The American Journal of Human Genetics, Volume *108*

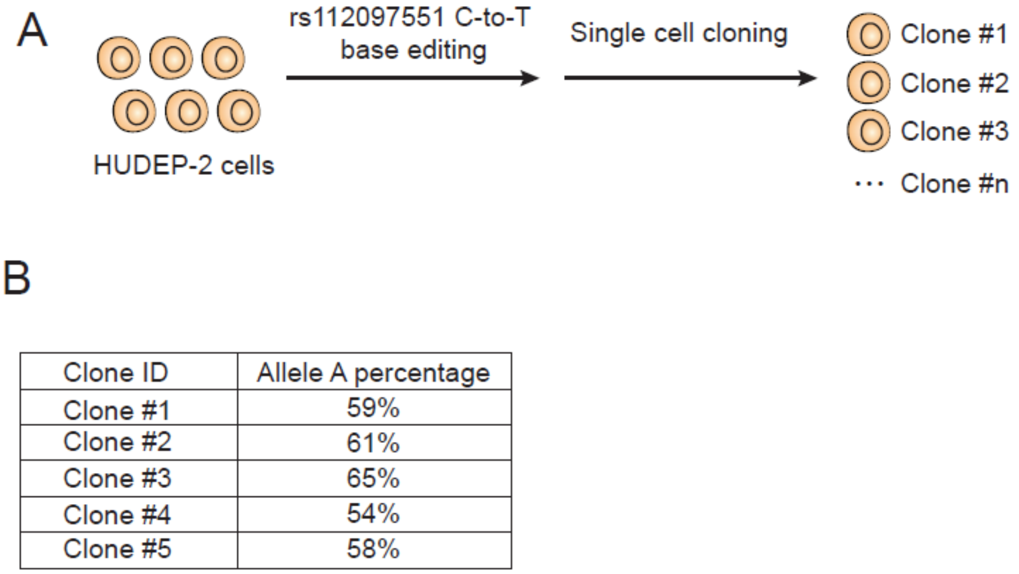
Supplemental information

Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program

Yao Hu, Adrienne M. Stilp, Caitlin P. McHugh, Shuquan Rao, Deepti Jain, Xiuwen Zheng, John Lane, Sébastian Méric de Bellefon, Laura M. Raffield, Ming-Huei Chen, Lisa R. Yanek, Marsha Wheeler, Yao Yao, Chunyan Ren, Jai Broome, Jee-Young Moon, Paul S. de Vries, Brian D. Hobbs, Quan Sun, Praveen Surendran, Jennifer A. Brody, Thomas W. Blackwell, Hélène Choquet, Kathleen Ryan, Ravindranath Duggirala, Nancy Heard-Costa, Zhe Wang, Nathalie Chami, Michael H. Preuss, Nancy Min, Lynette Ekunwe, Leslie A. Lange, Mary Cushman, Nauder Faraday, Joanne E. Curran, Laura Almasy, Kousik Kundu, Albert V. Smith, Stacey Gabriel, Jerome I. Rotter, Myriam Fornage, Donald M. Lloyd-Jones, Ramachandran S. Vasan, Nicholas L. Smith, Kari E. North, Eric Boerwinkle, Lewis C. Becker, Joshua P. Lewis, Goncalo R. Abecasis, Lifang Hou, Jeffrey R. O'Connell, Alanna C. Morrison, Terri H. Beaty, Robert Kaplan, Adolfo Correa, John Blangero, Eric Jorgenson, Bruce M. Psaty, Charles Kooperberg, Russell T. Walton, Benjamin P. Kleinstiver, Hua Tang, Ruth J.F. Loos, Nicole Soranzo, Adam S. Butterworth, Debbie Nickerson, Stephen S. Rich, Braxton D. Mitchell, Andrew D. Johnson, Paul L. Auer, Yun Li, Rasika A. Mathias, Guillaume Lettre, Nathan Pankratz, Cathy C. Laurie, Cecelia A. Laurie, Daniel E. Bauer, Matthew P. Conomos, Alexander P. Reiner, and and the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

Supplemental Figures and Legends

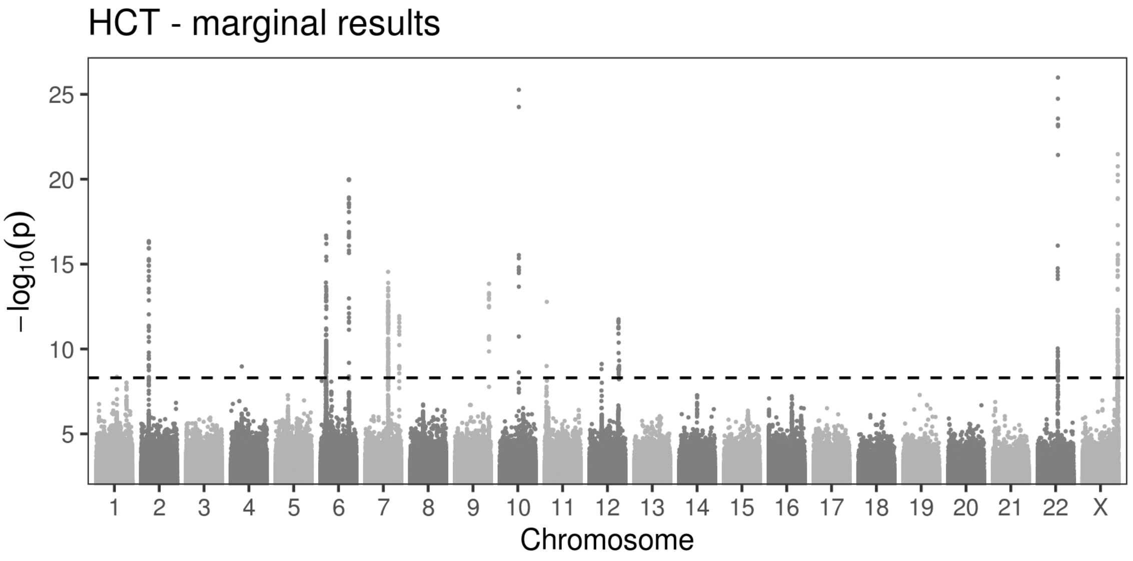
Figure S1. Rs112097551 C-to-T base editing and single cell cloning in HUDEP-2 cells. (A) Scheme of rs112097551 C-to-T base editing and FACS-based single cell separation. (B) Efficiency of rs112097551 C-to-T (G-to-A on opposing strand) base editing efficiency in all five clones. Since base editor and sgRNA are constitutively expressed, the frequency of C-to-T conversion may exceed 50% in heterozygous clones due to base editing after single cell cloning.



