SUPPLEMENTAL INFORMATION:

Mutation skew 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], Stephen M. Schwartz[3], Benjamin F. Voight[4], Roberto Elosua[5], Veikko Salomaa[6], Christopher J. O'Donnell[7,8], Geesje M. Dallinga-Thie[2], Sonia S. Anand[9], Salim Yusuf[9], Murray W. Huff[1] , Sekar Kathiresan[4] and Robert A. Hegele[1*]

[1]Robarts Research Institute, University of Western Ontario, London, Ontario N6A 5K8, Canada; [2]Departments of Experimental Vascular Medicine and Vascular Medicine, Academic Medical Center Amsterdam, Amsterdam 1105AZ, Netherlands; [3]Cardiovascular Health Research Unit and Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington 98101, USA; [4]Center for Human Genetic Research, Massachusetts General Hospital, Boston 02114, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge 02142, Massachusetts, USA; [5]Cardiovascular Epidemiology and Genetics, Institut Municipal D'investigacio Medica, and CIBER Epidemiologı´a y Salud Pu´ blica, Barcelona 08003, Spain; [6]Chronic Disease Epidemiology Unit, Department of Health Promotion and Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki 00300, Finland; [7]Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston 02114, Massachusetts, USA; [8]Framingham Heart Study and National, Heart, Lung, and Blood Institute, Framingham 01702, Massachusetts, USA; [9]Population Health Research Institute, Hamilton Health Sciences, and Departments of Medicine and Clinical Epidemiology, McMaster University, Hamilton, Ontario L8L 2X2, Canada.
[*]Corresponding author (hegele@robarts.ca).

Corresponding Author:
Robert A. Hegele MD FRCPC FACP
Blackburn Cardiovascular Genetics Laboratory
Robarts Research Institute, University of Western Ontario
London, Ontario, Canada, N6A 5K8
Tel: 519-931-5271; Fax: 519-931-5218; E-mail: hegele@robarts.ca

