Association of Rare Variants with Plasma HDL Cholesterol and Triglycerides

Gina M Peloso
MGH/Broad

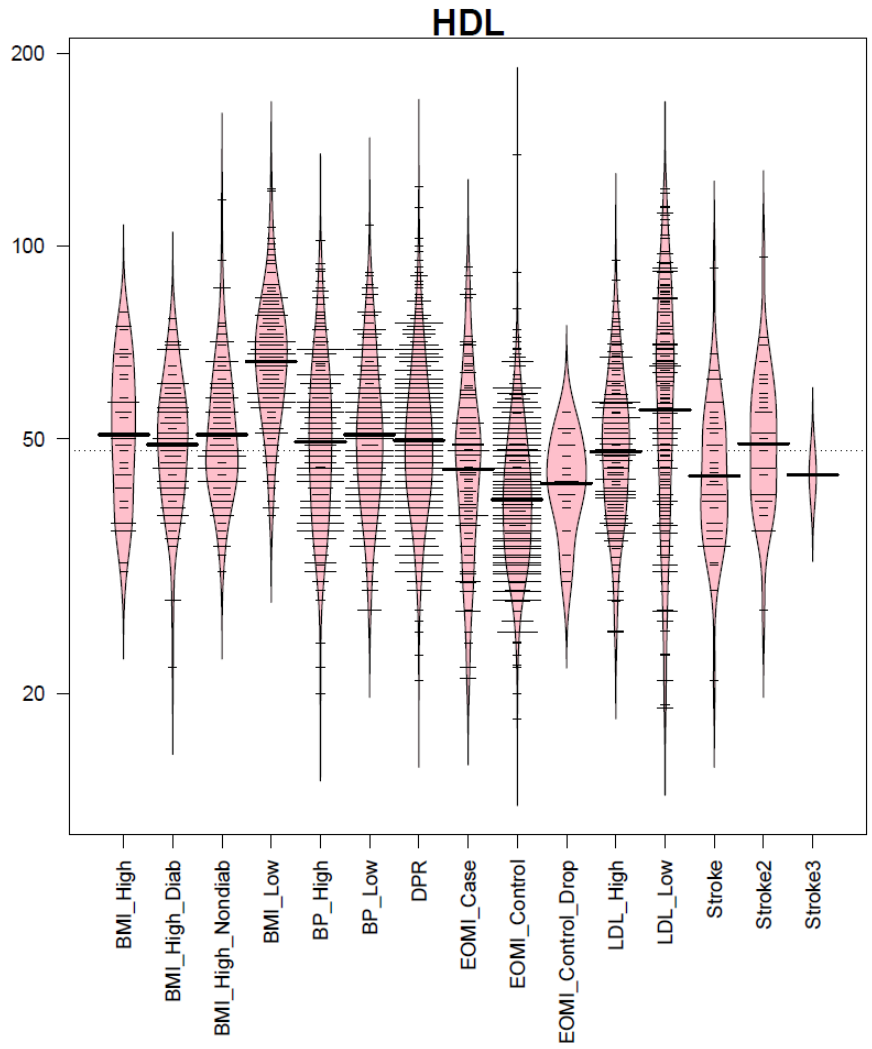
ESP In-person Meeting
March 28, 2012

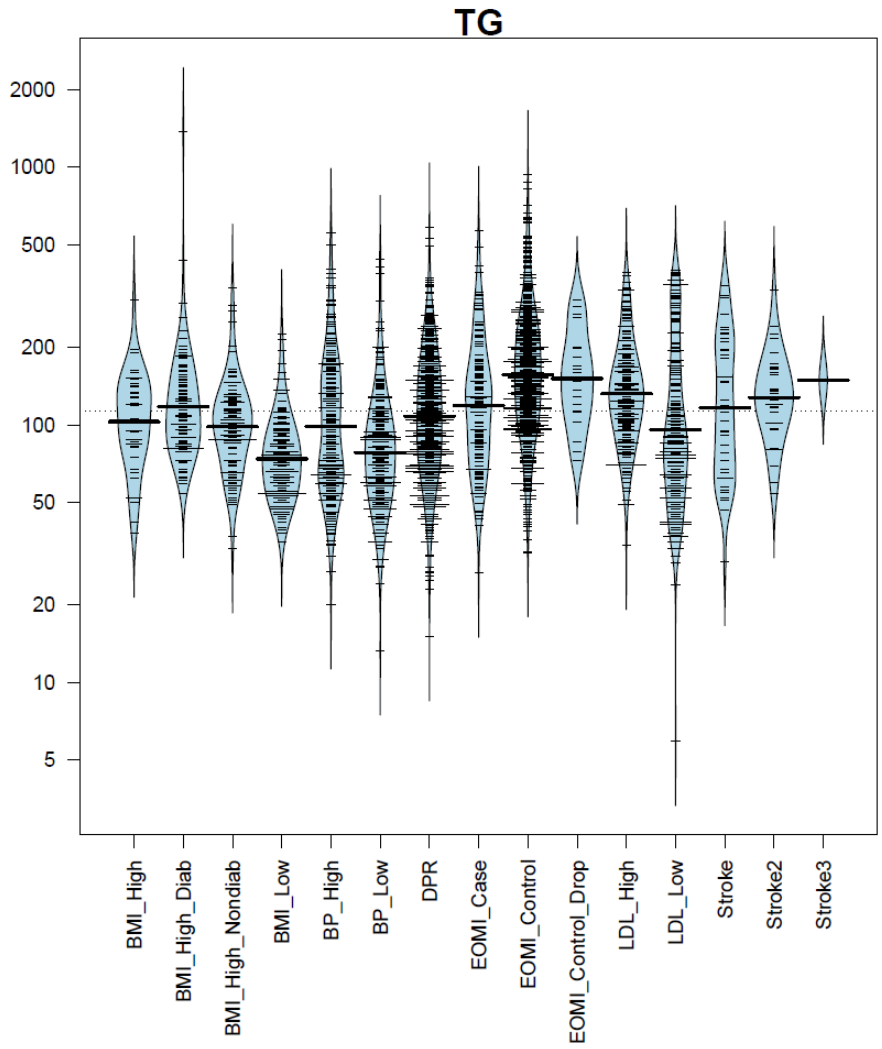
Data

- ESP5500 VCF
- 2012Feb01 phenotype file
- Subjects available for analysis (with genotype and phenotype data)

Group	AA	EA	Total
Broad	628	752	1380
UW	698	404	1102
Total	1326	1156	2482

Trait Distributions





Hypotheses

- Low frequency (0.5-5%) or common (>5%) variants associate with plasma HDL-C or TG.
  - Use single-variant analysis to test
- An aggregation of low-frequency and/or rare variants associate with plasma HDL-C or TG.
  - Use gene-based analyses to test

