Journal of Human Genetics (2017) 62, 175–184

www.nature.com/jhg

ORIGINAL ARTICLE

Genome-wide linkage and association analysis of cardiometabolic phenotypes in Hispanic Americans

Jacklyn N Hellwege[1,2], Nicholette D Palmer[1,2,3,4,5], Latchezar Dimitrov[1], Jacob M Keaton[1,2,5], Keri L Tabb[1,2,3], Satria Sajuthi[4,5,6], Kent D Taylor[7], Maggie CY Ng[1,2,4], Elizabeth K Speliotes[8,9], Gregory A Hawkins[1,4], Jirong Long[10], Yii-Der Ida Chen[7], Carlos Lorenzo[11], Jill M Norris[12], Jerome I Rotter[7], Carl D Langefeld[4,5,6], Lynne E Wagenknecht[4,13] and Donald W Bowden[1,2,3]

Linkage studies of complex genetic diseases have been largely replaced by genome-wide association studies, due in part to limited success in complex trait discovery. However, recent interest in rare and low-frequency variants motivates re-examination of family-based methods. In this study, we investigated the performance of two-point linkage analysis for over 1.6 million single-nucleotide polymorphisms (SNPs) combined with single variant association analysis to identify high impact variants, which are both strongly linked and associated with cardiometabolic traits in up to 1414 Hispanics from the Insulin Resistance Atherosclerosis Family Study (IRASFS). Evaluation of all 50 phenotypes yielded 83 557 000 LOD (logarithm of the odds) scores, with 9214 LOD scores $\geqslant 3.0$, 845 $\geqslant 4.0$ and 89 $\geqslant 5.0$, with a maximal LOD score of 6.49 (rs12956744 in the *LAMA1* gene for tumor necrosis factor-α (TNFα) receptor 2). Twenty-seven variants were associated with $P<0.005$ as well as having an LOD score >4, including variants in the *NFIB* gene under a linkage peak with TNFα receptor 2 levels on chromosome 9. Linkage regions of interest included a broad peak (31 Mb) on chromosome 1q with acute insulin response (max LOD = 5.37). This region was previously documented with type 2 diabetes in family-based studies, providing support for the validity of these results. Overall, we have demonstrated the utility of two-point linkage and association in comprehensive genome-wide array-based SNP genotypes.

Journal of Human Genetics (2017) **62,** 175–184; doi:10.1038/jhg.2016.103; published online 18 August 2016

INTRODUCTION

Family-based linkage analysis has largely been supplanted by genome-wide association studies, often using unrelated samples, following the limited success of linkage when applied to complex traits. Family-based analyses, however, have inherent strengths, which complement other approaches for identification of contributors to complex phenotypes.[1,2] Such analyses may be especially applicable to identifying low-frequency (minor allele frequency (MAF) 0.01–0.05) to rare (MAF < 0.01) alleles with high impact.[3–8] We have implemented approaches in parallel, which use simple two-point linkage analysis and conventional association analysis to search for genetic variants with meaningful contributions to phenotypic variance of traits. Two-point linkage analysis considers each variant independently, unlike multipoint analysis, which integrates the information from multiple variants simultaneously. Therefore, two-point linkage does not have the same issues with inflation because of linkage disequilibrium (LD) between markers and can be used to test putatively impactful variants for linkage directly. The combined two-point linkage and association approach has the advantage of being able to directly align single-nucleotide polymorphism (SNP) results for the two analyses, pinpointing variants that show evidence of both linkage and association at the single SNP level. In prior studies, this has been applied to exome chip data, thus focusing on coding variants[9] and characteristics of a functional SNP.[10]

Evaluation of association in the context of linkage has an extensive history,[11–13] with association typically used to determine whether genetic variants residing under the linkage peak explain the observed signal. We have observed that instances of strong linkage and association together at a single locus (e.g. *APOE* with ApoB levels, *CETP* (cholesterol ester transfer protein) with high-density lipoprotein

[1]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; [2]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; [3]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA; [4]Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC, USA; [5]Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, USA; [6]Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; [7]Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; [8]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA; [9]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA; [10]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [11]Department of Medicine, University of Texas Health Science Center, San Antonio, TX, USA; [12]Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA and [13]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
Correspondence: Dr DW Bowden, Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA.
E-mail: dbowden@wakehealth.edu
Received 8 March 2016; revised 1 July 2016; accepted 11 July 2016; published online 18 August 2016

(HDL) levels, *ADIPOQ* with adiponectin levels)[9,10] represent variants or loci that have a striking impact on phenotype, reflected as explanation of a high proportion of the variance of the trait (3–60%). We have also observed this across a range of minor allele frequencies (1–45%), indicating that this approach can be informative for a full range of genetic variation. Other groups have used combined metrics of linkage and association to identify variants with large impact;[11] however, that is a project currently undergoing evaluation separate from these analyses.

Here we have investigated the performance of these approaches in a contemporary genetic data set consisting of comprehensive genome-wide and exome chip data encompassing 1.6 million SNPs in 90 Hispanic families from the Insulin Resistance Atherosclerosis Family Study (IRASFS). Based on our prior work and recent evidence for the existence of high impact noncoding variants,[14] we hypothesize this family-based method is applicable to the search for such variants.

MATERIALS AND METHODS

Samples and phenotype data

The samples used in this study are from the Hispanic cohorts of the IRASFS.[15] Briefly, subjects were ascertained on the basis of large family size in San Luis Valley, Colorado and San Antonio, Texas. The sample consisted of 1425 individuals from 90 families, who were extensively phenotyped, including a frequently sampled intravenous glucose test, measures of blood lipids and inflammatory markers, anthropomorphic measures, as well as fat deposition measures by computed tomography and dual X-ray absorptiometry scans. Institutional Review Board approval was obtained at all clinical and analysis sites, and all participants provided informed consent.

Genotype data

SNP genotype data from three genotyping chips were used. Illumina OmniExpress and Illumina Omni1S chips were genotyped as part of the Genetics Underlying Diabetes in Hispanics (GUARDIAN) Consortium ($N=1034$ and 1038, respectively),[16] and the Illumina HumanExome Beadchip was genotyped on a larger subset ($N=1414$)[9] of the full IRASFS Hispanic cohorts. Genotyping of the Illumina HumanExome BeadChip v.1.0 ($N=552$) and v.1.1 ($N=862$) was performed at the Wake Forest Center for Genomics and Personalized Medicine Research, whereas the Illumina HumanOmniExpress BeadChip and Illumina Omni1S BeadChip were genotyped at the core genotyping laboratory at Cedars-Sinai Medical Center (Los Angeles, CA, USA). All genotypes were called separately by genotyping array using GenomeStudio (Illumina, San Diego, CA, USA). Sample and autosomal SNP call rates were $\geqslant 0.98$ (>0.99 SNP call rates for the OmniExpress and Omni1S chips), and Exome Chip SNPs with poor cluster separation (<0.35) were excluded. All data sets independently underwent Mendelian error checking using PedCheck[17] to detect genotypes discordant in families for Mendelian inheritance, with resolution by removing all inconsistent genotypes. The total number of unique SNPs available for analysis following quality control was as follows: 81 559 from the Exome Chip, 668 758 from OmniExpress and 920 823 from the Omni1S chip, for a total of 1 671 140 SNPs.

Imputation to the 1000 Genomes integrated reference panel (version 2) was performed using genotypes and samples from the OmniExpress data set ($N=634$K genotypes and 1034 individuals) using SHAPEIT[18] for phasing and IMPUTE2[19] for imputation.

Analyses

SNPs were evaluated for both two-point family-based linkage and single SNP association using Sequential Oligogenic Linkage Analysis Routines (SOLAR)[20] separately by genotyping platform. Both analyses used age, sex, body mass index (BMI) and study center as covariates. All phenotypes evaluated were transformed to approximate normality of the residuals if necessary (Supplementary Table 1). Additionally, because of the high impact of a low-frequency variant known to influence adiponectin levels in this population,[3,10] presence of the variant encoding the G45R missense mutation in *ADIPOQ* (rs200573126) was included as a covariate for analyses involving adiponectin. Visceral adipose tissue (VAT) area, visceral-to-subcutaneous tissue ratio (VSR), waist circumference and waist-to-hip ratio were run both with and without BMI as a covariate. However, subcutaneous adipose tissue area, percent body fat and body adiposity index were not adjusted for BMI. All association analyses included three admixture proportions as covariates. Existing admixture proportion estimates were available from previously genotyped exome chip data; estimates were computed by maximum-likelihood estimation of individual ancestries in ADMIXTURE[21] assuming five ancestral populations ($K=5$) from exome chip-wide SNP data after pruning for LD to produce admixture estimates for the greatest number of samples. Of the five variables considered, three variables were selected as representing the variation in these

Table 1 Demographic characteristics of the IRASFS Hispanic samples with selected phenotypes

