nature genetics

Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia

Christopher T Johansen[1], Jian Wang[1], Matthew B Lanktree[1], Henian Cao[1], Adam D McIntyre[1], Matthew R Ban[1], Rebecca A Martins[1], Brooke A Kennedy[1], Reina G Hassell[1], Maartje E Visser[2,3], Stephen M Schwartz[4], Benjamin F Voight[5,6], Roberto Elosua[7], Veikko Salomaa[8], Christopher J O'Donnell[9–11], Geesje M Dallinga-Thie[2,3], Sonia S Anand[12], Salim Yusuf[12], Murray W Huff[1,13], Sekar Kathiresan[5,6] & Robert A Hegele[1,13]

Genome-wide association studies (GWAS) have identified multiple loci associated with plasma lipid concentrations[1–5]. Common variants at these loci together explain <10% of variation in each lipid trait[4,5]. Rare variants with large individual effects may also contribute to the heritability of lipid traits[6,7]; however, the extent to which rare variants affect lipid phenotypes remains to be determined. Here we show an accumulation of rare variants, or a mutation skew, in GWAS-identified genes in individuals with hypertriglyceridemia (HTG). Through GWAS, we identified common variants in *APOA5*, *GCKR*, *LPL* and *APOB* associated with HTG. Resequencing of these genes revealed a significant burden of 154 rare missense or nonsense variants in 438 individuals with HTG, compared to 53 variants in 327 controls ($P = 6.2 \times 10^{-8}$), corresponding to a carrier frequency of 28.1% of affected individuals and 15.3% of controls ($P = 2.6 \times 10^{-5}$). Considering rare variants in these genes incrementally increased the proportion of genetic variation contributing to HTG.

GWAS have identified novel and known loci associated with population-based plasma lipid concentrations[1–5]. Despite the robustness of these associations, the proportion of variability explained by GWAS-identified loci is relatively modest, <10% in most studies[4,5]. Although vastly expanded study sample sizes continue to reveal new associations, each newly associated variant has an incrementally smaller effect size and contributes only marginally to the cumulative variation of each lipid phenotype[6]. This suggests that GWAS of population-based subjects may be reaching the limits of their ability to reveal genetic variation underlying complex traits. A question that has arisen is whether additional forms of genetic variation, such as rare variants with large individual effects, could contribute to the heritability of complex traits such as plasma lipid concentrations[6,7]. Although the mechanistic basis for the association between lipid traits and most of the common variants discovered in GWAS is still largely unknown, it remains possible that rare variants in GWAS-identified genes may contribute significantly to lipid phenotypes.

Studying subjects at the extremes of a quantitative phenotype distribution has proven useful in identifying functional rare variants[8–12]. Using missense-accumulation analysis in genes defined *a priori* as likely to contain rare variants, studies can statistically quantify a burden of mutations in subjects with severe phenotypes, before functional assessment of each variant. Primary HTG is one such complex polygenic disease, broadly defined by fasting plasma triglyceride concentrations above the ninety-fifth percentile[13]. Resequencing of triglyceride-modulating candidate genes has implicated both common and rare variants in HTG disease pathophysiology[9,14–16]; however, the majority of phenotypic variation underlying severe HTG remains unattributed[17]. Our objectives were (i) to perform an unbiased GWAS of individuals with HTG to identify common variants associated with HTG, and (ii) to resequence coding regions of candidate genes in loci reaching genome-wide significance to evaluate the burden of rare variants in individuals with HTG compared with controls. Here we show that loci found to be associated with HTG by GWAS also harbor a significant excess of rare variants.

In total, 555 individuals with HTG and 1,319 controls were included in two cohorts of the study: the GWAS cohort included 463 affected individuals and 1,197 controls, and the sequencing cohort included 438 affected individuals and 327 controls. Individuals with HTG were unrelated subjects diagnosed with Fredrickson hyperlipoproteinemia

[1]Department of Biochemistry, Robarts Research Institute, University of Western Ontario, London, Ontario, Canada. [2]Department of Experimental Vascular Medicine, Academic Medical Center Amsterdam, Amsterdam, The Netherlands. [3]Department of Vascular Medicine, Academic Medical Center Amsterdam, Amsterdam, The Netherlands. [4]Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. [5]Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [7]Cardiovascular Epidemiology and Genetics, Institut Municipal D'investigacio Medica, and CIBER Epidemiología y Salud Pública, Barcelona, Spain. [8]Chronic Disease Epidemiology Unit, Department of Health Promotion and Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [9]Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [10]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [11]Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA. [12]Population Health Research Institute, Hamilton Health Sciences, McMaster University, Hamilton, Ontario, Canada. [13]Department of Medicine, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada. Correspondence should be addressed to R.A.H. (hegele@robarts.ca).