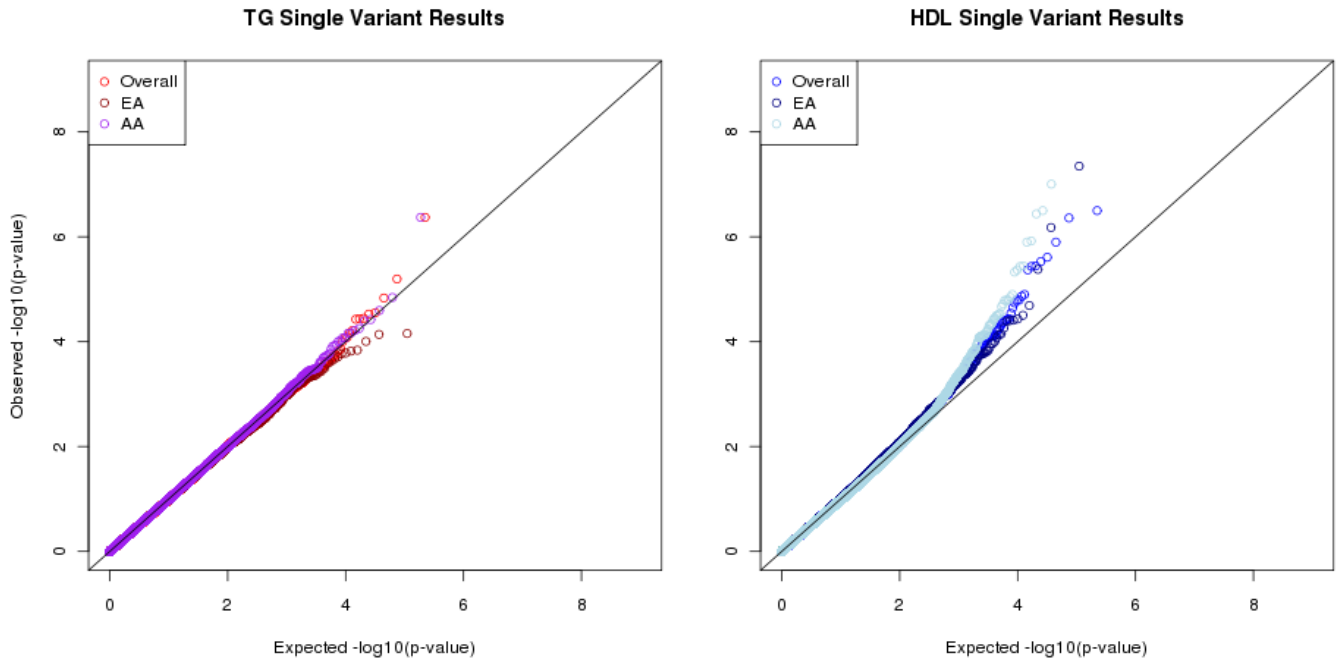
Hypotheses

- Low frequency (0.5-5%) or common (>5%) variants associate with plasma HDL-C or TG.
  - Use single-variant analysis to test

- An aggregation of low-frequency and/or rare variants associate with plasma HDL-C or TG.
  - Use gene-based analyses to test

Single Marker Analysis

- Trait = variant + age + sex + PCs
- Race and sequencing center specific analysis
- Sample size weighted meta-analysis to combine