Characteristic	*Exome Chip (81 559 variants)*		*Omni Express (668 758 variants)*		*Omni1S (920 823 variants)*	
Samples[a]		1414		1034		1038
Age (years)	1263	42.75 (18–81)	1034	40.63 (18–81)	1038	40.61 (18–81)
% Female	823	58.3	609	58.90	612	58.90
BMI (kg m^{-2})	1253	28.88 (16–58)	1027	28.28 (16–58)	1027	28.28 (16–58)
% T2D[b]	187	13.20	0	0	0	0
AIR (pmol ml^{-1} per min)	1035	761.86 (–80.9 to 313.7)	1034	760.29 (–80.9 to 4313.7)	1038	759.21 (–80.9 to 4313.7)
TNFα receptor 2 (ng ml^{-1})	982	7.05 (2.38–30.00)	821	6.79 (2.38–30.00)	824	6.79 (2.38–30.00)
Fibrinogen (mg dl^{-1})	1256	265.74 (113–591)	1032	259.37 (113–506)	1036	259.61 (113–506)
Cholesterol (mg dl^{-1})	1255	177.94 (74–348)	1031	176.12 (74–311)	1035	176.17 (74–311)
HDL (mg dl^{-1})	1254	43.82 (18–125)	1030	43.58 (18–100)	1034	43.60 (18–100)
LDL (mg dl^{-1})	1242	109.17 (31–218)	1022	109.04 (31–213)	1026	109.06 (31–213)
Triglycerides (mg dl^{-1})	1252	124.57 (18–836)	1030	118.30 (18–836)	1034	118.31 (18–836)
ACR (mg g^{-1})	1256	53.55 (1.63–3903.92)	1032	19.63 (1.93–1459.68)	1036	19.58 (1.93–1459.68)
Percent body fat	943	33.95 (10.10–55.03)	786	33.51 (10.10–51.78)	789	33.52 (10.10–51.78)
VAT (cm^2)	1206	114.02 (10.04–382.56)	994	106.56 (10.04–363.34)	998	106.52 (10.04–363.34)
VSR	1164	0.38 (0.07–1.63)	963	0.36 (0.07–1.56)	967	0.36 (0.07–1.56)

Abbreviations: ACR, albumin/creatinine ratio; AIR, acute insulin response; BMI, body mass index; HDL, high-density lipoprotein; IRASFS, Insulin Resistance Atherosclerosis Family Study; LDL, low-density lipoprotein; T2D, type 2 diabetes; TNFα, tumor necrosis factor-α; VAT, visceral adipose tissue; VSR, visceral-to-subcutaneous tissue ratio.
Data are presented as mean (range) or percent.
[a]From 90 pedigrees, not entirely overlapping.
[b]At baseline.

Hispanic samples, as inclusion of additional postulated ancestral populations began isolating individual pedigrees.

For validation of performance, genotypes imputed to the 1000 Genomes panel were also evaluated for linkage (and association) in two regions, which were selected for their linkage regions as well as being phenotypically of particular interest to our group: chromosome 1 for acute insulin response (AIR) to glucose and chromosome 7 for insulin sensitivity index (S_I). Best guess genotypes from the imputed data were used in the linkage analysis because methods that account for imputation uncertainty have not been developed for linkage. These analyses used the same covariates as previously mentioned.

RESULTS

The aim of this analysis was to test the utility of carrying out a combined linkage and association analysis in a contemporary data set made up of genome-wide association studies (GWAS) (Illumina OmniExpress and Omni1S) and exome chip data encompassing over 1.6 million SNPs. The combined performance was evaluated for a total of 50 quantitative traits from 7 phenotypic groups: glucose homeostasis, adiposity, lipids, biomarkers, hypertension, liver enzymes and liver fat, in 90 families from the IRASFS with an average family size of 15.4 individuals. Overall, 83 557 000 LOD (logarithm of the odds) scores and association *P*-values were calculated across the three genotyping sets.

Characteristics of the samples and genotyping are summarized in Table 1. The sample consisted of 1418 individuals from 90 families. Specifically, for the smallest genotyped sample (OmniExpress), sample sizes ranged from 786 (percent body fat) to 1034 (AIR), although larger sample sizes were available for SNPs present on the exome chip (up to 1256 for fibrinogen and albumin/creatinine ratio). Across all phenotypes, there were 9214 LOD scores $\geqslant 3$, 845 $\geqslant 4$ and 89 $\geqslant 5$. Of the variants with LOD scores $\geqslant 5.0$, 27 were linked to tumor necrosis factor-α (TNFα) receptor 2 levels, 13 to HDL levels, 24 to AIR, 13 to G45R-adjusted adiponectin levels and 3 to BMI-adjusted VSR. While a detailed summary of each trait analysis is impractical, following on our earlier observations,[9,10] we have initially focused on the patterns visible in linkage analysis followed by relating these results to association analysis results. In this report, we evaluated linkage and association with 50 cardiometabolic phenotypes (see Supplementary Table 1 for complete listing). Selected phenotypes, namely TNFα receptor 2 levels, HDL levels, AIR, adiponectin levels (adjusted for G45R, a high impact mutation identified previously in these samples[3,10]) and VSR are summarized in Table 1. Overall, 12 phenotypes (from four phenotype groups: glucose homeostasis, lipids, adiposity and biomarkers) were represented in this category of LOD scores >5.0 (results are summarized in Table 2), where highest LOD scores are grouped by phenotype and chromosome. A complete summary of LOD scores >5 is presented in Supplementary Table 2.

Evaluation of loci with high LOD scores

The overall maximal LOD score of 6.49 was observed with rs12956744 with the biomarker TNFα receptor 2 levels (Table 3 and Figure 1a). This SNP is located in intron 1 (nearer the 5′ end) of *LAMA1* (laminin subunit alpha-1 gene) on chromosome 18. Of note, three additional intronic variants in *LAMA1* were also linked to TNFα receptor 2 levels with LOD >6, and nine SNPs overall were linked with LOD >3 (Table 3). Notably, one SNP (rs28569884) was also associated with TNFα receptor 2 levels (P-value $=5.9\times10^{-4}$; LOD $=1.06$). The variant rs28569884 (in intron 56) is distal to the striking linkage signal (146 kb apart), although there was another LOD score over 4 (rs4395154; LOD $=4.47$) just 13 kb away at the 3′ end of the *LAMA1* gene (intron 62). *LAMA1* is a very large gene, with 63 exons and 245 SNPs analyzed. Of these, 11 (4.4%) had nominally significant association (P-value <0.05) with TNFα receptor 2 levels. Comparatively, 9 variants had LOD scores >3 (3.7%) and 23 variants had LOD >1 (9.4%).

Table 2 Summary of linkage results for phenotypes with at least one variant with LOD >4

Phenotype	*LOD > 5*	*LOD > 4*	*LOD > 3*
AIR	24	180	1335
S_I	1	17	247
DI		8	101
MCRI		6	100
Total cholesterol	1	16	269
HDL	13	129	1202
LDL	1	9	191
ApoB		9	291
Triglycerides	4	18	151
SBP		1	48
DBP		1	24
ACR		3	169
Adiponectin (adjusted)	13	96	621
CRP		5	84
Fibrinogen		16	341
TNF2	27	259	2458
RBP4		1	20
BMI	1	11	100
BAI		4	66
Percent body fat	1	18	159
Waist circumference		2	32
WHR		1	10
SAT	1	7	151
Visceral adipose tissue (adj. for BMI)		1	63
Visceral-to-subcutaneous ratio	1	8	138
Visceral-to-subcutaneous ratio (VSR; adj. for BMI)	3	9	141
Liver density		4	123
Inverse normalized liver		2	47
GGT		6	126

Abbreviations: ACR, albumin/creatinine ratio; AIR, acute insulin response; ApoB, apolipoprotein B; BAI, body adiposity index; BMI, body mass index; CRP, C-reactive protein; DBP, diastolic blood pressure; DI, disposition index; GGT, γ-glutamyl transpeptidase; LOD, logarithm of the odds; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MCRI, metabolic clearance rate of insulin; RBP4, retinol binding protein 4; SAT, subcutaneous adipose tissue; SBP, systolic blood pressure; S_I, insulin sensitivity index; TNF2, TNFα receptor 2; VAT, visceral adipose tissue; VSR, visceral-to-subcutaneous tissue ratio; WHR, waist-to-hip ratio.
Boldface indicates phenotypes with a LOD score >5.

A major focus of our laboratory is identifying genetic contributors to metabolic measures of glucose homeostasis. The top linkage result of LOD $=6.47$ (Table 4) for AIR was rs28479408, an intronic variant located in *SYCP2L* (synaptonemal complex protein 2-like gene) on chromosome 6 (Figure 1b). Although this variant was not associated with AIR (P-value $=0.71$), six other SNPs in this gene were also linked (rs4713044, LOD $=6.10$; rs12190237, LOD $=5.58$; rs12214063, LOD $=3.58$; rs1767771, LOD $=3.42$; rs2153159, LOD $=3.31$; rs1632103, LOD $=3.15$) but not associated (P-values >0.5) (Table 4).

