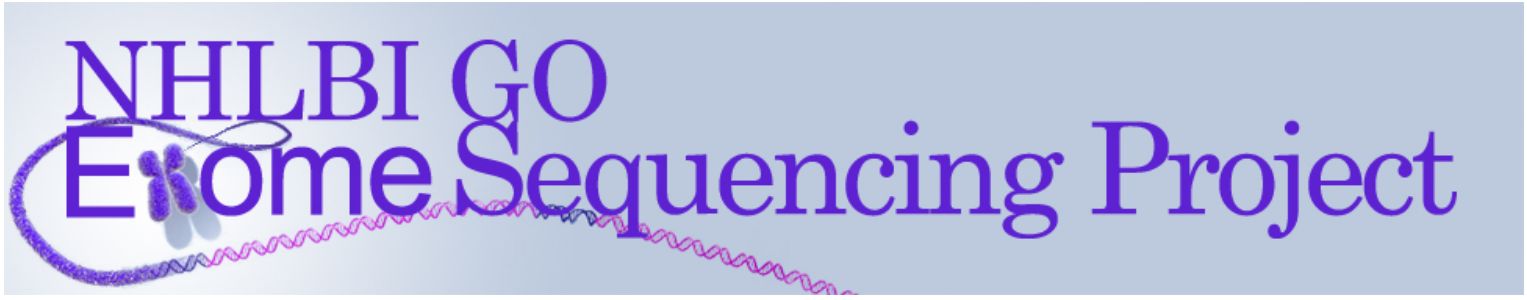
Common and rare frequency variants contribute to extreme LDL levels: findings from the NHLBI Exome Sequencing Project

Cristen Willer

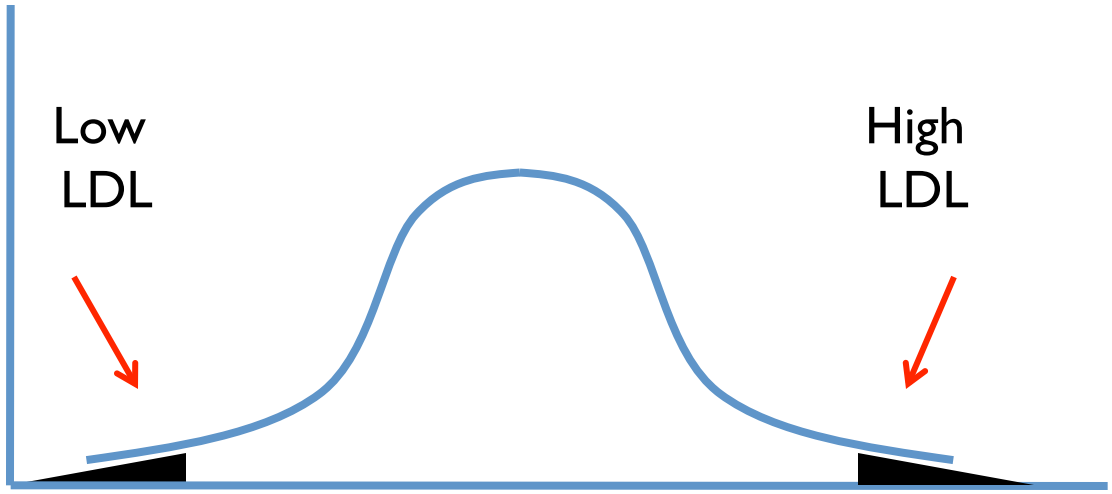
& Leslie Lange, Youna Hu



Exome Sequencing Project
LDL Extremes Sample Selection

Extremes of phenotype approach

Total samples sequenced (QC+): N = 412



African American	N = 93 (mean 49 mg/dl)	N = 102 (mean 242 mg/dl)
European American	N = 108 (mean 46 mg/dl)	N = 109 (mean 283 mg/dl)

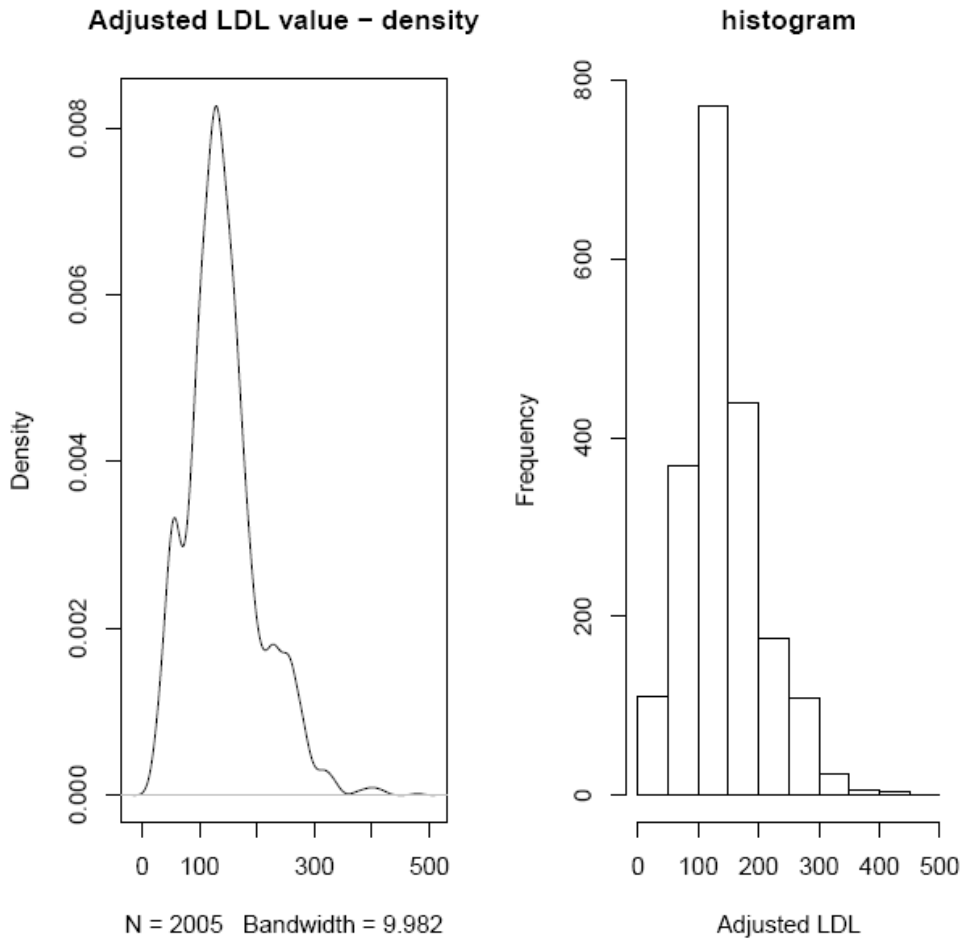
Samples are from ARIC, Cardiovascular Health Study, Framingham Heart Study, Jackson Heart Study

New since last presentation

- Analysis dataset of all ESP5500 individuals w/LDL
  - 412 extremes for LDL (< 2nd and > 98th percentile)
  - 1593 others selected for other traits
  - N = 2005 total

- QC of ESP LDL 2005 dataset
  - Y/X ratio – exclude 3 individuals
  - Excluded 31 individuals to avoid 1st and 2nd relatives
  - Genotypes filtered at depth 10 and 50% call rate

Med-adjusted LDL in ESPLDL2005



Analysis plan

- To deal with ascertainment, we flipped regression using LDL levels to predict genotype

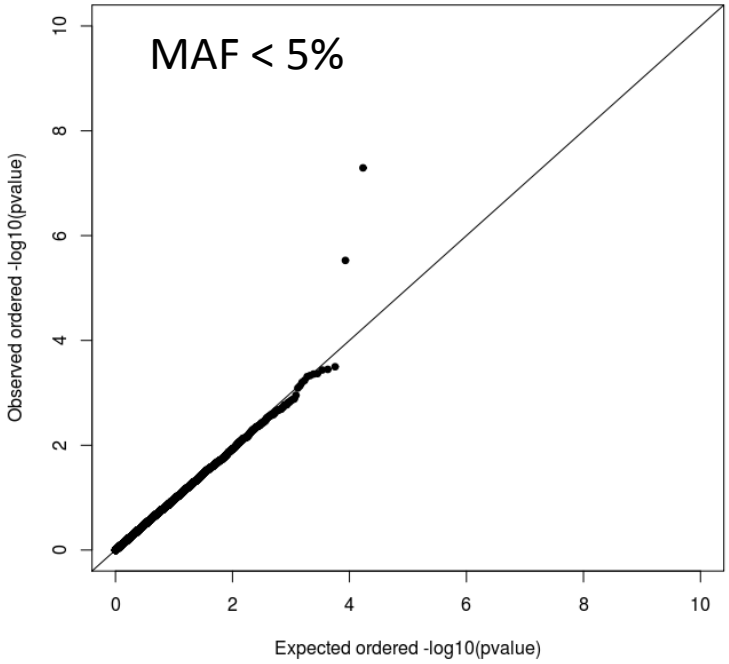
$$logit\big(P(burden\ score = 1)\big) = \beta_0 + \beta_1 LDL + \beta_2 age + \beta_3 sex + \beta_4 PC_1 + \beta_5 PC_2 + \beta_6 Center$$