Single Variant Results
Top Results

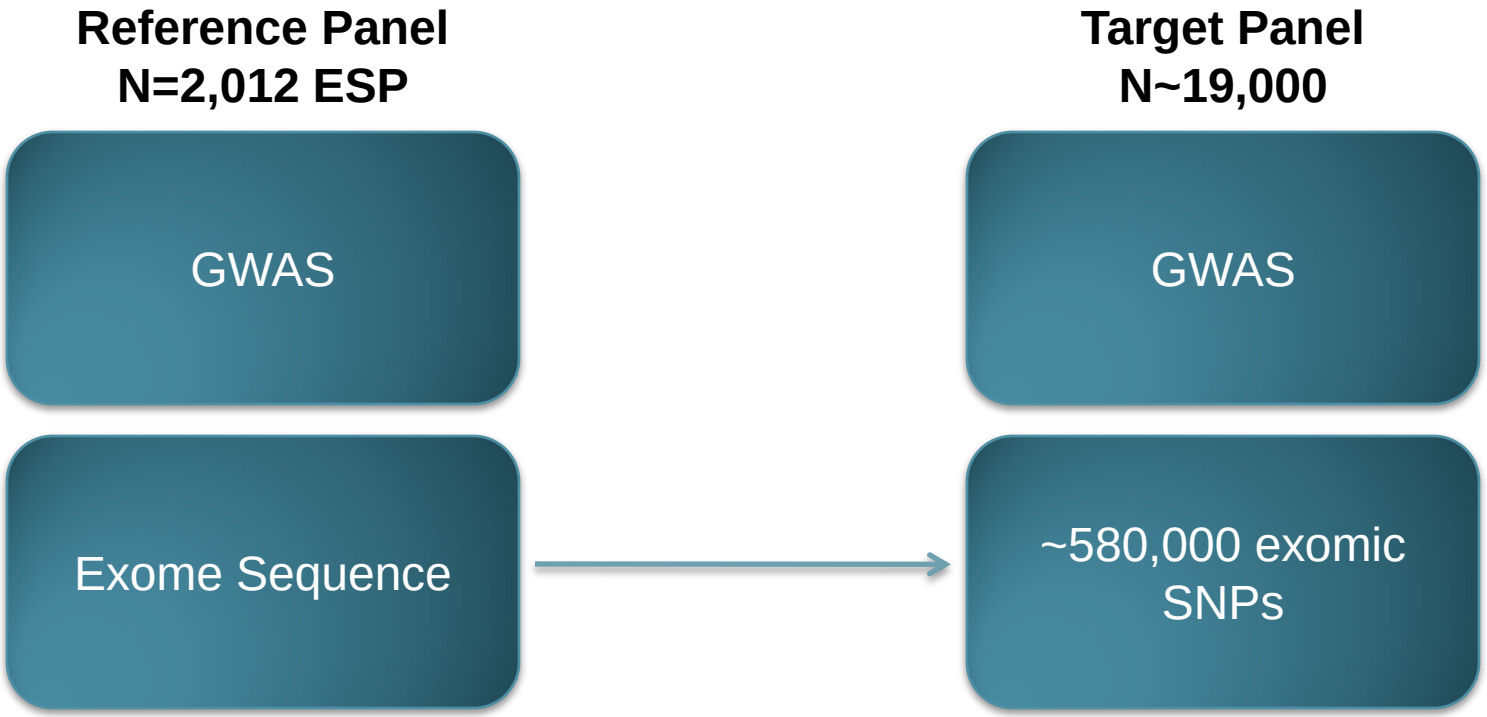
Trait	Gene	MarkerName	AF	EFFECT	On Exome Chip?	n	Direction	P-value
TG	CD209	chr19:7810421	0.1%	NON_SYN	No	1296	?-?-	4.26E-07
	APOA5	chr11:116662407	6.5%	NON_SYN	Yes	2454	++++	6.38E-06
	PSMB1	chr6:170862300	35.5%	NON_SYN	Yes	2455	----	1.47E-05
	LGI4	chr19:35617293	1.5%	NON_SYN	Yes	2051	++?+	2.81E-05
	CLVS1	chr8:62412077	0.1%	NON_SYN	No	1453	+??+	2.98E-05
HDL	DDX20	chr1:112302035	0.1%	NON_SYN	Yes	1296	?-?-	3.60E-06
	CD36	chr7:80300449	2.8%	STOP_GAINED	Yes	1296	?-?-	1.26E-05