Strikingly, chromosome 1 had a broad linkage peak for AIR, with a maximal LOD score of 6.37 (rs2252384) in the region between *FAM163A* and *TOR1AIP2* (located at ~179 Mb; 1q25.2; Figure 1b and Table 5). Chromosome 1 has a long history of linkage to diabetes, making this result all the more interesting.[22–25] Here, variants with LOD scores >3 spanned much of the proximal q arm of the

Table 3 Selected *LAMA1* results with TNFα receptor 2 protein levels (LOD > 1 and/or *P*-value < 0.01)

SNP	*Chr.*	*Position*	*Chip*	N	*MAF*	*LOD*	P-*value*	β-*Value*	*s.e.*	*Variance*
rs4395154	18	6 942 805	OmniExpress	820	0.46	**4.47**	0.25	0.016	0.014	0.001
rs2016639	18	6 943 264	OmniExpress	821	0.431	**3.46**	0.2	−0.018	0.014	0.002
rs17439137	18	6 951 060	OmniExpress	821	0.235	1.07	0.77	−0.005	0.016	0
rs8086875	18	6 951 710	Omni1S	821	0.208	1.15	0.36	0.015	0.017	0.0008
rs8088218	18	6 951 971	Omni1S	820	0.21	1.62	0.32	0.017	0.017	0.001
rs12454596	18	6 953 989	OmniExpress	821	0.446	1.85	0.73	0.005	0.014	0.0002
rs949215	18	6 955 676	OmniExpress	821	0.25	1.18	0.96	0.001	0.016	0
rs28569884	18	6 956 111	Omni1S	821	0.058	1.06	**5.94E−04**	−0.098	0.029	0.015
rs509497	18	6 957 193	OmniExpress	821	0.393	1.29	0.04	0.028	0.014	0.005
rs633691	18	6 967 089	OmniExpress	821	0.419	**3.18**	0.085	0.024	0.014	0.0044
rs11873205	18	6 979 621	Omni1S	818	0.13	1.54	**0.0072**	−0.055	0.021	0.0113
rs538815	18	6 982 443	OmniExpress	821	0.202	1.69	0.5	−0.011	0.017	0.0003
rs619106	18	7 011 413	OmniExpress	821	0.291	0.03	**0.042**	−0.032	0.015	0.009
rs67268419	18	7 013 648	Omni1S	820	0.077	1.74	0.74	−0.009	0.025	0.0006
rs541928	18	7 034 932	Omni1S	821	0.153	2.05	0.49	0.013	0.019	0
rs7240767	18	7 070 642	OmniExpress	821	0.468	0	**0.029**	−0.03	0.014	0.0058
rs7228959	18	7 076 464	OmniExpress	821	0.49	0	**0.044**	−0.027	0.014	0.0047
rs16951199	18	7 080 135	OmniExpress	815	0.068	0	**0.017**	−0.064	0.027	0.0081
rs11081298	18	7 085 706	Omni1S	820	0.466	2.91	0.94	−0.001	0.014	0.0001
rs12606163	18	7 096 977	OmniExpress	807	0.485	**4.78**	0.11	0.022	0.014	0.0038
rs972038	18	7 102 036	Omni1S	816	0.171	0.07	**0.046**	−0.036	0.018	0.0103
rs12955222	18	7 102 427	OmniExpress	821	0.482	**4.53**	0.13	0.02	0.013	0.0038
rs12956744	18	7 102 706	Omni1S	821	0.407	**6.49**	**0.03**	0.03	0.014	0.0071
rs12959835	18	7 103 146	Omni1S	820	0.408	**6.38**	**0.034**	0.029	0.014	0.0068
rs1462780	18	7 105 988	OmniExpress	820	0.019	0	**0.034**	−0.103	0.049	0.0072
rs34433741	18	7 108 999	Omni1S	820	0.415	**6.07**	0.089	0.023	0.014	0.0031
rs4798533	18	7 109 571	Omni1S	819	0.282	1.52	0.82	−0.003	0.015	0.0002
rs12454984	18	7 109 652	Omni1S	821	0.404	**6.02**	0.15	0.02	0.014	0.0019
rs984355	18	7 114 212	OmniExpress	821	0.217	2.55	0.36	0.016	0.017	0

Abbreviations: Chr., chromosome; LOD, logarithm of the odds; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; TNFα, tumor necrosis factor-α.
Boldface indicates LOD scores > 3 or *P*-values < 0.05.

chromosome, with the most concentrated linkage peak residing between 156 and 187 Mb, a region encompassing 357 RefSeq genes (1q22-31.1). Focusing on the peak LOD-1 substantially narrowed the region to a very narrow 1.57 Mb. Of the 343 variants within this region with LOD scores > 3, 73 of them had *P*-values < 0.05, with a best association signal occurring at rs6426957 (Chr1: 165 988 336; *P*-value = 6.34×10^{-4}, LOD = 3.09, MAF = 0.441; Supplementary Table 3). Notably, many variants within *RASAL2* (RAS protein activator like 2 gene) showed nominal evidence of association ($0.05 > P\text{-value} > 1.42 \times 10^{-3}$) in addition to linkage ($N = 45$ of 46 linked (LOD > 3) SNPs; Tables 5 and 6). LOD scores at this gene ranged from 3.00 to 5.38.

Additional linkage results of interest include regions on chromosomes 7 and 12, which were linked to insulin sensitivity index (S_I). Although these regions did not reach the magnitude seen for TNFα receptor 2 and AIR, the consistency of linkage in the region is compelling. On chromosome 7, the highest LOD score (5.11) was seen with rs1024591, an intergenic SNP over 300 kb from the nearest gene (a long intergenic noncoding RNA, *LINC01372*) (Supplementary Table 4). The linkage signal on chromosome 12 is made up of two distinct peaks (Figure 1c), one at ~ 53 Mb and the second at ~ 105 Mb (Supplementary Table 5). The LOD scores seen here are not as striking by magnitude (max LOD for each peak 4.27–4.28), but the consistency of LOD scores > 3 into tight peaks is notable (Supplementary Table 5). The first peak consists of 14 variants with LOD scores > 3, from 50.6 to 54.5 Mb, with multiple variants in the *KRT8* (keratin 8 gene) and *ESPL1* (extra spindle pole bodies like 1, separase) showing evidence for linkage, as well as single variants at the proximal end of the peak in *LIMA1* (LIM domain and actin binding 1 gene), *DIP2B* (disco-interacting protein 2 homolog B gene) and *SLC4A8* (solute carrier family 4, sodium bicarbonate cotransporter, member 8 gene). There was no evidence for association among linked variants at this linkage peak, although other, unlinked variants in the region showed nominal association (Supplementary Table 5).

The second linkage peak resides from 101 to 109 Mb on chromosome 12, and included 21 linked variants, which represented multiple signals from *CHST11* (carbohydrate (chondroitin 4) sulfotransferase 11 gene), *ACACB* (acetyl-CoA carboxylase beta gene) and *FOXN4* (forkhead box N4 gene), in addition to intergenic variants and genes implicated by a single variant, such as *CMKLR1* (chemerin chemokine-like receptor 1 gene) (Supplementary Table 5). One of these linked variants showed nominal evidence of association, with a *P*-value of 5.50×10^{-3} (rs11114094 in *SVOP* (SV2-related protein gene); Table 6 and Supplementary Tables 3 and 5), although like the prior peak, other unlinked variants in the linkage region also demonstrated evidence of association.

Variants with evidence of both linkage and association

Using the linkage results as a search tool and prioritizing those with any evidence of association identified 1076 variants with *P*-values < 0.05 as well as a LOD score ⩾ 3 (Supplementary Table 3).

