GENETICS/GENOMES/PROTEOMICS/METABOLOMICS

Recessive Genome-Wide Meta-analysis Illuminates Genetic Architecture of Type 2 Diabetes

Mark J. O'Connor,[1,2,3,4,5] Philip Schroeder,[3,4,5] Alicia Huerta-Chagoya,[6] Paula Cortés-Sánchez,[7] Silvía Bonàs-Guarch,[7] Marta Guindo-Martínez,[7] Joanne B. Cole,[4,5,8,9] Varinderpal Kaur,[3,4,5] David Torrents,[7,10] Kumar Veerapen,[8,11,12] Niels Grarup,[13] Mitja Kurki,[8,11,12] Carsten F. Rundsten,[13] Oluf Pedersen,[13] Ivan Brandslund,[14,15] Allan Linneberg,[16,17] Torben Hansen,[13] Aaron Leong,[1,2,3,4,5,8,18] Jose C. Florez,[1,2,3,4,5,8] and Josep M. Mercader[3,4,5,8]

Diabetes 2022;71:554–565 | https://doi.org/10.2337/db21-0545

Most genome-wide association studies (GWAS) of complex traits are performed using models with additive allelic effects. Hundreds of loci associated with type 2 diabetes have been identified using this approach. Additive models, however, can miss loci with recessive effects, thereby leaving potentially important genes undiscovered. We conducted the largest GWAS meta-analysis using a recessive model for type 2 diabetes. Our discovery sample included 33,139 case subjects and 279,507 control subjects from 7 European-ancestry cohorts, including the UK Biobank. We identified 51 loci associated with type 2 diabetes, including five variants undetected by prior additive analyses. Two of the five variants had minor allele frequency of <5% and were each associated with more than a doubled risk in homozygous carriers. Using two additional cohorts, FinnGen and a Danish cohort, we replicated three of the variants, including one of the low-frequency variants, rs115018790, which had an odds ratio in homozygous carriers of 2.56 (95% CI 2.05–3.19; $P = 1 \times 10^{-16}$) and a stronger effect in men than in women (for interaction, $P = 7 \times 10^{-7}$). The signal was associated with multiple diabetes-related traits, with homozygous carriers showing a 10% decrease in LDL cholesterol and a 20% increase in triglycerides; colocalization analysis linked this signal to reduced expression of the nearby *PELO* gene. These results demonstrate that recessive models, when compared with GWAS using the additive approach, can identify novel loci, including large-effect variants with pathophysiological consequences relevant to type 2 diabetes.

Type 2 diabetes affects nearly 1 in 12 adults globally (1), but its genetic architecture is still not fully understood. Over the past decade, large genome-wide association studies (GWAS) have used additive models to identify hundreds of associated loci (2–5). Additive models are most

[1] Department of Medicine, Massachusetts General Hospital, Boston, MA
[2] Endocrine Division, Massachusetts General Hospital, Boston, MA
[3] Diabetes Unit, Massachusetts General Hospital, Boston, MA
[4] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
[5] Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA
[6] Consejo Nacional de Ciencia y Tecnología (CONACYT), Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
[7] Barcelona Supercomputing Center (BSC), Barcelona, Spain
[8] Department of Medicine, Harvard Medical School, Boston, MA
[9] Center for Basic and Translations Obesity Research, Boston Children's Hospital, Boston, MA
[10] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[11] Stanley Center for Psychiatric Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA
[12] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA
[13] Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[14] Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark
[15] Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark
[16] Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
[17] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[18] Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA

Corresponding author: Josep M. Mercader, mercader@broadinstitute.org

Received 22 June 2021 and accepted 28 November 2021

A.L., J.C.F., and J.M.M. jointly directed this work.

This article contains supplementary material online at https://doi.org/10.2337/figshare.17099780

powerful when the effect of two copies of a risk allele is twice that of one copy. This model is computationally simple and statistically powerful, but it does not always match the pattern of inheritance of Mendelian disorders, including monogenic forms of diabetes, which can be transmitted in a dominant or recessive fashion (6). Variants with recessive effects, particularly low-frequency variants, can go undetected by additive models (7), suggesting that nonadditive models have the potential to generate new biological insights.

To date, recessive models have been used in a handful of studies to identify genetic associations with type 2 diabetes, but these have been limited by small sample sizes (7–9). Nevertheless, some promising findings have emerged. In a Greenlandic population, homozygous carriers of two copies of a nonsense mutation in the *TBC1D4* gene, which facilitates glucose transfer into skeletal muscle in the setting of insulin stimulation, had a 10-fold increase in diabetes risk compared with other individuals in the same population (8). Hyperglycemia due to this variant occurs postprandially, so the diagnosis of type 2 diabetes in homozygous carriers often requires an oral glucose tolerance test, creating an opportunity for precision medicine (10). More recently, members of our group conducted GWAS with nonadditive models for several age-related diseases (11) and identified multiple new loci, including one rare variant (rs77704739) associated with type 2 diabetes. This variant was also associated with reduced expression of the *PELO* gene, whose connection to diabetes is not well understood.

We have conducted potentially the largest GWAS meta-analysis reported to date using a recessive model for type 2 diabetes. Over the past few years, GWAS sample sizes have grown exponentially (12), and reference panels for imputation have improved, making it easier to ascertain low-frequency variants accurately (13). To take advantage of these developments, we combined data from seven discovery cohorts and two replication cohorts to conduct the recessive-model GWAS for type 2 diabetes or any other disease. We identified and replicated multiple variants missed by larger additive studies, confirmed and fine mapped the association near *PELO*, and conducted a phenome-wide association analysis to identify other affected traits to better understand the pathophysiology underlying this novel association.

RESEARCH DESIGN AND METHODS

Study Population and Outcome Definition

We used data from multiple European-ancestry cohorts (Supplementary Table 1) including the UK Biobank (14), five cohorts known collectively as 70K for T2D (4), and the Mass General Brigham (MGB) Biobank (15). The UK Biobank is a sample of approximately half a million people recruited in the United Kingdom between the ages of 40 and 69 years. The 70K for T2D cohort consists of 5 studies with publicly available data, and the MGB Biobank consists of ∼50,000 people recruited within a hospital system in the United States. We only considered individuals whose family relatedness was lower than that of third-degree relatives.

Definitions of type 2 diabetes varied according to cohort. In the UK Biobank, for example, we used a validated algorithm designed specifically to identify cases of diabetes in that cohort (16). In the MGB Biobank, type 2 diabetes was defined according to an algorithm developed by the Biobank team (17) to have 99% positive predictive value. In the UK and MGB Biobanks, which both have a relatively low prevalence of type 2 diabetes, we excluded control subjects younger than 55 years, because the mean age of onset for type 2 diabetes is ∼50 years (18).

Recessive Genome-Wide Meta-Analysis

Genotyping, phasing, and imputation, as well as sample and variant quality control, were conducted according to cohort-specific protocols (Supplementary Table 1). For the recessive analysis in each cohort, we controlled for age, sex, BMI, and principal components. For the UK Biobank, we also controlled for the genotyping platform, because two different genotyping arrays were used. For one of the five cohorts within 70K for T2D (6% of the cases in our discovery sample), age and BMI data were not available. In our models, we used the minor allele in Europeans as the recessive allele, not necessarily the nonreference allele, to maximize our chances of identifying variants missed by prior GWAS.

For the UK and MGB Biobanks, computations were conducted using Hail, version 0.2 (https://hail.is), and the 70K for T2D cohort was analyzed using the program SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html). After generating summary statistics using a recessive model for each cohort, we used the program METAL to meta-analyze the results (19), weighting cohorts by the inverse of the SE for each variant. Our threshold for genome-wide significance was $P = 5 \times 10^{-8}$, and we considered signals within 0.5 megabase pairs (Mb) to be part of the same locus. For comparison, we repeated our approach using an additive model. To visually inspect each genome-wide significant locus, we used the program LocusZoom (20). We estimated the power of our recessive and additive models to detect variants acting recessively across a range of allele frequencies and effect sizes, using a simulation-based approach, assuming a baseline case prevalence of 10%, similar to our case-control ratio.