Supplemental Methods

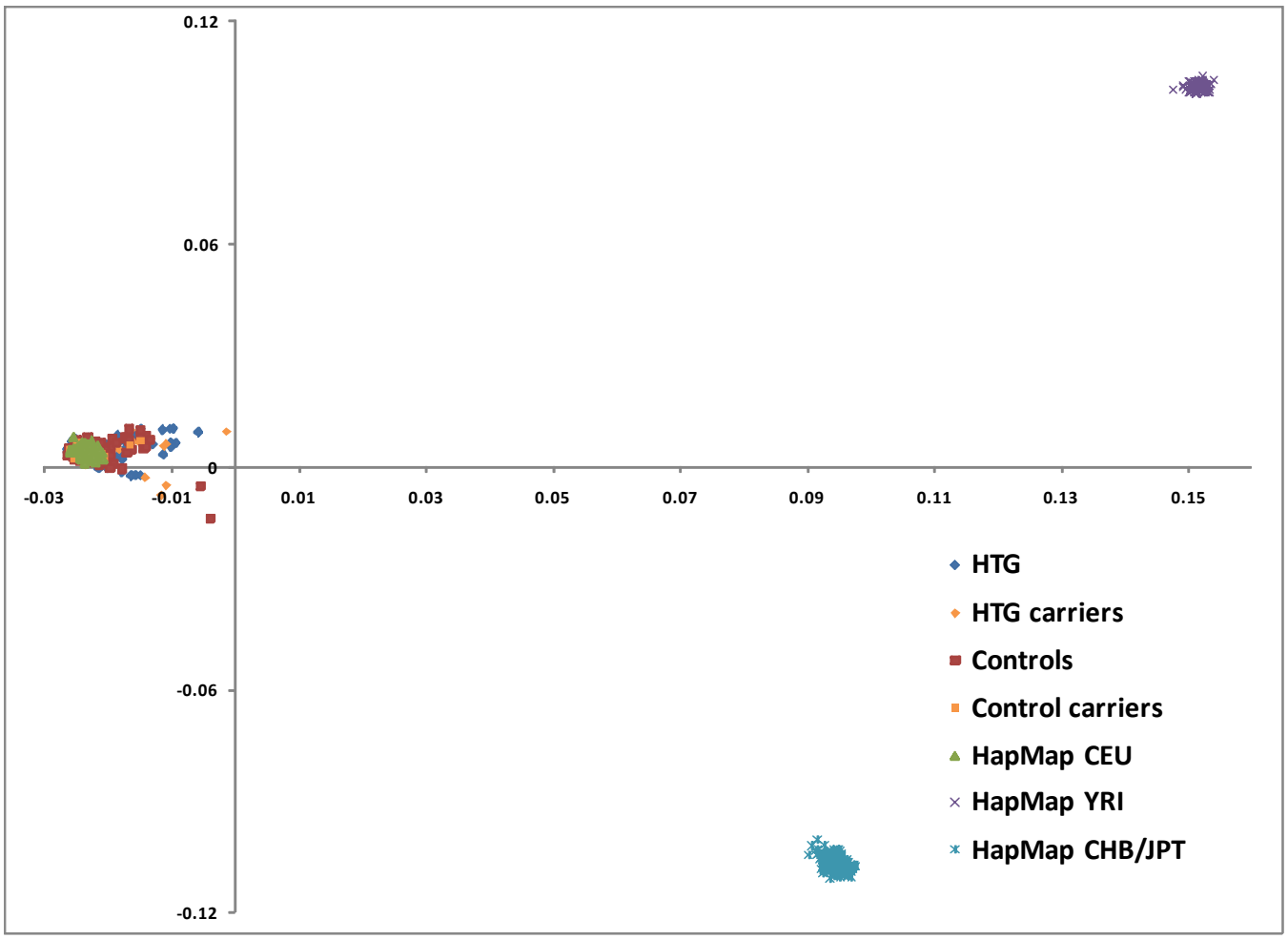
Study Subjects. This study was approved by the ethics boards at all institutions. All subjects provided informed consent for blood sampling, DNA analysis, and collection of clinical, biochemical and other demographic data. All subjects in this study were unrelated and of self-declared European ancestry. The GWAS included 463 HTG patients and 1197 low TG controls: HTG patients were predominantly obtained from a single tertiary referral lipid clinic (92% of patients) in London, Ontario, Canada, or from a tertiary referral lipid clinic in Amsterdam, Netherlands, and low TG 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[1] (18%) or the Myocardial Infarction Genetics Consortium[2] (78%). Subjects with familial hypercholesterolemia were included as negative controls only in the GWAS to counterbalance the increased cholesterol phenotype that is observed in patients with HTG. The resequencing cohort included 438 HTG patients and 327 low TG controls: HTG patients were obtained only from the lipid clinic in London, Ontario, Canada, and low TG controls included only normal healthy subjects from the Study of Health Assessment and Risk in Ethnic Groups.

Sequencing and Mutation Accumulation

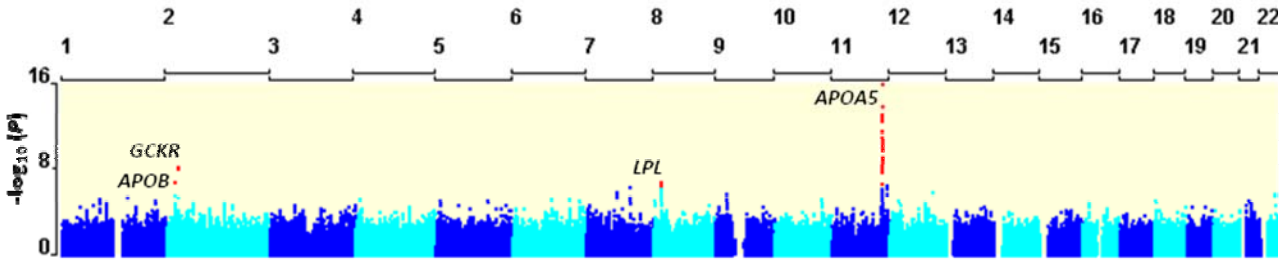
All subjects in the resequencing cohort were sequenced fully across the translated coding sequence of *APOA5*, *GCKR*, *LPL*, and exons 26 and 29 of *APOB* (67.8%). Subjects missing sequencing data in any one gene were removed prior to analysis. Rare variants were defined as having minor allele frequencies <1% in controls. Our intention was to identify rare missense and nonsense variants potentially