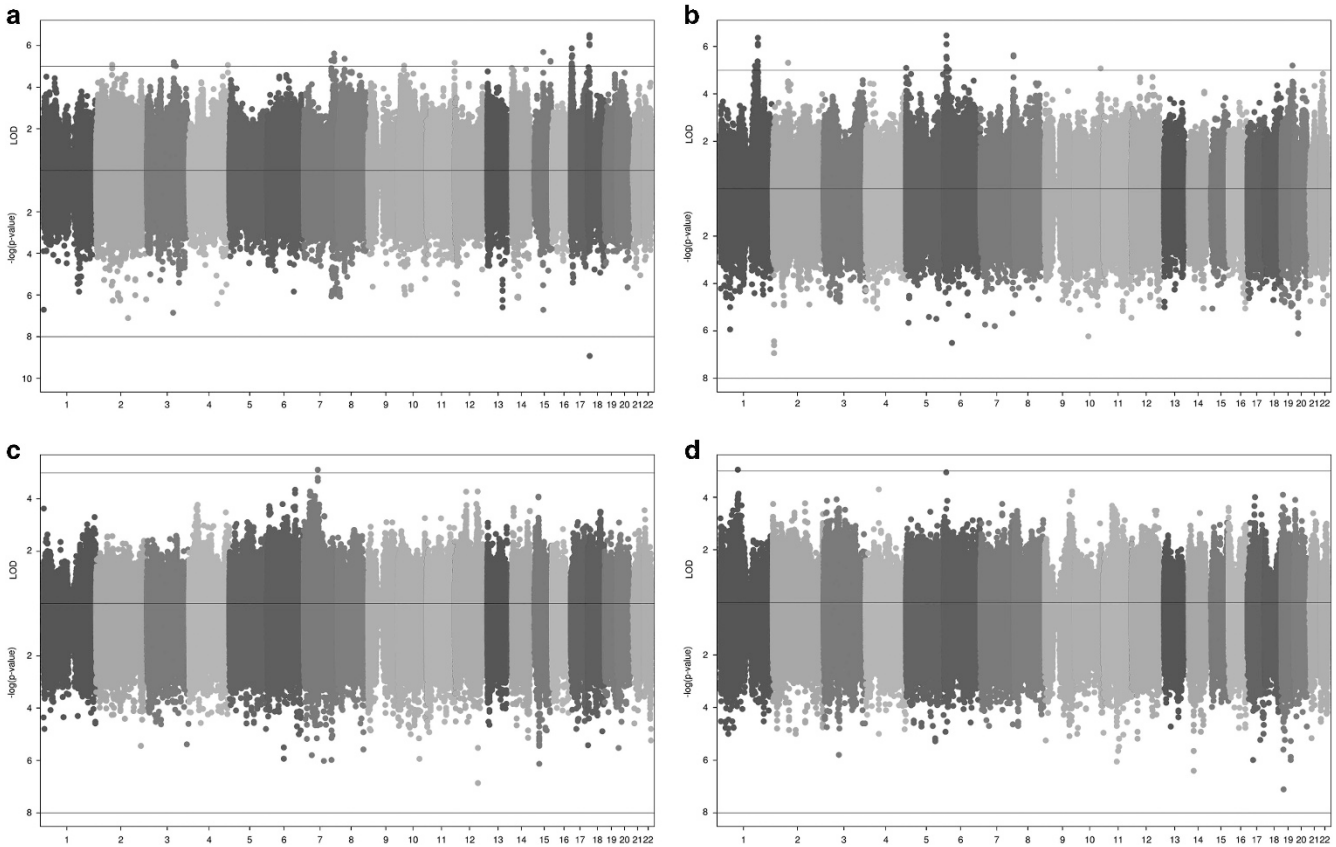

Figure 1 Opposed plots showing LOD (logarithm of the odds scores) from the two-point linkage (upper portion) and log-transformed *P*-values for association (lower portion) results across all arrays for (**a**) tumor necrosis factor-α (TNFα) receptor 2 levels, (**b**) acute insulin response (AIR) (note the broad linkage peak on chromosome 1, and the strong linkage also on chromosome 6), (**c**) insulin sensitivity index (S_I) (of particular note are the signals on chromosomes 7 and 12) and (**d**) low-density lipoprotein (LDL) levels (note the signals on chromosome 4, contributed by *LPHN3* and chromosome 19, which represents the *APOE* locus, evaluated in our previous publication with apolipoprotein B levels). A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Twenty-seven variants were associated with $P<0.005$, as well as having a LOD score >4 (Table 6). *NFIB* was the primary gene implicated under a linkage peak with TNFα receptor 2 levels on chromosome 9, where there was also evidence of nominal association (*P*-values on the order of 2×10^{-4}; Figure 1a and Supplementary Table 6). *NFIB*, which encodes nuclear factor I/B, is represented by 293 SNPs (135 from OmniExpress, 157 from Omni1S, 1 from exome chip), 289 of which were located in introns. Only one coding variant in this gene was polymorphic from the exome chip data set, this SNP (rs114558598; I24F) was not linked (LOD $=-0.005$) or associated (*P*-value $=0.08$). Ten common variants ($0.27<$MAF>0.49) within this gene (all intronic) had LOD scores >3. Overall, 68 *NFIB* variants had LOD scores >1, and 24 had LOD scores >2.

LPHN3 on chromosome 4 was a strong signal for LDL levels, with two intronic variants being both linked and associated (rs2343249; LOD $=4.30$; *P*-value $=1.00\times10^{-5}$ and rs9312078, LOD $=3.02$; *P*-value $=8.20\times10^{-5}$; Table 7 and Figure 1d). Both the linkage and association signals were confined to the gene region, with strong LD ($r^2>0.8$) between the two top SNPs. There was further support throughout the gene-encoding region for both modest linkage and association with diminishing LD (Supplementary Figure 1). The strongest association result among LOD scores $\geqslant3$ was with fibrinogen levels; rs1131878 from the OmniExpress chip, LOD $=3.08$ and *P*-value $=1.99\times10^{-6}$ (Supplementary Table 3). This SNP was located within the *UGT2B4* gene, which encodes UDP glucuronosyltransferase 2 family polypeptide B4.

DISCUSSION

This study evaluated the utility of combining two-point linkage with association analysis in a data set comprised of array-based SNP genotyping totaling 1.6 million noncoding and coding variants in a family-based sample of Hispanics with extensive phenotype information. The aim of the study was to evaluate whether GWAS data in the context of linkage adds insight into the genetic origins of cardiometabolic traits, while using association analysis as a follow-up to determine likely candidate loci. This builds upon our prior evaluation of combined linkage and association using exome chip data in this cohort.[9] Large-scale linkage analysis of SNP genotyping has been uncommon for complex phenotypes recently. To this end, we evaluated 50 phenotypes (46 distinct traits) related to glucose homeostasis, lipids, blood pressure, adiposity, liver fat and enzymes, and biomarkers. Given the breadth of genotypic data and the number of phenotypes, the results are extensive, but some noteworthy observations can be made. Broadly speaking, we believe the markedly denser genotypic data set reveals many insights into the genetic bases of the traits such as TNFα receptor 2, AIR and S_I when compared with our prior study using the more limited data from the exome chip.

Relatively dense genotyping data provides visual evidence of linkage similar to conventional multipoint methods. In addition, while exome

Table 4 Chromosome 6 AIR linkage peak with linked (LOD > 3) and/or associated (*P*-value < 0.05) variants

SNP	*Chr.*	*Position*	*Chip*	N	*MAF*	*Gene*	*LOD*	P-*value*	β-*Value*	*s.e.*	*Variance*
rs12208366	6	10 383 410	Omni1S	1034	0.146		**3.43**	0.578	0.39	0.701	0
rs480965	6	10 387 251	OmniExpress	1033	0.142		**3**	0.546	0.419	0.695	0
rs533558	6	10 395 572	OmniExpress	1033	0.406		**3.55**	0.122	−0.771	0.499	0.002
rs79025376	6	10 400 618	Omni1S	1033	0	*TFAP2A*	0	**5.06E−03**	−27.514	9.816	0.008
rs78497087	6	10 471 612	Omni1S	1032	0.356		**3.39**	0.813	0.123	0.518	0
rs491803	6	10 477 438	Omni1S	1033	0.331		**3.31**	0.885	0.075	0.521	0
rs9466917	6	10 606 584	Omni1S	1033	0.492	*GCNT2*	**3.32**	0.89	0.069	0.501	0
rs3798704	6	10 615 268	Omni1S	1034	0.494	*GCNT2*	**3.33**	0.923	0.048	0.5	0
rs1233887	6	10 739 432	OmniExpress	1033	0.36		**3.1**	0.714	−0.187	0.51	0
rs518954	6	10 791 859	OmniExpress	1029	0.278	*MAK*	**3.1**	0.184	0.727	0.546	0.003
rs12214063	6	10 855 738	Omni1S	1032	0.213	*SYCP2L*	**3.58**	0.753	−0.195	0.62	0
rs1767771	6	10 857 646	Omni1S	1034	0.473	*SYCP2L*	**3.42**	0.685	−0.203	0.499	0
rs1632103	6	10 862 649	Omni1S	1034	0.478	*SYCP2L*	**3.15**	0.558	−0.293	0.5	0
rs2153159	6	10 887 932	Omni1S	1033	0.36	*SYCP2L*	**3.31**	0.969	−0.02	0.506	0
rs4713044	6	10 911 282	OmniExpress	1033	0.182	*SYCP2L*	**6.1**	0.951	−0.039	0.63	0
rs28479408	6	10 912 131	Omni1S	1034	0.177	*SYCP2L*	**6.47**	0.712	−0.236	0.64	0
rs12190237	6	10 922 638	OmniExpress	1031	0.164	*SYCP2L*	**5.58**	0.775	0.188	0.66	0
rs6457131	6	11 227 328	OmniExpress	1029	0.207	*NEDD9*	**3.24**	0.919	0.061	0.604	0
rs55813531	6	11 238 023	Omni1S	1031	0.185	*NEDD9*	**5.14**	0.274	0.698	0.639	0.002
rs17496723	6	11 238 633	Omni1S	1031	0.413	*NEDD9*	1.2	**7.89E−03**	−1.323	0.498	0.004
rs9468690	6	11 239 119	OmniExpress	1033	0.455	*NEDD9*	0.86	**7.86E−03**	−1.316	0.495	0.005
rs9461574	6	11 239 518	OmniExpress	1033	0.492	*NEDD9*	1.94	**5.77E−03**	−1.354	0.49	0.006
rs12209631	6	11 242 203	OmniExpress	1028	0.175	*NEDD9*	**3.08**	0.0873	1.134	0.662	0.005
rs6908326	6	11 247 387	OmniExpress	1033	0.204	*NEDD9*	2.97	**5.11E−03**	1.683	0.6	0.009
rs10947066	6	11 253 969	Omni1S	1034	0.264	*NEDD9*	**4.34**	**0.0468**	1.117	0.562	0.007
rs10947067	6	11 253 990	Omni1S	1033	0.265	*NEDD9*	**4.25**	**0.0481**	1.113	0.563	0.006
rs6457197	6	11 254 692	Omni1S	1028	0.496	*NEDD9*	**3.72**	**0.0165**	−1.176	0.491	0.01
rs6457202	6	11 255 770	Omni1S	1033	0.445	*NEDD9*	**4.29**	**8.71E−03**	1.324	0.505	0.013
rs7766626	6	11 256 000	OmniExpress	1031	0.371	*NEDD9*	**3.73**	**0.0152**	1.206	0.496	0.01
rs210903	6	11 724 542	OmniExpress	1031	0.271	*C6orf105*	**3.93**	0.954	−0.032	0.561	0
rs4713831	6	11 726 626	OmniExpress	1014	0.298	*C6orf105*	**4.12**	0.726	0.189	0.541	0
rs210897	6	11 729 299	Omni1S	1034	0.282	*C6orf105*	**5.49**	0.893	0.075	0.557	0
rs114551218	6	11 736 145	Omni1S	1030	0.003	*C6orf105*	0	**3.48E−03**	13.077	4.476	0.014
rs210890	6	11 740 036	OmniExpress	1032	0.162	*C6orf105*	**3.13**	0.552	0.4	0.673	0
rs12204492	6	11 774 626	OmniExpress	1032	0.424	*C6orf105*	**3.62**	0.376	−0.431	0.487	0.001
rs2235384	6	11 776 631	OmniExpress	1031	0.205	*C6orf105*	**3.02**	0.481	0.419	0.594	0