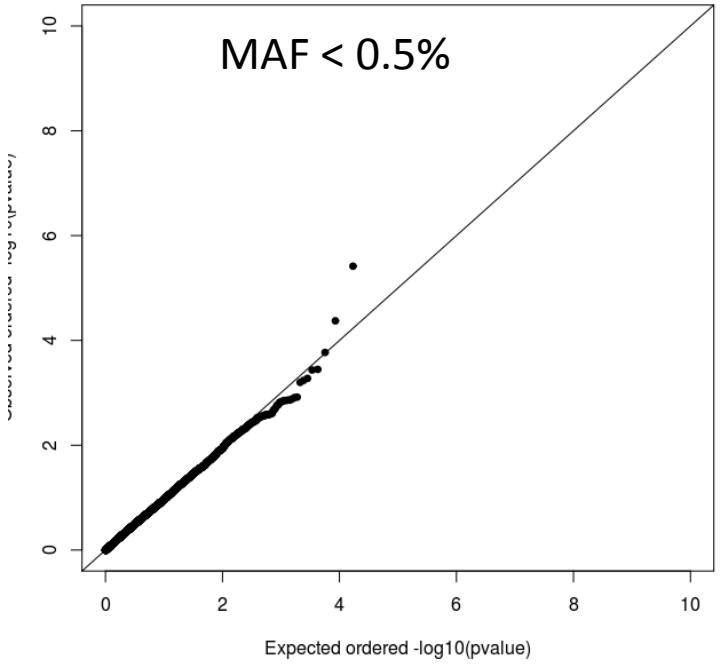
- Used variant frequency estimates from ESP5500 for determining inclusion in frequency-based burden tests
- Used multiple imputation for missing genotypes and averaged p-values across each imputation
- No ascertainment correction, vulnerable to indirect signals (gene -> ascertained trait -> LDL)

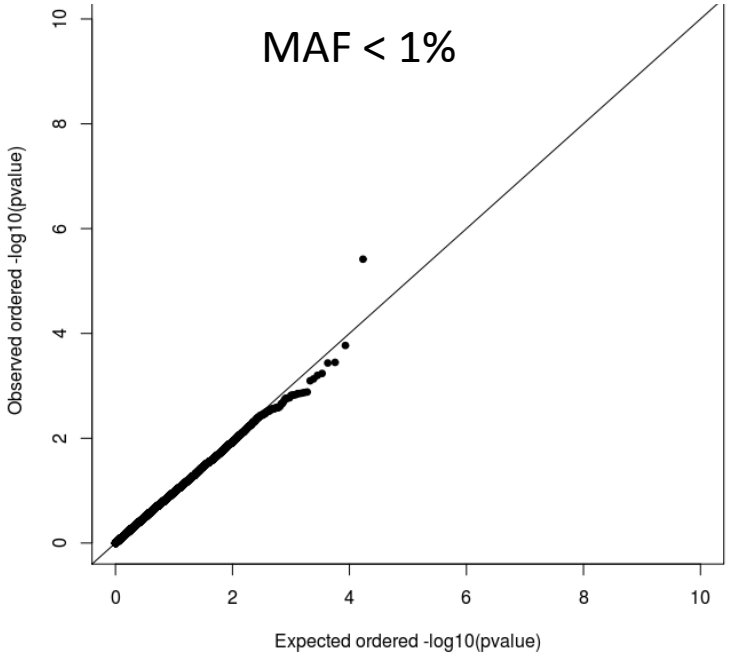
Q-Q plots of reverse regression burden test p-values

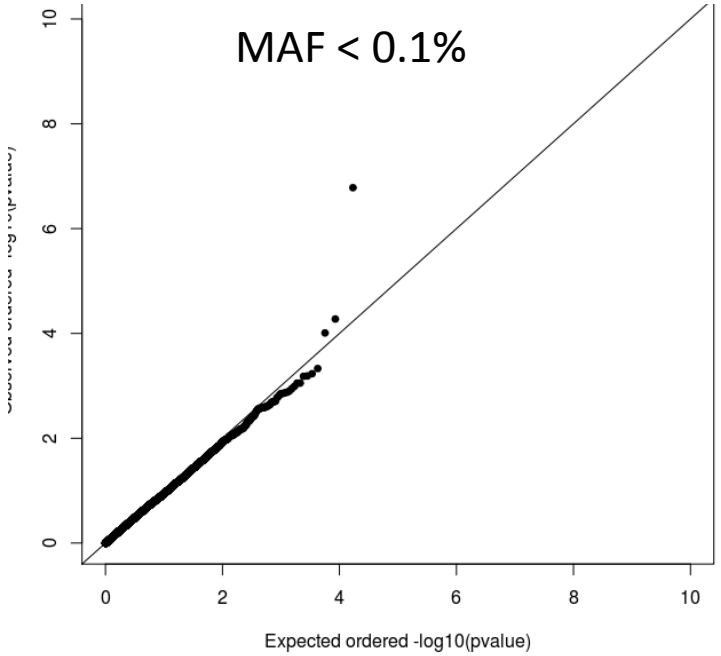
Analysis tests: Nonsynonymous variants with various frequency thresholds

LOF variants (nonsense, readthrough, splice) with MAF < 5%









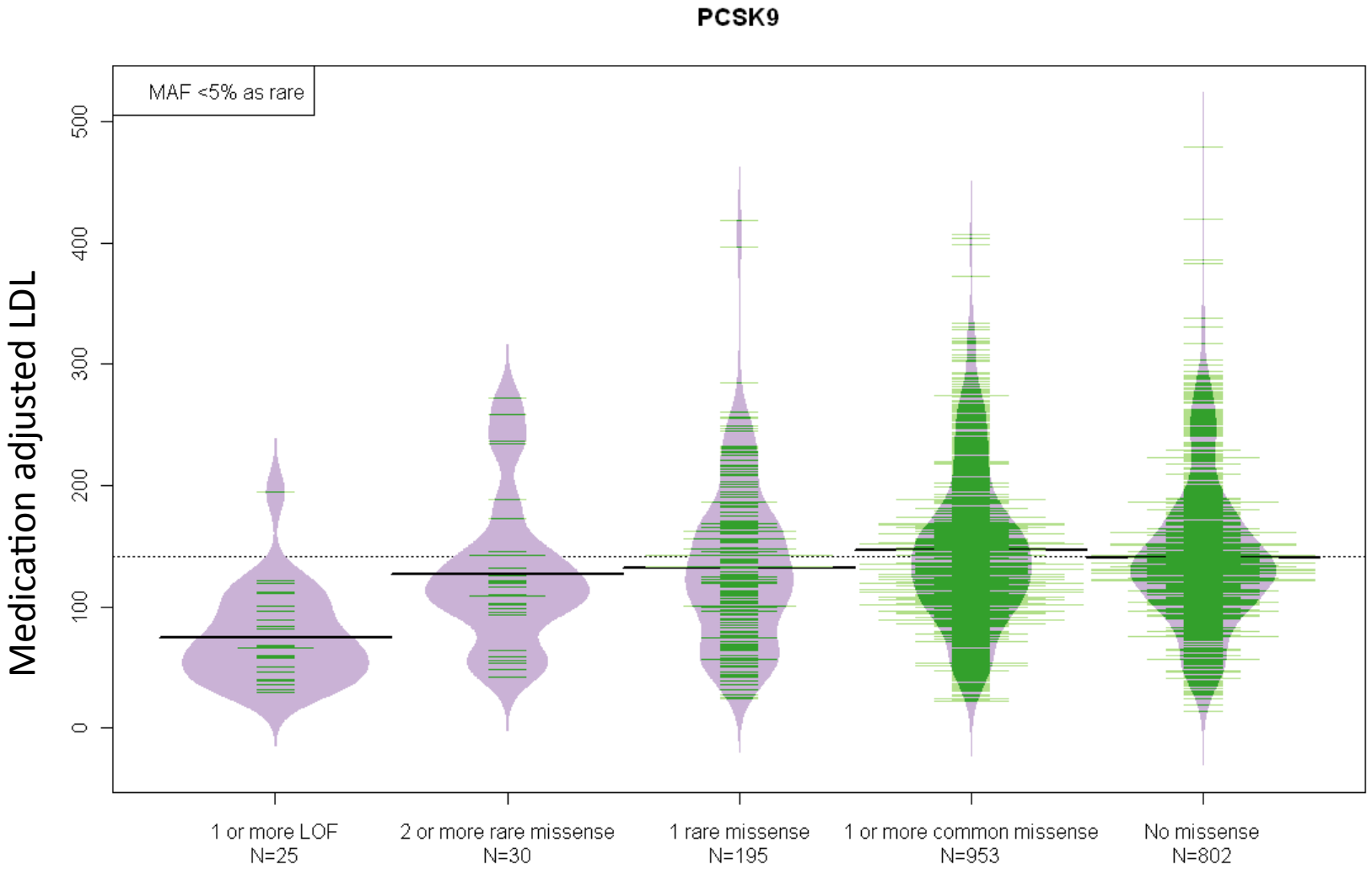
Results for previously known lipid genes in ESPLDL2005 dataset