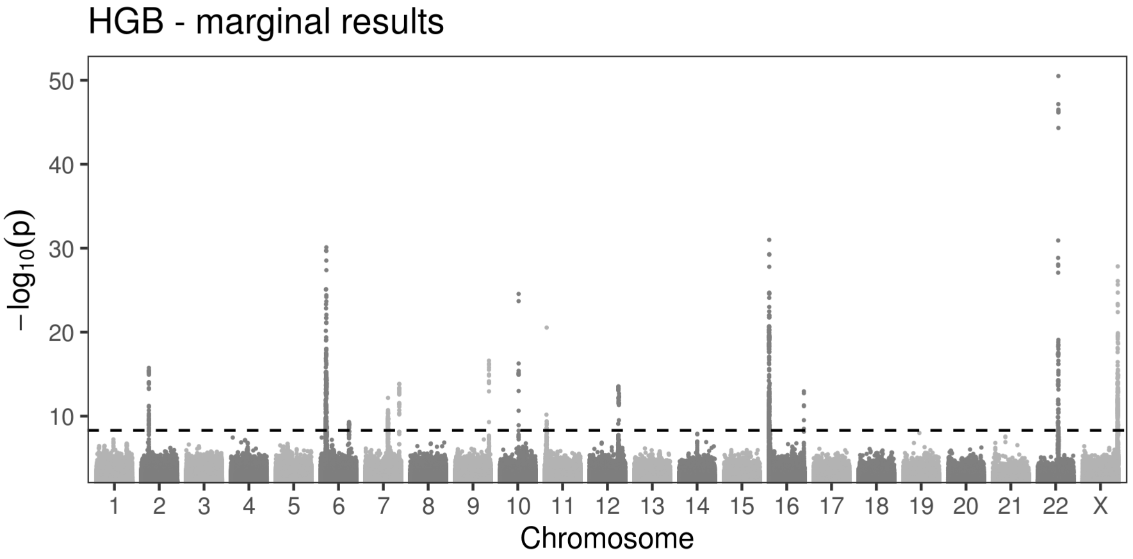
Clone ID	Allele A percentage
Clone #1	59%
Clone #2	61%
Clone #3	65%
Clone #4	54%
Clone #5	58%

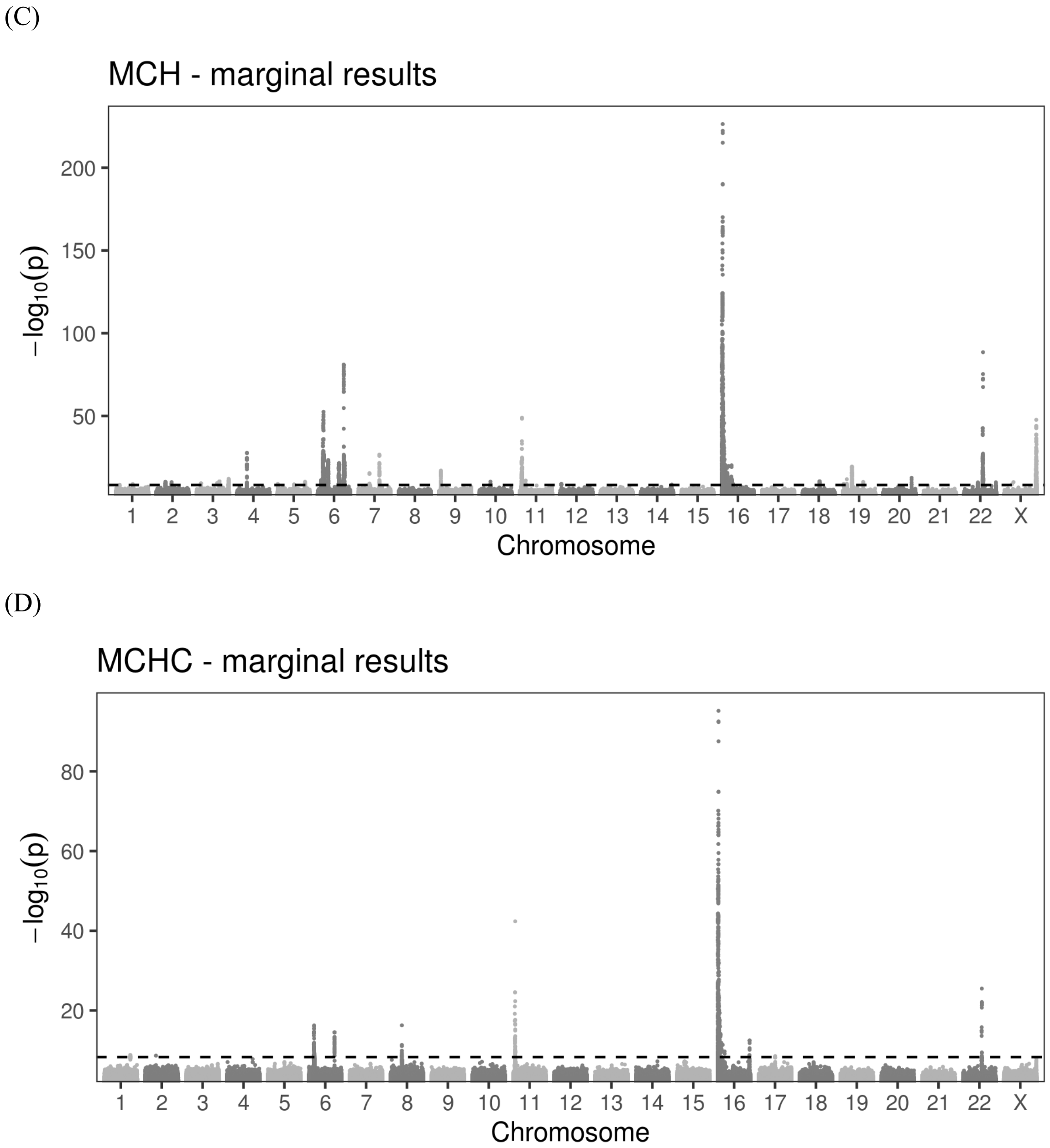
Figure S2. Manhattan plots of the marginal single-variant analyses in TOPMed. (A) HCT; (B) HGB; (C) MCH; (D) MCHC; (E) MCV; (F) RBC; (G) RDW.

(A)

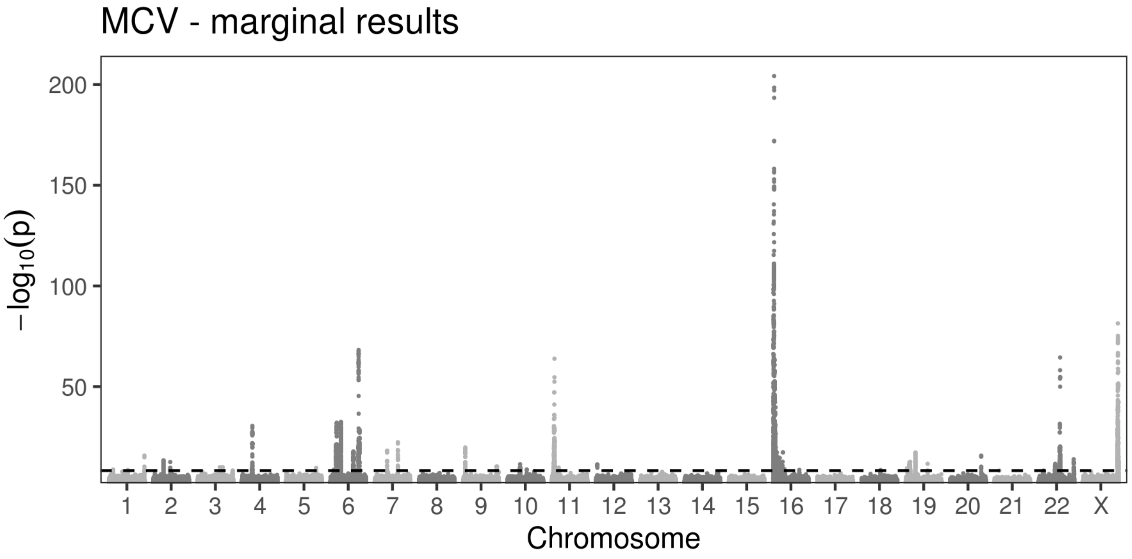


(B)