Abbreviations: AIR, acute insulin response; Chr., chromosome; LOD, logarithm of the odds; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
Boldface indicates LOD scores >3 or *P*-values <0.05.

chip analysis primarily targets models where functional variants are exonic, the GWAS data sets can potentially address other models such as high impact noncoding variants, especially through linkage analysis. Here we have observed few examples where evidence for both linkage and association are apparent. An example is *LPHN3* (Table 7 and Supplementary Figure 1), where LOD scores reached 4.30 with a *P*-value of 1.00×10^{-5}, suggesting a true impact on LDL levels. Given the actual low density of coverage in GWAS data sets, which are designed to cover genomic regions through LD relationships, it is unlikely to capture truly causal variants by chance. The ultimate test of whether this approach will be successful will require whole-genome sequencing data. Overall, these results incorporating two-point linkage and association analyses can identify meaningful signals that impact cardiometabolic traits, often in the absence of striking association alone. These conclusions are consistent with our prior work[9,10] in which we have shown that linkage evidence can be relatively strong, but association evidence only appears when the functional variant is also captured. The latter is unlikely in a GWAS data set. For these reasons, our main focus was on regions with evidence of linkage based on both the power of linkage methods and the 'far-sighted' ability of linkage to identify genetic relationships.[4–7,9,10]

As noted above, several genomic regions had relatively strong evidence of linkage, but limited association results. Based on our logic, this would suggest the possibility of underlying, as yet unidentified functional variants. Thus, for the strongest linkage with TNF2α receptor levels (LOD = 6.49), we would hypothesize that one or more high impact noncoding variants lie within the linkage region. *LAMA1* is similar to *LAMA5*, which has previously been related to TNFRSF1B expression,[26] making it plausible for *LAMA1* to be related to TNF2α receptor levels.

Analysis of traits of interest to our laboratory (AIR, S_I) also resulted in notable linkage peaks. It is tempting to scan these linked regions for biologically relevant genes. Genes located under a broad AIR linkage region on chromosome 1 (Figure 1b and Table 5) included *FAM163A*, also known as neuroblastoma-derived secretory protein (*NDSP*), *TOR1AIP2* and *RASAL2*. *FAM163A* (aka *NDSP*) has been associated in methylation analysis for borderline personality disorder[27] with

Table 5 Broad linkage region on chromosome 1 with acute insulin response: variants with LOD >4.5

SNP	*Chr.*	*Position*	*Chip*	N	*MAF*	*Gene*	*LOD*	P-*value*	β-*Value*	*s.e.*	*Variance*
rs12047043	1	164 625 696	OmniExpress	1029	0.225	*AX748175*	**4.95**	0.16	0.832	0.594	0.005
rs4657367	1	164 627 551	OmniExpress	1033	0.225	*AX748175*	**4.72**	0.15	0.857	0.591	0.005
rs4656475	1	166 004 063	Omni1S	1032	0.14	Intergenic	**4.66**	0.38	0.635	0.721	0.001
rs6662013	1	166 042 658	Omni1S	1034	0.247	*FAM78B*	**5.19**	0.71	−0.21	0.565	0
rs6680174	1	166 459 849	OmniExpress	1033	0.266	Intergenic	**4.73**	0.33	0.544	0.553	0.001
rs1476076	1	167 794 511	Omni1S	1031	0.467	*ADCY10*	**4.74**	0.81	−0.113	0.48	0
rs203849	1	167 849 414	OmniExpress	1033	0.484	*ADCY10*	**4.62**	0.43	−0.395	0.5	0.002
rs4656148	1	168 179 545	Omni1S	1031	0.273	Intergenic	**4.87**	0.42	0.427	0.535	0
rs11589732	1	168 585 289	OmniExpress	1033	0.228	Intergenic	**5.00**	0.86	0.106	0.582	0
rs7474070	1	171 050 589	OmniExpress	1033	0.22	Intergenic	**4.86**	0.11	−0.959	0.597	0.003
rs16863990	1	171 055 570	OmniExpress	1032	0.193	Intergenic	**5.09**	0.15	−0.929	0.644	0.003
rs12402693	1	171 057 312	OmniExpress	1032	0.193	Intergenic	**5.16**	0.14	−0.947	0.643	0.003
rs12404183	1	171 058 946	OmniExpress	1026	0.212	Intergenic	**4.59**	0.21	−0.754	0.603	0.002
rs1800822	1	171 076 935	OmniExpress	1029	0.201	*FMO3*	**4.62**	0.3	−0.637	0.613	0.002
rs2281002	1	171 080 629	OmniExpress	1033	0.189	*FMO3*	**4.79**	0.12	−1.005	0.646	0.004
rs909529	1	171 082 896	OmniExpress	1033	0.201	*FMO3*	**4.72**	0.078	−1.103	0.624	0.004
rs6659102	1	176 535 567	OmniExpress	1032	0.149	*PAPPA2*	**4.66**	0.91	−0.082	0.685	0
rs7540152	1	176 656 255	OmniExpress	1033	0.13	*PAPPA2*	**4.53**	0.73	0.256	0.731	0
rs791031	1	176 667 810	OmniExpress	1030	0.129	*PAPPA2*	**4.60**	0.82	0.165	0.734	0
rs11583320	1	178 042 145	OmniExpress	1029	0.221	Intergenic	**4.52**	**5.63E−03**	1.597	0.576	0.006
rs964993	1	178 062 359	OmniExpress	1033	0.188	*LOC100302401*	**4.67**	**1.89E-03**	1.988	0.639	0.007
rs10913506	1	178 092 233	OmniExpress	1033	0.186	*RASAL2*	**4.93**	**1.52E−03**	2.019	0.636	0.008
rs10798604	1	178 254 568	OmniExpress	1029	0.174	*RASAL2*	**4.96**	**0.033**	1.38	0.648	0.004
rs77603205	1	178 279 051	Omni1S	1033	0.173	*RASAL2*	**4.52**	**0.021**	1.504	0.652	0.005
rs10913550	1	178 408 795	OmniExpress	1033	0.174	*RASAL2*	**5.16**	**0.027**	1.435	0.65	0.004
rs9803679	1	178 410 425	OmniExpress	1033	0.174	*RASAL2*	**5.21**	**0.027**	1.435	0.65	0.004
rs2017349	1	178 419 417	OmniExpress	1033	0.259	*RASAL2*	**5.38**	0.07	1.034	0.57	0.004
rs12073428	1	178 427 933	OmniExpress	1030	0.157	*RASAL2*	**4.98**	**7.40E−03**	1.829	0.682	0.006
rs1008495	1	178 458 708	OmniExpress	1029	0.19	Intergenic	**4.73**	0.065	1.134	0.613	0.004
rs2252384	1	179 785 891	OmniExpress	1033	0.242	Intergenic	**6.37**	0.095	−0.937	0.561	0.004
rs2794579	1	179 787 027	OmniExpress	1033	0.243	Intergenic	**6.12**	0.09	−0.965	0.568	0.004
rs1148821	1	179 795 505	OmniExpress	1033	0.24	Intergenic	**6.05**	0.095	−0.945	0.566	0.004
rs2804699	1	181 322 837	Omni1S	1026	0.351	Intergenic	**4.91**	0.49	0.353	0.515	0
rs2804694	1	181 331 833	Omni1S	1033	0.332	Intergenic	**4.55**	0.53	0.333	0.531	0.001

Abbreviations: Chr., chromosome; LOD, logarithm of the odds; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
Boldface indicates LOD scores > 3 or *P*-values < 0.05.

overexpression observed in neuroblastoma.[28,29] *TOR1AIP2* encodes torsin A-interacting protein 2, which is involved in the nuclear envelope.[30,31] Mutations in *TOR1AIP1* have been shown to cause muscular dystrophy.[32] *RASAL2* (RAS protein activator like 2) has been implicated as an obesity susceptibility gene in both Chinese[33] and Mexican populations,[34] as well as having a role in the susceptibility of many cancers, including liver,[35] thyroid,[36] ovarian,[37] breast[37,38] and lung.[39]

Genes under the S_I linkage peaks also included interesting candidates. On chromosome 12, the most relevant gene with linkage in the distal linkage peak was *CMKLR1* (chemerin chemokine-like receptor 1), which is believed to have a role in glucose homeostasis,[40–42] obesity[41,43,44] and diabetes development.[45] Of note, a strong association signal (P-value $= 1 \times 10^{-7}$) was also seen within this linkage peak in *WSCD2* (WSC domain containing 2; 100 Mb from *CMKLR1*) (Figure 1c).