Gene	Test (most significant p-value)	P-value	Number of variants N/S/R/M*	Singleton Variants N/S/R/M*	% non-singleton alleles/% alleles on exome chip	% High/ Low EA cases with variant	% High/ Low AA cases with variant
PCSK9	NS <5%	5.1×10^{-8}	3/1/0/36	1/1/0/21	304 (92.1)	2.8/13.9	17.5/35.5
LDLR	NS <0.1%	1.7×10^{-7}	2/2/0/34	2/2/0/24	57 (49.1)	9.2/0	6.8/1.1
ABCG5	NS <5%	3.2×10^{-4}	3/0/0/35	3/0/0/20	276 (93.1)	6.5/3.7	31.1/19.4
APOB	LOF <5%	4.4×10^{-4}	5/3/0/-	5/3/0/-	8 (0)	0/5.6	0/1.1
NPC1L1	NS <0.5%	5.3×10^{-4}	2/1/0/59	2/1/0/26	261 (88.1)	2.8/9.2	13.6/24.7

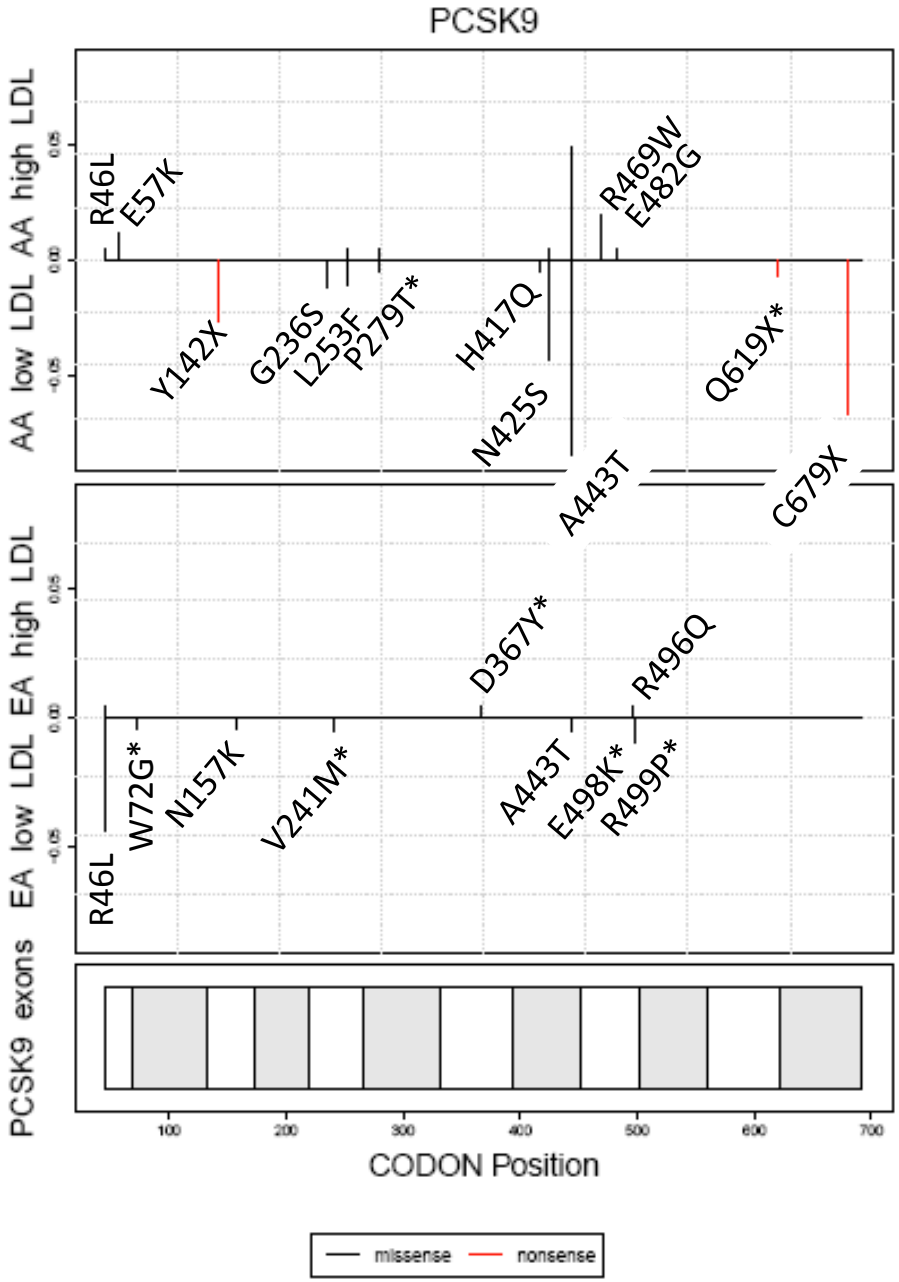
- Variant Categories: Nonsense, Splice, Readthrough, Missense
- 2 Genes at genome-wide significance of $p < 3 \times 10^{-6}$

Distribution of LDL values by mutation type in the *PCSK9* gene

ESPLDL2005 dataset



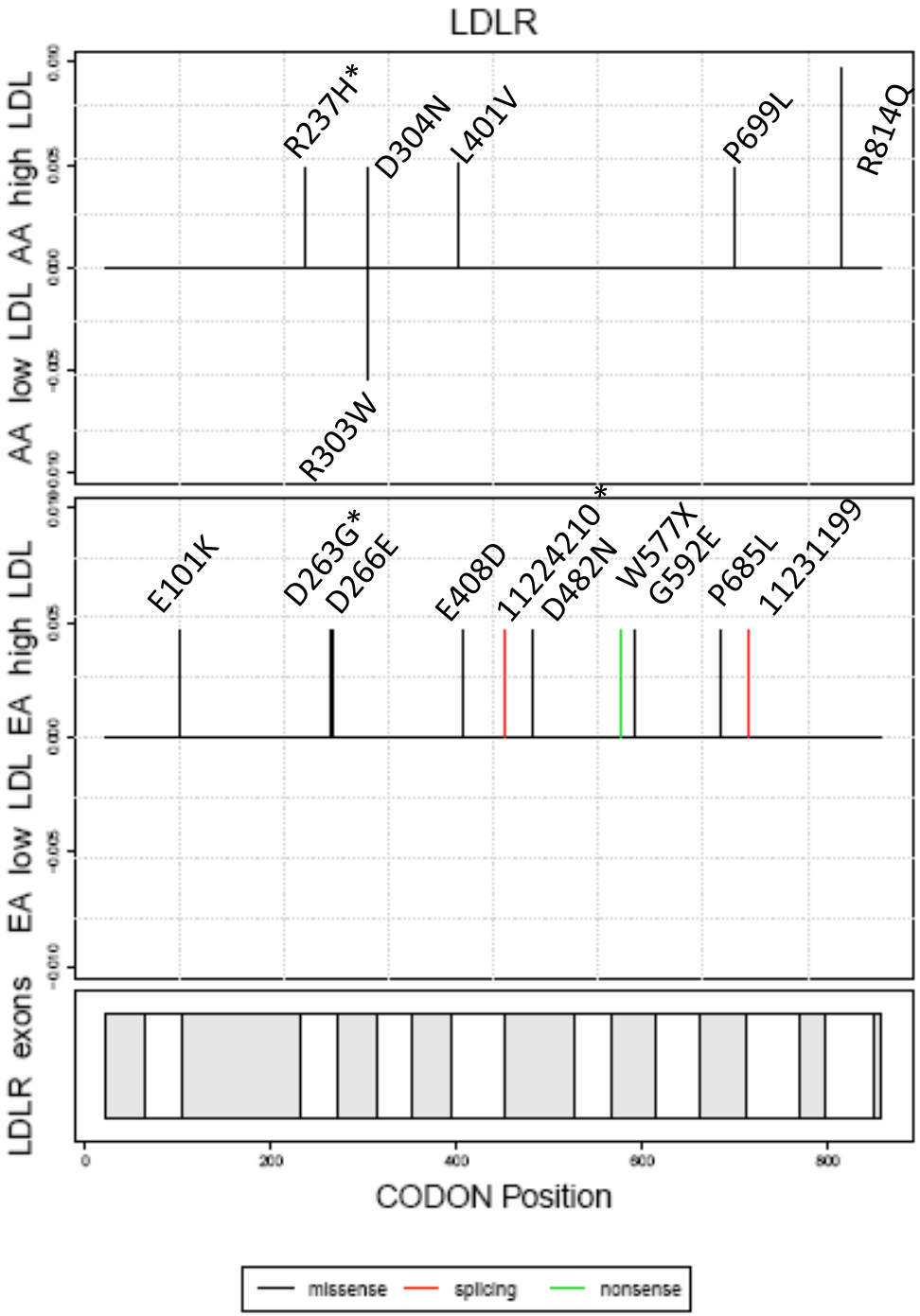
Mutations in PCSK9, which mediates degradation of LDLR, can cause familial hypobetalipoproteinemia or hypocholesterolemia (Abifadel 2003, Shioji 2004, Cohen 2005)



