First, we assessed colocalization through two criteria of LD and conditional analysis (11), requiring both high pairwise LD ($r^2 > 0.8$) between the lead GWAS variant and the lead eSNP, as well as attenuation of the eQTL signal in conditional analysis (conditional $P > 2.4 \times 10^{-4}$, see Materials and Methods). Then, we assessed colocalization using the Bayesian test COLOC, defining colocalization based on a high posterior probability that a single shared variant is responsible for both signals (PP4 > 0.8) (22).

Colocalization of primary eQTL signals with WHRadjBMI GWAS signals

Among the 71 pairs of GWAS and primary eQTL signals, 20 were classified as colocalized by LD and conditional analysis (Table 1). Only 15 of the 20 signal pairs were reported in our previous study (11) due to differences in software to identify eQTL. Similarly, based on LD and subcutaneous adipose tissue eQTL data from other studies (10), the GIANT consortium had described 9 of the 20 signal pairs as colocalized. New colocalized

Table 1. Colocalization of primary adipose eQTL signals with WHRadjBMI GWAS signals

GWAS locus	eQTL gene	GWAS variant association with expression level						Primary eQTL signal						
		GWAS variant	A1/A[a]	EAF	β[b]	P[b]	Lead eSNP	A1/A2[a]	EAF	β[c]	P[c]	LD r^2[d]	Conditional P-value[e]	COLOC (PP4)[f]
Evidence of GWAS and lead eQTL signal colocalization in conditional analysis and COLOC (PP4 > 0.8)														
CMIP	*CMIP*	rs2925979	T/C	0.32	0.44	1.7E–16	rs2925979	T/C	0.32	0.44	1.7E–16	1.0[g]	1.00	1.00
VEGFA	*VEGFA*	rs998584	A/C	0.51	−0.29	2.6E–09	rs998584	A/C	0.51	−0.3	2.6E–09	1.0[g]	1.00	1.00
HOXC4–6	*HOXC4*	rs2071449	A/C	0.43	0.54	3.4E–28	rs2071449	A/C	0.43	0.54	3.4E–28	1.0[g]	1.00	0.99
JUND	*JUND*	rs12608504	A/G	0.35	−0.25	2.2E–06	rs12608504	A/G	0.35	−0.3	2.2E–06	1.0[g]	1.00	0.96
WARS2	*TBX15*	rs2645294	T/C	0.72	0.30	1.1E–07	rs2645294	T/C	0.72	0.30	1.1E–07	1.0[g]	1.00	0.92
CPEB4	*CPEB4*	rs7705502	A/G	0.48	1.2	3.1E–167	rs6861681	A/G	0.48	1.2	3.1E–167	1.00	0.71	0.91
PBRM1	*GNL3*	rs13083798	A/G	0.53	0.42	1.1E–16	rs2590838	G/A	0.53	0.42	8.6E–17	1.00	0.47	0.92
TNFAIP8	*TNFAIP8*	rs1045241	C/T	0.67	−0.48	6.2E–20	rs7703744	C/G	0.67	−0.5	5.7E–20	1.00	0.43	0.97
PBRM1	*NEK4*	rs13083798	A/G	0.53	0.34	1.3E–11	rs2590838	G/A	0.53	0.35	8.4E–12	1.00	0.15	0.91
ADAMTS9	*ADAMTS9*	rs2371767	G/C	0.78	−0.49	1.1E–15	rs4616635	C/G	0.78	−0.5	6.3E–16	0.99	0.32	0.96
MSC	*EYA1*	rs12679556	G/T	0.27	0.46	1.1E–16	rs4738141	G/A	0.27	0.47	3.4E–17	0.98	0.10	0.92
ZNRF3	*ZNRF3*	rs2294239	A/G	0.59	−0.36	2.0E–12	rs12321	G/C	0.59	−0.4	3.2E–13	0.98	0.03	0.83
NT5DC2	*NT5DC2*	rs12489828	T/G	0.54	−0.50	4.0E–24	rs6778735	C/T	0.55	−0.5	1.8E–24	0.97	0.19	0.86
GORAB	*PRRX1*	rs10919388	C/A	0.75	−0.45	4.5E–14	rs6427242	G/C	0.74	−0.5	3.7E–14	0.93	0.25	0.93
LEKR1	*TIPARP*	rs17451107	T/C	0.70	−0.40	5.4E–13	rs10049090	G/A	0.68	−0.4	7.1E–14	0.92	0.05	0.82
Evidence of GWAS and lead eQTL signal colocalization in conditional analysis but not COLOC (PP4 < 0.8)														
ZNF664	*ZNF664*	rs863750	T/C	0.55	−0.63	5.5E–37	rs10773049	T/C	0.55	−0.63	5.5E–37	1.00	0.75	0.22
SNX10	*SNX10*	rs1534696	C/A	0.46	−1.1	3.4E–150	rs1534696	C/A	0.46	−1.1	3.4E–150	1.00	1.00	0.00
SNX10	*CBX3*	rs1534696	C/A	0.46	−0.39	1.1E–13	rs1534696	C/A	0.46	−0.39	1.1E–13	1.00	1.00	0.00
NT5DC2	*C3orf78*	rs12489828	T/G	0.54	−0.44	4.4E–19	rs9853056	C/T	0.55	−0.47	1.2E–20	0.96	0.003	0.66
NKX3	*STC1*	rs7830933	A/G	0.78	0.35	1.7E–08	rs6983481	G/T	0.80	0.38	3.1E–09	0.94	0.08	0.76

GWAS and primary eQTL signals that show significant evidence of colocalization between pairs based on conditional analysis (LD $r^2 > 0.8$ between the lead GWAS variant, conditional $P > 2.4 \times 10^{-4}$) and strength of evidence of colocalization from COLOC (PP4).

[a]A1/A2: the effect/non-effect alleles; the allele associated with increasing WHRadjBMI level was labeled as the effect allele (A1).

[b]The effect size and *P*-value of the association between the lead GWAS variant and the gene expression level.

[c]The effect size and *P*-value of the association between the lead eSNP and the gene expression level.

[d]The pairwise LD r^2 between the lead GWAS variant and the lead eSNP, calculated from the 770 METSIM individuals included in eQTL analysis.

[e]The conditional *P*-value of the association for the lead eSNP after accounting for the lead GWAS variant.

[f]The posterior probabilities computed from COLOC; PP4 > 0.8 suggests that one shared variant is responsible for the GWAS and eQTL signals.

[g]The lead GWAS variant and the lead eSNP are the same variant.

GWAS–eQTL signals identified here include a GWAS signal near *VEGFA* colocalized with an eQTL for *VEGFA*, a GWAS signal at *HOXC4–HOXC6* colocalized with an eQTL for *HOXC4*, and a GWAS signal near *PBRM1* colocalized with eQTL for both *GNL3* and *NEK4* (Table 1).