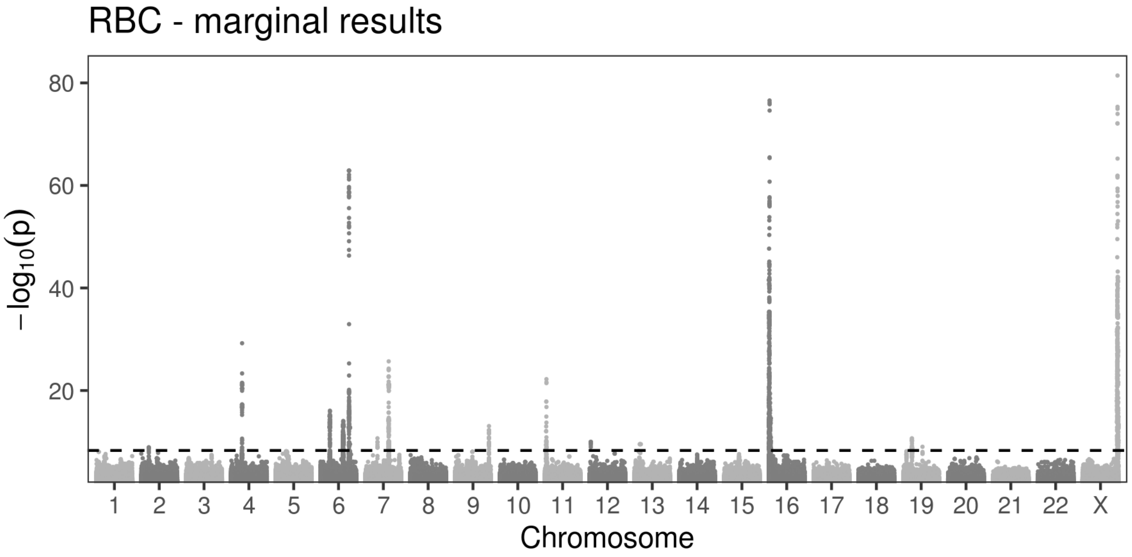



(C)
MCH - marginal results
−log10(p)
200
150
100
50
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Chromosome
(D)
MCHC - marginal results
−log10(p)
80
60
40
20
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Chromosome

(E)


MCV - marginal results
−log10(p)
200
150
100
50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

(F)


RBC - marginal results
−log10(p)
80
60
40
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

(G)

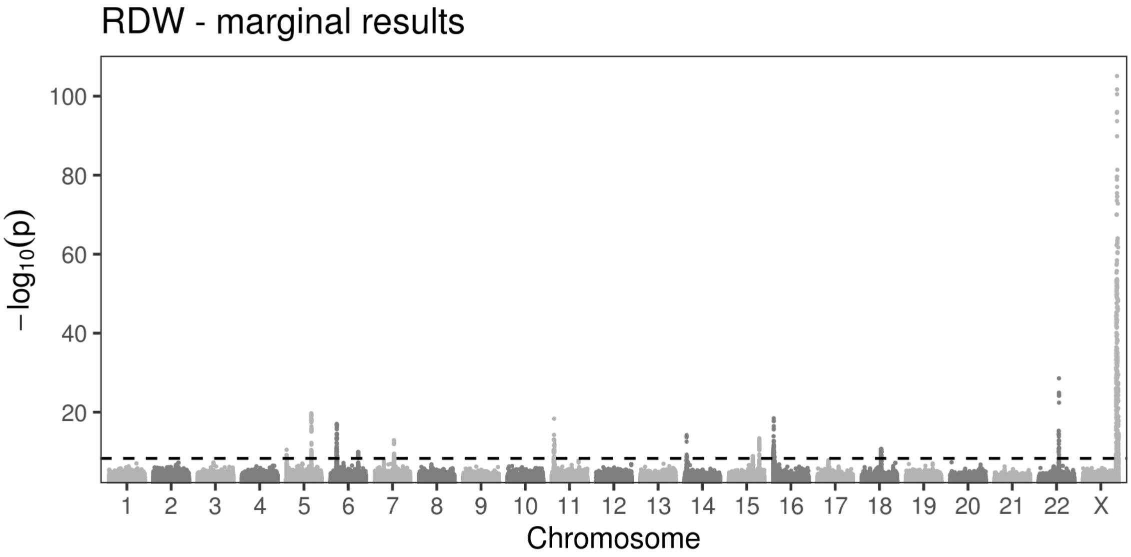
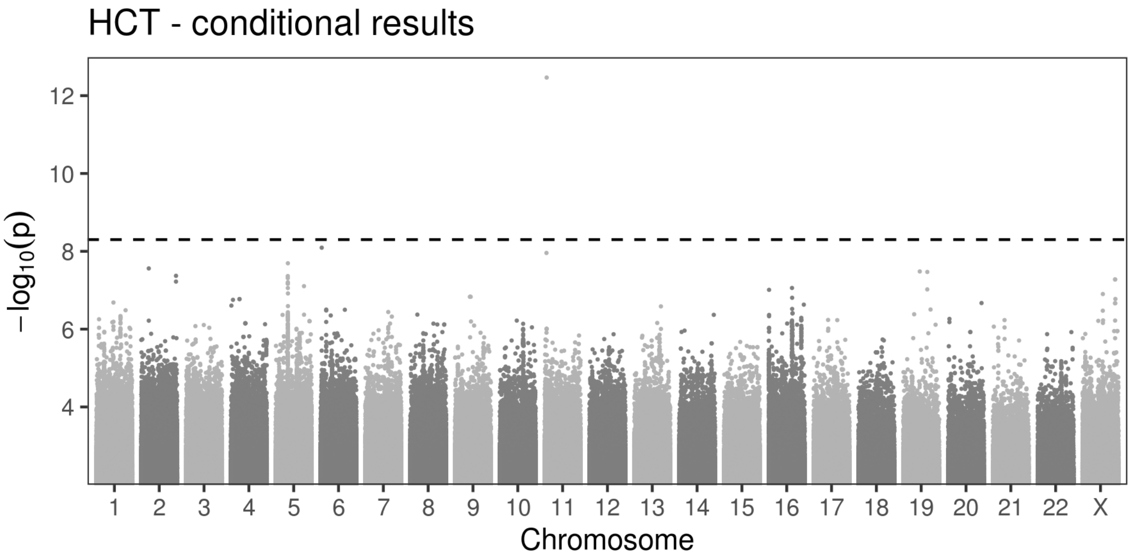
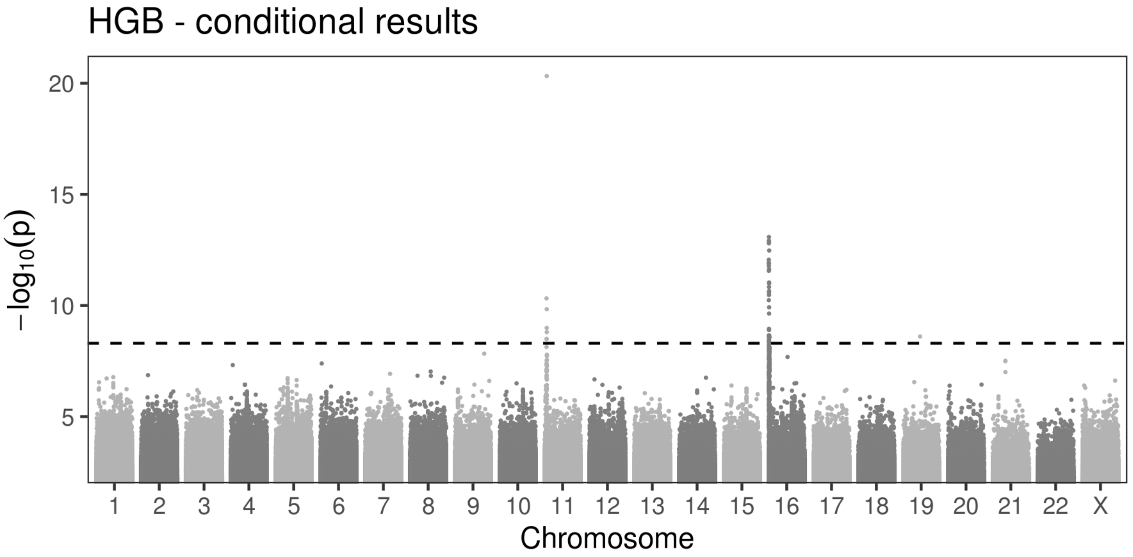