Variants observed in the ESPLDL412 extreme individuals in *PCSK9*

57 in low LDL vs 24 in high LDL
Top gene in 4 of 5 burden tests
Nonsyn MAF < 5% $p = 5 \times 10^{-8}$
LOF MAF < 5% $p = 3 \times 10^{-7}$
Nonsyn MAF < 1%, $p = 4 \times 10^{-6}$
Nonsyn MAF < 0.5% $p = 4 \times 10^{-6}$

Although R46L & A443T, reported by Cohen and Hobbs in 2005, are excluded in tests for MAF < 1%, we still see $p = 4 \times 10^{-6}$



Variants observed in the ESPLDL412 extreme individuals in *LDLR*

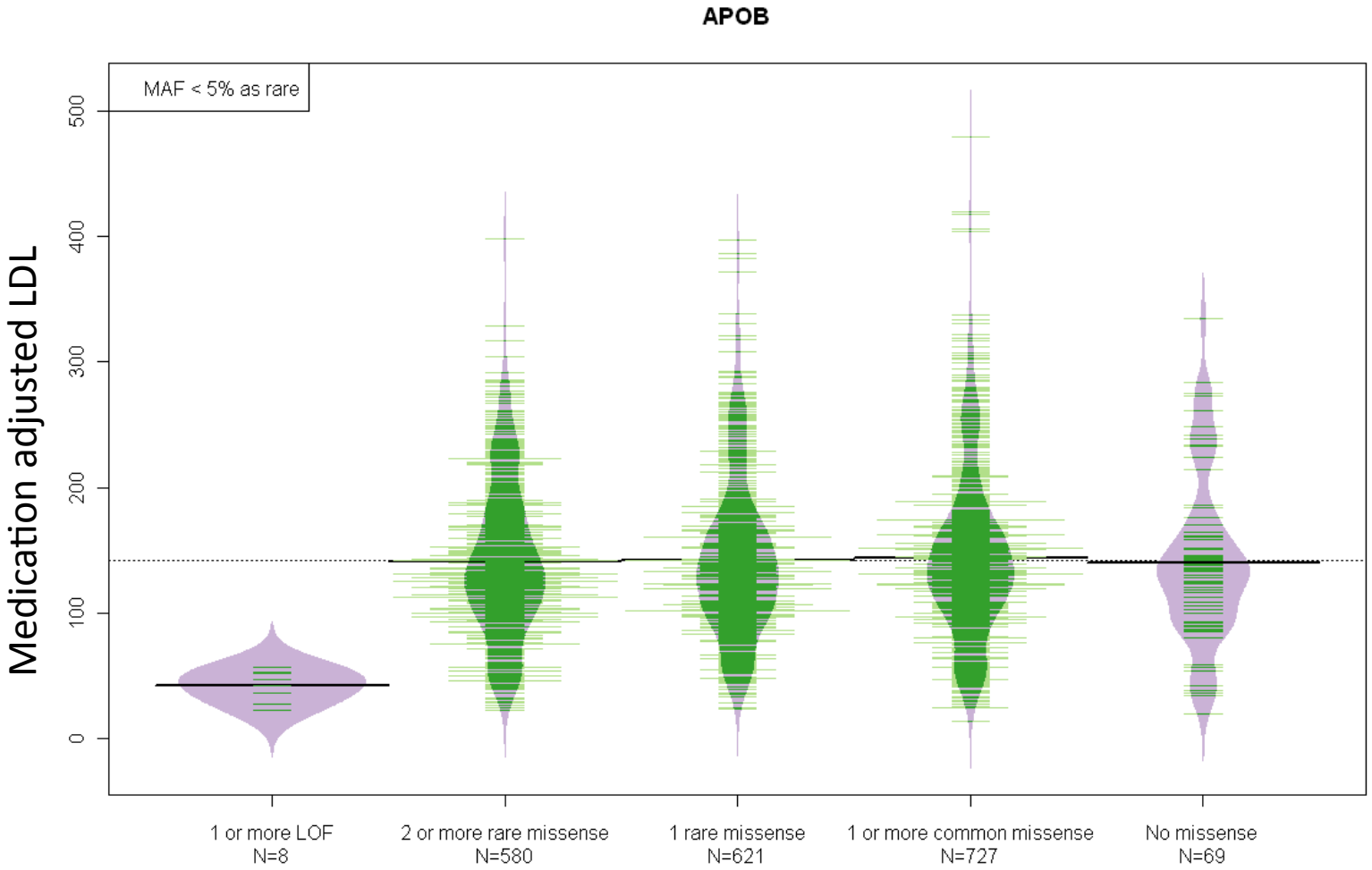
17 in high LDL vs 1 in low LDL
Nonsyn MAF < 0.1% p = 2×10^{-7}
LOF MAF < 5% p = 1×10^{-3}

16 variants identified in 412 extremes

- 3 are novel sites (E408D)
- 1 is novel amino acid change

1122 previously reported variants

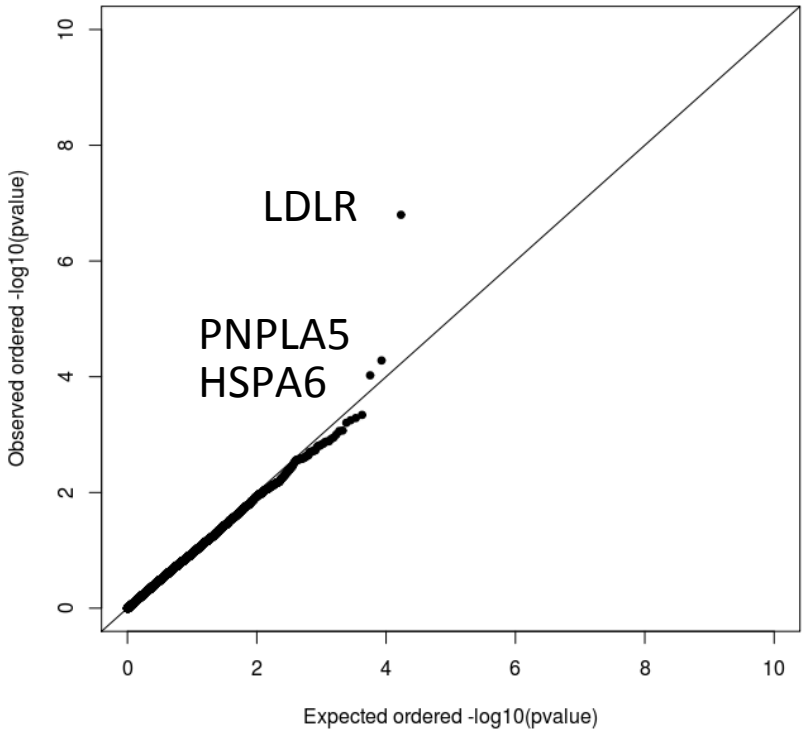
Distribution of LDL values by mutation type in the *APOB* gene