Received 1 April; accepted 22 June; published online 25 July 2010; doi:10.1038/ng.628

Table 1 Baseline clinical attributes of the study sample

		n	Female (%)	T2D (%)	Age (years)	BMI (kg m^{-2})	TC (mmol l^{-1})	HDL-C (mmol l^{-1})	LDL-C (mmol l^{-1})	TG (mmol l^{-1})
GWAS cohort	HTG	463	30.7	25.7	50.9 ± 13.0	29.9 ± 4.9	8.2 ± 3.9	0.9 ± 0.3	–	14.3 ± 19.8
	Control	1,197	40.4	0.4	47.8 ± 11.1	26.4 ± 4.6	5.3 ± 1.3	1.4 ± 0.4	3.4 ± 1.2	1.1 ± 0.7
Resequencing cohort	HTG	438	33.2	28.1	51.3 ± 13.1	30.0 ± 4.9	8.7 ± 4.3	0.9 ± 0.3	–	14.2 ± 19.0
	Control	327	50.5	4.1	49.9 ± 15.1	26.8 ± 4.5	4.9 ± 0.8	1.3 ± 0.4	3.2 ± 0.9	1.2 ± 0.4

BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides; T2D, type 2 diabetes. Both cohorts have in common 346 individuals with HTG and 205 healthy controls. Lipid measurements were conducted after a 12-h fasting period. Values are means ± s.d. LDL cholesterol is not accurately calculated using the Friedewald equation for affected individuals in whom plasma triglyceride concentration exceeds 4.5 mmol l^{-1}.

phenotypes 2B (MIM 144250), 3 (MIM 107741), 4 (MIM 144600) or 5 (MIM 144650), ascertained primarily from a single tertiary referral lipid clinic. The mean plasma triglyceride concentration of individuals with HTG was 14.3 mmol l^{-1}. We chose controls with maximum recorded fasting plasma triglyceride concentrations <2.3 mmol l^{-1} to exclude undiagnosed HTG. All study subjects were of self-declared European ancestry; subjects deviating from European ancestry as determined by multidimensional scaling using whole-genome SNP data were removed from sequencing analysis (**Supplementary Fig. 1**). As expected, clinical characteristics of individuals with HTG were less favorable than those of controls, with worse lipid profiles and an increased prevalence of type 2 diabetes (**Table 1**).

The HTG phenotype was tested for association with >2.1 million SNPs using an additive multivariate logistic regression model (**Supplementary Fig. 2**). This model appropriately adjusted for sex, body-mass index, diabetes status and ten principal components of ancestry (**Supplementary Fig. 3**). Four loci were associated with HTG ($P < 5 \times 10^{-7}$): *APOA5*, *GCKR*, *LPL* and *APOB* (**Table 2**). Most associations with HTG were mediated by the same genomic loci associated with fasting plasma triglyceride concentration in population-based GWAS[5]: *APOA5* and *GCKR* were associated at the same lead SNP, and *LPL* was associated with the same haplotype block. In contrast, the HTG-associated SNPs in *APOB* were ~123 kb upstream of the gene, which would be consistent with the involvement of regulatory elements in the overexpression of triglyceride-rich lipoproteins in HTG pathophysiology. Investigation of subthreshold association signals did not provide any additional insight into novel HTG-associated genes.

Next, we tested the hypothesis that common genetic variants in remaining known triglyceride-associated loci are similarly associated with HTG[5]. Only three loci were replicated at a Bonferroni-corrected significance threshold of $P < 0.005$: *MLXIPL*, *TRIB1* and *ANGPTL3* (**Table 2**). Positive replication of these triglyceride-associated loci, combined with trends toward significance at *FADS1-FADS2-FADS3* ($P = 0.05$) and *NCAN-CILP2-PBX4* ($P = 0.07$), suggest that additional triglyceride-modulating loci may also be involved in HTG pathophysiology; however, smaller effect sizes probably limit their detection.