Defining Novel Recessive Signals

We compared our results with those of the largest additive GWAS with available summary statistics (2,3), and we defined signals as novel if they were not in significant linkage disequilibrium (LD) with a known signal ($r^2 < 0.3$). This analysis was conducted with R, version 3.6 (https://www.R-project.org) and the R package *LDlinkR* (21,22). The LD information was calculated using a British reference panel (1000 Genomes Project). For each signal, we used PLINK, version 1.9 (23) to calculate dominance deviation *P* values (24) using data from the UK Biobank and GERA, the largest cohort within 70K for T2D, and then we meta-analyzed the results. Signals were deemed to be nonadditive if this *P* was <0.05. To ensure that signals near the major histocompatibility complex (MHC) region were not due to contamination of our cases with cases of type 1 diabetes, which is known to be heavily associated with haplotypes in the MHC region, we performed conditional analysis in the UK Biobank sample, adjusting for MHC haplotypes relevant to type 1 diabetes (25). We excluded variants that lost significance by more than one order of magnitude.

Replication

We attempted to replicate our novel findings in two cohorts: FinnGen and a Danish cohort (Supplementary Table 2). FinnGen is a study based in Finland that combines genotyping with digital health data of >100,000 people, and the Danish cohort consists of >20,000 individuals (22% cases) from Denmark. The program SNPTEST was used to analyze both cohorts. We meta-analyzed the results from the replication cohorts with our initial results, using the R package *rmeta*.

Credible Sets

For each novel variant, we identified the set of variants with 99% probability of containing the causal variant. We used a Bayesian refinement approach (26), considering variants in LD with the lead variant ($r^2 > 0.1$). Each credible set is akin to a CI for the true causal variant. Within a locus, each variant is assigned an approximate Bayes factor (ABF) on the basis of the following equation:

$$ABF = \sqrt{1 - r}e^{rz^2/2}$$

where $r = 0.04/(SE^2+0.04)$ and $z = \beta/SE$. The β and SE values are the estimated effect size and corresponding SE, respectively, from the recessive-model logistic regression. This calculation assumes a Gaussian prior with mean of 0 and variance of 0.04. The posterior probability for a variant is equal to its ABF divided by the sum of all ABF values for the locus. Variants are ranked by ABF in decreasing order, and the cumulative probability is calculated starting at the top of the list and stopping when the value exceeds 99%.

Colocalization with Gene Expression

To shed light on variants' functional consequences, we used the Genotype-Tissue Expression (GTEx) project, version 8 (27). This database links genetic variants with tissue-specific gene expression, enabling identification of expression quantitative trait loci (eQTLs). For our most significant variant, which was an eQTL for the gene *PELO*, we performed colocalization analysis to confirm that our GWAS signal matched the signal influencing gene expression. Colocalization analysis compares *P* values for two traits across a locus to generate a posterior probability for the hypothesis that both traits are being influenced by the same variant. Because of the rarity of homozygous carriers of our variants, we used additive summary statistics from GTex for this analysis. We used the R package "coloc" (28) and considered a window of 1 Mb around our leading signal. We also investigated our most significant variant in the Translational Human Pancreatic Islet Genotype Tissue-Expression Resource (https://tiger.bsc.es), a genotype-expression resource with data from >500 pancreatic islets (29).

Phenome- and Metabolome-Wide Association Study

For the variant located near the *PELO* gene, we performed a phenome-wide association study (PheWAS) in the UK Biobank, which provides detailed information about each participant's health, dietary habits, and lifestyle characteristics. Phenotypes were curated and transformed using the Phenome Scan Analysis Tool (30). As in our GWAS, we used a recessive model. We used logistic regression for binary phenotypes and linear regression for continuous phenotypes. We controlled for age, sex, 10 principal components, and the genotyping platform. Limiting our binary phenotypes to those with more than five cases among homozygotes for the risk variant, we analyzed 1,731 binary phenotypes, 30 biomarkers such as cholesterol levels, and 1,345 other continuous phenotypes. We also analyzed a subset of UK Biobank participants (90,644 participants) with metabolomic data (*n* = 249 metabolites) generated by nuclear magnetic resonance by Nightingale Health (31). To illustrate our metabolomic results, we generated volcano plots using the R package *EnhancedVolcano* and a heat map using the R package *ggplot2*. For significant associations, we used colocalization analysis to quantify the probability that the phenotype shared the same causal variant as type 2 diabetes. We used the R package *mediation* for mediation analysis (32). We also performed a PheWAS in the Danish cohort; we looked at 16 glycemic traits, using the same covariates as we did when analyzing the UK Biobank data.

Sex-Stratified Analysis

To test whether the genetic effects of the variant near the *PELO* gene differed by sex, we performed a sex-stratified analysis within the UK Biobank for the biomarkers in our data set and also for type 2 diabetes itself. We assessed the significance of the difference between sexes by including an interaction term in our regression model. We then confirmed sex-specific differences for type 2 diabetes in our two replication cohorts.

Data and Resource Availability

The complete summary statistics from this study will be deposited and made available at the Common Metabolic Diseases Knowledge Portal (https://cmdkp.org/).

RESULTS

Genome-Wide Meta-Analysis Using a Recessive Model

Our discovery sample consisted of 33,139 case subjects with type 2 diabetes and 279,507 control subjects from seven cohorts. We meta-analyzed 11,634,328 variants and fitted additive and recessive models to compare the results. We identified 51 loci (Supplementary Table 3) that reached genome-wide significance in the recessive model, and 121 loci using the additive model (Fig. 1). Of the 51 signals identified with the recessive model, 33% deviated from additivity (for dominance deviation, $P < 0.05$), and of these, five were distinct from the set of previously reported additive signals (Table 1).

The strongest recessive signal (rs115018790) was located within an intron of the *PELO* and *ITGA1* genes on chromosome 5 (Fig. 2) and was in complete LD ($r^2 = 1$) with the lead variant (rs77704739) that was previously identified in the GERA cohort (11), one of the discovery cohorts in this study. With minor allele frequency (MAF) 0.04, rs115018790 had an odds ratio (OR) for homozygous