Of the 20 GWAS–eQTL signal pairs classified as colocalized by LD and conditional analysis, 15 were also classified as colocalized by COLOC (PP4 $\geq$ 0.8), but 5 were not (PP4: 0–0.76, Table 1). COLOC did not identify any additional colocalizations that LD and conditional analysis did not also classify as colocalizations. Two of the five signals COLOC did not classify as colocalized had marginal PP4 values (0.76 at the *NKX3* GWAS locus for an *STC1* eQTL and 0.66 for a secondary GWAS signal at the *NT5DC* locus for a *C3orf78* eQTL). Since prior probabilities can play an important role in the posterior expectations in COLOC (23) and our priors were conservative, we carried out sensitivity analysis to address whether altering the priors could lead to different posterior probabilities. When we increased the prior probability that a shared causal variant influences both WHRadjBMI and gene expression level from 1×10^{-6} to 5×10^{-6}, the PP4 posterior probabilities increased from 0.66 to 0.91 for *C3orf78* and from 0.76 to 0.94 for *STC1*, respectively (Supplementary Material, Table S2). As expected, the 15 colocalized signals discovered with the conservative priors showed stronger Bayesian evidence of colocalization as the priors became less stringent.

Three remaining putative colocalizations (based on LD and conditional analysis) had low PP4 values (PP4 < 0.8) even with more lenient priors. These colocalizations were found at two GWAS loci that each consists of more than one distinct GWAS signal, *CCDC92–ZNF664* and *NFE2L3–SNX10* (Supplementary Material, Table S3) (10); (24). Since the presence of more than one causal signal within a GWAS locus is expected to limit COLOC's power to detect colocalizations (23), we used a tool for Genome-wide Complex Trait Analysis (GCTA) conditional analysis to estimate residual GWAS association statistics for each signal, conditioning on the effect of the other nearby signals (see Materials and Methods) (24). We then provided COLOC the GCTA residual summary statistics, which should mitigate the impact of multiple significant signals in the region. Using GCTA's approximate conditional summary statistics of the GWAS data at these three loci, COLOC identified the same three colocalizations with eQTL signals detected by conditional analysis (Table 2). At the *CCDC92–ZNF664* locus, the secondary (rs863750) GWAS signal was colocalized with the primary eQTL for *ZNF664* (PP4 = 0.98, Table 2, Fig. 1). At the *NFE2L3–SNX10* locus, the secondary (rs1534696) GWAS signal was colocalized with the primary eQTL for both *SNX10* (PP4 = 0.99) and *CBX3* (PP4 = 0.99, Table 2). Overall, COLOC and conditional analysis had high agreement in colocalization classification, with some differences that could be due to the assigned priors in the Bayesian test and/or thresholds to define colocalization.

Given that METSIM study participants are all men, we examined whether six colocalizations between METSIM eQTL signals and WHRadjBMI GWAS signals that had exhibited sex heterogeneity can be detected in other eQTL data from men-only and women-only analyses. We compared evidence of colocalization of the six signals with the lead adipose eQTL signals from men and women in the Diabetes Epidemiology: Collaborative analysis of Diagnostic criteria in Europe (deCODE) study and the female-only TwinsUK Multiple Tissue Human Expression Resource (MuTHER) study (10); (25). As shown in Supplementary Material, Table S4, four of the signal pairs colocalized with the eQTL in women-only data [the lead eSNPs for all four are in strong LD ($r^2 \geq 0.91$) with the lead GWAS variants], showing

Table 2. Adipose eQTL colocalization at loci with multiple GWAS signals

GWAS locus	eQTL gene	GWAS variant association with expression level					Primary eQTL signal						COLOC (PP4)		
		2nd signal lead GWAS variant	A1/A2[a]	EAF	β[b]	P[b]	Lead eSNP	A1/A2[a]	EAF	β[c]	P[c]	LD r^2[d]	Conditional P-value[e]	Original GWAS summary statistics[f]	GCTA residual summary statistics[g]
CCDC92–ZNF664	*ZNF664*	rs863750	T/C	0.55	−0.63	5.5E−37	rs10773049	T/C	0.55	−0.63	5.5E−37	1.00	0.75	0.22	0.98
NFE2L3–SNX10	*SNX10*	rs1534696	C/A	0.46	−1.12	3.4E−150	rs1534696	C/A	0.46	−1.12	3.4E−150	1.0[h]	1.00	0.00	0.99
NFE2L3–SNX10	*CBX3*	rs1534696	C/A	0.46	−1.12	3.4E−150	rs1534696	C/A	0.46	−0.39	1.1E−13	1.0[h]	1.00	0.00	0.99

Discrepancies in colocalization classification between COLOC and conditional analysis are resolved at loci with multiple GWAS signals when COLOC is provided residual summary statistics to account for the effect of nearby signals.

[a]A1/A2: the effect/non-effect alleles; the allele associated with increasing WHRadjBMI level was used as the effect allele (A1).

[b]The effect size and *P*-value of the association between the GWAS variant and the gene expression level.

[c]The effect size and *P*-value of the association between the lead eSNP and the gene expression level.

[d]The pairwise LD r2 between the lead GWAS variant and the lead eSNP, which was calculated from the 770 METSIM individuals included in eQTL analysis.

[e]The conditional *P*-value of the association for the lead eSNP after accounting for the lead GWAS variant.

[f]The posterior probabilities computed from COLOC, using the original GWAS summary statistics from GIANT; PP4 > 0.8 suggests that one shared variant is responsible for the GWAS and eQTL signals.

[g]The posterior probabilities computed from COLOC using the GCTA residual approximate conditional summary statistics of GIANT data; PP4 > 0.8 suggests that one shared variant is responsible for the GWAS and eQTL signals.

[h]The lead GWAS variant and the lead eSNP are the same variant.