responsible for HTG disease causation, accordingly we excluded intronic variants, untranslated region variants, and synonymous variants from mutation accumulation analyses. Exclusive variants were defined as rare variants found exclusively in HTG patients or controls (not both), deliberately excluding variants previously reported without demonstrated functional compromise. Mutation accumulation analyses compared either the number of observed rare alleles versus reference alleles, or the number of rare variant carriers versus non-carriers, in HTG patients and controls. Carriers were defined as having ≥1 rare variant. Association between mutation accumulation and HTG phenotype was tested using a two-tailed Fisher's exact test, with nominal significance defined as $P < 0.05$.

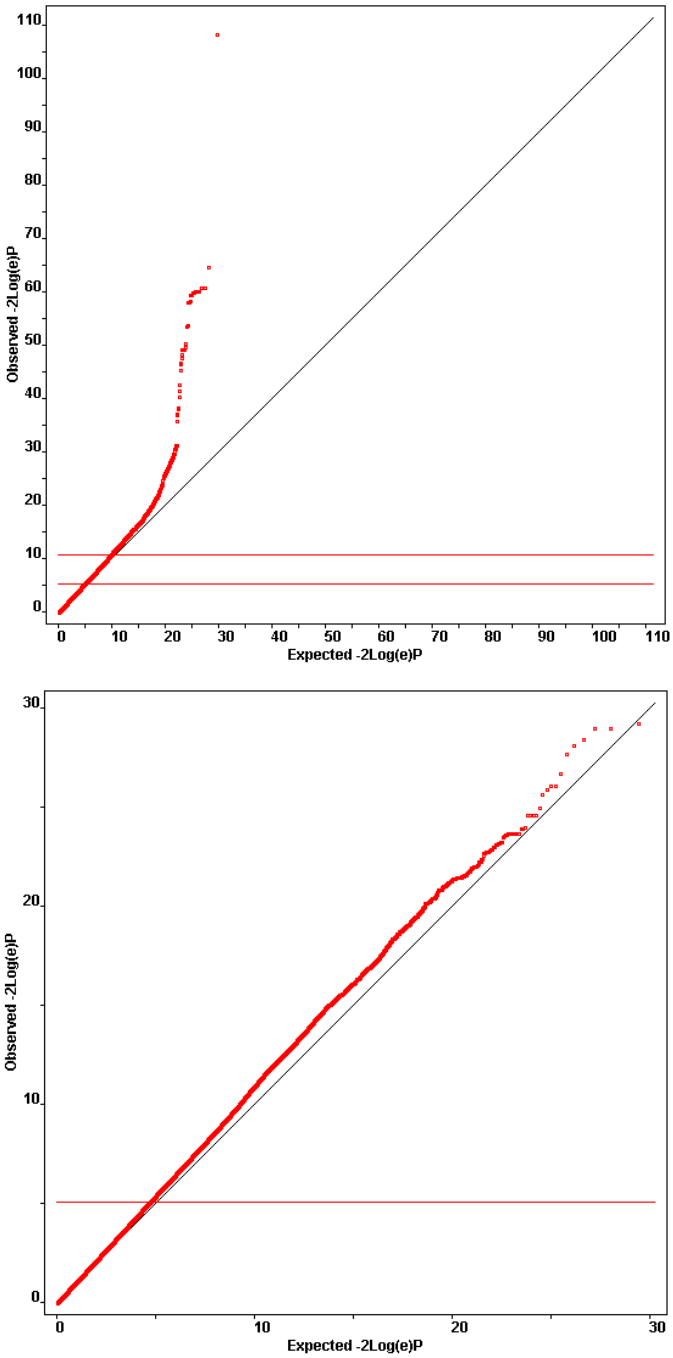
Explained Variation. Subjects included in this analysis were common to both GWAS and re-sequencing cohorts. Explained variation was calculated from the residuals of a multivariate logistic regression model, using discrete case-control status as the dependent variable. Independent variables included clinical 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 7 HTG-associated loci, and rare variants as a continuous variable including the number of rare variants carried by each subject. The calculation of explained variation was generated using a published SAS v9.2 macro written for this purpose[3]. In logistic regression, the proportion of explained variation [as measured by the coefficient of determination (R^2) in linear regression] does not exist in an interpretable straightforward manner. The macro used here produces a metric comparable to R^2 calculated from the residuals of the logistic regression model[3].