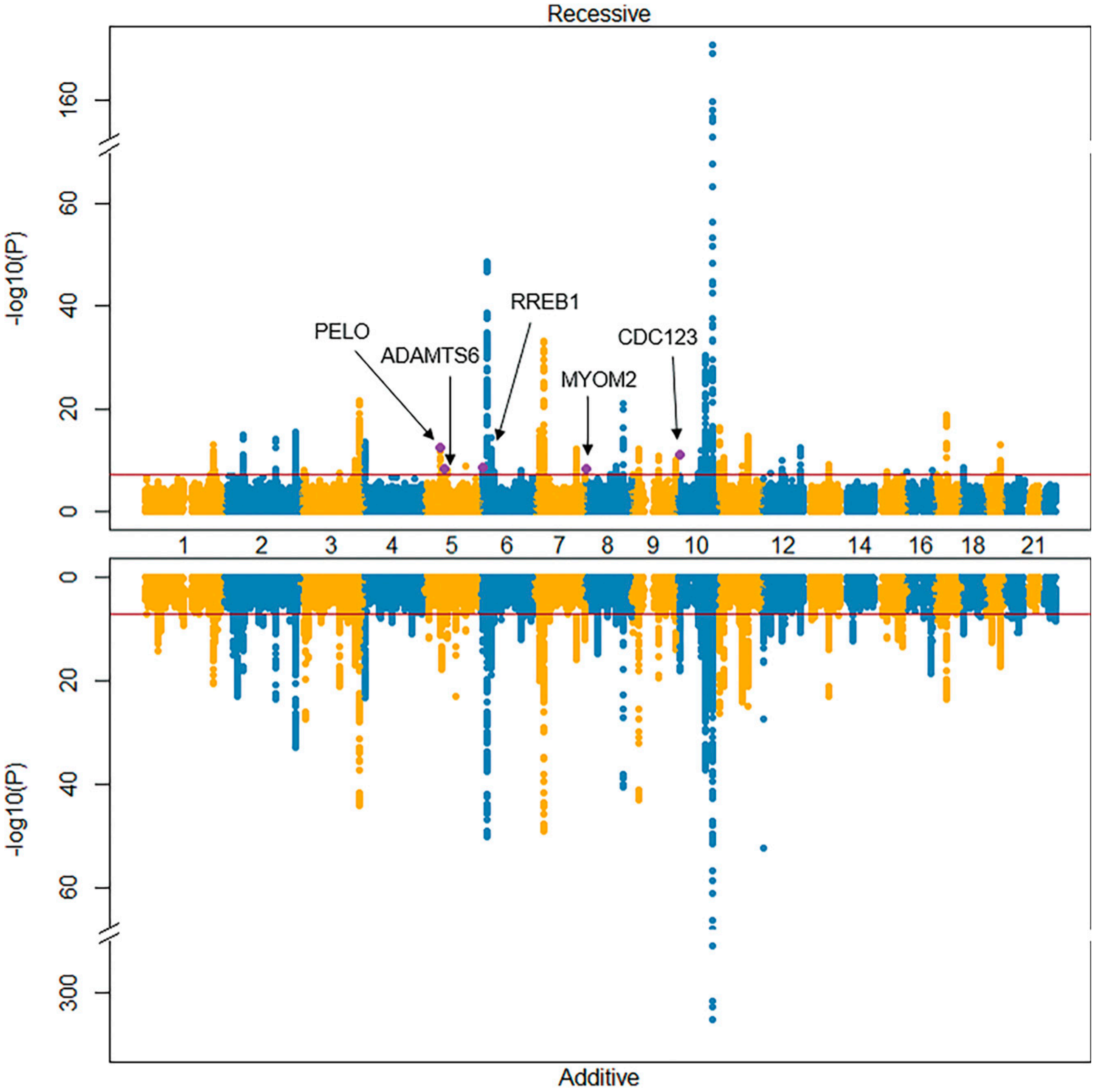


Figure 1—Miami plot comparing recessive and additive results. Nonadditive signals are purple and labeled. The dark red line is the threshold for genome-wide significance.

Table 1—Novel recessively acting variants

Variant (position)*	Major allele	Minor allele	MAF	Nearest gene	OR (95% CI), *P* value				Dominance deviation *P* value
					Discovery		Replication	Combined	
					Additive	Recessive	Recessive	Recessive	
rs115018790 (5:52088271)	G	A	0.04	*PELO*	1.07 (1.02–1.12), 0.004	2.63 (2.03–3.41), 3×10^{-13}	2.37 (1.55–3.62), 6×10^{-5}	2.56 (2.05–3.19), 1×10^{-16}	3×10^{-11}
rs140453320 (5:64485239)	T	C	0.01	*ADAMTS6*	1.03 (0.95–1.11), 0.49	6.94 (3.63–13.27), 5×10^{-9}	1.18 (0.23–6.07), 0.84	5.46 (2.99–9.98), 3×10^{-8}	5×10^{-8}
rs2714337 (6:7240577)	A	T	0.35	*RREB1*	1.06 (1.04–1.08), 1×10^{-10}	1.12 (1.08–1.16), 3×10^{-9}	1.10 (1.06–1.15), 5×10^{-6}	1.11 (1.08–1.14), 7×10^{-14}	0.02
rs755900673 (8:2008956)	TC	T	0.36	*MYOM2*	1.04 (1.02–1.06), 3×10^{-5}	1.13 (1.08–1.17), 5×10^{-9}	1.01 (0.96–1.06), 0.64	1.08 (1.05–1.12), 2×10^{-6}	6×10^{-4}
rs33932777 (10:12311465)	G	A	0.50	*CDC123*	1.06 (1.04–1.08), 8×10^{-11}	1.11 (1.07–1.14), 9×10^{-12}	1.07 (1.03–1.11), 4×10^{-5}	1.09 (1.07–1.12), 4×10^{-15}	0.01

*Position is from genome assembly GRCh37 (hg19). For replication, variant rs140453320 was only assessed in the FinnGen cohort, because there were only three homozygotes in the Danish cohort. Dominance deviation *P* values were calculated in the UK Biobank and GERA cohorts and meta-analyzed.

carriers of 2.63 (95% CI 2.03–3.41), much greater than the additive-model OR of 1.07 (95% CI 1.02–1.12). The *P* value for the recessive model ($P = 3 \times 10^{-13}$) was 10 orders of magnitude more significant than the additive one, and the dominance deviation test confirmed the variant's recessive nature ($P = 3 \times 10^{-5}$). This variant was near known additive signals (rs17261179, rs3811978, and rs62357230) associated with type 2 diabetes (2), but it was not in strong LD with any of these previously identified variants (maximum $r^2 = 0.08$).

We identified another nonadditive, low-frequency variant (rs140453320) with large effect size on chromosome 5. This variant (MAF 0.01; OR 6.94 [95% CI 3.63–13.27]; $P = 5 \times 10^{-9}$) lies within an intron of the gene *ADAMTS6*. The additive *P* value was 0.48, leading to highly significant dominance deviation ($P = 4 \times 10^{-9}$). This signal was more than 1 Mb away from any previously known signal associated with type 2 diabetes.

The other three novel nonadditive signals were significantly more common, each with MAF >30%. Two of the three were located <0.5 Mb from known additive loci, but these signals were in weak LD with previously reported associations, with maximum r^2 between 0.1 and 0.3 (Supplementary Table 4). The third (rs755900673) was an insertion-deletion (OR 1.13 [95% CI 1.08–1.17]; $P = 5 \times 10^{-9}$) on chromosome 8 located within an intron of the *MYOM2* gene, more than 7 Mb away from any locus additively associated with type 2 diabetes.

We performed power simulations for our top variant (rs115018790; MAF 0.04; OR 2.63) and found that a genome-wide association study with an additive model with our case-control ratio would need ∼1.8 million participants to have 80% power to detect a genome-wide significant signal, whereas a recessive model would only need 160,000 participants. At higher allele frequencies, the benefits of the recessive model become much less pronounced (Supplementary Figure 1).

Replication

Our two replication cohorts consisted of 28,336 case subjects and 62,253 control subjects. Of the four nonadditive signals for which we had sufficient power, three replicated, and one did not (Table 1, Supplementary Figure 2). Variant rs115018790 replicated in both cohorts (meta-analysis OR 2.56 [95% CI 2.05–3.19]; $P = 1 \times 10^{-16}$). Two of the other three variants for which we had power also replicated. The insertion-deletion near *MYOM2*, rs755900673, did not replicate ($P = 0.84$) and showed high heterogeneity ($P = 0.008$).

Our power to replicate the rare variant near *ADAMTS6* was limited. The lack of replication was likely related to the small number of homozygous carriers in our replication samples ($n = 10$ in FinnGen; $n = 3$ in the Danish cohort) as opposed to poor imputation, as the *info score* in FinnGen was 0.98. Nevertheless, the signal retained genome-wide significance when we meta-analyzed the discovery and replication cohorts ($P = 3 \times 10^{-8}$).