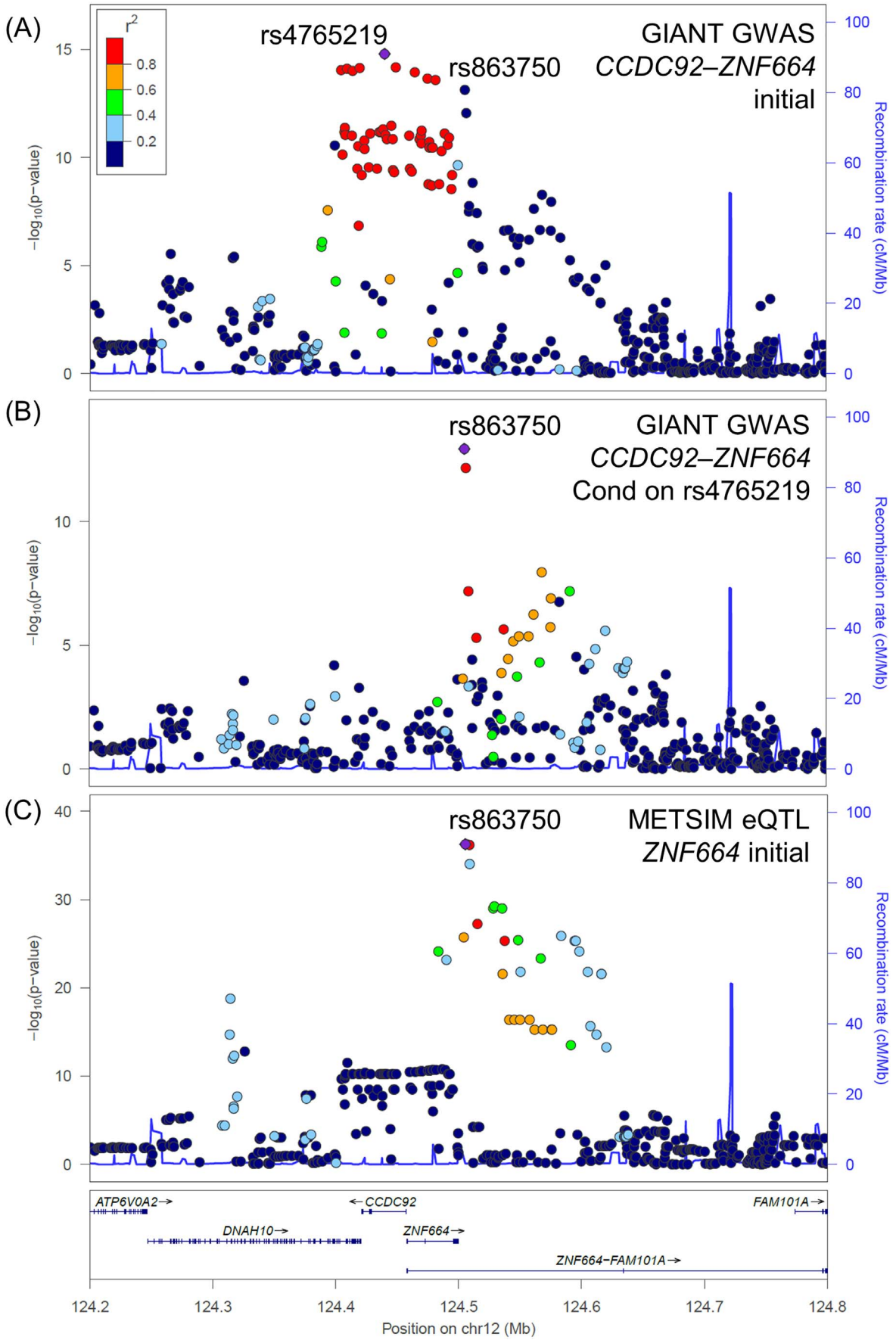

Figure 1. Colocalization of the *ZNF664* eQTL with the 2nd signal (rs863750) of the two independent GWAS signals at *CCDC92–ZNF664*. (**A**) Initial variant association with WHRadjBMI in GIANT. Color indicates LD (r^2) with GWAS variant of 1st signal (rs4765219). (**B**) Residual association with WHRadjBMI after accounting for the effect of the 1st GWAS signal in GIANT; the GWAS variant rs863750 (representing the 2nd signal) was plotted as the lead variant. (**C**) Variant association with ZNF664 expression levels in METSIM. The GWAS variant rs863750 representing the secondary signal was plotted as the lead variant.

that the primary eQTL signals for *TNFAIP8*, *ADAMTS9*, *SNX10* and *CBX3* are not sex-specific. Our results are consistent with existing literature: previously, WHRadjBMI signals have shown largely similar evidence of colocalization with adipose eQTL signals from men and women, even for GWAS loci with significant evidence of sex heterogeneity (10).

Colocalization of secondary eQTL signals with WHRadjBMI GWAS signals

We next tested for secondary eQTL signals at the 71 genes that were associated with WHRadjBMI GWAS signals. We performed association analyses with gene expression level while including the lead eSNP as a covariate in the regression model, and defined the new variant that exhibited the strongest association in the conditional analysis as the secondary eSNP. We restricted the analysis to secondary signals, rather than testing for further signals (that is, tertiary and beyond), due to limited power to detect these smaller effects.

After adjusting for the lead eSNP, lead variants for 37 conditionally distinct GWAS signals were associated with expression level of 51 genes (FDR < 1%, $P < 2.4 \times 10^{-4}$) (Supplementary Material, Table S5). Conditional analysis classified four GWAS signals as colocalized with a secondary eQTL signal, but not the primary eQTL signal for the following four genes: the *FAM13A* GWAS signal was colocalized with the secondary eQTL for *FAM13A* ($r^2 = 1.00$, conditional $P \geq 0.37$), the *GRB14–COBLL1* GWAS signal was colocalized with the secondary eQTL for *GRB14* ($r^2 = 0.93$, conditional $P \geq 0.39$), the *LEKR1* GWAS signal was colocalized with the secondary eQTL for *SSR3* ($r^2 = 0.94$, conditional $P \geq 0.43$) and the *GORAB* GWAS signal was colocalized with the secondary eQTL for *FMO1* ($r^2 = 1.00$, conditional $P = 1.00$) (Table 3). As with colocalization of primary eQTL signals, COLOC did not identify additional signals as colocalized that were not also discovered by LD and conditional analysis. COLOC did classify the first three GWAS signals as colocalized with the secondary eQTL (PP4 ≥ 0.92); at the *GORAB* locus, the PP4 value very narrowly missed the colocalization classification threshold (PP4 = 0.79). We next describe these four loci in further detail.

At the *FAM13A* WHRadjBMI locus, the lead GWAS variant rs9991328 was associated with the expression level of *FAM13A* ($P = 1.0 \times 10^{-32}$ in primary eQTL analysis, Table 3 and Fig. 2). While COLOC found suggestive evidence for colocalization between the lead GWAS variant and the lead eSNP, rs10024506 (PP4 = 0.73), the two variants were in very weak LD ($r^2 = 0.02$). After controlling for the effect of the *FAM13A* lead eSNP (rs10024506), we identified a secondary eQTL signal represented by rs2290782 (eQTL $P_{uncond} = 1.0 \times 10^{-32}$, $P_{cond} = 2.6 \times 10^{-29}$ when adjusting for lead eSNP rs10024506). In contrast to the primary eSNP, the secondary eSNP was in strong LD with the lead GWAS variant ($r^2 = 1.00$), and both COLOC and conditional analysis provided strong evidence of colocalization (COLOC PP4 = 1.00, both conditional $P > 0.37$). Our findings suggest that *FAM13A* might be a functional gene mediating the genetic influence of this GWAS locus on body fat distribution. This conclusion is bolstered by recent studies that linked the *FAM13A* and *Fam13a* genes to adipose morphology and adipose tissue function in human and mice (26).