	C12orf45	chr12:105388357	0.4%	STOP_GAINED	No	2209	++++	1.36E-05
	NCAM1	chr11:113102355	0.8%	NON_SYN	No	2373	++++	2.26E-05
	FDXACB1	chr11:111747725	1.5%	START_GAINED	Yes	2227	++++	5.41E-05
	OMP	chr11:76814268	1.1%	NON_SYN	Yes	1155	?+?+	7.70E-05
	ABCC2	chr10:101604107	1.2%	NON_SYN	Yes	1700	?+++	8.30E-05

Single Marker Replication

- Analysis of rare variants on Exome Chip for association with plasma HDL-C and TG

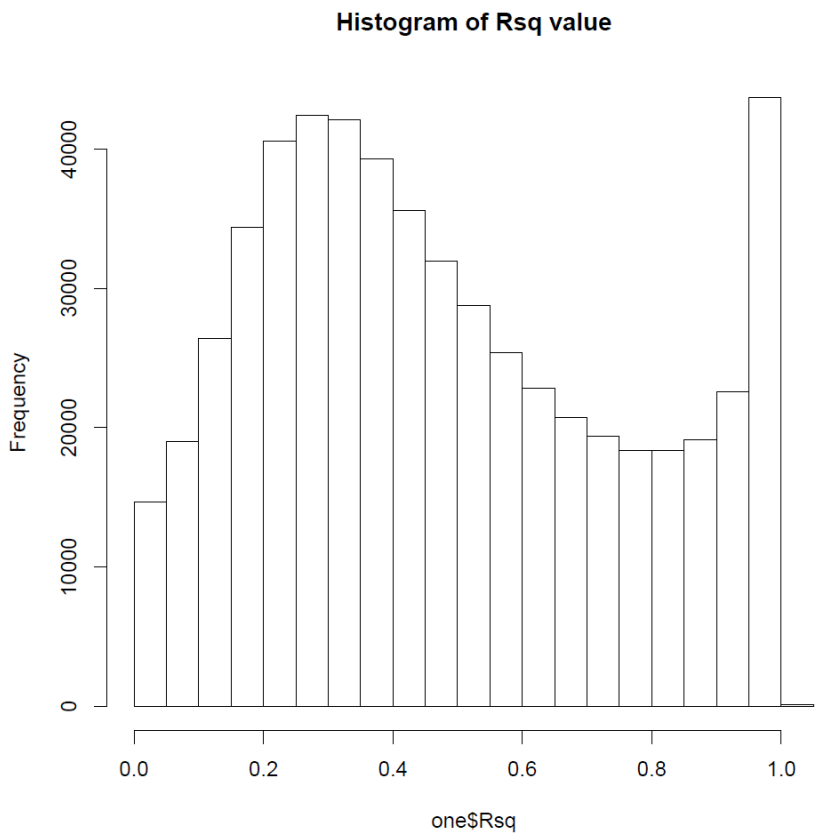
- Imputation into GWAS Caucasian Cohorts
  - Will describe in next few slides