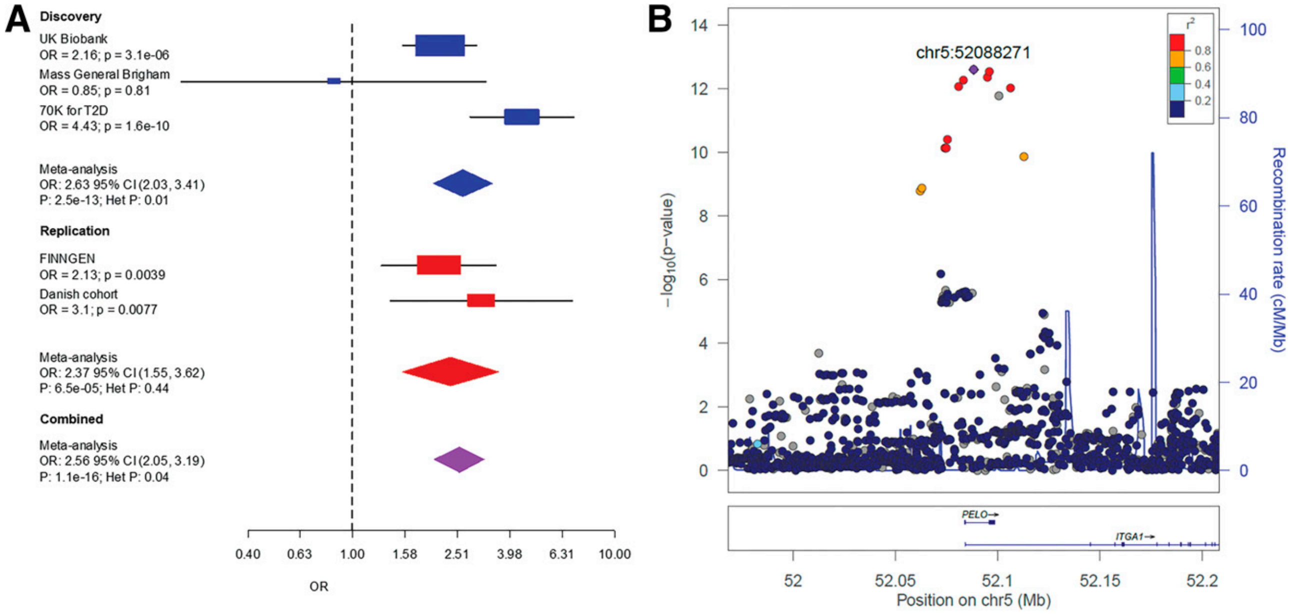


Figure 2—Replication of variant rs115018790. (*A*) A forest plot of the discovery and replication cohorts. Cohort-specific odds ratios are denoted by boxes proportional to the size of the cohort, and error bars represent the 95% CI. (*B*) Discovery GWAS *P* values at the *PELO* locus. Each dot represents a variant, with its genomic position (hg19) on the *x*-axis and its *P* value (−log10) on the *y*-axis. Nearby genes are shown at the bottom of the plot. chr5, chromosome 5; Het, heterogeneity.

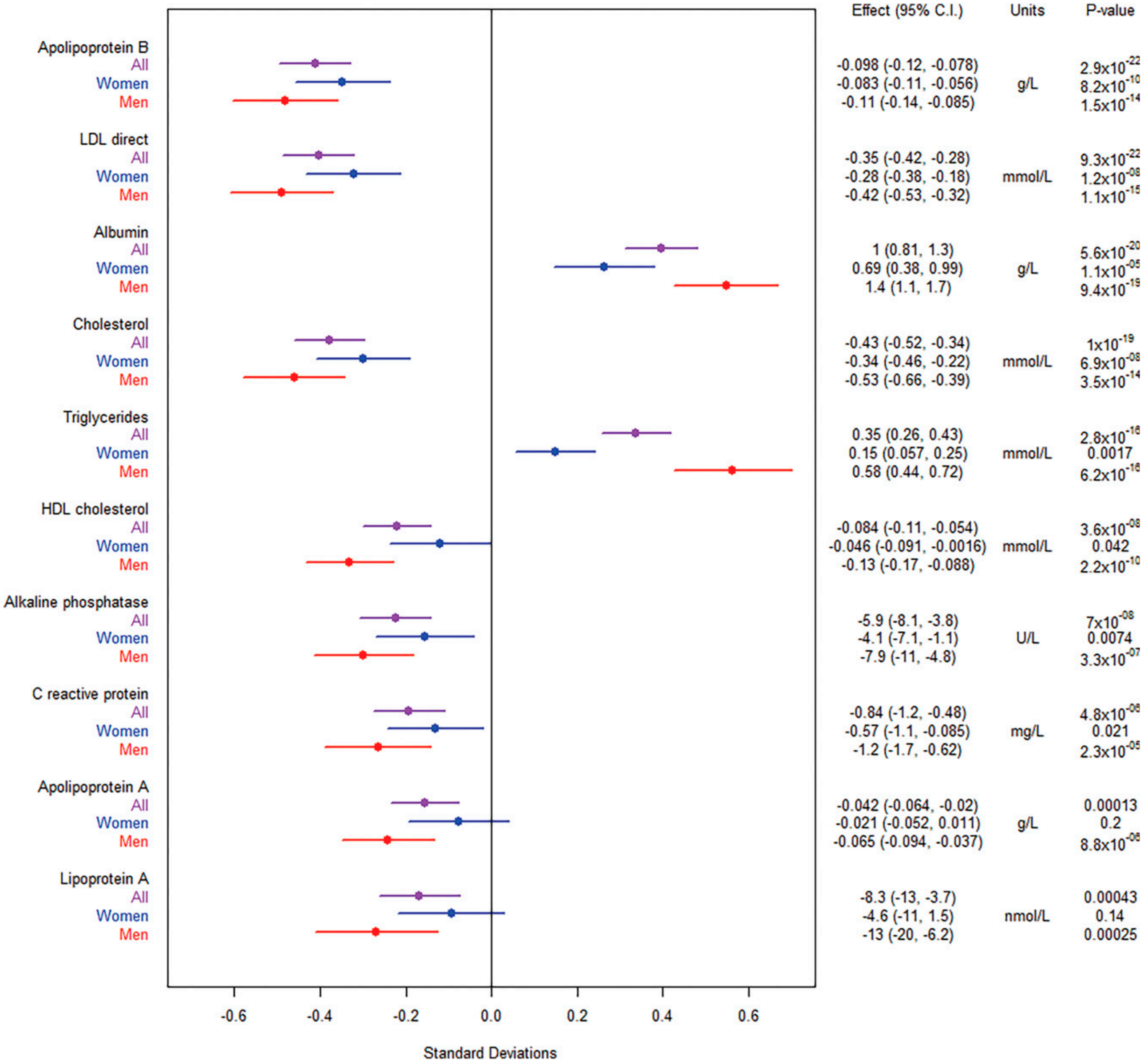
Gene Expression Colocalization Analysis

Using GTEx data, we found that rs115018790 was associated with reduced *PELO* expression in multiple tissues, although it was not an eQTL in pancreatic islets, according to the Translational Human Pancreatic Islet Genotype Tissue-Expression Resource database. Colocalization analysis, which tests the hypothesis that traits are associated and share a single causative variant, confirmed the link between rs115018790 and reduced *PELO* expression in many tissues, including subcutaneous adipose tissue (posterior probability 0.99; n = 581), skeletal muscle (posterior probability 0.99, n = 706), and the pancreas (posterior probability 0.96, n = 305). Colocalization plots (Supplementary Figure 3) comparing the recessive *P* values for the association with type 2 diabetes with additive *P* values from the gene expression data set showed a high degree of correlation between the two sets of *P* values, visually confirming the rs115018790s connection to reduced *PELO* gene expression. The signal’s 99% credible set (Supplementary Table 5) contains rs185240714 (posterior probability 0.23), which is located in the 5′ untranslated region of the *PELO* transcription start site, further supporting a causal link, although it is difficult to tell which variant is causal, because the 99% credible set at this locus contains six variants in near-complete LD.