At the *GRB14–COBLL1* WHRadjBMI locus, the lead GWAS variant rs10195252 was associated with the expression level of *GRB14* ($P = 3.8 \times 10^{-6}$ in primary eQTL analysis), but neither LD nor COLOC supported colocalization with the primary eQTL signal rs1474249 (PP4 = 0.25, LD $r^2 = 0.00$) (Table 1; Supplementary Material, Fig. S1). After controlling for the effect of the primary eSNP,

Table 3. Colocalization of secondary adipose eQTL with WHRadjBMI GWAS loci

GWAS locus	eQTL gene	GWAS variant association with expression level					Primary and secondary eQTL signals								
		GWAS variant	A1/A2[a]	EAF	β[b]	P[b]	Lead eSNP	A1/A2[a]	EAF	β[c]	P[c]	β cond	P cond[d]	LD r^2[e]	Conditional P-value[f]	COLOC (PP4)[g]
FAM13A	*FAM13A*	rs9991328	T/C	0.53	0.57	1.00E–32	rs10024506 (1st)	G/C	0.64	0.62	3.1E–34	-	-	0.02	3.2E–31	0.73
							rs2290782 (2nd)	C/T	0.53	0.57	1.0E–32	0.50	2.6E–29	1.00	0.37	0.95
GRB14–COBLL1	*GRB14*	rs10195252	T/C	0.68	0.26	3.80E–06	rs1474249 (1st)	C/T	0.73	0.31	4.1E–08	-	-	0.00	1.4E–07	0.25
							rs1128249 (2nd)	G/T	0.70	0.26	5.5E–06	0.25	4.7E–06	0.93	0.72	0.92
LEKR1	*SSR3*	rs17451107	T/C	0.70	−0.36	3.90E–11	rs10936027 (1st)	G/T	0.14	−0.65	3.4E–18	-	-	0.02	5.2E–16	0.00
							rs900400 (2nd)	T/C	0.30	−0.37	2.9E–11	−0.31	4.2E–09	0.98	0.43	0.94
GORAB	*FMO1*	rs10919388	C/A	0.75	0.29	1.40E–06	rs16864302 (1st)	A/G	0.88	0.50	4.9E–11	-	-	0.02	1.3E–09	0.00
							rs4471313 (2nd)	T/G	0.75	0.29	1.4E–06	0.24	4.0E–05	1.00	1	0.79

Additional eQTL associations for genes identified by secondary eQTL analysis of GIANT WHRadjBMI loci. GWAS and secondary eQTL signals that show significant evidence of colocalization.
[a]A1/A2: the effect/non-effect alleles; the allele associated with increasing WHRadjBMI level was used as the effect allele (A1).
[b]The effect size and P-value of the association between the GWAS variant and the gene expression level.
[c]The effect size and P-value of the association between the lead eSNP (1st or 2nd eQTL signal) and the gene expression level.
[d]The effect size and P-value of the residual association between the secondary eQTL signal and gene expression level when conditional on the primary eQTL signal (lead eSNP).
[e]The pairwise LD r^2 between the lead GWAS variant and the eSNP representing the 1st or 2nd eQTL signal, which was calculated from the 770 METSIM individuals included in eQTL analysis.
[f]The conditional P-value of the association for the lead eSNP after accounting for the lead GWAS variant.
[g]The posterior probabilities computed from COLOC; PP4 > 0.8 suggests that one shared variant is responsible for the GWAS and eQTL signals.