RDW - marginal results
−log10(p)
100
80
60
40
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

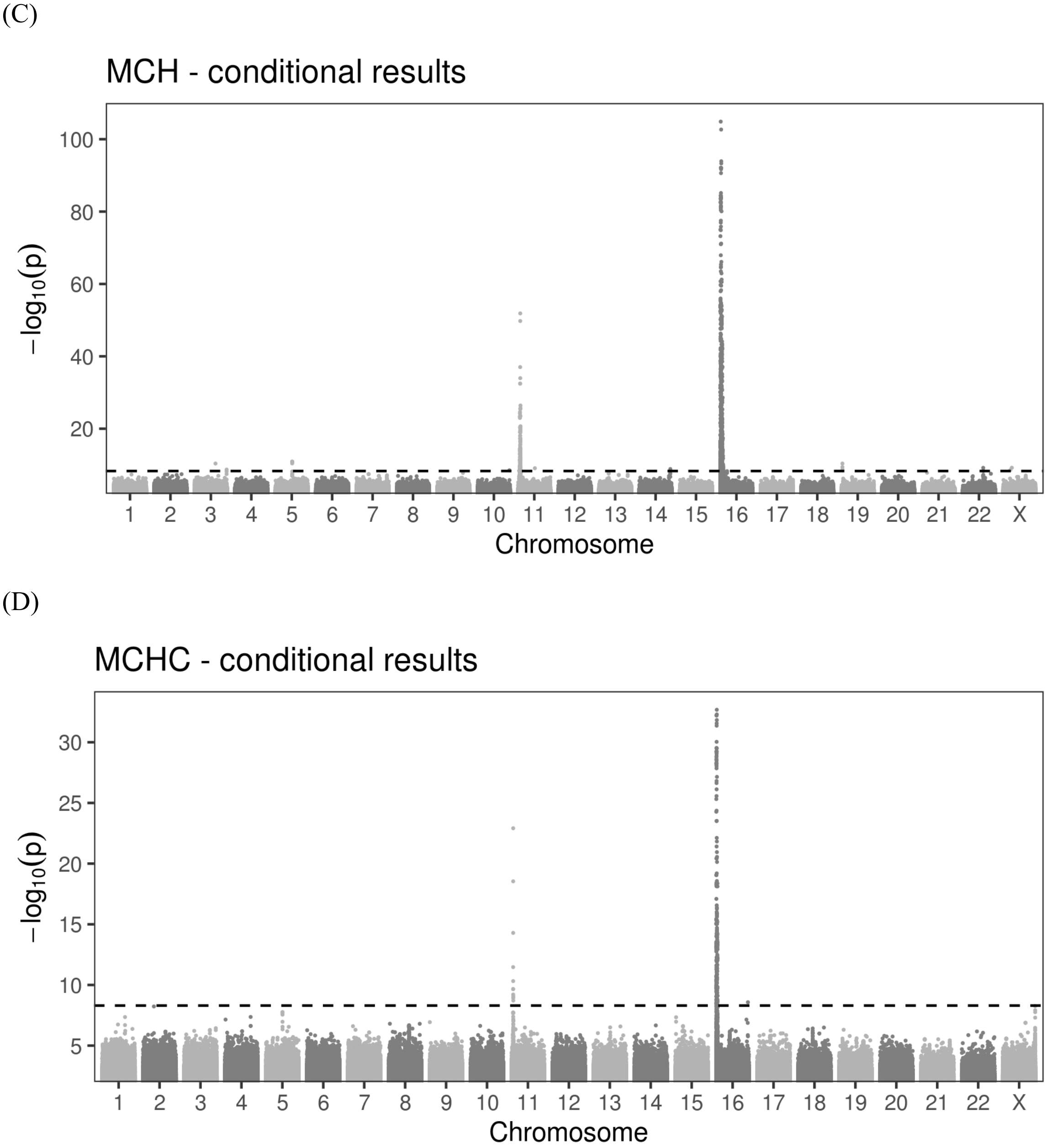
Figure S3. Manhattan plots of the trait-specific conditional single-variant analyses in TOPMed. (A) HCT; (B) HGB; (C) MCH; (D) MCHC; (E) MCV; (F) RBC; (G) RDW.