Phenome-Wide Association Study

Using a recessive model, we found that multiple biomarkers (Fig. 3) were associated with rs115018790 in the UK Biobank. Homozygotes for the risk allele had significantly higher levels of triglycerides and lower levels of LDL, HDL, and total cholesterol. Effect sizes were large. For example, being a homozygous carrier of the risk allele was associated with a 0.35 mmol/L (14 mg/dL) decrease in LDL level (10% change relative to the mean) and a 0.35 mmol/L (31 mg/dL) increase in levels of triglycerides (20%). These associations, particularly for triglycerides, were less significant using an additive model (Supplementary Table 6), suggesting that rs115018790 acts in a recessive manner for these traits as well. Colocalization analysis (Supplementary Figure 3) confirmed that these lipid associations are the result of a single shared variant (posterior probability $>$ 0.99 for each trait). It did not appear that medication use was responsible for the observed effects on lipids, because homozygotes for the risk allele were less likely (OR 0.66 [95% CI 0.49–0.88]; P = 0.005) to be using LDL-lowering therapy. Other biomarkers associated with rs115018790 included albumin and C-reactive protein. There was also a nominally significant (P = 0.01) association with estradiol. The other novel variants did not have comparably significant and numerous biomarker associations (Supplementary Table 7).

When we examined nonbiomarker phenotypes (Supplementary Table 8) using a recessive model, we found that the variant near *PELO* was associated with a variety of hematologic features (eg, decreased blood cell count, increased reticulocyte count and percentage, increased mean corpuscular hemoglobin and volume, and decreased red blood cell distribution width) as well as increased alcohol intake frequency. None of the binary phenotypes reached a strict Bonferroni-corrected significance threshold of 1.5×10^{-5}, but the top two phenotypes were metformin use (OR



		Effect (95% C.I.)	Units	P-value
Apolipoprotein B	All	-0.098 (-0.12, -0.078)	g/L	$2.9x10^{-22}$
	Women	-0.083 (-0.11, -0.056)		$8.2x10^{-10}$
	Men	-0.11 (-0.14, -0.085)		$1.5x10^{-14}$
LDL direct	All	-0.35 (-0.42, -0.28)	mmol/L	$9.3x10^{-22}$
	Women	-0.28 (-0.38, -0.18)		$1.2x10^{-08}$
	Men	-0.42 (-0.53, -0.32)		$1.1x10^{-15}$
Albumin	All	1 (0.81, 1.3)	g/L	$5.6x10^{-20}$
	Women	0.69 (0.38, 0.99)		$1.1x10^{-05}$
	Men	1.4 (1.1, 1.7)		$9.4x10^{-19}$
Cholesterol	All	-0.43 (-0.52, -0.34)	mmol/L	$1x10^{-19}$
	Women	-0.34 (-0.46, -0.22)		$6.9x10^{-08}$
	Men	-0.53 (-0.66, -0.39)		$3.5x10^{-14}$
Triglycerides	All	0.35 (0.26, 0.43)	mmol/L	$2.8x10^{-16}$
	Women	0.15 (0.057, 0.25)		0.0017
	Men	0.58 (0.44, 0.72)		$6.2x10^{-16}$
HDL cholesterol	All	-0.084 (-0.11, -0.054)	mmol/L	$3.6x10^{-08}$
	Women	-0.046 (-0.091, -0.0016)		0.042
	Men	-0.13 (-0.17, -0.088)		$2.2x10^{-10}$
Alkaline phosphatase	All	-5.9 (-8.1, -3.8)	U/L	$7x10^{-08}$
	Women	-4.1 (-7.1, -1.1)		0.0074
	Men	-7.9 (-11, -4.8)		$3.3x10^{-07}$
C reactive protein	All	-0.84 (-1.2, -0.48)	mg/L	$4.8x10^{-06}$
	Women	-0.57 (-1.1, -0.085)		0.021
	Men	-1.2 (-1.7, -0.62)		$2.3x10^{-05}$
Apolipoprotein A	All	-0.042 (-0.064, -0.02)	g/L	0.00013
	Women	-0.021 (-0.052, 0.011)		0.2
	Men	-0.065 (-0.094, -0.037)		$8.8x10^{-06}$
Lipoprotein A	All	-8.3 (-13, -3.7)	nmol/L	0.00043
	Women	-4.6 (-11, 1.5)		0.14
	Men	-13 (-20, -6.2)		0.00025

Figure 3—Biomarker associations for variant rs115018790. The figure to the left shows effect sizes normalized by each trait's SD. The error bars in the figure to the left represent 95% CIs.

2.27 [95% CI 1.56–3.31]; $P = 2 \times 10^{-5}$) and diabetes diagnosed by a doctor (OR 1.87 [95% CI 1.39–2.54]; $P = 6 \times 10^{-5}$). We did not detect significant associations with cardiovascular phenotypes such as heart attack or stroke, after correcting for multiple testing.

We investigated the variant's effect on subcutaneous adipose tissue stores, as measured by impedance, and we did find that the effect was nominally associated with reduced fat mass in the lower extremities (Supplementary Table 9), raising the possibility of a lipodystrophy-like phenotype, although we did not observe sex-specific effects for this locus. We could not examine phenotypes such as liver fat content (based on MRI), because small sample sizes precluded the use of a recessive model.

In the Danish cohort, our power to detect recessive associations with glycemic traits was limited due to the low number of homozygous carriers (Supplementary Table 10). None of the traits were recessively associated with the variant. In an additive analysis, the variant was associated with increased insulin and C-peptide levels at the 120-min time point of an oral glucose tolerance test.

Sex-Stratified Analysis

Because the variant near *PELO* was nominally associated with estradiol, we performed an analysis stratified by sex and, in the case of women, by menopause status. We found that the effect of rs115018790 on estradiol was only significant in premenopausal women, with homozygotes for the risk allele having higher estradiol levels (174 pmol/L [95% CI 49–300]; $P = 0.006$) than other premenopausal women. For other biomarkers such as cholesterol and triglycerides, effects were stronger in men than in women (Fig. 3, Supplementary Table 6). For example, the recessive association with triglycerides was >12 orders of magnitude more significant in men ($P = 3 \times 10^{-16}$) than in women ($P = 0.002$).

We also performed a sex-stratified analysis for type 2 diabetes itself. In the UK Biobank, we found that the association was limited to men (OR 3.05 [95% CI 2.10–4.43]; $P = 5 \times 10^{-9}$) as opposed to women (OR 0.89 [95% CI 0.42–1.87]; $P = 0.75$), with an interaction P value of 7×10^{-7}. We replicated this finding in our replication cohorts (Supplementary Table 11). Meta-analysis across cohorts confirmed a large effect in men (OR 2.99 [95% CI 2.18–4.10]; $P = 1 \times 10^{-11}$) and little to no effect in women (OR 1.41 [95% CI 0.87–2.28]; $P = 0.15$).

Metabolome-Wide Association Study

Our metabolome-wide association study showed widespread changes in lipid-related phenotypes associated with rs115018790, using a recessive model (Supplementary Table 12), consistent with the aforementioned effects on lipid biomarkers. The associations were generally stronger in men than women and for the recessive model compared with the additive model (Fig. 4*A*). A total of 120 metabolite associations, many of them correlated with each other (Fig. 4*B*), were significant after correcting for multiple testing. The variant was particularly associated with decreased cholesterol percentage (normalized $\beta = -1.9$ [95% CI −2.1, −1.7]; $P = 6 \times 10^{-57}$) and increased triglyceride percentage (normalized $\beta = 1.8$ [95% CI 1.5–2.0]; $P = 1 \times 10^{-49}$) in large LDL particles in men. We performed a mediation analysis to test whether the effect of rs115018790 on the risk of type 2 diabetes in men was mediated by either of these two correlated metabolites, and we found evidence for causal mediation ($P < 1 \times 10^{-4}$) via both metabolomic parameters (Supplementary Table 13).