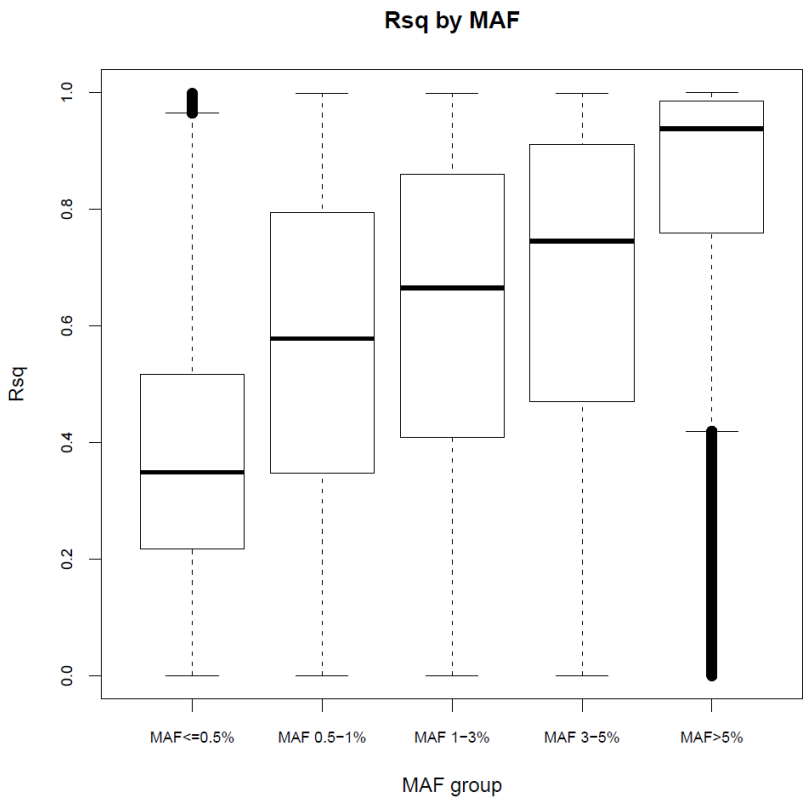
Statistical Imputation



Statistically infer exome variants into imputation target panel based on haplotype structure

Imputation Quality





Preliminary Imputation Results Positive Control Variants

Trait	Gene	MarkerName	MAF	n	Direction	P-value
TG	APOA5	chr11:116662407 p.Ser19Trp	6.4%	19,134	+++	1.35E-32
	GCKR	chr2:27730940 p.Leu446Pro	3.8%	19,134	+++	4.83E-35
HDL	LPL	chr8:19819724 p.S474*	9.5%	19,172	---	1.41E-29
	CETP	chr16:57017319 p.R468Q	3.1%	19,172	---	3.04E-20
	LIPG	chr18:47109955 p.N396S	1.2%	19,172	---	4.07E-15

- Top coding results
  - TG: APOA5, GCKR, ZNF259, LPL, APOC3
  - HDL: CETP, LPL, LIPG, APOC3