We next hypothesized that HTG-associated genes would harbor rare variants related to HTG disease causation. The protein-coding sequences of *APOA5*, *GCKR*, *LPL* and exons 26 and 29 (67.8%) of *APOB* were resequenced in individual subjects as the regions most likely to harbor protein-compromising mutations. Across the four genes, 80 distinct rare variants were identified with minor allele frequencies <1% in controls (**Fig. 1** and **Supplementary Table 1**). A significant accumulation of rare variants was identified in individuals with HTG (**Table 3**), including 154 total variants in 438 HTG diploid genomes, compared to 53 total variants in 327 control diploid genomes ($P = 6.2 \times 10^{-8}$), corresponding to a significantly higher carrier frequency of 28.1% in individuals with HTG compared to 15.3% in controls ($P = 2.6 \times 10^{-5}$). A more restricted analysis of rare variants found exclusively in either individuals with HTG or controls, deliberately removing all reported variants without demonstrated functional compromise, similarly revealed a significant burden of 47 variants in affected individuals, compared to nine variants in controls ($P = 2.4 \times 10^{-5}$); this corresponds to a significantly higher carrier frequency of 10.3% in individuals with HTG compared to 2.8% in controls ($P = 4.4 \times 10^{-5}$). The fasting plasma triglyceride concentrations of carriers with HTG ranged from 3.10 to 88.5 mmol/l, whereas those of control carriers ranged from 0.45 to 1.93 mmol/l. No associations were discerned between such attributes as the gene, mutation type or mutation position and the plasma triglyceride concentration or HTG phenotype.

The strength of association between HTG and genomic loci did not predict the mutation accumulation observed in the resequenced genes. *LPL* harbored the largest relative proportion of rare variants, followed by *GCKR*, *APOB* and *APOA5*; these had, respectively, 30.9, 10.7, 9.3 and 4.5 rare variants per kilobase of coding sequence in individuals with HTG, and 5.6, 2.7, 4.3 and 0.9 rare variants per kilobase of coding sequence in controls. The burden of rare variants found in individuals with HTG is highly suggestive of phenotype causation, an idea supported by several truncation mutations, *in silico* predictions of deleterious effects, and bona fide characterized deleterious mutations (**Supplementary Table 1**). The majority

Table 2 Genetic loci associated with HTG

Locus	SNP	Chr.	Position	Minor allele	HTG MAF	Control MAF	OR (95% CI)	*P*
APOA5	rs964184	11	116.2	G	0.33	0.14	3.28 (2.61–4.14)	5.4×10^{-24}
GCKR	rs1260326	2	2.8	T	0.52	0.41	1.75 (1.45–2.12)	6.5×10^{-9}
LPL	rs7016880	8	19.9	C	0.03	0.10	0.32 (0.21–0.49)	2.0×10^{-7}
APOB	rs4635554	2	21.2	G	0.39	0.31	1.67 (1.38–2.02)	2.0×10^{-7}
MLXIPL	rs714052	7	72.5	G	0.07	0.13	0.44 (0.31–0.62)	0.000003
TRIB1	rs2954029	8	126.6	T	0.37	0.46	0.71 (0.59–0.86)	0.0004
ANGPTL3	rs10889353	1	62.9	C	0.27	0.32	0.73 (0.59–0.89)	0.002
NCAN	rs17216525	19	19.5	T	0.07	0.09	0.71 (0.50–1.00)	0.05
FADS	rs174547	11	61.3	C	0.40	0.33	1.20 (0.99–1.44)	0.07
XKR6	rs7819412	8	11.1	G	0.46	0.50	0.87 (0.72–1.05)	0.14
PLTP	rs7679	20	44.0	C	0.20	0.19	1.17 (0.94–1.47)	0.16

Chr., chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio. Association was tested using an additive multivariate logistic regression model, with sex, body-mass index, diabetes status and ten principal components of ancestry entered as covariates. The top four loci surpassed a prespecified threshold for genome-wide significance of $P < 5 \times 10^{-7}$. Remaining loci were replicated from GWAS meta-analysis of population-based triglyceride concentrations[5].