Additional genes included *LIMA1* (LIM domain and actin binding 1, also known as *EPLIN* and *SREPB3*), a tumor suppressor; *DIP2B* (disco-interacting protein 2 homolog B), replicated as a susceptibility locus for colorectal cancer;[46] and *SLC4A8*, a sodium bicarbonate transporter, which may have a role in regulation of blood pressure with some variants in this gene having been previously implicated.[47,48] Further, *KRT8* (keratin 8, type II), which is overexpressed in human liver disease, resides under the linkage peak on 12q.[49] The linkage region on chromosome 7 contained only one putative gene, *LOC102723427*, about which there is no known information.

The most intriguing signal lies in *LPHN3* and was both linked and associated with LDL levels at two separate variants. This gene encodes latrophilin 3 (recently renamed as *ADGRL3*;[50] adhesion G-protein-coupled receptor L3), which is related to latrotoxin, the toxin produced by the black widow spider.[51] There is evidence suggesting a role for latrophilin 3 (among other latrophilins) in binding to fibronectin leucine-rich transmembrane (FLRT) family members, which has been shown to promote the development of glutamatergic synapses.[52,53] Additionally, genetic variants in *LPHN3* have been associated reproducibly with attention deficit hyperactivity disorder and other psychiatric conditions.[54–56] *LPHN3* is also being investigated as a pharmacogenetic target.[57] Despite the lack of biological evidence directly supporting the link between *LPHN3* variants and LDL cholesterol levels, cholesterol is crucially important in the brain, and further study may elucidate a

Table 6 Variants with LOD score >4 and *P*-value <0.005

SNP	*Chr.*	*Position*	N	*MAF*	*Trait*	*Gene*	*Variant*	*LOD*	P-*value*	β-*Value*	*Variance*
rs17109504	1	83 468 851	965	0.2363	ApoB		Unknown	4.08	3.99E−03	0.182	0.005
rs10919343	1	170 224 982	1032	0.205	AIR		Unknown	4.32	0.003	1.86	0.012
rs10494510	1	178 074 581	1030	0.187	AIR	*RASAL2*	Intron	4.08	0.002	1.98	0.007
rs6670912	1	178 082 410	1033	0.187	AIR	*RASAL2*	Intron	4.28	0.0014	2.04	0.007
rs4440820	1	178 088 698	1034	0.186	AIR	*RASAL2*	Intron	4.18	0.0015	2.03	0.008
rs12071903	1	178 095 804	1034	0.187	AIR	*RASAL2*	Intron	4.22	0.0014	2.041	0.007
rs10798597	1	178 108 248	1032	0.185	AIR	*RASAL2*	Intron	4.01	0.0019	1.996	0.007
rs10157702	1	178 109 045	1033	0.186	AIR	*RASAL2*	Intron	4.28	0.0019	1.99	0.007
rs10913513	1	178 135 941	1034	0.186	AIR	*RASAL2*	Intron	4.08	0.0018	2.002	0.007
rs2343249	4	62 419 426	1017	0.3033	LDL	*LPHN3*	Intron	4.3	1.00E−05	−0.324	0.027
rs13245847	7	38 596 983	821	0.431	TNF2	*AMPH*	Intron	4.14	5.20E−05	−0.056	0.019
rs723968	9	14 154 231	820	0.2701	TNF2	*NFIB*	Intron	4.11	1.28E−03	−0.05	0.012
rs7044402	9	14 157 468	821	0.2966	TNF2	*NFIB*	Intron	4.19	9.05E−04	−0.049	0.012
rs16931436	9	14 185 939	821	0.2716	TNF2	*NFIB*	Intron	4.09	1.58E−03	−0.049	0.013
rs10756748	9	16 327 712	1029	0.313	HDL		Unknown	4.1	0.0027	−0.039	0.013
rs1939523	11	132 599 003	821	0.2954	TNF2	*OPCML*	Intron	4.01	3.13E−03	−0.046	0.006
rs73202582	12	92 044 537	954	0.138	Adiponectin		Unknown	4.15	0.0019	−0.091	0.02
rs9596564	13	33 508 797	1029	0.2755	Triglycerides	*PDS5B (243392)-KL (81403)*	Unknown	4.13	4.68E−03	−0.08	0.011
rs11158243	14	20 473 910	821	0.316	TNF2		Unknown	4.92	0.0037	−0.046	0.014
rs11643893	16	16 285 847	784	0.425	Percent fat	*ABCC6*	Intron	4.03	0.0034	−0.891	0.018
rs11076039	16	54 450 940	1024	0.466	HDL		Unknown	5.43	0.0011	−0.039	0.007
rs11645463	16	54 456 353	1030	0.47	HDL		Unknown	5.06	0.0049	−0.033	0.004
rs5882	16	57 016 092	1020	0.46	HDL	*CETP*	Missense V422I	4.29	4.91E−04	0.042	0.012
rs12602333	17	10 169 293	821	0.1681	TNF2	*GAS7 (245974)-MYH13 (34889)*	Unknown	4.65	3.32E−03	−0.051	0.012
rs17745091	17	52 938 797	785	0.498	Percent fat		Unknown	5.01	1.80E−04	1.156	0.014
rs2332308	17	52 944 373	784	0.4802	Percent fat	*–TOM1L1 (33678)*	Unknown	4.03	2.44E−04	1.141	0.01
rs75500748	22	48 739 692	819	0.093	TNF2		Unknown	4.21	2.70E−04	0.084	0.022

Abbreviations: Chr., chromosome; LOD, logarithm of the odds; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

mechanism by which genetic variants in *LPHN3* impact plasma LDL levels.

We previously reported *CETP* linkage and association with HDL levels in exome chip data from this Hispanic sample.[9] Linkage of *CETP* in this data set was stronger with LOD scores of up to 5.43, an increase of 1.14 over the previous top signal (Table 6 and Supplementary Table 2). The addition of GWAS data implicated additional linked variants (LOD > 5, $N = 4$) proximal to the coding region, perhaps occluding interpretation of the functional impact of this linkage result.

Here we assessed the impact of SNP density to provide insight into linkage relationships with the conclusion that dense SNP maps do reveal additional insight. We have extended this query further by evaluation of imputed genotype data in regions of particular interest because of evidence of strong linkage with glucose homeostasis-related phenotypes. Three regions were selected based on substantial linkage evidence and a particular interest in glucose homeostasis: chromosome 1 with AIR and chromosomes 7 and 12 with S_I. Utilization of imputed data increases the number of markers capturing the region by 22-fold (18 411 directly genotyped markers, 406 K imputed markers). The maximal LOD score from the imputed AIR region was 6.45 at rs2252384 (the same SNP implicated in the directly genotyped data; Supplementary Figure 2). The slight increase in LOD score (6.37–6.45) can likely be attributed to more complete information following imputation of missing genotypes. For chromosome 7 with S_I, a new best SNP rs2530421 had the maximum LOD score of 5.53 (compared with the prior best LOD of 5.11 at rs1024591). The imputed best SNP lies very near the original peak linkage, providing little additional guidance in refining the causal variant(s), given the high degree of correlation between the top-linked SNPs ($r^2 = 0.937$). Evaluation of another linked region (chromosome 12 with S_I) also showed some limited improvement in linkage signals, but linkage signals were only modestly increased, as could be expected because of the information carried by these imputed markers being wholly derived from the genotyped markers, which had already been informative. Thus, inclusion of imputed genotypes marginally improved the maximal LOD scores when evaluated in this small number of examples. However, the improvements did not further refine the regions of interest (Supplementary Figure 2).

In conclusion, we have built upon our previous analysis of combined two-point linkage and association[9] and evaluated utility of the approach in a data set comprised of comprehensive genome-wide array-based SNP genotypes. As seen previously, there were few examples in these data where linkage and association both provided striking evidence at the same locus, which, based on our prior analysis,[10] would implicate a likely ungentoyped causal variant. However, the GWAS plus exome chip design identified multiple additional regions of linkage, which were not seen in exome chip analysis alone. Positive, strong evidence of association with SNPs was not observed, suggesting that functional variants, if they are indeed captured by the linkage signal, have not been identified. To truly test the broad utility of this approach, whole-genome sequencing data will be necessary, which will incorporate the full spectrum of variant frequencies.