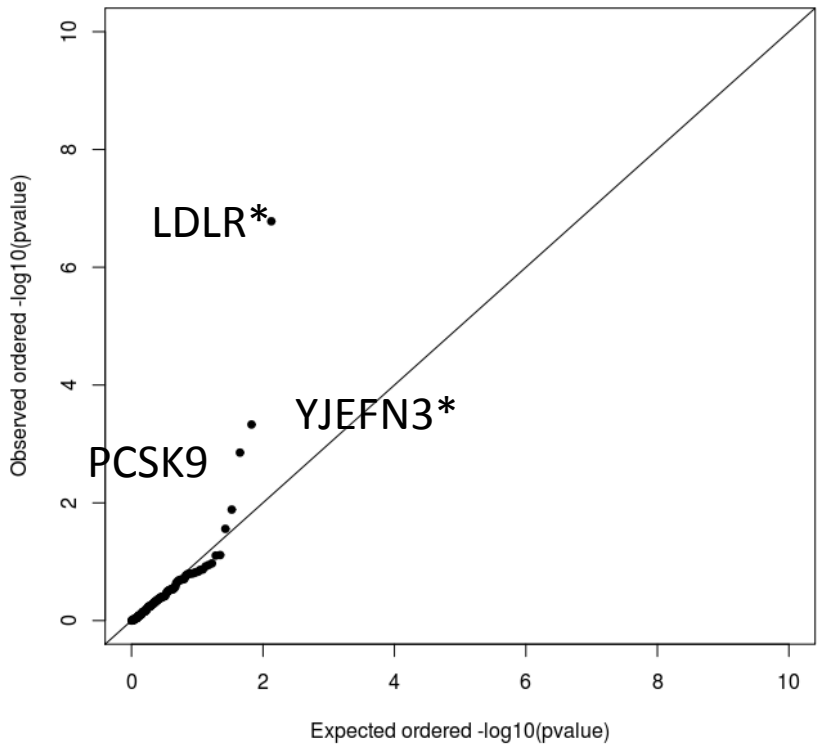
Truncating mutations cause familial hypobetalipoproteinemia (low LDL) wherease missense mutations in receptor binding domain of ApoB cause familial defective apo B (Rahalkar 2008)

Q-Q plot
Burden test of nonsynonymous < 0.1%

All genes tested (N = 16,992)



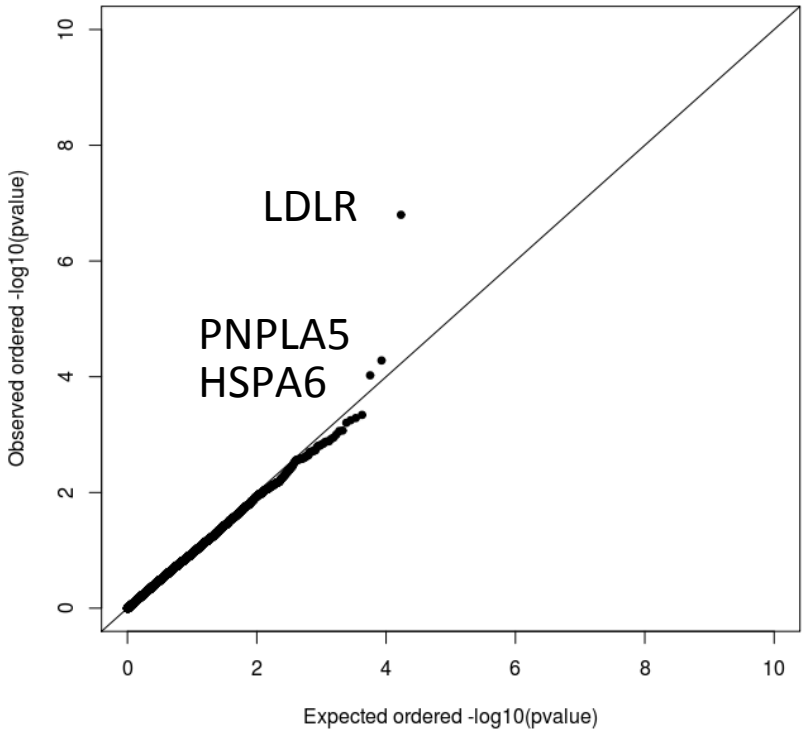
Genes in LDL GWAS loci (N = 133)



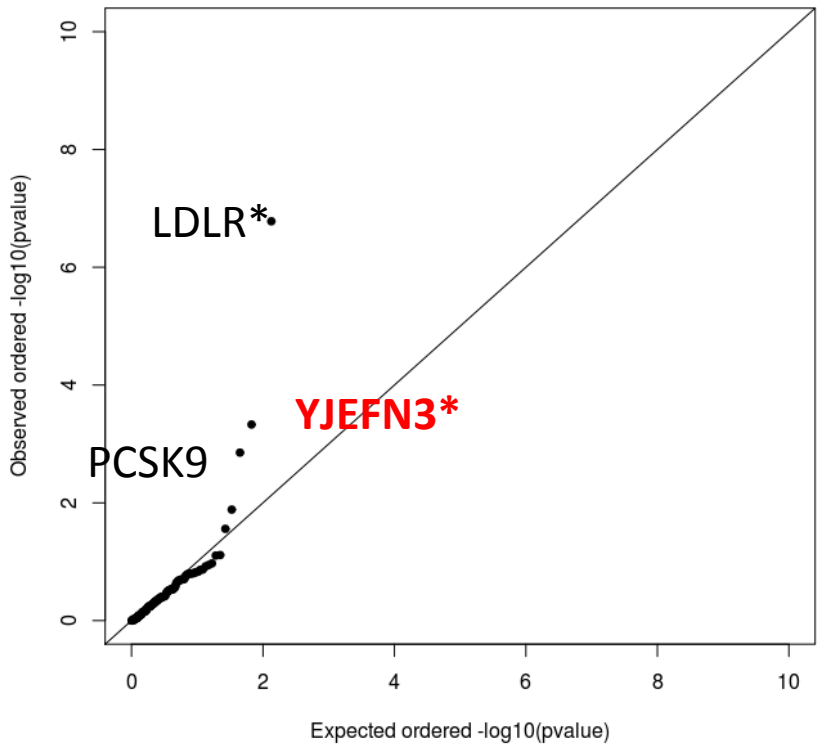
* Significant after correcting for 133 genes tested

Q-Q plot
Burden test of nonsynonymous < 0.1%

All genes tested (N = 16,992)