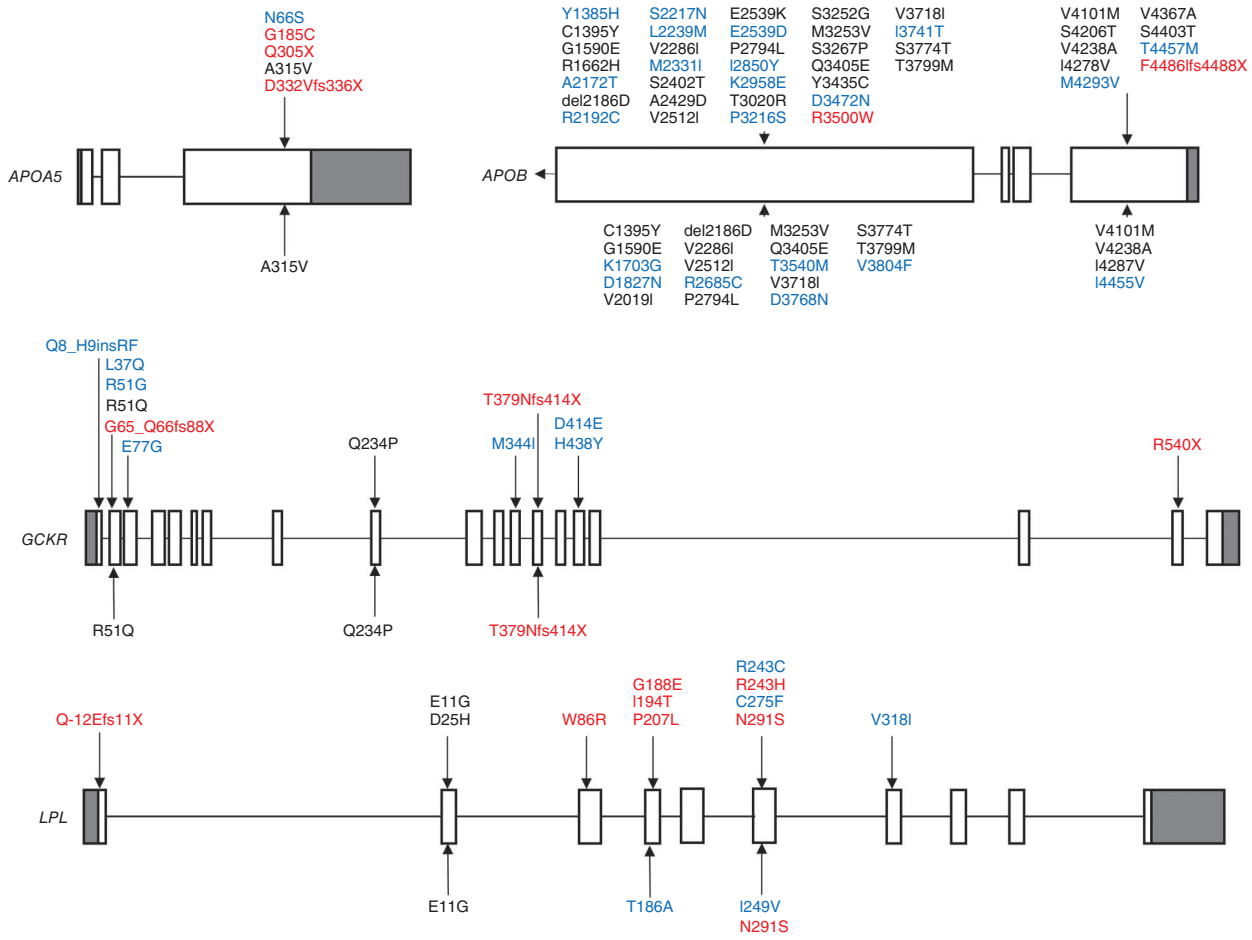

Figure 1 Rare variants identified by resequencing GWAS-identified genes in individuals with HTG and controls. Variants listed above gene maps were identified in affected individuals, and those below gene maps were identified in controls. Rare variants are colored as follows: black, identified in control subjects or previously identified in subjects of unknown clinical status; blue, exclusive to affected individuals or controls; red, proven biological dysfunction or truncation. Nomenclature for variants refers to functional protein sequences. Only exons 26 and 29 were resequenced in *APOB*. Gene maps are roughly to scale, although they differ in scale between genes.