Supplemental Figure 1. Subjects in the resequencing cohort have confirmed European ancestry. Identity-by-state and multidimensional scaling was performed on resequencing subjects with whole-genome SNP data to ensure they were of European ancestry, and that rare variants identified by resequencing cannot be attributed to differences in ancestry.



Supplemental Figure 2. Manhattan plot of regression *P*-values. SNPs were independently tested for association with HTG using multivariate logistic regression entering sex, body mass index, diabetes status and 10 principal components of ancestry as covariates. A threshold of $P < 5 \times 10^{-7}$ was considered genome-wide significant. Red data points represent SNPs surpassing genome-wide significance, as visualized using WGAViewer[4]. Genome-wide associated loci are labeled. *APOA5* reached a maximum association statistic of 5.4×10^{-24}, however the y-axis scale is truncated for better visualization of other results.



Supplemental Figure 3. Quantile-quantile plot of regression *P*-values. Deviation of *P*-values from the null is caused predominantly by significant associations with hypertriglyceridemia (A), which is eliminated when these loci are removed from analysis (B). Both plots show some residual inflation of association test statistics as visualized using WGAViewer[4]. Genomic control inflation factor was calculated as $\lambda = 1.07$ in PLINK[5].

Supplemental Table 1. Annotation of rare variants found in GWAS-identified genes.

	Mutation	New/Known	HTG	Controls	Damaging?*	Published Dysfunction
APOA5						
	p.N66S	New	1	0	Possibly	
	p.G185C	Known	1	0	Possibly	Reduced LPL activation by ~23%[6]
	p.Q305X	New	1	0	Truncation	
	p.A315V	Known	1	1	Benign	
	p.D332Vfs336X	New	1	0	Truncation	
GCKR						
	p.Q8_H9insRF	New	1	0	Insertion	
	p.L37Q	New	1	0	Benign	
	p.R51G	New	1	0	Possibly	
	p.R51Q	New	1	1	Benign	
	p.G65_Q66fs88X	New	1	0	Truncation	
	p.E77G	New	1	0	Possibly	
	p.Q234P	New	6	3	Possibly	
	p.M344I	New	1	0	Benign	
	p.T379Nfs414X	New	4	1	Truncation	
	p.D414E	New	1	0	Possibly	
	p.H438Y	New	1	0	Possibly	
	p.R540X	New	1	0	Truncation	
LPL						
	p.Q-12Efs11X	Known	1	0	Truncation	
	p.E11G	New	1	1	Probably	
	p.D25H	Known	1	0	Probably	
	p.W86R	Known	1	0	Probably	Causative of chylomicronemia, LPL activity <3%[7]
	p.T186A	New	0	1	Probably	
	p.G188E	Known	10	0	Possibly	Causative of chylomicronemia, LPL activity < 1%[8]
	p.I194T	Known	1	0	Probably	Causative of chylomicronemia, LPL activity < 1%[9]
	p.P207L	Known	1	0	Probably	Causative of chylomicronemia, LPL activity < 1%[10]
	p.R243C	Known	1	0	Probably	
	p.R243H	Known	1	0	Probably	Abolished enzyme activty[11]
	p.I249V	New	0	1	Benign	
	p.C275F	New	1	0	Probably	
	p.N291S	Known	24	5	Benign	LPL activity ~60%[12]
	p.V318I	New	1	0	Benign	
APOB	p.Y1385H	New	1	0	Possibly	
	p.C1395Y	Known	1	1	Benign	
	p.G1590E	New	1	1	Probably	
	p.R1662H	Known	2	0	Benign	