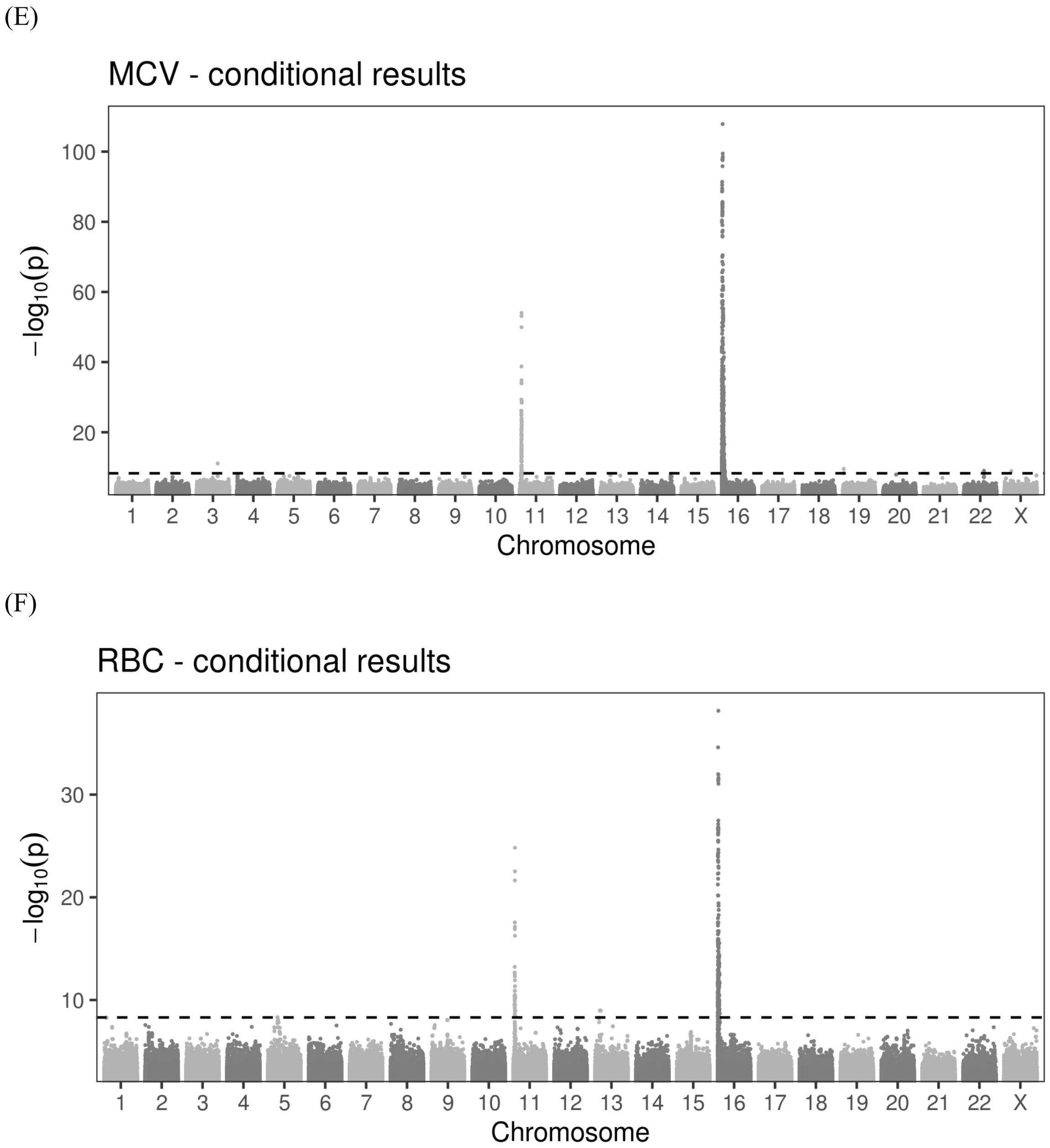
(A)



(B)




(C)
MCH - conditional results
100
80
60
40
20
−log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome
(D)
MCHC - conditional results
30
25
20
15
10
5
−log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome


(E)
MCV - conditional results
−log10(p)
100
80
60
40
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome
(F)
RBC - conditional results
−log10(p)
30
20
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

(G)

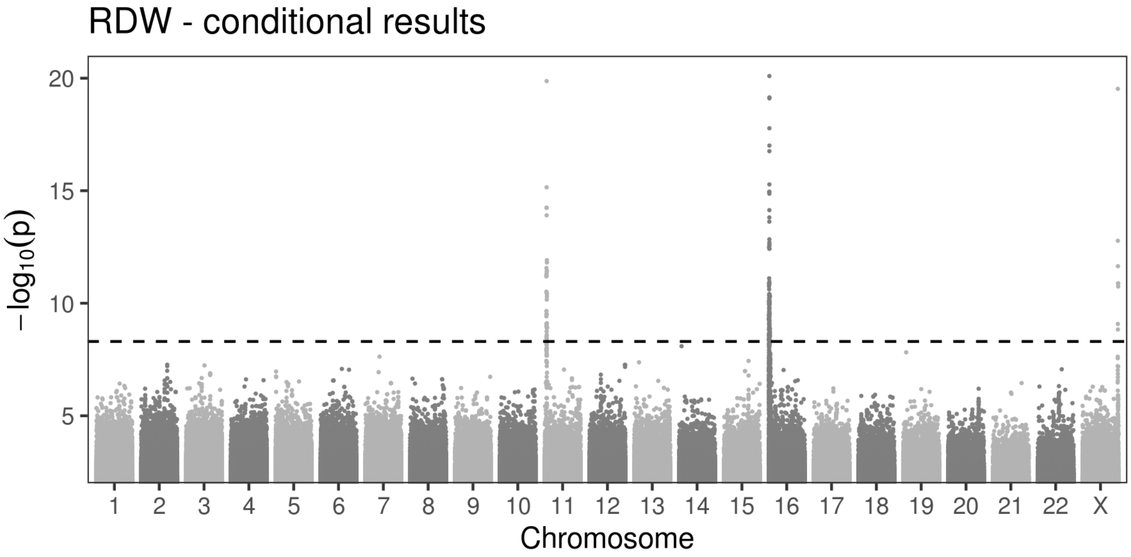
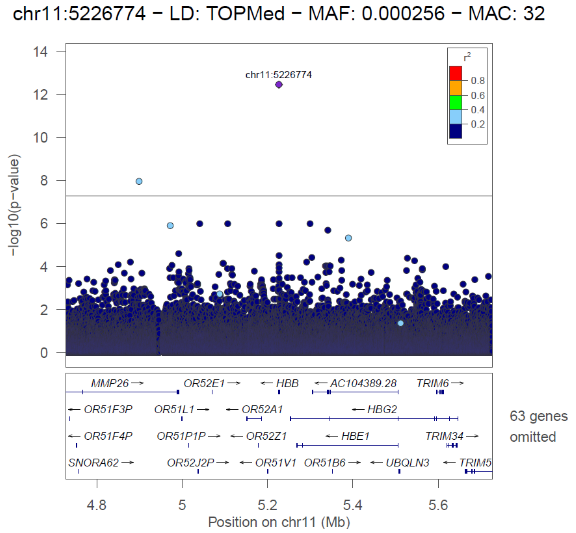
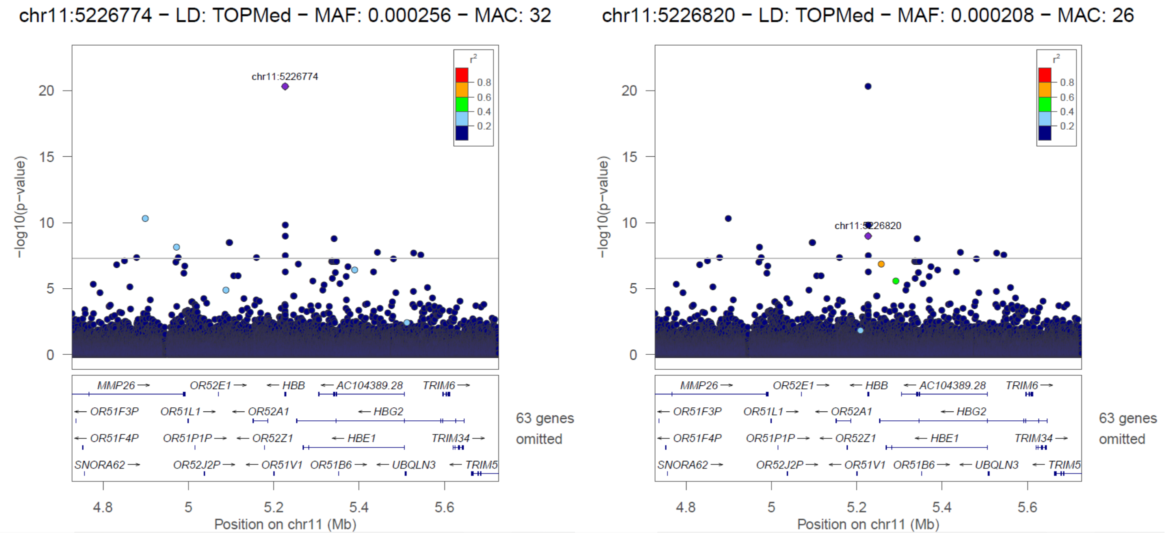