of subjects carried only one rare variant; however, subjects with multiple rare variants were also significantly over-represented among individuals with HTG (6.6% of carriers among affected individuals compared with 0.9% of carriers among controls; $P = 3.7 \times 10^{-5}$). Any given rare variant in affected individuals is not necessarily sufficient to cause HTG, but rare variants probably contribute to the biochemical heterogeneity observed among affected people. For instance, the APOB R3500W variant causes hypercholesterolemia[18], but we found it in an individual with Fredrickson hyperlipoproteinemia phenotype 2B, defined by both plasma triglyceride and total cholesterol in excess of the ninety-fifth percentile. For this individual, APOB R3500W is more likely contributing to the elevated total cholesterol phenotype, but the mutation is a part of his genetic background that led to his ascertainment through the lipid clinic. This person exemplifies our working hypothesis that both common and rare genetic determinants in triglyceride-associated genes together contribute to the phenotypic heterogeneity underlying HTG.

Finally, we assessed the contribution of genetic and clinical variables to the total variation in HTG diagnosis, using the individuals included in both the GWAS and resequencing cohorts. A comprehensive logistic regression model including clinical variables and both common and rare genetic variants explained 41.6% of total variation in HTG diagnosis: clinical variables explained 19.7%, common genetic variants in seven HTG-associated loci explained 20.8%, and rare genetic variants in four HTG-associated loci explained 1.1%. These data suggest that rare variants found in four GWAS-identified genes incrementally contribute to the unexplained genetic variation contributing to HTG pathophysiology.

In summary, we performed a GWAS and resequencing of HTG-associated genes and found a significant accumulation of missense and nonsense mutations that contribute to the unexplained genetic component of HTG. Our results suggest that a complex genetic architecture of both common and rare variants in a spectrum of triglyceride-associated genes is responsible for HTG. Future studies using high-throughput next-generation sequencing are required to determine whether these associations extend to additional HTG-associated genes, including *MLXIPL*, *TRIB1* and *ANGPTL3*, and to triglyceride-associated genes identified by epidemiological-scale GWAS of population-based samples. It also remains possible that rare variants in triglyceride-modulating genes that have not yielded signals in GWAS, such as *GPIHBP1* or *LMF1*, will further contribute to HTG phenotypes[19,20]. Functional analyses may more accurately define the extent of dysfunction of rare variants identified in individuals with HTG and their role in disease causation, and higher-level analyses, including studies of gene-gene and gene-environment interactions,

Table 3 Rare variant accumulation in individuals with HTG and controls

		All mutations		Missense/Indels		Nonsense	
		HTG	Controls	HTG	Controls	HTG	Controls
	Total alleles	876	654	876	654	876	654
All variants	*APOA5*	5	1	3	1	2	0
	GCKR	20	5	14	4	6	1
	LPL	44	8	43	8	1	0
	APOB	85	39	84	39	1	0
	Total	154	53	146	52	9	1
		$P = 6.2 \times 10^{-8}$		$P = 3.2 \times 10^{-7}$		$P = 0.051$	
Exclusive variants	*APOA5*	4	0	2	0	2	0
	GCKR	9	0	7	0	2	0
	LPL	19	2	18	2	1	0
	APOB	15	7	14	7	1	0
	Total	47	9	42	9	5	0
		$P = 2.4 \times 10^{-5}$		$P = 1.4 \times 10^{-4}$		$P = 0.075$	

Exclusive variants refer to rare variants found exclusively in individuals with HTG or exclusively in healthy controls; previously reported variants without characterized functional compromise are deliberately excluded. Fisher's exact test was used to calculate the significance of rare-variant accumulation in individuals with HTG, with nominal statistical significance defined as a two-sided $P < 0.05$. Mutation counts and annotations are found in **Supplementary Table 1**.

will determine the combined impact of multiple genetic variants on plasma triglyceride concentration in individuals with HTG. Our study shows that an accumulation of rare variants is present in GWAS-identified genes and that these contribute to the heritability of complex traits among individuals at the extreme of a lipid phenotype.