Comparison With Known Lipid-Associated Variants

To put effect sizes into context, we compared rs115018790 with previously described lipid-related variants of large effect size (33) using UK Biobank data. The LDL-lowering effect (0.35 mmol/L [14 mg/dL]) of rs115018790 in homozygotes was comparable to the effect (0.34 mmol/L [13 mg/dL]) of carrying one copy of a well-known protective variant (rs11 591147) associated with the *PCSK9* gene. The triglyceride-increasing effect (0.35 mmol/L [31 mg/dL]) of rs115018790 in homozygotes was larger than the change (−0.29 mmol/L [−26 mg/dL]) seen in homozygotes for a known variant (rs1569209) linked to the lipoprotein lipase (*LPL*) gene, known to be involved in triglyceride metabolism. For men, the size of rs115018790s effect (0.58 mmol/L [51 mg/dL]) was almost double that of the *LPL* variant.

DISCUSSION

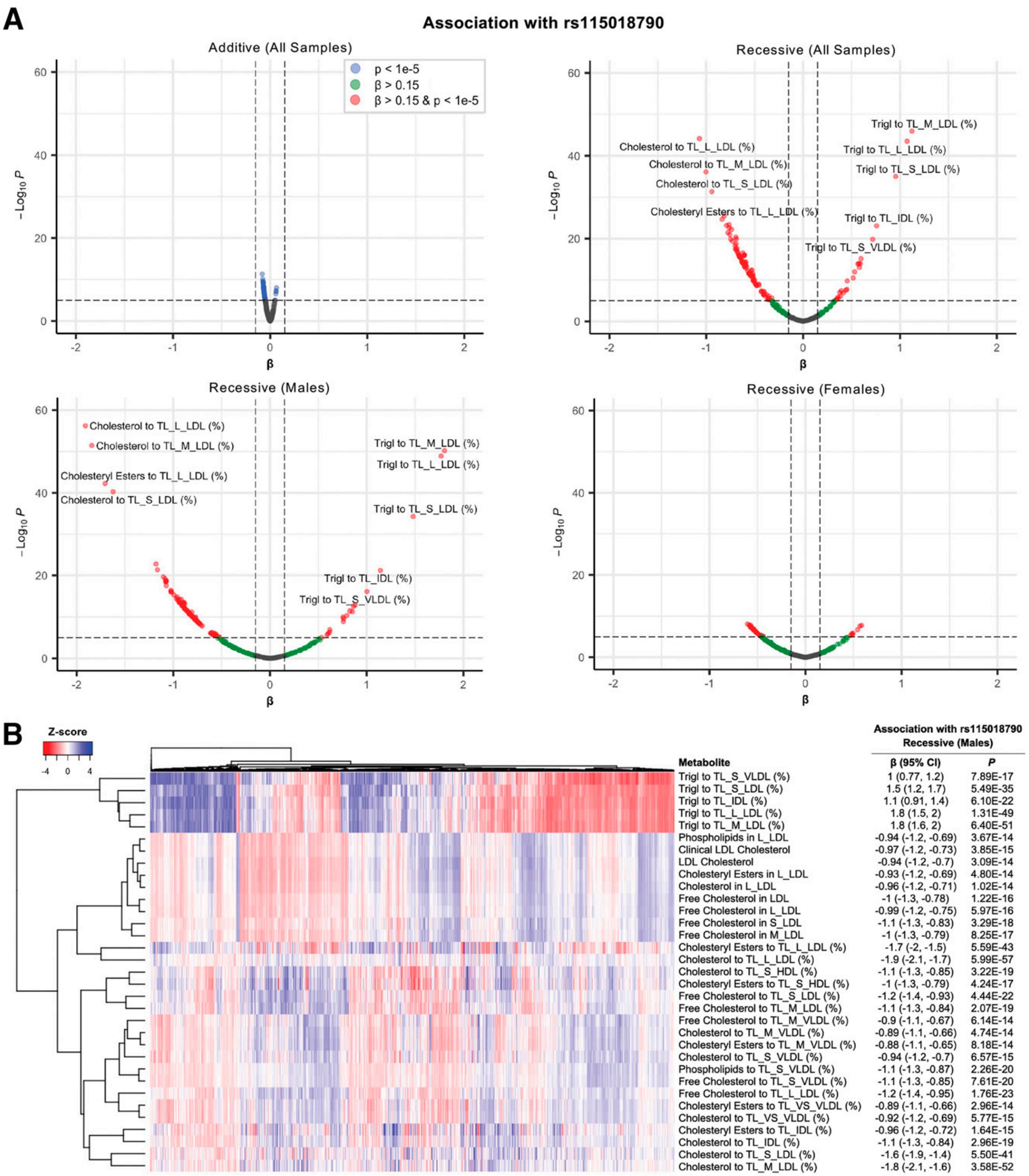
Type 2 diabetes is a highly polygenic trait, and hundreds of loci associated with the disease have been identified, mostly via large GWAS meta-analyses conducted under additive genetic models (2,3). This prior work has produced useful results, identifying potential therapeutic targets and also enabling the creation of polygenic scores capable of quantifying one's genetic risk (34). A sizeable fraction of the heritability of type 2 diabetes, however, remains unexplained by loci identified using additive models. Recessive modeling offers a way to identify new associations, creating opportunities for discovery and improved genetic risk stratification.

Our work takes advantage of the increasing number of genetic data sets now available, and, to our knowledge, it is currently the largest GWAS using a recessive model yet reported for type 2 diabetes or any other complex disease. We were able to identify multiple variants acting recessively, including two low-frequency variants of large effect size. Most of the variants identified via additive analyses have ORs of less than 1.1, but the most significant variant we identified had an OR of 2.56 in homozygous carriers. Our minimum sample size to detect this variant was 10 times smaller because we used a recessive, not an additive, model.

This variant was located near the *PELO* gene, and one of the six variants in the 99% credible set was in the gene's upstream 5′ untranslated region, suggesting a role for this variant in gene expression regulation, a link we confirmed across multiple tissues using colocalization approaches. Members of our group first identified this association in one of our cohorts while conducting recessive-model GWAS for multiple age-related diseases (11). In this study, we confirmed the association with a larger sample size, fine-mapped the region, and used the power of the UK Biobank to demonstrate that the phenotypic effects of this variant are not limited to type 2 diabetes.

Homozygous carriers of the *PELO* variant exhibit significantly different circulating triglyceride and cholesterol levels compared with other individuals. These effects were most pronounced in men but were also seen in women, and the effect sizes were clinically relevant and comparable to previously discovered genetic variants that revealed novel therapeutic targets. The reduction in LDL level associated with rs115018790 was ~10% (given an average LDL level of 3.62 mmol/L [140 mg/dL]), whereas statins, the most commonly used LDL-lowering medications, typically lower LDL levels by 30% to 60% (35). As would be expected for carriers of an LDL-lowering variant, homozygotes for the minor allele at rs115018790 were less likely to be taking statin medication. For triglycerides, the effect size (20%) was even larger.