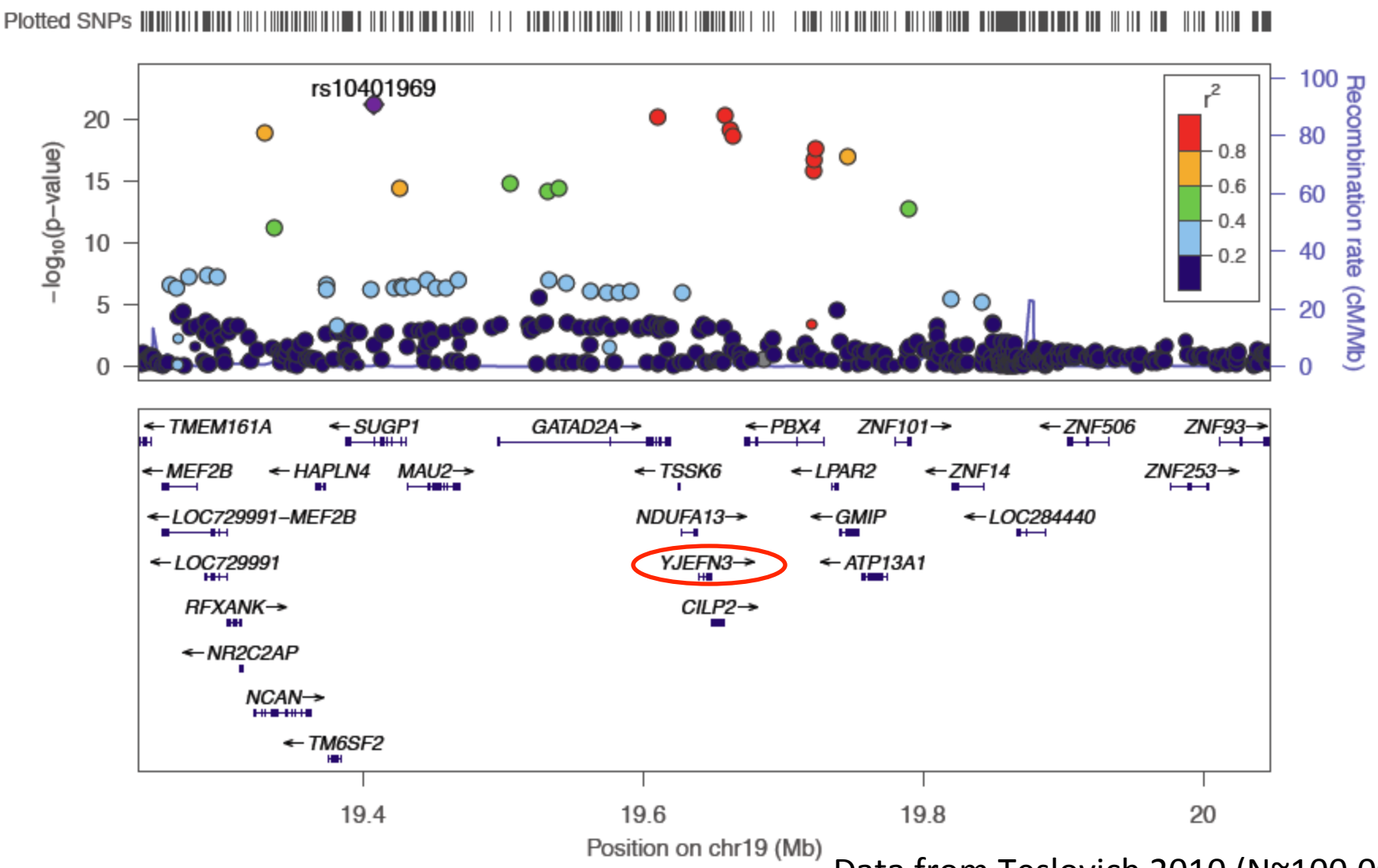


Genes in LDL GWAS loci (N = 133)



* Significant after correcting for 133 genes tested

GWAS LDL locus at NCAN/CILP2



Data from Teslovich 2010 (N~100,000)

LDL-associated locus identified by GWAS; Willer et al. NG 2008 & Kathiresan et al. NG 2008

Biological candidates among top ranked genes
We need replication before we can 'claim'

Gene	Test (most significant p-value)	P-value	Number of variants N/S/R/M*	Singleton Variants N/S/R/M*	N alleles/% alleles on exome chip	% High/ Low EA cases with variant	% High/ Low AA cases with variant
SLC35A5	NS < 5%	$3.0x10^{-6}$	0/1/0/19	0/1/0/11	148 (92.6)	0/2.8	7.8/25.8
HSPA6	NS <0.1%	$9.8x10^{-5}$	0/0/0/13	0/0/0/7	21 (80.4)	5.6/0.9	0/0
ANAPC10	NS <5%	$3.6x10^{-4}$	0/0/0/4	0/0/0/2	10 (90.0)	1.8/0	1.0/0
YJEFN3	NS <0.1%	$4.7x10^{-4}$	1/0/0/9	1/0/0/5	15 (60.1)	1.8/0	2.9/1.1
FADS6	NS <0.1%	$6.5x10^{-4}$	0/0/0/11	0/0/0/6	17 (58.8)	0/2.8	0/2.2

SLC35A5 – probable UDP-sugar transporter protein
HSPA6 – heat shock protein induced by ox-LDL IC
ANAPC10 – anaphase promoting complex that binds LDL receptor family proteins
YJEFN3 – homology to ApoA1-binding protein, in GWAS LDL signal (NCAN/CILP2)
FADS6 – fatty acid desaturase (FADS1-2-3 locus identified by GWAS)

* Variant Categories: Nonsense, Splice, Readthrough, Missense

Top 10 results for each burden test

Single Variant Test	Nonsyn MAF < 0.1%	Nonsyn MAF < 0.5%	Nonsyn MAF < 1%	Nonsyn MAF < 5%	LOF MAF < 5%
TOMM40/APOE $p=6.0\times10^{-8}$	LDLR* $p=1.7\times10^{-7}$	PCSK9 $p=3.8\times10^{-6}$	PCSK9 $p=3.8\times10^{-6}$	PCSK9* $p=5.1\times10^{-8}$	PCSK9* $p=2.6\times10^{-7}$
KRTAP5-5 $p=9.2\times10^{-8}$	PNPLA5 $p=5.3\times10^{-5}$	FAM59A $p=4.2\times10^{-5}$	TRIM4 $p=1.7\times10^{-4}$	SLC35A5* $p=3.0\times10^{-6}$	APOB $p=4.4\times10^{-4}$
JAK3 $p=1.7\times10^{-6}$	HSPA6 $p=9.8\times10^{-5}$	TRIM4 $p=1.7\times10^{-4}$	ANAPC10 $p=3.6\times10^{-4}$	ABCG5 $p=3.2\times10^{-4}$	OLFM4 $p=1.1\times10^{-3}$
ZNF587 $p=3.8\times10^{-6}$	YJEFN3 $p=4.7\times10^{-4}$	ANAPC10 $p=3.6\times10^{-4}$	AZIN1 $p=3.7\times10^{-4}$	ANAPC10 $p=3.6\times10^{-4}$	LDLR $p=1.1\times10^{-3}$
SERPINB11 $p=1.1\times10^{-5}$	ST18 $p=5.9\times10^{-4}$	AZIN1 $p=3.7\times10^{-4}$	CHURC1 $p=5.8\times10^{-4}$	AZIN1 $p=3.7\times10^{-4}$	HMCN1 $p=2.2\times10^{-3}$
MUC3A $p=2.1\times10^{-5}$	FADS6 $p=6.5\times10^{-4}$	NPC1L1 $p=5.3\times10^{-4}$	YJEFN3 $p=6.3\times10^{-4}$	PPM1J $p=4.3\times10^{-4}$	DEF8 $p=2.5\times10^{-3}$
PERP $P=2.4\times10^{-5}$	LEFTY2 $p=6.6\times10^{-4}$	CHURC1 $p=5.8\times10^{-4}$	SLC35A5 $p=7.4\times10^{-4}$	KLK2 $p=4.4\times10^{-4}$	ART3 $p=2.6\times10^{-3}$
RBM15 $p=3.4\times10^{-5}$	GLI3 $p=8.9\times10^{-4}$	YJEFN3 $p=6.3\times10^{-4}$	SLC5A6 $p=8.0\times10^{-4}$	SHD $p=4.7\times10^{-4}$	SLFN11 $p=2.9\times10^{-3}$
OR2W3 $p=3.5\times10^{-5}$	SULT1E1 $p=8.9\times10^{-4}$	RNLS $p=0.0012$	CLNS1A $p=0.0013$	SLC12A5 $p=4.9\times10^{-4}$	OR4M2 $p=4.1\times10^{-3}$
FKBP15 $p=3.7\times10^{-5}$	TCERG1L $p=0.0010$	REXO1 $p=0.0012$	KRTAP5-8 $p=0.0013$	CHURC1 $p=5.8\times10^{-4}$	CCDC113 $p=6.5\times10^{-3}$

* Indicates genome-wide significance at $p < 3 \times 10^{-6}$ (Correcting for ~16,000 genes tested

Results Summary from ESP-LDL

- 2 or 3 genome-wide significant genes; *PCSK9, LDLR, SLC35A5?* (3×10^{-6})

Amongst top-ranked genes;

- Rare variants in *LDLR* and intermediate frequency variants in *ABCG5* are associated with increased LDL
- Rare and intermediate frequencies variants in *PCSK9* and rare variants in *NPC1L1* are associated with decreased LDL
- LOF variants in *APOB* associated with decreased LDL
- Signal in some genes (*LDLR, APOB*) driven by singletons – need sequencing-based follow-up
- Rare variants in *YJEFN3* lead to functional candidate in gene rich LDL-associated locus

Conclusions from ESP-LDL

- Well controlled type 1 error, known genes reach significance
- Novel variants even in heavily sequenced genes (LDLR, PCSK9)
- Genetic architecture quite different at associated genes
  - Very rare variants in LDLR
  - Very rare loss-of-function variants in APOB
  - Mix of 1-2% and rare missense and nonsense variants in PCSK9
  - Common variants in APOE
- Variants in different genes associated with very high and very low LDL
- Variants in the same gene associated with high and low LDL (PCSK9)
- Sometimes the same variant in these genes observed in extreme highs and extreme lows
- Even these “Mendelian-like” mutations observed in extremes appear to be complex

Moving forward with ESP-LDL

- First manuscript draft based on current data
- Planning on menu of alternative analyses strategies to ensure findings are robust
- Can we use sequencing or exome chip to confirm YJEFN3 and 4 other strong biological candidates