Imputation As Replication for Top ESP Results

Trait	Gene	MarkerName	On Exome Chip?	Discovery P-value	MAF	n	Replication P-value
TG	CD209	chr19:7810421	No	4.26E-07	NA		
	APOA5	chr11:116662407	Yes	6.38E-06	6.4%	19,134	1.35E-32
	PSMB1	chr6:170862300	Yes	1.47E-05	38.6%	19,134	0.82
	LGI4	chr19:35617293	Yes	2.81E-05	NA		
	CLVS1	chr8:62412077	No	2.98E-05	NA		
HDL	DDX20	chr1:112302035	Yes	3.60E-06	0.05%	11,623	0.74
	CD36	chr7:80300449	Yes	1.26E-05	3.0%	11,623	0.43
	C12orf45	chr12:105388357	No	1.36E-05	0.4%	16,682	0.003
	NCAM1	chr11:113102355	No	2.26E-05	0.8%	16,682	0.003
	FDXACB1	chr11:111747725	Yes	5.41E-05	1.4%	19,172	0.85
	OMP	chr11:76814268	Yes	7.70E-05	0.9%	11,623	0.26
	ABCC2	chr10:101604107	Yes	8.30E-05	0.7%	19,172	0.30

Imputation As Replication for Top ESP Results

Trait	Gene	MarkerName	On Exome Chip?	Discovery P-value	MAF	n	Replication P-value
TG	CD209	chr19:7810421	No	4.26E-07	NA		
	APOA5	chr11:116662407	Yes	6.38E-06	6.4%	19,134	1.35E-32
	PSMB1	chr6:170862300	Yes	1.47E-05	38.6%	19,134	0.82
	LGI4	chr19:35617293	Yes	2.81E-05	NA		
	CLVS1	chr8:62412077	No	2.98E-05	NA		
HDL	DDX20	chr1:112302035	Yes	3.60E-06	0.05%	11,623	0.74
	CD36	chr7:80300449	Yes	1.26E-05	3.0%	11,623	0.43
	C12orf45	chr12:105388357	No	1.36E-05	0.4%	16,682	0.003
	NCAM1	chr11:113102355	No	2.26E-05	0.8%	16,682	0.003
	FDXACB1	chr11:111747725	Yes	5.41E-05	1.4%	19,172	0.85
	OMP	chr11:76814268	Yes	7.70E-05	0.9%	11,623	0.26
	ABCC2	chr10:101604107	Yes	8.30E-05	0.7%	19,172	0.30

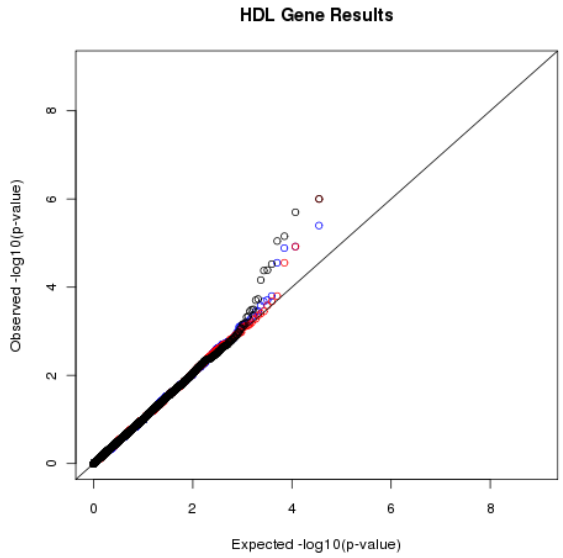
Hypotheses

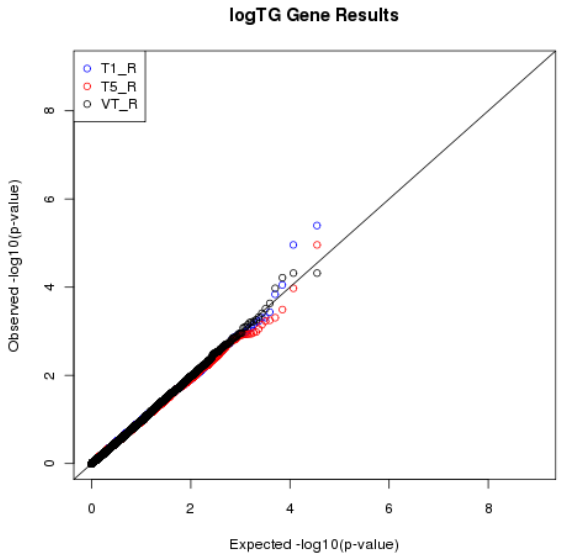
- Low frequency (0.5-5%) or common (>5%) variants associate with plasma HDL-C or TG.
  - Use single-variant analysis to test