RDW - conditional results
20
15
10
5
−log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

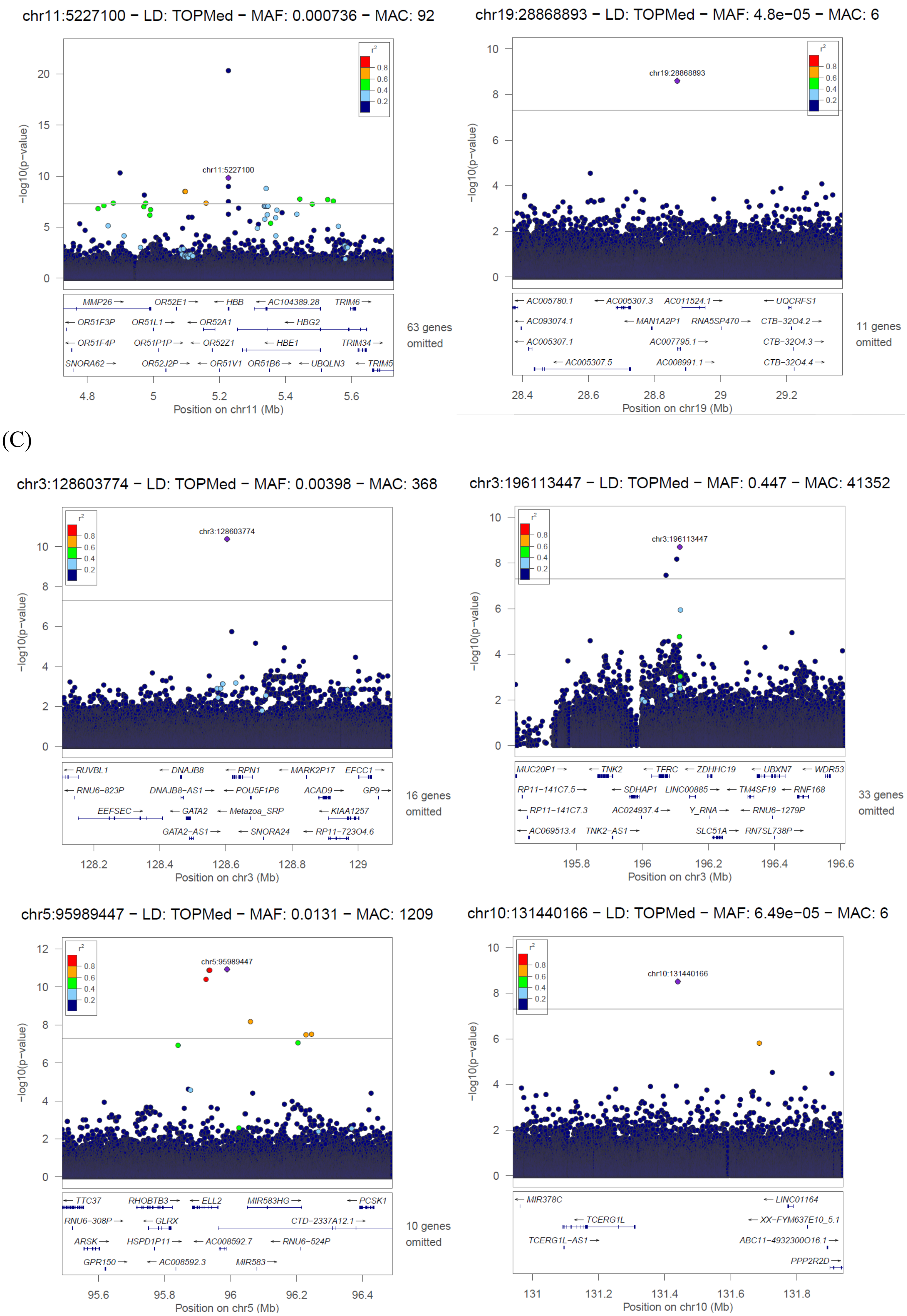
Figure S4. Locuszoom plots of the 12 novel variants and conditionally independent variants identified in TOPMed. (A) HCT; (B) HGB; (C) MCH; (D) MCHC; (E) MCV; (F) RBC; (G) RDW

(A)

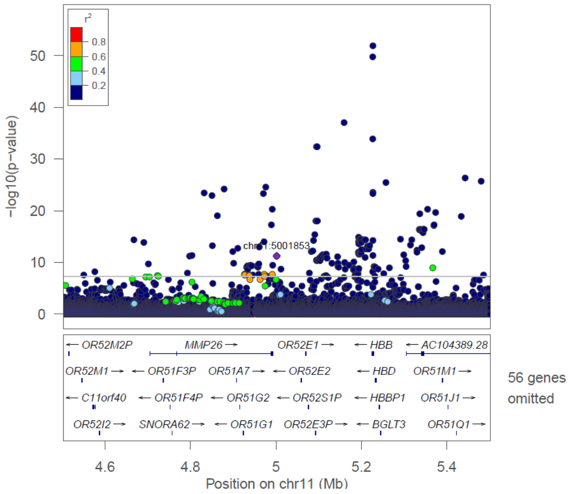


(B)