Acknowledgements - ESP

ESP LDL Project team
Leslie Lange
Alex Reiner
Charles Kooperberg
Ron Do
Danyu Lin
Zhengzheng Tang
Maja Barbalic
Nora Franceschini
Wendy Post
Christina Wassel
Chris Bizon
Gail Jarvik
Ethan Lange
Chris O'Donnell
Christie Ballantyne
Bruce Psaty
Sek Kathiresan
Adrienne Cupples
Goncalo Abecasis
Kari North
Steve Rich
Deborah Nickerson
Russ Tracy

SeattleGO lab team
Mark Rieder and many others

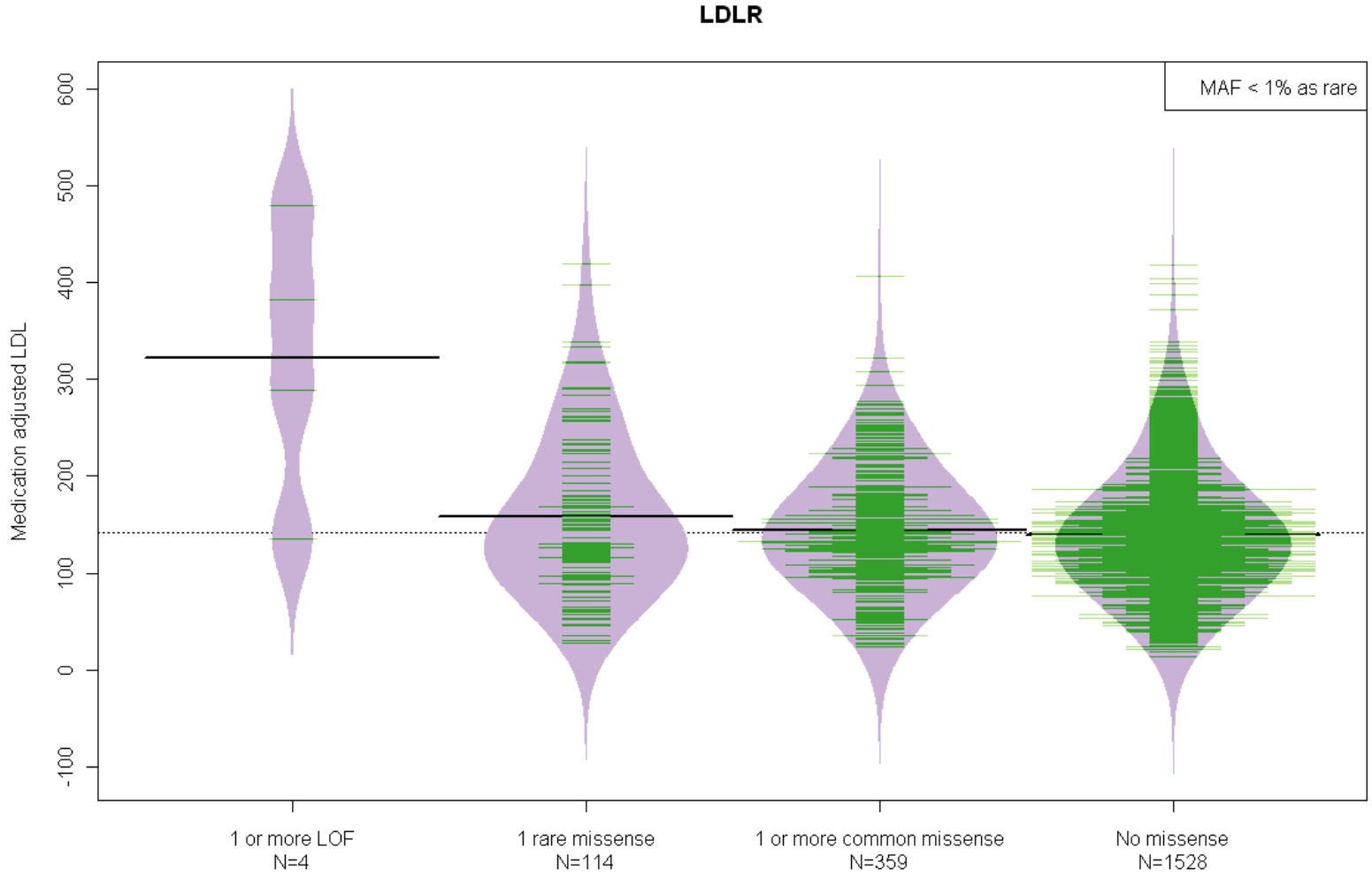
U. of Vermont lab
Peter Durda

ESP StatGen team
Suzanne Leal

U. of Michigan team
Youna Hu
Goo Jun
Chenyi Xue
Hyun Min Kang
Goncalo Abecasis

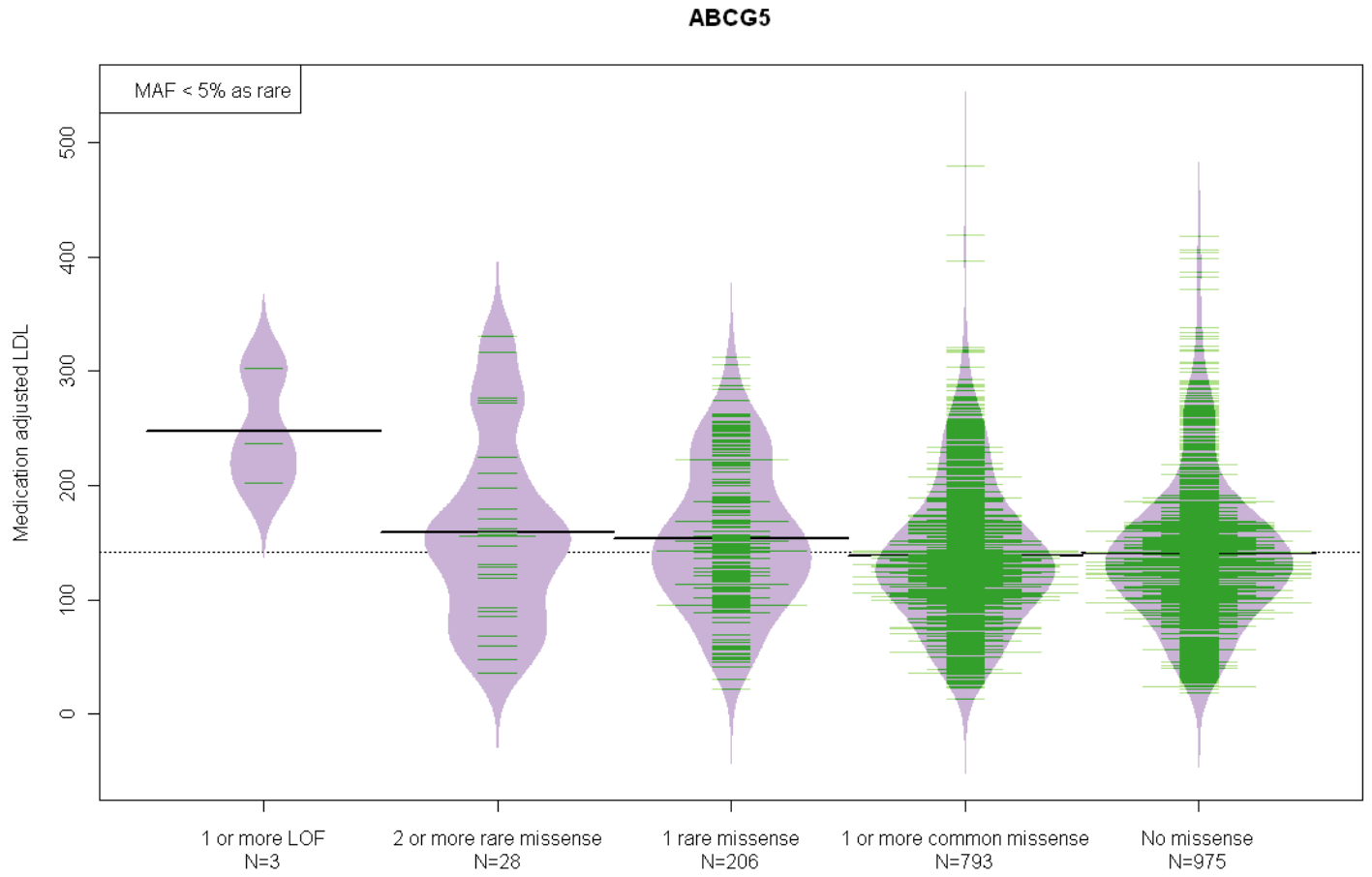
HeartGO cohort representatives
Alanna Morrison
Kerri Wiggins
Nancy Heard-Costa
Craig Johnson
Shih-Jen Hwang
Sharina Person