URLs. Broad Institute, http://www.broadinstitute.org/; London Regional Genomics Center, http://www.lrgc.ca/; PolyPhen, http://genetics.bwh.harvard.edu/pph/; SHARCNET, http://www.sharcnet.ca/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the London Regional Genomics Centre (D. Carter, G. Barbe and K. Kang) for their dedication to this project, and the Myocardial Infarction Genetics Consortium (MIGen) study for the use of their genotype data as control data in our study. The MIGen study was funded by the US National Institutes of Health through the National Heart, Lung, and Blood Institute's STAMPEED genomics research program (R01 HL087676) and the National Center for Research Resources (U54 RR020278). This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET). C.T.J. is supported by a Canadian Institutes of Health Research (CIHR) Banting and Best Canada Graduate Scholarship, a Heart and Stroke Foundation of Ontario Program Grant and a CIHR Vascular Research Fellowship. V.S. was supported by the Sigrid Juselius Foundation and by the Finnish Academy (grant 129494). S.S.A. is supported by the Michael G. DeGroote Heart and Stroke Foundation of Ontario Chair and the Eli Lilly May Cohen Chair in Women's Health Research at McMaster University. R.A.H. is supported by the Jacob J. Wolfe Distinguished Medical Research Chair, the Edith Schulich Vinet Canada Research Chair in Human Genetics (Tier I), the Martha G. Blackburn Chair in Cardiovascular Research and operating grants from the CIHR (MOP-13430, MOP-79523, CTP-79853), the Heart and Stroke Foundation of Ontario (NA-6059, T-6018, PRG-4854), the Pfizer Jean Davignon Distinguished Cardiovascular and Metabolic Research Award and Genome Canada through the Ontario Genomics Institute.

AUTHOR CONTRIBUTIONS

Manuscript and experiment conceptualization, C.T.J. and R.A.H.; project management, C.T.J. and J.W.; GWAS and statistical analysis: C.T.J. and M.B.L.; sequencing, J.W., H.C., A.D.M., R.A.M., R.G.H. and C.T.J.; biochemical analysis, M.W.H.; clinical database management, M.R.B. and B.A.K.; study sample contributions, R.A.H., S.S.A., S.Y., M.E.V., G.M.D.-T., S.M.S., B.F.V., R.E., V.S., C.J.O. and S.K.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).
2. Willer, C.J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
3. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**, 35–46 (2009).
4. Aulchenko, Y.S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
5. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).
6. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
7. Kryukov, G.V., Pennacchio, L.A. & Sunyaev, S.R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–739 (2007).
8. Plomin, R., Haworth, C.M. & Davis, O.S. Common disorders are quantitative traits. *Nat. Rev. Genet.* **10**, 872–878 (2009).
9. Wang, J. *et al.* Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler. Thromb. Vasc. Biol.* **27**, 2450–2455 (2007).
10. Nejentsev, S. *et al.* Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
11. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
12. Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat. Genet.* **39**, 513–516 (2007).
13. Yuan, G., Al-Shali, K.Z. & Hegele, R.A. Hypertriglyceridemia: its etiology, effects and treatment. *CMAJ* **176**, 1113–1120 (2007).
14. Hegele, R.A. *et al.* A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Hum. Mol. Genet.* **18**, 4189–4194 (2009).
15. Wang, J. *et al.* Polygenic determinants of severe hypertriglyceridemia. *Hum. Mol. Genet.* **17**, 2894–2899 (2008).
16. Wang, J. *et al.* APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. *Nat. Clin. Pract. Cardiovasc. Med.* **5**, 730–737 (2008).
17. Hegele, R.A. Plasma lipoproteins: genetic influences and clinical implications. *Nat. Rev. Genet.* **10**, 109–121 (2009).
18. Gaffney, D. *et al.* Independent mutations at codon 3500 of the apolipoprotein B gene are associated with hyperlipidemia. *Arterioscler. Thromb. Vasc. Biol.* **15**, 1025–1029 (1995).
19. Beigneux, A.P. *et al.* Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. *Arterioscler. Thromb. Vasc. Biol.* **29**, 956–962 (2009).
20. Péterfy, M. *et al.* Mutations in LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. *Nat. Genet.* **39**, 1483–1487 (2007).