The overall consequences of the effect of variant rs115018790 on lipid levels remain unclear. Low LDL concentration is known to protect against cardiovascular events. High levels of triglycerides and low HDL, on the



Association with rs115018790 Recessive (Males)

Metabolite	β (95% CI)	*P*
Trigl to TL_S_VLDL (%)	1 (0.77, 1.2)	7.89E-17
Trigl to TL_S_LDL (%)	1.5 (1.2, 1.7)	5.49E-35
Trigl to TL_IDL (%)	1.1 (0.91, 1.4)	6.10E-22
Trigl to TL_L_LDL (%)	1.8 (1.5, 2)	1.31E-49
Trigl to TL_M_LDL (%)	1.8 (1.6, 2)	6.40E-51
Phospholipids in L_LDL	-0.94 (-1.2, -0.69)	3.67E-14
Clinical LDL Cholesterol	-0.97 (-1.2, -0.73)	3.85E-15
LDL Cholesterol	-0.94 (-1.2, -0.7)	3.09E-14
Cholesteryl Esters in L_LDL	-0.93 (-1.2, -0.69)	4.80E-14
Cholesterol in L_LDL	-0.96 (-1.2, -0.71)	1.02E-14
Free Cholesterol in LDL	-1 (-1.3, -0.78)	1.22E-16
Free Cholesterol in L_LDL	-0.99 (-1.2, -0.75)	5.97E-16
Free Cholesterol in S_LDL	-1.1 (-1.3, -0.83)	3.29E-18
Free Cholesterol in M_LDL	-1 (-1.3, -0.79)	8.25E-17
Cholesteryl Esters to TL_L_LDL (%)	-1.7 (-2, -1.5)	5.59E-43
Cholesterol to TL_L_LDL (%)	-1.9 (-2.1, -1.7)	5.99E-57
Cholesterol to TL_S_HDL (%)	-1.1 (-1.3, -0.85)	3.22E-19
Cholesteryl Esters to TL_S_HDL (%)	-1 (-1.3, -0.79)	4.24E-17
Free Cholesterol to TL_S_LDL (%)	-1.2 (-1.4, -0.93)	4.44E-22
Free Cholesterol to TL_M_LDL (%)	-1.1 (-1.3, -0.84)	2.07E-19
Free Cholesterol to TL_M_VLDL (%)	-0.9 (-1.1, -0.67)	6.14E-14
Cholesterol to TL_M_VLDL (%)	-0.89 (-1.1, -0.66)	4.74E-14
Cholesteryl Esters to TL_M_VLDL (%)	-0.88 (-1.1, -0.65)	8.18E-14
Cholesterol to TL_S_VLDL (%)	-0.94 (-1.2, -0.7)	6.57E-15
Phospholipids to TL_S_VLDL (%)	-1.1 (-1.3, -0.87)	2.26E-20
Free Cholesterol to TL_S_VLDL (%)	-1.1 (-1.3, -0.85)	7.61E-20
Free Cholesterol to TL_L_LDL (%)	-1.2 (-1.4, -0.95)	1.76E-23
Cholesteryl Esters to TL_VS_VLDL (%)	-0.89 (-1.1, -0.66)	2.96E-14
Cholesterol to TL_VS_VLDL (%)	-0.92 (-1.2, -0.69)	5.77E-15
Cholesteryl Esters to TL_IDL (%)	-0.96 (-1.2, -0.72)	1.64E-15
Cholesterol to TL_IDL (%)	-1.1 (-1.3, -0.84)	2.96E-19
Cholesterol to TL_S_LDL (%)	-1.6 (-1.9, -1.4)	5.50E-41
Cholesterol to TL_M_LDL (%)	-1.8 (-2.1, -1.6)	3.58E-52

Figure 4—Metabolite associations for variant rs115018790. (*A*) Volcano plots with normalized effect size on the *x*-axis and *P* value (−log10) on the *y*-axis for each metabolite using an additive model and recessive models in the entire sample, in men, and in women. The scale is the same in all plots. (*B*) A heat map of the top 33 metabolites in the male samples with the standardized recessive-model effect of variant rs115018790 on each metabolite shown to the right of each row. IDL, intermediate-density lipoprotein; L, large; M, medium; S, small; TL, total lipids; Trigl, triglycerides; VLDL, very-low-density lipoprotein; VS, very small.

other hand, are associated with cardiovascular disease, although for these two lipid particles, it is not clear whether the relationship is causal (36,37). For homozygotes at rs115018790, the protective effects of lower LDL levels may be offset by the high triglyceride and lower HDL levels, meaning that the net effect on cardiovascular risk could be beneficial, harmful, or neutral. Our PheWAS did not reveal associations with cardiovascular events such as myocardial infarction or stroke. This lack of association, however, must be interpreted with caution in the setting of limited power, automatically curated phenotypes, and "healthy volunteer" selection bias in the UK Biobank (38).

The mechanism by which *PELO* affects diabetes risk is not clear. The gene is ubiquitously expressed, and its genetic deletion in mice leads to embryonic lethality (39). It is evolutionarily conserved and plays a role in rescuing stalled ribosomes, thus affecting the translation of multiple mRNA transcripts (40). It has a known role in sustaining protein synthesis in developing blood cells and platelets (41). Results of a recent CRISPR loss-of-function screen in human pancreatic β cells suggest that *PELO* may also play a role in insulin secretion (42). The sex-specific effect on diabetes risk in our study was striking, and more investigation is needed to determine what factors underlie the increased risk in men.

It is possible that the effect of *PELO* on diabetes risk is mediated, at least in part, by lower LDL concentration itself, or by lower cholesterol content within LDL, as suggested by our mediation analyses. Statin therapy, which lowers LDL levels, is known to be associated with new-onset diabetes (43), and mendelian randomization studies have suggested a causal link between lower LDL level and type 2 diabetes (44,45). Homozygous carriers of variant rs115018790 have reduced *PELO* expression in several tissues, including the liver, suggesting that the *PELO* gene could be contributing to a mechanism linking lower LDL levels and diabetes risk.

Our metabolome-wide association study adds detail to our understanding of the effect of *PELO* on lipids. Comparing our results with those of a prior prospective study of almost 12,000 individuals followed for 8 to 15 years (31), we found that many of the metabolomic pathways associated with increased diabetes risk in that study were also associated with variant rs115018790. For example, our most strongly associated metabolite (cholesterol percentage in large LDL) was also associated with type 2 diabetes risk in the prospective study, with the same direction of effect. Other metabolites known to be linked to diabetes risk, such as monounsaturated versus polyunsaturated fat percentages, were also significant in our results. More studies will be needed to better understand the clinical relevance of the lipid percentages in each of the different lipoprotein classes.

One limitation of our study is the restriction of the analyses to participants of European ancestry. Estimation of recessive effects requires large sample sizes, because homozygous carriers of low-frequency variants are rare. Progress has been made in terms of recruiting diverse participants for genetic studies, but people of European ancestry still make up the bulk of available data sets. In the future, nonadditive methods may yield new insights when applied to non-European populations—work that could be particularly fruitful given the increased genetic diversity of these populations (46) and the increasing availability of assembling multiethnic cohorts (47).

It is worth noting that most of the associations detected in our recessive analysis had already been uncovered in prior additive GWAS, and some of our novel signals deviated only slightly from additivity. Indeed, three of the five seemingly recessive signals were common variants, and they deviated only slightly from additivity, with only a nominally significant *P* value, raising the possibility that these variants have still an additive effect. This observation matches our power simulations comparing additive and recessive models. In these simulations, the benefit of the recessive model was significant at the low end of the allele-frequency spectrum, whereas both models had similar power to detect high-frequency variants with recessive effects.

Our work illustrates the value of performing nonadditive analyses to uncover low-frequency recessive variants. By conducting what is currently the largest GWAS using a recessive model for type 2 diabetes, we confirmed that a variant linked to reduced *PELO* gene expression appears to have significant effects not just on diabetes but also on lipid metabolism. Recessive models of type 2 diabetes and glycemic traits as part of larger and more diverse genetic discovery efforts are likely to provide additional associations that, in turn, will provide a better understanding of diabetes pathophysiology and possibly enhance the predictive power of polygenic scores.

Acknowledgments. We acknowledge Bianca C. Porneala (Massachusetts General Hospital) for technical assistance in the collection and curation of the genotype and phenotype data from the MGB Biobank. We also acknowledge Josee Dupuis and Peitao Wu (Boston University) for their valuable feedback.