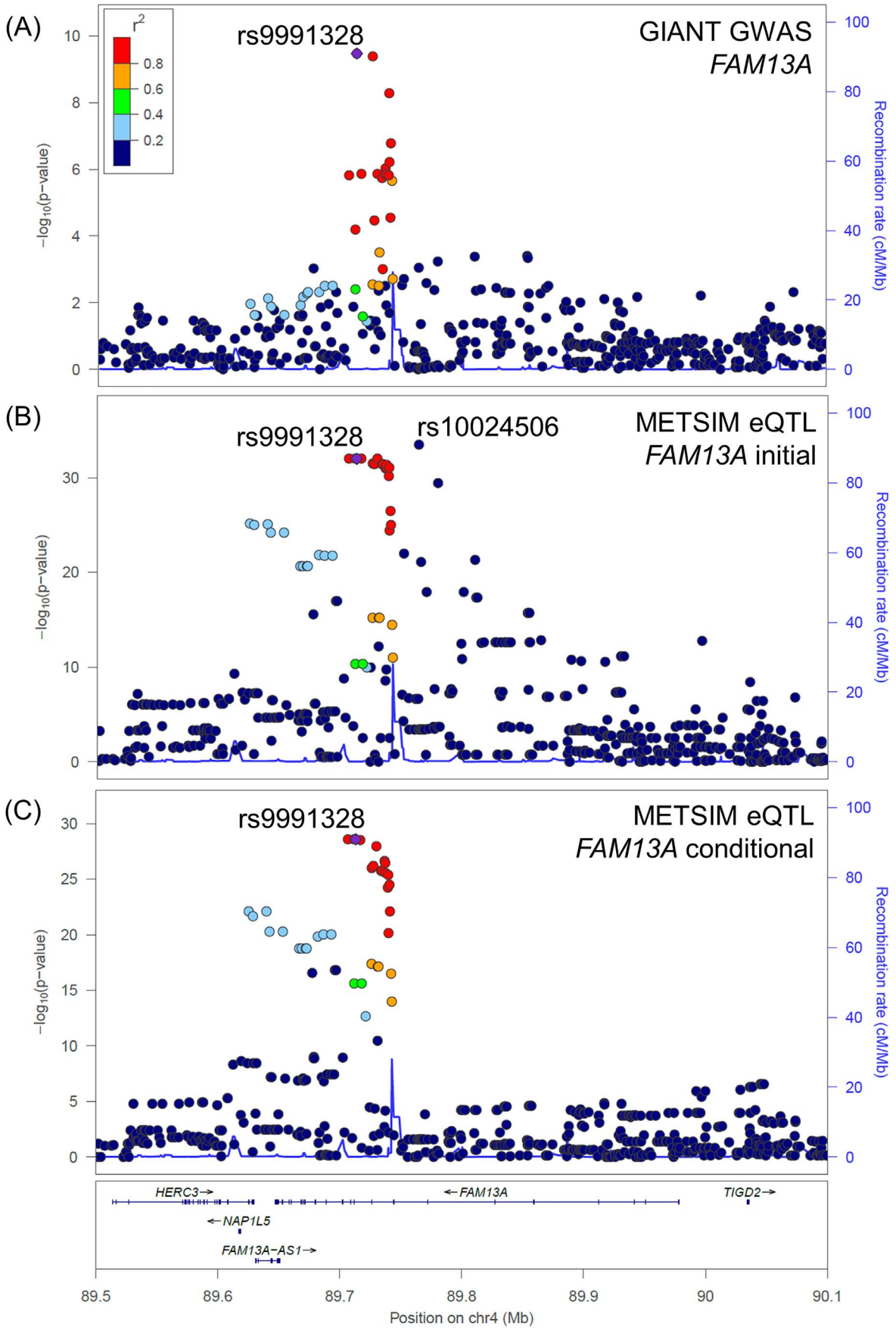

Figure 2. Colocalization of GWAS locus *FAM13A* with the secondary but not the primary eQTL signal for *FAM13A* gene. (**A**) Regional variant association with WHRadjBMI in GIANT, lead GWAS variant rs9991328; (**B**) Initial association with expression level of *FAM13A*; and (**C**) residual eQTL association after accounting for the lead eSNP rs10024506 in METSIM. The GWAS variant rs9991328 was plotted as the lead variant for all plots.

we observed a secondary eQTL signal for *GRB14* at rs1128249 ($P_{cond} = 4.7 \times 10^{-6}$, Table 3). The secondary eSNP was in high LD with the lead GWAS variant ($r^2 = 0.93$), and both COLOC and conditional analysis provided strong evidence of colocalization (COLOC PP4 = 0.92, conditional $P = 0.72$). To further explore the relationship between *GRB14* and WHRadjBMI, we tested for the association between cardiometabolic traits and gene expression level in METSIM. While higher expression level of *GRB14* was not significantly associated with WHRadjBMI ($P = 0.07$), it was significantly associated with several related traits, including higher fasting plasma insulin ($P = 5.9 \times 10^{-6}$), higher HOMA-β ($P = 1.1 \times 10^{-5}$), lower Matsuda index ($P = 1.4 \times 10^{-4}$) and higher fasting plasma proinsulin ($P = 1.5 \times 10^{-4}$) (Supplementary Material, Table S6). These findings are consistent with previous observations of improved glucose homeostasis and enhanced insulin signaling in *Grb14*-deficient mice (27), and prioritize *GRB14* as a candidate gene potentially mediating the WHRadjBMI association at this locus.

The third and fourth examples of WHRadjBMI GWAS signals that colocalized with secondary eQTL signals identified different genes than those that colocalized with primary eQTL signals. The GWAS signal at *LEKR1* (lead variant rs17451107) colocalized with the primary eQTL signal for *TIPARP* (Table 1; Supplementary Material, Fig. S2), as well as the secondary eQTL signal for *SSR3*, located ~500 kb away (PP4 = 0.94; LD $r^2 = 0.98$; Table 3; Supplementary Material, Fig. S2). Similarly, the GWAS signal at *GORAB* (lead variant rs10919388) colocalized with a primary eQTL at *PRRX1* (Table 1; Supplementary Material, Fig. S3) and with the secondary eQTL for *FMO1* (PP4 = 0.79; LD $r^2 = 1.00$; Table 3; Supplementary Material, Fig. S3). These results suggest that these GWAS loci may be mediated through altered expression levels of either or both genes.

Discussion

We show examples of how colocalization between GWAS and eQTL signals can be influenced by the presence of multiple GWAS signals at a locus or multiple eQTL signals for the same gene. In our study of 49 GWAS loci for WHRadjBMI and primary eQTL from subcutaneous adipose tissue from the METSIM study, we describe 20 colocalized signal pairs. At two loci with multiple GWAS signals, COLOC initially was unable to identify signal pairs as colocalized, despite complete LD between the secondary lead GWAS variant and lead eSNP ($r^2 = 1.00$). Because the presence of multiple signals violates COLOC's assumptions and likely reduces power to detect a true colocalization, we provided the program with estimated residual GWAS association summary statistics, conditioning on the neighboring signal. Using approximate conditional summary statistics of the GWAS data, COLOC identified the signal pairs as colocalized. In addition, analyses of secondary eQTL signals in the METSIM study identified four colocalized eQTL that were not detected in analyses of primary eQTL signals. At loci with multiple eQTL and/or GWAS signals, comparing the signals separately after conditional analysis led to more robust evidence of colocalization. Dissecting the allelic heterogeneity provided insight into how GWAS loci might influence WHRadjBMI through gene expression.

At least three of the genes detected using secondary eQTL signals, *FAM13A*, *GRB14* and *FMO1*, have other evidence suggesting that they may influence WHRadjBMI. *FAM13A* expression level increases during adipocyte differentiation and is associated with adipocyte hyperplasia, consistent with alleles associated with both higher WHRadjBMI and higher *FAM13A* (26). *Grb14*-deficient mice showed improved glucose homeostasis and enhanced insulin signaling (27), and *Fmo1*-deficient mice were leaner and stored fewer triglycerides in white adipose tissue than wild-type mice (28), both consistent with our observations that alleles associated with higher expression level are associated with higher WHRadjBMI (Table 3). The secondary eQTL signal for *FMO1* is colocalized with the same GWAS signal as a primary eQTL signal for *PRRX1*, which has been shown to inhibit adipogenesis (29), suggesting that the GWAS signal may act through both genes to influence WHRadjBMI. The allele associated with higher WHRadjBMI was associated with lower *PRRX1*, corresponding to more adipogenesis and higher *FMO1*, corresponding to more storage of triglyceride in adipose (28), (29). The fourth gene detected using a secondary eQTL signal, *SSR3*, contributes to the formation of a vascular network in murine placenta (30), and may reflect the blood component of adipose tissue; this GWAS signal colocalized with a primary eQTL for *TIPARP*, which positively regulates liver X receptor, which can impair adipose expansion (31). While colocalized GWAS and eQTL signals do not provide causal mechanisms, they can suggest candidate genes for further investigation.