- An aggregation of low-frequency and/or rare variants associate with plasma HDL-C or TG.
  - Use gene-based analyses to test

Gene-Based Analysis

- Using Score-Seq
- Trait = burden + age + sex + PCs
- Non-synonymous/nonsense/splice variants with MAF < 0.05





http://www.bios.unc.edu/~lin/software/SCORE-Seq/

Gene-Based Results
"Positive Control" Genes

Trait	Gene	VT p-value
HDL	ABCA1	1.00E-06
	CETP	2.00E-06
	LIPG	0.31
	LPL	0.39
TG	APOC3	4.80E-05
	APOB	0.05
	APOA5	0.11
	GCKR	0.20
	LPL	0.70

Gene-Based Results
Top Results

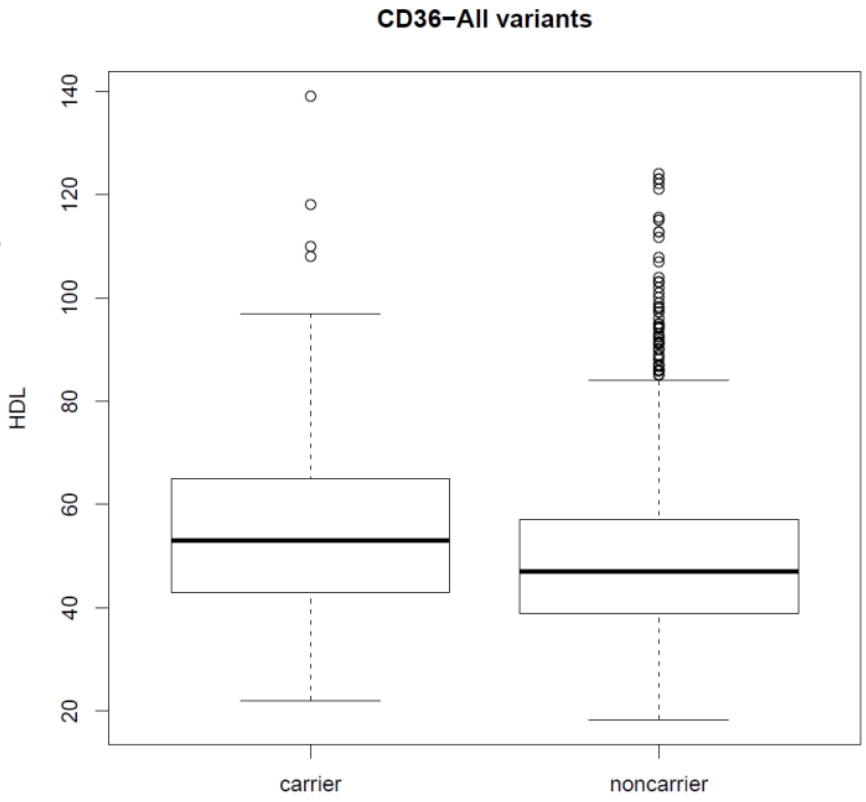
Trait	GENE	T5 p-value	VT p-value
HDL	ABCA1	3.97E-04	1.00E-06
	CETP	0.02	2.00E-06
	CD36	1.00E-06	7.00E-06
	ATG13	0.64	9.00E-06
	DNAI2	5.26E-04	3.00E-05
	LCN2	0.05	4.10E-05
	CSRP2	4.40E-04	4.20E-05
	APOC3	1.20E-05	6.90E-05
	PSMD5	2.80E-05	1.86E-04
	SOCS6	1.6E-04	6.43E-04
TG	APOC3	1.10E-05	4.80E-05
	FILIP1	3.22E-04	4.80E-05
	HIGH	0.02	6.10E-05
	AP1S1	0.003	1.06E-04
	ATP2C2	0.002	3.91E-04
	MED8	1.06E-04	5.48E-04