Funding. M.J.O. was supported by National Institutes of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) award T32 DK110919. J.M.M. is supported by American Diabetes Association Innovative and Clinical Translational award 1-19-ICTS-068 and by National Human Genome Research Institute (NHGRI) grant U01HG011723. S.B.-G. was supported by an FI-Directorate-General for Research (DGR) Fellowship from FI-DGR 2013 from Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR; Generalitat de Catalunya) and by a "Juan de la Cierva" postdoctoral fellowship (Ministry of Economy, Industry and Competitiveness FJCI-2017-32090). J.C.F. is supported by NIDDK grant K24 DK110550. This project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 667191. This work was also supported by grant SEV-2011-00067 of the Severo Ochoa Program, awarded by the Spanish Government. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available at www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. This study also makes use of data generated by the UK10K Consortium, derived from samples from UK10K Cohort Imputation (EGAS00001000713). A full list of the investigators who contributed to the generation of the data is available at www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310. Work in the UK Biobank was done under application numbers 27892 and 31063. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk). The GTEx Project was supported by the Common Fund of the Office of the Director of the NIH, and by the National Cancer Institute; NHGRI; National Heart, Lung, and Blood Institute; National Institute on Drug Abuse; National Institute of Mental Health; and National Institute of Neurological Disorders and Stroke.

Duality of Interest. J.C.F. has received consulting honoraria from Goldfinch Bio and AstraZeneca, and speaking honoraria from Novo Nordisk, AstraZeneca, and Merck for research presentations over which he had full control of content. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. J.M.M. and A.L. planned the analysis. M.J.O. performed the genome-wide association study in the UK Biobank and MGB Biobank. A.H.-C., J.B.C., and V.K. assisted M.J.O. with these cohorts. P.C.S., S.B.-G., M.G.-M., D.T., and J.M.M. analyzed the 70K for T2D cohort. K.V. performed and M.K. supervised replication analysis in FinnGen, and C.F.R. performed the analysis in the Danish cohort supervised by N.G. and T.H., with assistance from O.P., I.B., and A.L., who all contributed to assembling the data. M.J.O. performed the meta-analysis and the phenome-wide association study. P.S. performed the metabolomic analyses. M.J.O. and J.M.M. wrote the manuscript. J.M.M., A.L., and J.C.F. supervised the work. J.M.M. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this work were presented in abstract form at the American Diabetes Association's 80th Scientific Sessions (virtual), 12–16 June 2020.

References

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14:88–98

2. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505–1513

3. Vujkovic M, Keaton JM, Lynch JA, et al.; HPAP Consortium; Regeneron Genetics Center; VA Million Veteran Program. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 2020;52:680–691

4. Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I, et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat Commun 2018;9:321

5. Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature 2020;582:240–245

6. Riddle MC, Philipson LH, Rich SS, et al. Monogenic diabetes: from genetic insights to population-based precision in care. reflections from a *Diabetes Care* editors' expert forum. Diabetes Care 2020;43:3117–3128

7. Grarup N, Moltke I, Andersen MK, et al. Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. Diabetologia 2018;61:2005–2015

8. Moltke I, Grarup N, Jørgensen ME, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 2014; 512:190–193

9. Wood AR, Tyrrell J, Beaumont R, et al.; GIANT consortium. Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. Diabetologia 2016;59:1214–1221

10. Manousaki D, Kent JW Jr, Haack K, et al. Toward precision medicine: TBC1D4 disruption is common among the Inuit and leads to underdiagnosis of type 2 diabetes. Diabetes Care 2016;39:1889–1895

11. Guindo-Martínez M, Amela R, Bonàs-Guarch S, et al.; FinnGen Consortium. The impact of non-additive genetic associations on age-related complex diseases. Nat Commun 2021;12:2436

12. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol 2019;2:9

13. Kowalski MH, Qian H, Hou Z, et al; NHLBI Trans-Omics for Precision Medicine Consortium, TOPMed Hematology & Hemostasis Working Group. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. PLoS Genet 2019;15:e1008500.

14. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779

15. Karlson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the Partners HealthCare Biobank at Partners Personalized Medicine: informed consent, return of research results, recruitment lessons and operational considerations. J Pers Med 2016;6:E2

16. Eastwood SV, Mathur R, Atkinson M, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. PLoS One 2016;11:e0162388

17. Yu S, Liao KP, Shaw SY, et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. J Am Med Inform Assoc 2015;22:993–1000

18. Koopman RJ, Mainous AG 3rd, Diaz VA, Geesey ME. Changes in age at diagnosis of type 2 diabetes mellitus in the United States, 1988 to 2000. Ann Fam Med 2005;3:60–63

19. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191

20. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336–2337

21. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31:3555–3557

22. Myers TA, Chanock SJ, Machiela MJ. *LDlinkR*: an R package for rapidly calculating linkage disequilibrium statistics in diverse populations. Front Genet 2020;11:157

23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575

24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7

25. Nguyen C, Varney MD, Harrison LC, Morahan G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. Diabetes 2013;62:2135–2140

26. Wellcome Trust Case Control Consortium, Maller JB, McVean G, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 2012;44:1294–1301

27. Carithers LJ, Ardlie K, Barcus M, et al.; GTEx Consortium. A novel approach to high-quality postmortem tissue procurement: the GTEx Project. Biopreserv Biobank 2015;13:311–319

28. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383

29. Alonso L, Piron A, Morán I, et al.; MAGIC. TIGER: the gene expression regulatory variation landscape of human pancreatic islets. Cell Rep 2021;37:109807

30. Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K. Software application profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. Int J Epidemiol 2018;47:29–35

31. Ahola-Olli AV, Mustelin L, Kalimeri M, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia 2019;62:2298–2309

32. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. J Stat Softw 2014;59:1–38

33. Klarin D, Damrauer SM, Cho K, et al.; Global Lipids Genetics Consortium; Myocardial Infarction Genetics (MIGen) Consortium; Geisinger-Regeneron DiscovEHR Collaboration; VA Million Veteran Program. Genetics of blood lipids among ∼300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet 2018;50:1514–1523

34. Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic risk scores for diabetes diagnosis and precision medicine. Endocr Rev 2019;40:1500–1520

35. Jones P, Kafonek S, Laurora I, Hunninghake D. Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (the CURVES study). Am J Cardiol 1998;81:582–587

36. Sarwar N, Danesh J, Eiriksdottir G, et al. Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. Circulation 2007;115:450–458

37. Rosenson RS. The high-density lipoprotein puzzle: why classic epidemiology, genetic epidemiology, and clinical trials conflict? Arterioscler Thromb Vasc Biol 2016;36:777–782

38. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 2017;186:1026–1034

39. Nyamsuren G, Kata A, Xu X, et al. Pelota regulates the development of extraembryonic endoderm through activation of bone morphogenetic protein (BMP) signaling. Stem Cell Res (Amst) 2014;13:61–74

40. Liakath-Ali K, Mills EW, Sequeira I, et al. An evolutionarily conserved ribosome-rescue pathway maintains epidermal homeostasis. Nature 2018;556:376–380

41. Mills EW, Wangen J, Green R, Ingolia NT. Dynamic regulation of a ribosome rescue pathway in erythroid cells and platelets. Cell Rep 2016;17:1–10

42. Grotz AK, Navarro-Guerrero E, Bevacqua RJ, et al. 2021. A genome-wide CRISPR screen identifies regulators of beta cell function involved in type 2 diabetes risk. medRxiv 28 May 2021 [preprint]. doi: 10.1101/2021.05.28.445984.

43. Ko MJ, Jo AJ, Kim YJ, et al. Time- and dose-dependent association of statin use with risk of clinically relevant new-onset diabetes mellitus in primary prevention: a nationwide observational cohort study. J Am Heart Assoc 2019;8:e011320

44. Pan W, Sun W, Yang S, et al. LDL-C plays a causal role on T2DM: a Mendelian randomization analysis. Aging (Albany NY) 2020;12:2584–2594

45. Schmidt AF, Swerdlow DI, Holmes MV, et al.; LifeLines Cohort study group; UCLEB consortium. PCSK9 genetic variants and risk of type 2 diabetes: a Mendelian randomisation study. Lancet Diabetes Endocrinol 2017;5:97–105

46. Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68–74

47. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" research program. N Engl J Med 2019;381:668–676