Although the results of COLOC and the LD and conditional analysis approach were largely concordant, conditional analysis did identify five GWAS/eQTL pairs as colocalized that COLOC initially did not (Table 1). Two GWAS/eQTL variant pairs that were classified as colocalized only by conditional analysis had marginal COLOC PP4 probabilities (PP4 0.66 for *C3orf78* and 0.76 for *STC1*). However, we expect that the priors we selected for COLOC will be conservative; when we increased the prior probability that a variant is causal to both GWAS and eQTL to COLOC's default, PP4 posterior probabilities increased sufficiently for COLOC to also classify the pairs as colocalized (PP4 0.95 and 0.97, respectively). The remaining three pairs of GWAS/eQTL signals with inconsistent results between COLOC and conditional analysis were secondary GWAS signals at multi-signal loci. Since COLOC assumes that the trait is associated with at most one causal variant per locus (22); (23), the presence of multiple association signals could lead to missed colocalized signals. After accounting for multiple signals by using GCTA, COLOC also classified the three GWAS/eQTL variant pairs as colocalized. While COLOC's conclusions in this study ultimately align with LD and conditional analysis, our findings do highlight the care needed to properly implement COLOC such that important colocalizations are not missed. Overall, COLOC and conditional analysis had high agreement in colocalization classification; differences can be attributed to be unaccounted for multiple signals per locus or can be reconciled through changes to assigned priors.

Our ability to colocalize signals might have been affected by the limitations of this study. First, the GWAS loci were identified by GIANT based on HapMap-imputed genotypes (10). If the METSIM lead eSNP or its LD proxies imputed from the higher density Haplotype Reference Consortium (HRC) reference panel better represented an underlying signal, we might fail to capture the colocalized signals. Compared to the HapMap Project, more recent studies have expanded the coverage of human genetic variation (32) and enhanced the ability of GWAS to fine-map complex traits (33). Second, we expect to have missed some colocalized signals due to the statistical power of our eQTL analysis. Although the METSIM eQTL study ($n = 770$) had a reasonable sample size to identify initial eQTL signals (11), larger studies will better differentiate between variants in moderate LD with each other (34) and better detect allelic heterogeneity at eQTL. With the increasing availability of eQTL studies, the integration of GWAS data with eQTL results from larger studies or meta-

analyses of multiple eQTL datasets (35); (36) will increase the opportunity to detect additional GWAS-relevant eQTL. Third, we were unable to assess potential sexual dimorphism on gene expression level and could have missed some colocalizations because all samples analyzed in METSIM were from men. While the GIANT data demonstrated that 19 of the 49 WHRadjBMI loci had stronger genetic effects in women, 5 of these 19 loci showed evidence of colocalized eQTL in men in METSIM, and none of the remaining loci exhibited evidence of *cis*-eQTL in the MuTHER eQTL study of women (36), suggesting that these loci do not display strong sex-specific effects on gene expression levels. The credibility of the colocalizations reported at these 5 GWAS loci is bolstered by the observation that many variants which exhibit sex heterogeneity in WHRadjBMI, including effects observed exclusively in women, have a similar effect on body fat percentage in men as well as women (37). Fourth, the effect of an eQTL can vary across tissue and time (35); (38). Identification of colocalizing signals may be dependent on measuring expression at the appropriate time and in a trait-relevant tissue. Fifth, signal colocalization is dependent on haplotype structure. Haplotypes can differ by ancestry, even between individuals of broad European descent and specifically from Finland, and may result in false negative and false positive colocalizations.

Identifying additional signals in eQTL data has the potential to reveal previously undiscovered colocalizations with GWAS loci. However, as we demonstrated, care must be taken when either the GWAS or the eQTL study have multiple signals within a locus. Conditional analyses appear to separate multi-signal loci well, although additional assessments using simulated data are warranted. Other analytic methods, such as COLOC, might fail to detect a colocalization when more than one signal is present in the region. However, providing COLOC with residual statistics that account for the effects of other signals within a locus might be a solution. As eQTL studies grow in size, the ability to detect allelic heterogeneity will increase, thus complicating tests of colocalization. Testing for colocalization at every distinct signal, and selection of colocalization analytic pipelines that are robust to the presence of multiple signals can reveal colocalizations we otherwise would miss.

Materials and Methods

GIANT consortium data for WHRadjBMI

We obtained GIANT consortium 2015 GWAS results for WHRadjBMI (10) from www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. The downloaded files included dbSNP name, effect/non-effect alleles (an effect allele is the WHRadjBMI-increasing allele in the sex-combined analysis), effect allele frequency, beta, standard error, *P*-value and sample size for each variant. At the 29 loci with no evidence of sexual dimorphism, we used GIANT association statistics from the sex-combined meta-analysis. For the locus *GDF5*, which showed a male-specific effect on WHRadjBMI, we used the male-only GWAS results. At the WHRadjBMI loci (*PLXND1*, *NMU*, *FAM13A*, *MAP3K1*, *HMGA1*, *NKX2–6*, *SFXN2*, *MACROD1–VEGFB*, *CMIP*, *BCL2*, *SNX10*, *LYPLAL1*, *GRB14–COBLL1*, *PPARG*, *ADAMTS9*, *TNFAIP8*, *VEGFA* and *RSPO3*) with significantly larger genetic estimated effects on trait variation in women than men, we used the association results from the female-only meta-analysis. We used the results from the European-ancestry meta-analysis for all loci except the locus *SNX10*, for which the all-ancestry meta-analysis data were used because *SNX10* achieved genome-wide significance only in the all-ancestry analysis, with no evidence of heterogeneity across ancestries (10).

METSIM subcutaneous adipose eQTL data

In Civelek *et al.* (11), we described in detail the subcutaneous adipose eQTL data from the METSIM study. Briefly, METSIM is a population-based study of 10 197 men, aged from 45 to 73 years at time of enrollment, randomly selected from the population register of Kuopio, Eastern Finland, and examined in 2005–2010 (39). The University of Kuopio and Kuopio University Hospital ethics committee approved the study, and all study participants provided informed consent. We used the Illumina HumanOmniExpress BeadChip and the Illumina HumanCoreExome array to obtain genotypes, and imputed based on haplotypes from the HRC (version 1) (www.haplotype-reference-consortium.org/) (40) by using Minimac3 on the Michigan Imputation Server (imputationserver.sph.umich.edu/index.html). A total of 7.8 million imputed variants passed the eQTL mapping inclusion criteria of MAF ≥ 0.01 and imputation quality $r^2 > 0.3$.

As previously described, we isolated total ribonucleic acid (RNA) from subcutaneous adipose tissue from 770 METSIM participants, and used a Robust Multi-Array Average (RMA) approach to normalize expression data measured using the Affymetrix Human Genome U219 Array for 43 145 probesets corresponding to 18 250 unique genes (11). We applied probabilistic estimation of expression residuals (PEER) (41) to account for complex non-genetic factors influencing gene expression levels. Based on METSIM adipose eQTL PEER factor observations described previously (11), and to better match the previous results, we adjusted for 35 PEER-inferred confounding factors. We then inverse normal transformed the PEER-processed residuals.