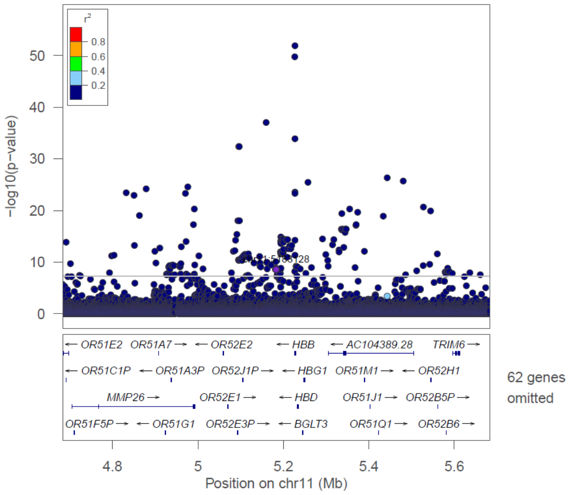


chr11:5227100 − LD: TOPMed − MAF: 0.000736 − MAC: 92
chr11:5227100
r2
0.8
0.6
0.4
0.2
−log10(p−value)
0
5
10
15
20
MMP26
OR52E1
HBB
AC104389.28
TRIM6
OR51F3P
OR51L1
OR52A1
HBG2
OR51F4P
OR51P1P
OR52Z1
HBE1
TRIM34
SNORA62
OR52J2P
OR51V1
OR51B6
UBQLN3
TRIM5
63 genes omitted
4.8
5
5.2
5.4
5.6
Position on chr11 (Mb)
chr19:28868893 − LD: TOPMed − MAF: 4.8e−05 − MAC: 6
chr19:28868893
−log10(p−value)
0
2
4
6
8
10
AC005780.1
AC005307.3
AC011524.1
UQCRFS1
AC093074.1
MAN1A2P1
RNA5SP470
CTB-32O4.2
AC005307.1
AC007795.1
CTB-32O4.3
AC005307.5
AC008991.1
CTB-32O4.4
11 genes omitted
28.4
28.6
28.8
29
29.2
Position on chr19 (Mb)
(C)
chr3:128603774 − LD: TOPMed − MAF: 0.00398 − MAC: 368
chr3:128603774
−log10(p−value)
0
2
4
6
8
10
RUVBL1
DNAJB8
RPN1
MARK2P17
EFCC1
RNU6−823P
DNAJB8−AS1
POU5F1P6
ACAD9
GP9
EEFSEC
GATA2
Metazoa_SRP
KIAA1257
GATA2−AS1
SNORA24
RP11−723O4.6
16 genes omitted
128.2
128.4
128.6
128.8
129
Position on chr3 (Mb)
chr3:196113447 − LD: TOPMed − MAF: 0.447 − MAC: 41352
chr3:196113447
−log10(p−value)
0
2
4
6
8
10
MUC20P1
TNK2
TFRC
ZDHHC19
UBXN7
WDR53
RP11−141C7.5
SDHAP1
LINC00885
TM4SF19
RNF168
RP11−141C7.3
AC024937.4
Y_RNA
RNU6−1279P
AC069513.4
TNK2−AS1
SLC51A
RN7SL738P
33 genes omitted
195.8
196
196.2
196.4
196.6
Position on chr3 (Mb)
chr5:95989447 − LD: TOPMed − MAF: 0.0131 − MAC: 1209
chr5:95989447
−log10(p−value)
0
2
4
6
8
10
12
TTC37
RHOBTB3
ELL2
MIR583HG
PCSK1
RNU6−308P
GLRX
CTD−2337A12.1
ARSK
HSPD1P11
AC008592.7
RNU6−524P
GPR150
AC008592.3
MIR583
10 genes omitted
95.6
95.8
96
96.2
96.4
Position on chr5 (Mb)
chr10:131440166 − LD: TOPMed − MAF: 6.49e−05 − MAC: 6
chr10:131440166
−log10(p−value)
0
2
4
6
8
10
MIR378C
LINC01164
TCERG1L
XX−FYM637E10_5.1
TCERG1L−AS1
ABC11−4932300O16.1
PPP2R2D
131
131.2
131.4
131.6
131.8
Position on chr10 (Mb)

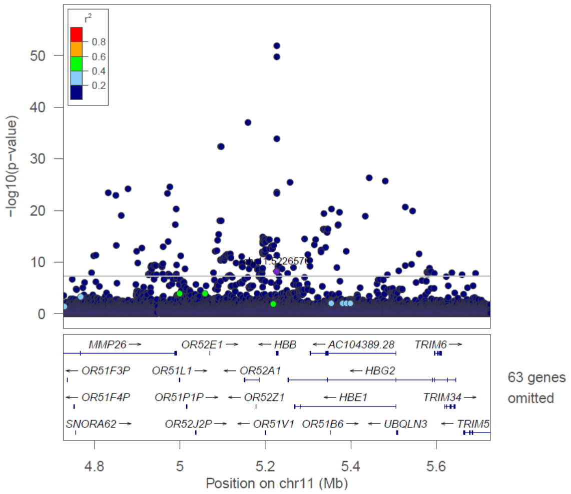
chr11:5001853 − LD: TOPMed − MAF: 0.0044 − MAC: 407



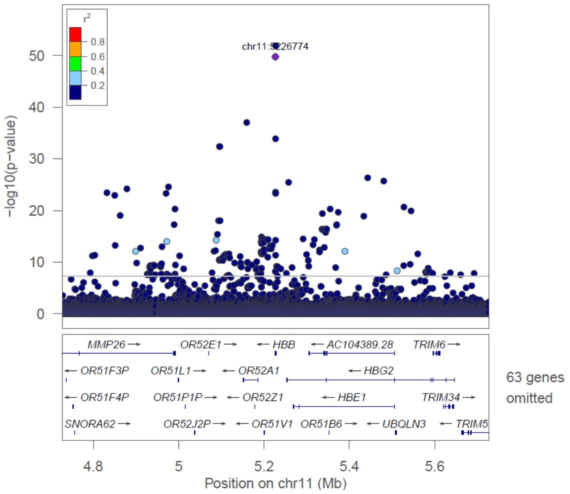
chr11:5183128 − LD: TOPMed − MAF: 5.41e−05 − MAC: 5



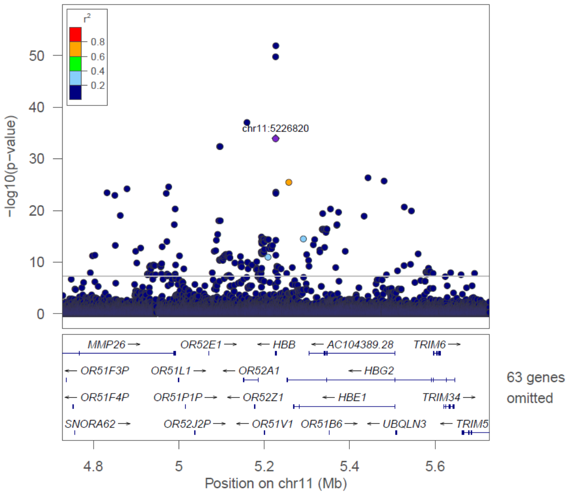
chr11:5226576 − LD: TOPMed − MAF: 7.57e−05 − MAC: 7