Gene-Based Results
Top Results

Trait	GENE	T5 p-value	VT p-value
HDL	ABCA1	3.97E-04	1.00E-06
	CETP	0.02	2.00E-06
	CD36	1.00E-06	7.00E-06
	ATG13	0.64	9.00E-06
	DNAI2	5.26E-04	3.00E-05
	LCN2	0.05	4.10E-05
	CSRP2	4.40E-04	4.20E-05
	APOC3	1.20E-05	6.90E-05
	PSMD5	2.80E-05	1.86E-04
	SOCS6	1.6E-04	6.43E-04
TG	APOC3	1.10E-05	4.80E-05
	FILIP1	3.22E-04	4.80E-05
	HIGH	0.02	6.10E-05
	AP1S1	0.003	1.06E-04
	ATP2C2	0.002	3.91E-04
	MED8	1.06E-04	5.48E-04

CD36 Association With HDL

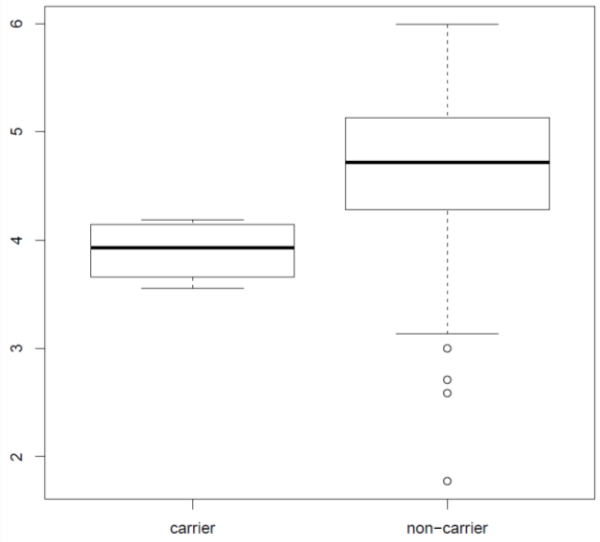
- P-value ~ 7.00E-06
- 65 variants in gene with HDL values
  - 48 singletons
  - 8 doubletons
  - 9 with MAC ≥ 3
- 29 Variants on exome chip



APOC3 Association With TG

- P-value ~ 4.80E-05

POS	codon	protein	AC	Exome Chip	EFFECT	TG Levels
chr11:116701284			1		SPLICE SITE ACCEPTOR	
chr11:116701326	c.G28A	p.A10T	2		NON SYNONYMOUS	67
chr11:116701353	C55T	R19X	1	Y	STOP GAINED	
chr11:116701354	c.55+1G>A		14	Y	SPLICE SITE DONOR	43,35,66, 60,68
chr11:116701560	c.G127A	p.A43T	11	Y	NON SYNONYMOUS	42,79,51, 36,23,381, 73,58
chr11:116701613	c.179+1G>T		5	Y	SPLICE SITE DONOR	
chr11:116703493	c.G193A	p.D65N	2		NON SYNONYMOUS	415
chr11:116703532	c.A232G	p.K78E	1		NON SYNONYMOUS	



Gene Based Replication

- Need re-sequencing to confirm genes
- Exome chip
  - Will work for some genes
- Collaborate with CHARGE-S

Summary

- For both single marker and gene-based analyses, known variants and genes are among top associations.

Summary

- For both single marker and gene-based analyses, known variants and genes are among top associations.

- Genotyping and sequencing-based replication is needed to make new rare-variant discoveries

Summary

- For both single marker and gene-based analyses, known variants and genes are among top associations.
- Genotyping and sequencing-based replication is needed to make new rare-variant discoveries
- Primary genotyping follow-up will be Exome Chip (ESP Exome Chip samples with lipids, possibly CHARGE Exome Chip)

Summary

- For both single marker and gene-based analyses, known variants and genes are among top associations.
- Genotyping and sequencing-based replication is needed to make new rare-variant discoveries
- Primary genotyping follow-up will be Exome Chip (ESP Exome Chip samples with lipids, possibly CHARGE Exome Chip)
- Primary sequencing-based follow-up will be CHARGE-S exomes

Acknowledgements

MGH/Broad
- Sekar Kathiresan
- Ron Do
- Nathan Stitziel

University of Texas
- Jacy Crosby
- Eric Boerwinkle
- Alanna Morrison

University of Washington
- David Crosslin
- Gail Jarvik
- Paul Auer

University of Michigan
- Cristen Willer
- Goncalo Abecasis

UNC
- Leslie Lange
- Kari North