We define a *cis*-eQTL as an eQTL located within 1 Mb of a gene transcript. Association tests for *cis*-eQTL were carried out for variant–probeset pairs with a distance between the variant and the closer boundary of the gene <1 Mb using EPACTS-multi, in which EMMAX accounted for family relatedness (genome.sph.umich.edu/wiki/EPACTS) (42). We selected an FDR < 0.01 (equivalent $P < 2.4 \times 10^{-4}$) as the significance threshold for a *cis*-eQTL and defined the lead eSNP as the variant for which the association with gene expression level resulted in the smallest *P*-value for that gene.

Conditional analyses on GWAS and eQTL data

For METSIM adipose eQTL data, we carried out conditional analyses by including the lead eSNP in the linear regression model then testing for evidence that other variants are associated with that gene's expression level.

For GIANT WHRadjBMI loci identified as containing more than one association signal by the GCTA joint model (10); (24), we used GCTA to run approximate conditional analyses on the GIANT GWAS summary statistics, using LD data from HapMap-imputed METSIM genotype data on 10 070 Finnish men (24); (43). This method estimates residual association statistics after conditioning on the lead GWAS variant in the region, allowing for identification and effect size estimation of secondary signals. Approximate analysis based on summary statistics was required because individual-level data were not available.

Colocalization of GWAS and eQTL signals

First, we assessed the relationship between gene expression level and the lead GWAS variants associated with WHRadjBMI provided by the GIANT consortium (10). Next, for GWAS variants

that were associated with gene expression level (at FDR < 1%, $P < 2.4 \times 10^{-4}$), we tested whether the GWAS variant was colocalized with the lead eSNP. To do so, we extracted the summary statistics for variants located within 1 Mb of the lead GWAS variants from the GIANT 2015 GWAS results for WHRadjBMI and the METSIM subcutaneous adipose eQTL dataset.

We first tested for colocalization using a conditional analysis approach similar to that implemented in Civelek *et al.* (11) Specifically, we calculated pairwise LD r^2 between the lead GWAS and the lead eSNP that had the strongest evidence of association with the corresponding probesets. LD estimates were calculated from the HRC-imputed genotypes of the 770 METSIM individuals. For variant pairs with LD $r^2 > 0.8$, we examined the changes of the eQTL association for the lead eSNP when conditioned on the lead GWAS variant. Following Civelek *et al.* (11), we applied two criteria and defined GWAS-colocalized eQTL by requiring lead variant pairwise $r^2 > 0.8$ and change in the eSNP *P*-value to be no longer significant after conditional analysis ($P > 2.4 \times 10^{-4}$ corresponding to FDR > 0.01) for the lead eSNP. Of note, the LD criterion helps prevent signals from being erroneously defined as colocalized based on small variation around the threshold.

We compared results to those obtained using a Bayesian test for colocalization, COLOC (22); (23). We applied COLOC using the Approximate Bayes Factor computations on the intersection of variants available in both the GIANT WHRadjBMI and METSIM eQTL datasets. We used default priors that a random variant in the region is associated with either GWAS or eQTL individually (prior probabilities = 1×10^{-4}), and set the prior probability that the random variant is causal to both GWAS and eQTL (prior probability = 1×10^{-6}). We selected a more conservative prior probability than COLOC's default for this last scenario (default prior = 1×10^{-5}) because we treat COLOC as a confirmatory test of results discovered by conditional analysis. However, in sensitivity analyses, we raised the prior probability that a random variant is causal to both GWAS and eQTL from 1×10^{-6} to 5×10^{-6} and 1×10^{-5}. As recommended by the authors of the method, we defined the variants as colocalized when the posterior probability of a colocalized signal (PP4) was >0.8. Bayesian colocalization analyses were conducted by using the R package 'coloc' (cran.r-project.org/web/packages/coloc) (22).

Association of expression level with cardiometabolic traits

We conducted regression analyses to evaluate the association for gene expression level with 16 cardiometabolic traits as follows: waist–hip ratio, Matsuda index, insulin, BMI, HOMA-β, proinsulin, triglycerides, total fatty acids, waist circumference, fat-free mass, free fatty acids, total cholesterol, glucose, LDL-C, HDL-C and adiponectin in up to 770 METSIM participants. Of the 770 participants, 27 had type 2 diabetes at their baseline visit; diabetics were included in all regression analyses. The RMA-normalized expression levels were inverse normal transformed after accounting for 35 PEER-inferred factors. All cardiometabolic-related traits were adjusted for age and BMI before inverse normal transformation except BMI, which was only adjusted for age. Traits were adjusted for BMI to be comparable to recent GWAS analyses of cardiometabolic traits (10); (44)–(46).

Supplementary Material. Supplementary Material is available at *HMG* online.

Conflict of Interest statement

A.H. and C.T. are employees and shareholders of Bristol–Myers Squibb.

Funding

National Institutes of Health (R01DK093757, U01DK105561, R01DK072193, R00HL121172, P01HL28481, U01DK062370, R01DK098032, 1-ZIA-HG000024, R01HL095056, F31HL127921, T32HL129982, T32GM067553); Academy of Finland (77299, 124243); Finnish Heart Foundation; Finnish Diabetes Foundation; Finnish Funding Agency for Technology and Innovation (1510/31/06); Commission of the European Community (HEALTH-F2-2007-201681).

Acknowledgements

We thank the participants of the METSIM study, Peter Chines for contributions to genotyping, Niyas Saleem for RNA isolation and Peter Gargalovic and Todd Kirchgessner for contributions in generating the expression data. Bristol–Myers Squibb supported the generation of METSIM microarray data.

References

1. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. and Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, **10**, 184–194.
2. Cheung, V.G. and Spielman, R.S. (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat. Rev. Genet.*, **10**, 595–604.
3. Nica, A.C. and Dermitzakis, E.T. (2008) Using gene expression to investigate the genetic basis of complex disorders. *Hum. Mol. Genet.*, **17**, R129–R134.
4. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. *et al.* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.
5. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
6. Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I. *et al.* (2014) Genome-wide *trans*-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.*, **46**, 234–244.
7. Dobbyn, A., Huckins, L.M., Boocock, J., Sloofman, L.G., Glicksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y. *et al.* (2018) Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. *Am. J. Hum. Genet.*, **102**, 1169–1184.
8. Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M. *et al.* (2018) GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat. Commun.*, **9**, 5141.
9. Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J. *et al.* (2019) Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.*, **51**, 230–236.

10. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E. *et al.* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, **518**, 187–196.
11. Civelek, M., Wu, Y., Pan, C., Raulerson, C.K., Ko, A., He, A., Tilford, C., Saleem, N.K., Stancakova, A., Scott, L.J. *et al.* (2017) Genetic regulation of adipose gene expression and cardio-metabolic traits. *Am. J. Hum. Genet.*, **100**, 428–443.
12. Hormozdiari, F., Zhu, A., Kichaev, G., Ju, C.J., Segre, A.V., Joo, J.W.J., Won, H., Sankararaman, S., Pasaniuc, B., Shifman, S. *et al.* (2017) Widespread allelic heterogeneity in complex traits. *Am. J. Hum. Genet.*, **100**, 789–802.
13. Wu, Y., Waite, L.L., Jackson, A.U., Sheu, W.H., Buyske, S., Absher, D., Arnett, D.K., Boerwinkle, E., Bonnycastle, L.L., Carty, C.L. *et al.* (2013) *Trans*-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet.*, **9**, e1003379.
14. Horikoshi, M., Mgi, R., van de Bunt, M., Surakka, I., Sarin, A.P., Mahajan, A., Marullo, L., Thorleifsson, G., Hgg, S., Hottenga, J.J. *et al.* (2015) Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. *PLoS Genet.*, **11**, e1005230.
15. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hagg, S., Athanasiu, L. *et al.* (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.*, **51**, 404–413.
16. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M. *et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.*, **27**, 3641–3649.
17. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
18. Jansen, R., Hottenga, J.J., Nivard, M.G., Abdellaoui, A., Laport, B., de Geus, E.J., Wright, F.A., Penninx, B. and Boomsma, D.I. (2017) Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.*, **26**, 1444–1451.
19. Wood, A.R., Hernandez, D.G., Nalls, M.A., Yaghootkar, H., Gibbs, J.R., Harries, L.W., Chong, S., Moore, M., Weedon, M.N., Guralnik, J.M. *et al.* (2011) Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Hum. Mol. Genet.*, **20**, 4082–4092.
20. Pischon, T., Boeing, H., Hoffmann, K., Bergmann, M., Schulze, M.B., Overvad, K., van der Schouw, Y.T., Spencer, E., Moons, K.G., Tjonneland, A. *et al.* (2008) General and abdominal adiposity and risk of death in Europe. *N. Engl. J. Med.*, **359**, 2105–2120.
21. Yusuf, S., Hawken, S., Ounpuu, S., Bautista, L., Franzosi, M.G., Commerford, P., Lang, C.C., Rumboldt, Z., Onen, C.L., Lisheng, L. *et al.* (2005) Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet*, **366**, 1640–1649.
22. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
23. Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A. and Wallace, C. (2015) Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.*, **24**, 3305–3313.
24. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
25. Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K. *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.*, **7**, e1002003.
26. Dahlman, I., Ryden, M., Brodin, D., Grallert, H., Strawbridge, R.J. and Arner, P. (2016) Numerous genes in loci associated with body fat distribution are linked to adipose function. *Diabetes*, **65**, 433–437.
27. Cooney, G.J., Lyons, R.J., Crew, A.J., Jensen, T.E., Molero, J.C., Mitchell, C.J., Biden, T.J., Ormandy, C.J., James, D.E. and Daly, R.J. (2004) Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. *EMBO J.*, **23**, 582–593.
28. Veeravalli, S., Omar, B.A., Houseman, L., Hancock, M., Gonzalez Malagon, S.G., Scott, F., Janmohamed, A., Phillips, I.R. and Shephard, E.A. (2014) The phenotype of a flavin-containing monooyxgenase knockout mouse implicates the drug-metabolizing enzyme FMO1 as a novel regulator of energy balance. *Biochem. Pharmacol.*, **90**, 88–95.
29. Du, B., Cawthorn, W.P., Su, A., Doucette, C.R., Yao, Y., Hemati, N., Kampert, S., McCoin, C., Broome, D.T., Rosen, C.J. *et al.* (2013) The transcription factor paired-related homeobox 1 (Prrx1) inhibits adipogenesis by activating transforming growth factor-beta (TGFbeta) signaling. *J. Biol. Chem.*, **288**, 3036–3047.
30. Yamaguchi, Y.L., Tanaka, S.S., Oshima, N., Kiyonari, H., Asashima, M. and Nishinakamura, R. (2011) Translocon-associated protein subunit Trap-gamma/Ssr3 is required for vascular network formation in the mouse placenta. *Dev. Dyn.*, **240**, 394–403.
31. Dong, Y., Xu, Z., Zhang, Z., Yin, X., Lin, X., Li, H. and Zheng, F. (2017) Impaired adipose expansion caused by liver X receptor activation is associated with insulin resistance in mice fed a high-fat diet. *J. Mol. Endocrinol.*, **58**, 141–154.
32. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
33. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C. *et al.* (2015) A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.*, **47**, 1121–1130.
34. Spain, S.L. and Barrett, J.C. (2015) Strategies for fine-mapping complex traits. *Hum. Mol. Genet.*, **24**, R111–R119.
35. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
36. Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A. *et al.* (2012) Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.*, **44**, 1084–1089.
37. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y. *et al.* (2019) Meta-analysis of genome-wide association studies

for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.*, **28**, 166–174.
38. Bryois, J., Buil, A., Ferreira, P.G., Panousis, N.I., Brown, A.A., Vinuela, A., Planchon, A., Bielser, D., Small, K., Spector, T. *et al.* (2017) Time-dependent genetic effects on gene expression implicate aging processes. *Genome Res.*, **27**, 545–552.
39. Stancakova, A., Javorsky, M., Kuulasmaa, T., Haffner, S.M., Kuusisto, J. and Laakso, M. (2009) Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes*, **58**, 1212–1221.
40. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
41. Stegle, O., Parts, L., Piipari, M., Winn, J. and Durbin, R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.*, **7**, 500–507.
42. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C. and Eskin, E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.*, **42**, 348–354.
43. Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusis, A.J., Collins, F.S., Mohlke, K.L. and Boehnke, M. (2017) The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid Res.*, **58**, 481–493.
44. Walford, G.A., Gustafsson, S., Rybin, D., Stancakova, A., Chen, H., Liu, C.T., Hong, J.Y., Jensen, R.A., Rice, K., Morris, A.P. *et al.* (2016) Genome-wide association study of the modified Stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes*, **65**, 3200–3211.
45. Strawbridge, R.J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J.R., Travers, M.E., Bouatia-Naji, N., Dimas, A.S. *et al.* (2011) Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes*, **60**, 2624–2634.
46. Zillikens, M.C., Demissie, S., Hsu, Y.H., Yerges-Armstrong, L.M., Chou, W.C., Stolk, L., Livshits, G., Broer, L., Johnson, T., Koller, D.L. *et al.* (2017) Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. *Nat. Commun.*, **8**, 80.