ONLINE METHODS

Subjects. The project was approved by the University of Western Ontario Institutional Review Board (protocol 07920E) and by ethics boards at collaborating institutions. All subjects provided informed consent for blood sampling, DNA analysis and collection of clinical, biochemical and other demographic data. All subjects in this study were unrelated and of self-declared European ancestry. The GWAS included 463 individuals with HTG and 1,197 controls. Affected individuals were obtained predominantly from a single tertiary referral lipid clinic (92% of affected individuals) in London, Ontario, Canada, or from a tertiary referral lipid clinic in Amsterdam. Controls were subjects with familial hypercholesterolemia (4% of controls) obtained from a single tertiary referral lipid clinic in London, Ontario, Canada, or normal healthy controls obtained from population-based studies, including the Study of Health Assessment and Risk in Ethnic Groups[21] (18%) and the Myocardial Infarction Genetics Consortium[22] (78%). We chose controls with plasma triglyceride concentration <2.3 mmol l^{-1} to exclude potentially undiagnosed HTG, but controls were otherwise not phenotypically selected. Subjects with familial hypercholesterolemia were included as negative controls only in the GWAS. The resequencing cohort included 438 individuals with HTG and 327 healthy controls: affected individuals were obtained only from the lipid clinic in London, Ontario, Canada, and healthy controls included only healthy subjects from the Study of Health Assessment and Risk in Ethnic Groups. Biochemical analyses were conducted separately in each cohort, as previously described[14,21,22].

Genome-wide association study. All subjects were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0 according to protocols of the London Regional Genomics Centre or the Broad Institute. Imputation was conducted using HapMap CEU phased haplotypes in MACH[23]. All genotypes were filtered for minor allele frequency >1%, Hardy Weinberg $P > 0.0001$, and 95% call rate or imputation quality $r^2 > 0.4$. Identity-by-state calculations, multidimensional scaling and association testing were conducted in PLINK[24]. Genome-wide significance was prespecified as $P < 5 \times 10^{-7}$; nominal significance for replication of known triglyceride-associated SNPs was a Bonferroni-corrected threshold $P < 0.005$. Covariates entered into all analyses included sex, body-mass index, diabetes status and ten principal components of ancestry as generated by Eigenstrat[25,26].

Sequencing and mutation accumulation. All genes were bidirectionally sequenced in individual samples using an ABI 3730 Automated DNA Sequencer and called using automated software (Applied Biosystems). Rare variants were manually curated, confirmed by repeat analysis and annotated *in silico* for functional effects using PolyPhen. Rare variants were defined as having minor allele frequencies <1% in controls. Carriers were defined as having at least one rare variant. Rare-variant accumulation was compared between individuals with HTG and controls using Fisher's exact test, with nominal significance defined as a two-sided $P < 0.05$. All subjects in the resequencing cohort were sequenced fully across the translated coding sequences of *APOA5* (NCBI NG_015894.1 and NP_443200.2), *GCKR* (NM_001486.3 and NP_001477.2), *LPL* (NG_008855.1 and NP_000228.1), and exons 26 and 29 of *APOB* (NG_011793.1 and NP_000375.2) (67.8%). Subjects missing sequencing data in any one gene were removed before analysis. Our intention was to identify rare missense and nonsense variants potentially responsible for HTG disease causation; accordingly, we excluded intronic variants, UTR variants and synonymous variants from mutation-accumulation analyses. Controls were population-based and not phenotypically selected on the basis of extremely low plasma triglyceride concentration; thus, they do not represent a supernormal control group in which protective rare variants would be enriched upon resequencing. Exclusive variants were defined as rare variants found exclusively in individuals with HTG or in controls (not both), with deliberate exclusion of variants previously reported without demonstrated functional compromise. Mutation-accumulation analyses compared either the number of observed rare alleles with the number of reference alleles, or the number of rare-variant carriers with the number of noncarriers, in individuals with HTG and controls.

Genetic variation explained. Subjects included in this analysis were common to both GWAS and resequencing cohorts. The proportion of genetic variation explained was calculated from the residuals of a multivariate logistic regression model, using discrete case-control status as the dependent variable, with a published SAS version 9.2 macro written for this purpose[27]. Independent variables included the clinical covariates age, sex, body-mass index and diabetes status as either continuous or discrete variables, common variants as continuous variables of HTG risk–associated alleles at each of the seven HTG-associated loci, and rare variants as a continuous variable including the number of rare variants carried by each subject.

21. Anand, S.S. *et al.* Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). *Lancet* **356**, 279–284 (2000).
22. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
23. Li, Y. & Abecasis, G.R. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.* **S79**, 2290 (2006).
24. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
25. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
26. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
27. Mittlböck, M. & Schemper, M. Computing measures of explained variation for logistic regression models. *Comput. Methods Programs Biomed.* **58**, 17–24 (1999).