Table 7 *LPHN3* linkage and association with LDL levels

SNP	*Chr.*	*Position*	*Chip*	N	*MAF*	*LOD*	P-*value*	β-*Value*	*s.e.*	*Variance*
rs17828264	4	62 079 015	Omni1S	1021	0.5	1.17	0.44	−0.051	0.066	0
rs17090416	4	62 098 937	OmniExpress	1022	0.279	1.42	0.65	−0.034	0.074	0
rs1505682	4	62 111 856	OmniExpress	1022	0.315	1.44	0.22	−0.089	0.073	0.001
rs1505670	4	62 115 243	Omni1S	1021	0.475	1.26	0.69	−0.027	0.067	0
rs13140257	4	62 128 750	Omni1S	999	0.321	1.66	**0.037**	−0.152	0.073	0.003
rs11723103	4	62 128 825	Omni1S	1019	0.375	1.33	0.052	−0.137	0.07	0.004
rs1505663	4	62 132 090	OmniExpress	1022	0.229	0.15	**7.90E−03**	0.213	0.08	0.003
rs1505664	4	62 132 345	OmniExpress	1020	0.371	1.42	**0.05**	−0.137	0.07	0.004
rs67050759	4	62 135 455	Omni1S	1019	0.496	1.49	0.12	−0.105	0.068	0.003
rs74329144	4	62 136 292	Omni1S	1022	0.055	1.02	0.076	0.263	0.148	0.002
rs77082869	4	62 254 565	Omni1S	1021	0.015	0.00	**1.77E−03**	0.896	0.287	0.013
rs10008278	4	62 366 666	OmniExpress	1018	0.092	1.28	0.096	0.2	0.12	0.003
rs904243	4	62 406 445	OmniExpress	1021	0.164	0.75	**6.49E−04**	−0.312	0.091	0.018
rs7656189	4	62 411 676	OmniExpress	1020	0.408	0.74	**4.07E−03**	0.2	0.069	0.013
rs9312078	4	62 412 292	OmniExpress	1015	0.331	**3.02**	**8.20E−05**	−0.282	0.071	0.022
rs56905501	4	62 413 961	Omni1S	1018	0.392	0.69	**2.98E−03**	0.207	0.07	0.014
rs7688741	4	62 416 470	Omni1S	1022	0.383	1.46	**2.11E−04**	−0.262	0.071	0.019
rs2132074	4	62 416 499	OmniExpress	1021	0.392	0.64	**1.86E−03**	0.216	0.069	0.014
rs2343249	4	62 419 426	OmniExpress	1017	0.303	**4.30**	**1.00E−05**	−0.324	0.073	0.027
rs958862	4	62 434 848	OmniExpress	1018	0.341	1.87	**3.60E−04**	−0.258	0.072	0.02
rs10018746	4	62 445 246	Omni1S	1021	0.5	0.97	**4.19E−03**	0.192	0.067	0.013
rs11941524	4	62 446 484	Omni1S	1022	0.5	0.86	**4.17E−03**	0.192	0.067	0.013
rs2172802	4	62 453 209	Exome	1012	0.45	0.50	**6.37E−03**	0.184	0.067	0.01
rs17239080	4	62 455 462	OmniExpress	1022	0.374	2.02	**2.32E−03**	−0.212	0.069	0.014
rs11131334	4	62 457 454	OmniExpress	1017	0.379	2.11	**4.84E−03**	−0.195	0.069	0.011
rs1497901	4	62 461 940	OmniExpress	1021	0.359	2.07	**1.77E−03**	−0.221	0.07	0.013
rs2343250	4	62 472 682	Omni1S	1022	0.36	2.09	**1.59E−03**	−0.224	0.071	0.013
rs10001410	4	62 474 229	OmniExpress	1019	0.47	0.91	**3.89E−03**	−0.199	0.069	0.016
rs1497921	4	62 526 281	OmniExpress	1022	0.356	0.64	**3.19E−03**	−0.204	0.069	0.014
rs66614141	4	62 550 335	Omni1S	1022	0.326	1.45	**1.35E−04**	−0.268	0.07	0.02
rs6843311	4	62 568 688	OmniExpress	1022	0.363	0.61	**5.25E-03**	–0.194	0.069	0.014
rs11734607	4	62 693 692	OmniExpress	1021	0.453	0.24	**2.44E−03**	0.204	0.067	0.015
rs4860106	4	62 850 522	OmniExpress	1021	0.422	1.13	0.71	0.025	0.068	0
rs1510921	4	62 895 592	OmniExpress	1017	0.241	0.26	**4.00E−03**	0.223	0.077	0.007
rs6827266	4	62 902 162	Omni1S	1020	0.437	0.08	**5.00E−03**	0.188	0.067	0.004
rs62306380	4	62 908 281	Omni1S	1022	0.239	0.23	**3.55E−03**	0.225	0.077	0.007

Abbreviations: Chr., chromosome; LDL, low-density lipoprotein; LOD, logarithm of the odds; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
Boldface indicates LOD scores >3 or *P*-values <0.05.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Grants R01 HG007112 (to DWB and CDL) and R01 DK087914 (to MCYN). The GUARDIAN study, which contributed the IRASFS GWAS genotypes to this project is supported by Grant R01 DK085175 (to LEW), and the IRASFS study was supported by HL060944, HL061019 and HL060919. The provision of GWAS genotyping data was supported, in part, by UL1TR000124 (CTSI), and DK063491 (DRC). The provision of exome chip data was supported, in part, by the Department of Internal Medicine at University of Michigan, the Doris Duke Medical Foundation and R01 DK106621 (to EKS). Computational support was provided, in part, by the Center for Public Health Genomics at Wake Forest School of Medicine.

1 Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nat. Rev. Genet.* **12**, 465–474 (2011).
2 Speed, D. & Balding, D. J. Relatedness in the post-genomic era: is it still useful? *Nat. Rev. Genet.* **16**, 33–44 (2015).
3 Bowden, D. W., An, S. S., Palmer, N. D., Brown, W. M., Norris, J. M., Haffner, S. M. *et al.* Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Hum. Mol. Genet.* **19**, 4112–4120 (2010).
4 Bowden, D. W. Will family studies return to prominence in human genetics and genomics? Rare variants and linkage analysis of complex traits. *Genes Genomics* **33**, 1–8 (2011).
5 Wang, Q., Lu, Q. & Zhao, H. A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing. *Front. Genet.* **6**, 149 (2015).
6 Teare, M. D. & Santibanez Koref, M. F. Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. *Brief. Funct. Genomics* **13**, 378–383 (2014).
7 Ott, J., Wang, J. & Leal, S. M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat. Rev. Genet.* **16**, 275–284 (2015).
8 Saad, M. & Wijsman, E. M. Power of family-based association designs to detect rare variants in large pedigrees using imputed genotypes. *Genet. Epidemiol.* **38**, 1–9 (2014).
9 Hellwege, J. N., Palmer, N. D., Raffield, L. M., Ng, M. C., Hawkins, G. A., Long, J. *et al.* Genome-wide family-based linkage analysis of exome chip variants and cardiometabolic risk. *Genet. Epidemiol.* **38**, 345–352 (2014).
10 Hellwege, J. N., Palmer, N. D., Brown, W. M., Ziegler, J. T., Sandy An, S., Guo, X. *et al.* Empirical characteristics of family-based linkage to a complex trait: the ADIPOQ region and adiponectin levels. *Hum. Genet.* **134**, 203–213 (2015).