p.K1703T	New	0	1	Possibly	
p.D1827N	New	0	1	Possibly	
p.V2019I	Known	0	1	Possibly	
p.A2172T	New	1	0	Benign	
p. del2186D	New	3	4	Deletion	
p.R2192C	New	1	0	Probably	
p.S2217N	New	1	0	Benign	
p.L2239M	New	1	0	Benign	
p.V2286I	Known	2	1	Benign	
p.M2331I	New	1	0	Possibly	
p.S2402T	Known	1	0	Benign	
p.A2429D	Known	1	0	Benign	
p.V2512I	Known	2	3	Benign	
p.E2539K	Known	5	0	Benign	
p.E2539D	New	1	0	Benign	
p.R2685C	New	0	1	Possibly	
p.P2794L	Known	6	2	Probably	
p.I2850Y	New	1	0	Possibly	
p.K2958E	New	1	0	Benign	
p.T3020R	Known	1	0	Benign	
p.P3216S	New	1	0	Benign	
p.S3252G	Known	4	0	Possibly	
p.M3253V	New	1	1	Possibly	
p.S3267P	Known	3	0	Possibly	
p.Q3405E	Known	5	3	Benign	
p.Y3435C	Known	1	0	Probably	
p.D3472N	New	1	0	Benign	
p.R3500W	Known	1	0	Probably	Causative of familial defective apolipoprotein B-100[13]
p.T3540M	New	0	1	Benign	
p.V3718I	New	1	1	Benign	
p.I3741T	New	1	0	Possibly	
p.D3768N	New	0	1	Possibly	
p.S3774T	New	3	3	Benign	
p. T3799M	Known	1	1	Possibly	
p.V3804F	New	0	1	Benign	
p.V4101M	Known	3	3	Benign	
p.S4206T	Known	1	0	Benign	
p.V4238A	New	8	3	Benign	
p.I4287V	Known	6	4	Benign	
p.M4293V	New	1	0	Benign	
p.V4367A	Known	1	0	Benign	
p.S4403T	Known	1	0	Benign	
p.I4455V	New	0	1	Benign	
p.T4457M	Known	6	0	Possibly	
p.F4486Ifs4488X	New	1	0	Truncation	

*Polyphen (http://genetics.bwh.harvard.edu/pph/) was used only to predict the deleterious nature of non-synonymous variants, other mutation types are indicated.

Supplemental References

1. 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**, 279-84 (2000).
2. Kathiresan, S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
3. Mittlbock, M. & Schemper, M. Computing measures of explained variation for logistic regression models. *Comput Methods Programs Biomed* **58**, 17-24 (1999).
4. Ge, D. et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* **18**, 640-3 (2008).
5. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
6. Dorfmeister, B. et al. Effects of six APOA5 variants, identified in patients with severe hypertriglyceridemia, on in vitro lipoprotein lipase activity and receptor binding. *Arterioscler Thromb Vasc Biol* **28**, 1866-71 (2008).
7. Ishimura-Oka, K. et al. A missense mutation (Trp86----Arg) in exon 3 of the lipoprotein lipase gene: a cause of familial chylomicronemia. *Am J Hum Genet* **50**, 1275-80 (1992).
8. Emi, M. et al. Missense mutation (Gly----Glu188) of human lipoprotein lipase imparting functional deficiency. *J Biol Chem* **265**, 5910-6 (1990).
9. Henderson, H.E. et al. Amino acid substitution (Ile194----Thr) in exon 5 of the lipoprotein lipase gene causes lipoprotein lipase deficiency in three unrelated probands. Support for a multicentric origin. *J Clin Invest* **87**, 2005-11 (1991).
10. Ma, Y. et al. A mutation in the human lipoprotein lipase gene as the most common cause of familial chylomicronemia in French Canadians. *N Engl J Med* **324**, 1761-6 (1991).
11. Gotoda, T. et al. Heterogeneous mutations in the human lipoprotein lipase gene in patients with familial lipoprotein lipase deficiency. *J Clin Invest* **88**, 1856-64 (1991).
12. Busca, R. et al. The mutant Asn291-->Ser human lipoprotein lipase is associated with reduced catalytic activity and does not influence binding to heparin. *FEBS Lett* **367**, 257-62 (1995).
13. Gaffney, D. et al. Independent mutations at codon 3500 of the apolipoprotein B gene are associated with hyperlipidemia. *Arterioscler Thromb Vasc Biol* **15**, 1025-9 (1995).