Distribution of LDL values by mutation type in the *LDLR* gene



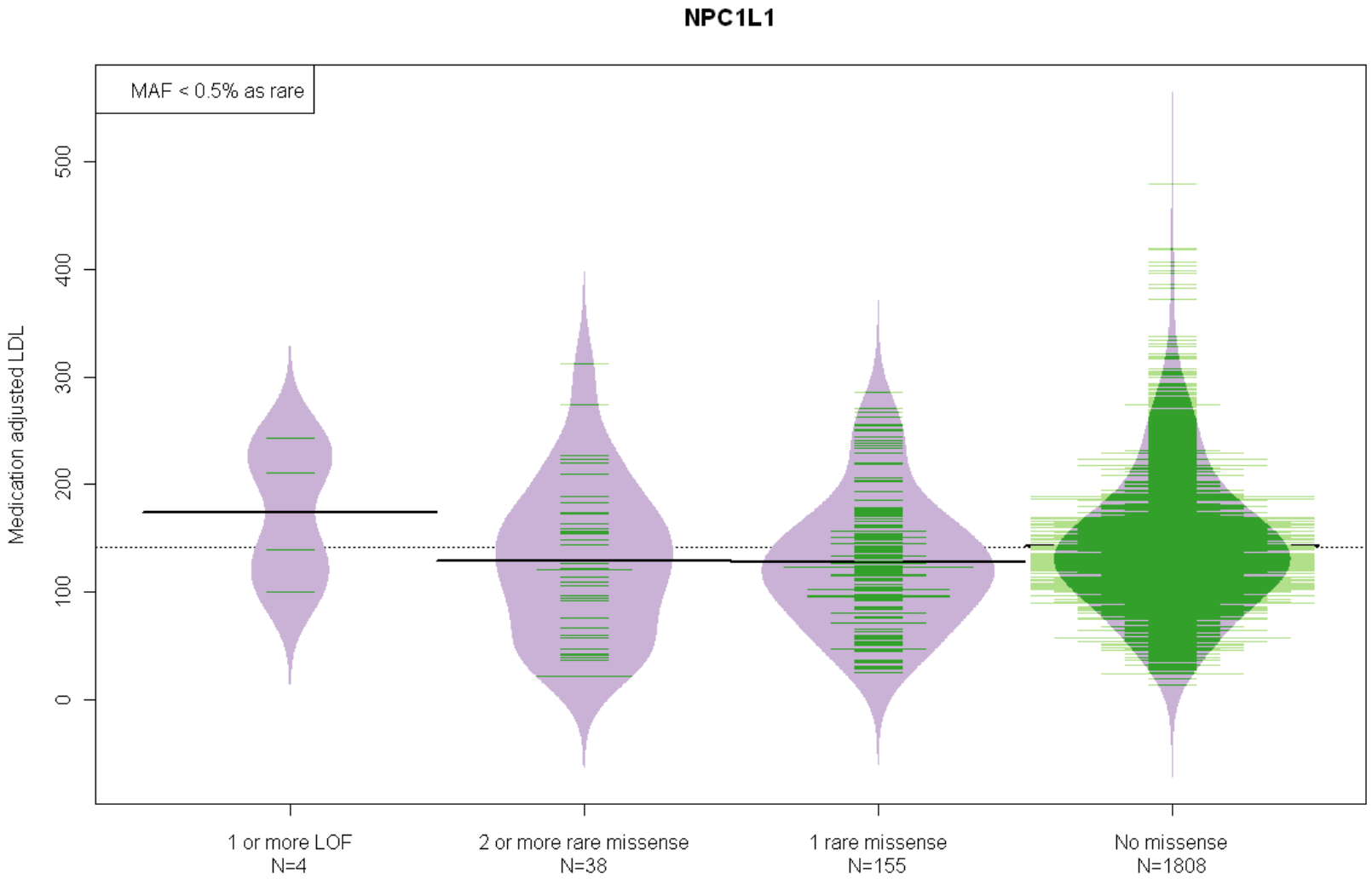
Mutations in LDLR are most common cause of familial hypercholesterolemia (Hobbs 1986), 1122 mutations identified to date – binds ApoB for uptake of LDL particles into hepatocytes

Distribution of LDL values by mutation type in the *ABCG5* gene



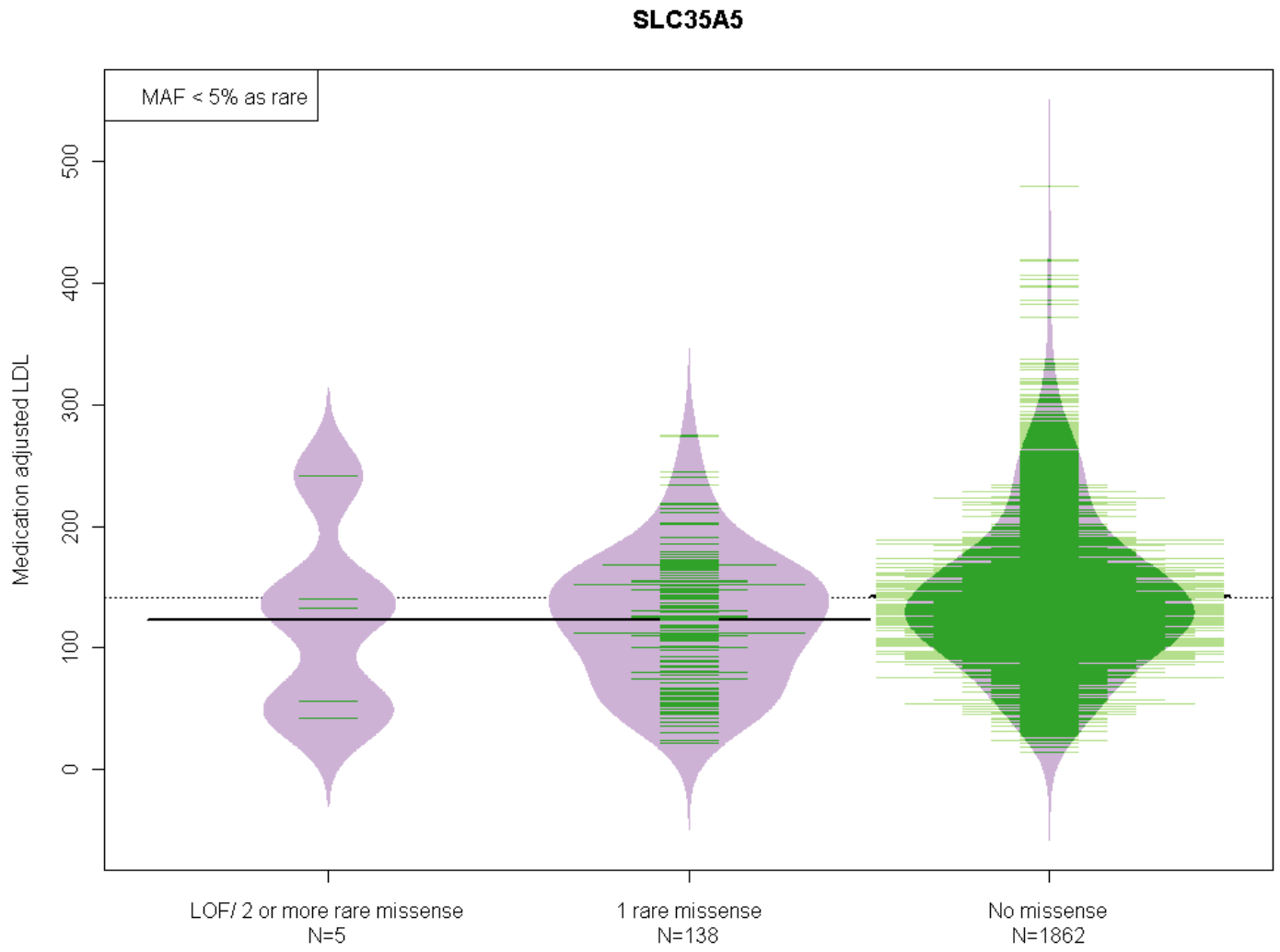
Homozygous mutations cause phytosterolemia (Lee 2001)– phenotype similar to heterozygous LDLR mutations (increased LDL)

Distribution of LDL values by mutation type in the *NPC1L1* gene



Excess of rare nonsyn variants assoc with low LDL chol in AA - campesterol:lathesterol ratio used to selected extremes for low cholesterol absorption (Hobbs et al. PNAS 2006)

Distribution of LDL values by mutation type in *SLC35A5*



No compelling biology – gene is (almost/barely) genome-wide significant (p 3 x 10^{-6})