11 Voruganti, V. S., Kent, J. W. Jr, Debnath, S., Cole, S. A., Haack, K., Göring, H. H. *et al.* Genome-wide association analysis confirms and extends the association of SLC2A9 with serum uric acid levels to Mexican Americans. *Front. Genet.* **4**, 279 (2013).
12 Nemesure, B. B., He, Q. & Mendell, N. Integration of linkage analyses and disease association studies. *Genet. Epidemiol.* **12**, 653–658 (1995).
13 Heo, M., Leibel, R. L., Fontaine, K. R., Boyer, B. B., Chung, W. K., Koulu, M. *et al.* A meta-analytic investigation of linkage and association of common leptin receptor (LEPR) polymorphisms with body mass index and waist circumference. *Int. J. Obesity Relat. Metab. Disord.* **26**, 640–646 (2002).
14 Weedon, M. N., Cebola, I., Patch, A. M., Flanagan, S. E., De Franco, E., Caswell, R. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
15 Henkin, L., Bergman, R. N., Bowden, D. W., Ellsworth, D. L., Haffner, S. M., Langefeld, C. D. *et al.* Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS Family Study design and methods. *Ann. Epidemiol.* **13**, 211–217 (2003).
16 Palmer, N. D., Goodarzi, M. O., Langefeld, C. D., Wang, N., Guo, X., Taylor, K. D. *et al.* Genetic variants associated with quantitative glucose homeostasis traits translate to type 2 diabetes in Mexican Americans: the GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium. *Diabetes* **64**, 1853–1866 (2014).
17 O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
18 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
19 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
20 Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
21 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
22 Das, S. K., Hasstedt, S. J., Zhang, Z. & Elbein, S. C. Linkage and association mapping of a chromosome 1q21–q24 type 2 diabetes susceptibility locus in northern European Caucasians. *Diabetes* **53**, 492–499 (2004).
23 Langefeld, C. D., Wagenknecht, L. E., Rotter, J. I., Williams, A. H., Hokanson, J. E., Saad, M. F. *et al.* Linkage of the metabolic syndrome to 1q23–q31 in Hispanic families: the Insulin Resistance Atherosclerosis Study Family Study. *Diabetes* **53**, 1170–1174 (2004).
24 Wiltshire, S., Hattersley, A. T., Hitman, G. A., Walker, M., Levy, J. C., Sampson, M. *et al.* A genomewide scan for loci predisposing to type 2 diabetes in a UK population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.* **69**, 553–569 (2001).
25 Vionnet, N., Hani, E. H., Dupont, S., Gallina, S., Francke, S., Dotte, S. *et al.* Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21–q24. *Am. J. Hum. Genet.* **67**, 1470–1480 (2000).
26 Adair-Kirk, T. L., Atkinson, J. J., Kelley, D. G., Arch, R. H., Miner, J. H. & Senior, R. M. A chemotactic peptide from laminin alpha 5 functions as a regulator of inflammatory immune responses via TNF alpha-mediated signaling. *J. Immunol. (Baltimore, Md: 1950)* **174**, 1621–1629 (2005).
27 Prados, J., Stenz, L., Courtet, P., Prada, P., Nicastro, R., Adouan, W. *et al.* Borderline personality disorder and childhood maltreatment: a genome-wide methylation analysis. *Genes Brain Behav.* **14**, 177–188 (2015).
28 Vasudevan, S. A., Russell, H. V., Okcu, M. F., Burlingame, S. M., Liu, Z. J., Yang, J. *et al.* Neuroblastoma-derived secretory protein messenger RNA levels correlate with high-risk neuroblastoma. *J. Pediatr. Surg.* **42**, 148–152 (2007).
29 Vasudevan, S. A., Shang, X., Chang, S., Ge, N., Diaz-Miron, J. L., Russell, H. V. *et al.* Neuroblastoma-derived secretory protein is a novel secreted factor overexpressed in neuroblastoma. *Mol. Cancer Ther.* **8**, 2478–2489 (2009).
30 Goodchild, R. E., Kim, C. E. & Dauer, W. T. Loss of the dystonia-associated protein torsinA selectively disrupts the neuronal nuclear envelope. *Neuron* **48**, 923–932 (2005).
31 Vander Heyden, A. B., Naismith, T. V., Snapp, E. L., Hodzic, D. & Hanson, P. I. LULL1 retargets TorsinA to the nuclear envelope revealing an activity that is impaired by the DYT1 dystonia mutation. *Mol. Biol. Cell* **20**, 2661–2672 (2009).
32 Kayman-Kurekci, G., Talim, B., Korkusuz, P., Sayar, N., Sarioglu, T., Oncel, I. *et al.* Mutation in TOR1AIP1 encoding LAP1B in a form of muscular dystrophy: a novel gene related to nuclear envelopathies. *Neuromusc. Disord.* **24**, 624–633 (2014).
33 Cheung, C. Y., Tso, A. W., Cheung, B. M., Xu, A., Ong, K. L., Fong, C. H. *et al.* Obesity susceptibility genetic variants identified from recent genome-wide association studies: implications in a Chinese population. *J. Clin. Endocrinol. Metab.* **95**, 1395–1403 (2010).
34 Leon-Mimila, P., Villamil-Ramirez, H., Villalobos-Comparan, M., Villarreal-Molina, T., Romero-Hidalgo, S., López-Contreras, B. *et al.* Contribution of common genetic variants to obesity and obesity-related traits in Mexican children and adults. *PLoS ONE* **8**, e70640 (2013).
35 Stefanska, B., Cheishvili, D., Suderman, M., Arakelian, A., Huang, J., Hallett, M. *et al.* Genome-wide study of hypomethylated and induced genes in patients with liver cancer unravels novel anticancer targets. *Clin. Cancer Res.* **20**, 3118–3132 (2014).
36 Xu, Y., Deng, Y., Ji, Z., Liu, H., Liu, Y., Peng, H. *et al.* Identification of thyroid carcinoma related genes with mRMR and shortest path approaches. *PLoS ONE* **9**, e94022 (2014).
37 Feng, M., Bao, Y., Li, Z., Li, J., Gong, M., Lam, S. *et al.* RASAL2 activates RAC1 to promote triple-negative breast cancer progression. *J. Clin. Invest.* **124**, 5291–5304 (2014).
38 Huang, Y., Zhao, M., Xu, H., Wang, K., Fu, Z., Jiang, Y. *et al.* RASAL2 down-regulation in ovarian cancer promotes epithelial-mesenchymal transition and metastasis. *Oncotarget* **5**, 6734–6745 (2014).
39 Li, N. & Li, S. RASAL2 promotes lung cancer metastasis through epithelial-mesenchymal transition. *Biochem. Biophys. Res. Commun.* **455**, 358–362 (2014).
40 Rourke, J. L., Muruganandan, S., Dranse, H. J., McMullen, N. M. & Sinal, C. J. Gpr1 is an active chemerin receptor influencing glucose homeostasis in obese mice. *J. Endocrinol.* **222**, 201–215 (2014).
41 Ernst, M. C., Haidl, I. D., Zuniga, L. A., Dranse, H. J., Rourke, J. L., Zabel, B. A. *et al.* Disruption of the chemokine-like receptor-1 (CMKLR1) gene is associated with reduced adiposity and glucose intolerance. *Endocrinology* **153**, 672–682 (2012).
42 Sell, H., Laurencikiene, J., Taube, A., Eckardt, K., Cramer, A., Horrighs, A. *et al.* Chemerin is a novel adipocyte-derived factor inducing insulin resistance in primary human skeletal muscle cells. *Diabetes* **58**, 2731–2740 (2009).
43 Gruben, N., Aparicio Vergara, M., Kloosterhuis, N. J., van der Molen, H., Stoelwinder, S., Youssef, S. *et al.* Chemokine-like receptor 1 deficiency does not affect the development of insulin resistance and nonalcoholic fatty liver disease in mice. *PLoS ONE* **9**, e96345 (2014).
44 Kim, S. H., Lee, S. H., Ahn, K. Y., Lee, D. H., Suh, Y. J., Cho, S. G. *et al.* Effect of lifestyle modification on serum chemerin concentration and its association with insulin sensitivity in overweight and obese adults with type 2 diabetes. *Clin Endocrinol (Oxf)* **80**, 825–833 (2014).
45 Roman, A. A., Parlee, S. D. & Sinal, C. J. Chemerin: a potential endocrine link between obesity and type 2 diabetes. *Endocrine* **42**, 243–251 (2012).
46 Closa, A., Cordero, D., Sanz-Pamplona, R., Solé, X., Crous-Bou, M., Paré-Brunet, L. *et al.* Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis* **35**, 2039–2046 (2014).
47 Guo, L., Liu, F., Chen, S., Yang, X., Huang, J., He, J. *et al.* Common variants in the Na-coupled bicarbonate transporter genes and salt sensitivity of blood pressure: the GenSalt study. *J. Hum. Hypertens.* **30**, 543–548 (2015).
48 Aalkjaer, C., Boedtkjer, E., Choi, I. & Lee, S. Cation-coupled bicarbonate transporters. *Compr. Physiol.* **4**, 1605–1637 (2014).
49 Guldiken, N., Usachov, V., Levada, K., Trautwein, C., Ziol, M., Nahon, P. *et al.* Keratins 8 and 18 are type II acute-phase responsive genes overexpressed in human liver disease. *Liver Int.* **35**, 1203–1212 (2015).
50 Hamann, J., Aust, G., Arac, D., Engel, F. B., Formstone, C., Fredriksson, R. *et al.* International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors. *Pharmacol. Rev.* **67**, 338–367 (2015).
51 Martinez, A. F., Muenke, M. & Arcos-Burgos, M. From the black widow spider to human behavior: Latrophilins, a relatively unknown class of G protein-coupled receptors, are implicated in psychiatric disorders. *Am. J. Med. Genet. B* **156b**, 1–10 (2011).
52 Jackson, V. A., del Toro, D., Carrasquero, M., Roversi, P., Harlos, K., Klein, R. *et al.* Structural basis of latrophilin-FLRT interaction. *Structure (London, England: 1993)* **23**, 774–781 (2015)
53 O'Sullivan, M. L., de Wit, J., Savas, J. N., Comoletti, D., Otto-Hitt, S., Yates, J. R. III *et al.* FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. *Neuron* **73**, 903–910 (2012).
54 Fallgatter, A. J., Ehlis, A. C., Dresler, T., Reif, A., Jacob, C. P., Arcos-Burgos, M. *et al.* Influence of a latrophilin 3 (LPHN3) risk haplotype on event-related potential measures of cognitive response control in attention-deficit hyperactivity disorder (ADHD). *Eur. Neuropsychopharmacol.* **23**, 458–468 (2013).
55 Ribases, M., Ramos-Quiroga, J. A., Sanchez-Mora, C., Bosch, R., Richarte, V., Palomar, G. *et al.* Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study. *Genes Brain Behav.* **10**, 149–157 (2011).
56 Arcos-Burgos, M., Jain, M., Acosta, M. T., Stanescu, H., Wallis, D., Domené, S. *et al.* A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Mol. Psychiatry* **15**, 1053–1066 (2010).
57 Bruxel, E. M., Salatino-Oliveira, A., Akutagava-Martins, G. C., Tovo-Rodrigues, L., Genro, J. P., Zeni, C. P. *et al.* LPHN3 and attention-deficit/hyperactivity disorder: a susceptibility and pharmacogenetic study. *Genes Brain Behav* **14**, 419–427 (2015).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)