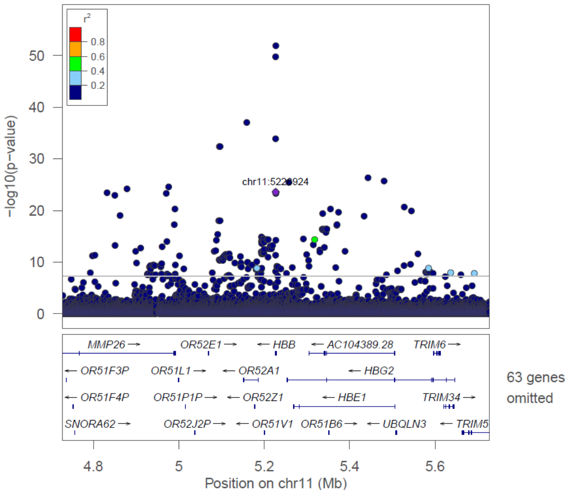
chr11:5226774 − LD: TOPMed − MAF: 0.000292 − MAC: 27



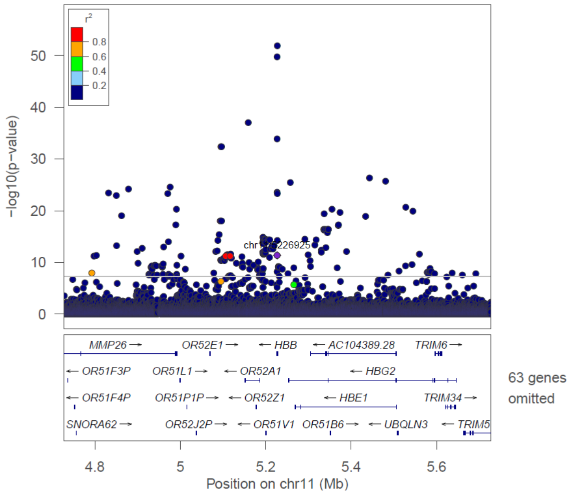
chr11:5226820 − LD: TOPMed − MAF: 0.000216 − MAC: 20



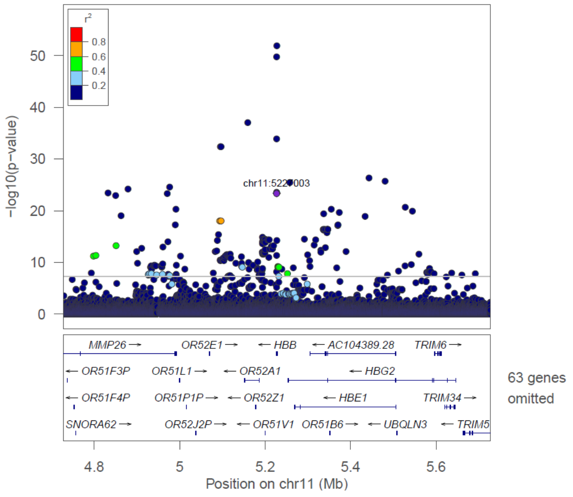
chr11:5226924 − LD: TOPMed − MAF: 0.000151 − MAC: 14



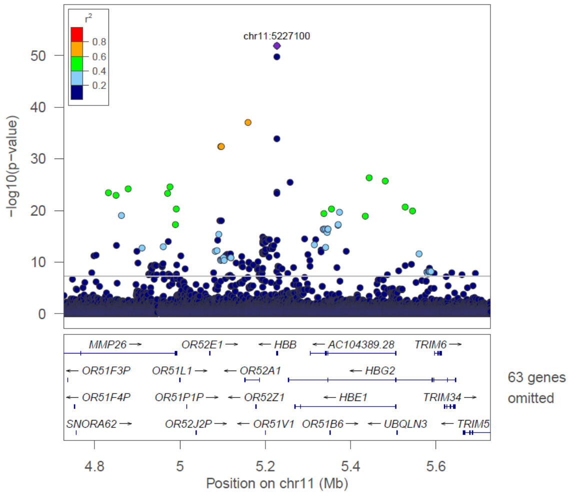
chr11:5226925 − LD: TOPMed − MAF: 0.000108 − MAC: 10



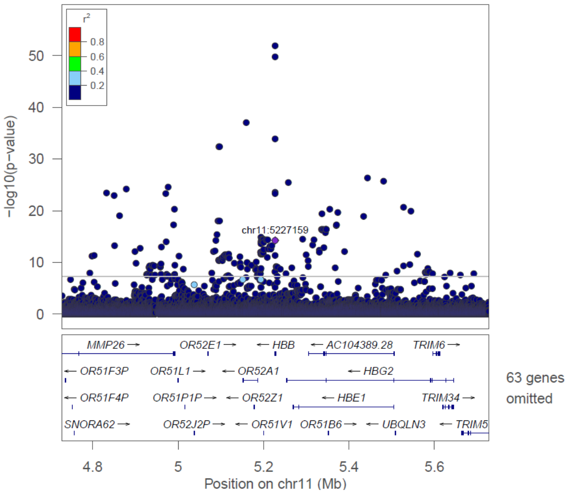
chr11:5227003 − LD: TOPMed − MAF: 0.00408 − MAC: 377



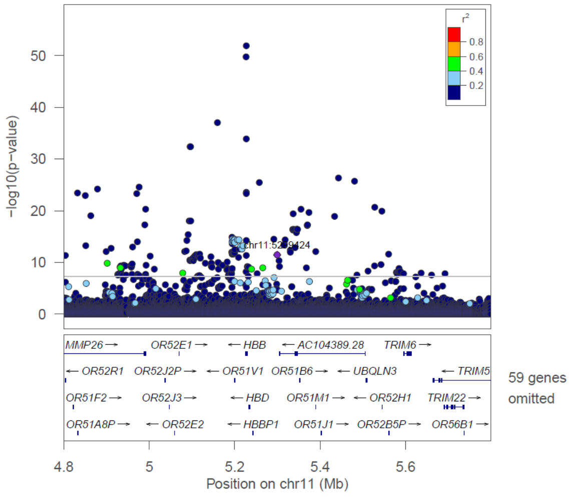
chr11:5227100 − LD: TOPMed − MAF: 0.000833 − MAC: 77



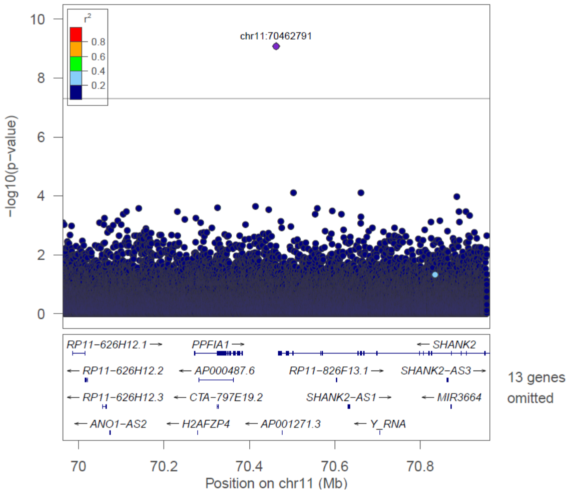
chr11:5227159 − LD: TOPMed − MAF: 0.000249 − MAC: 23



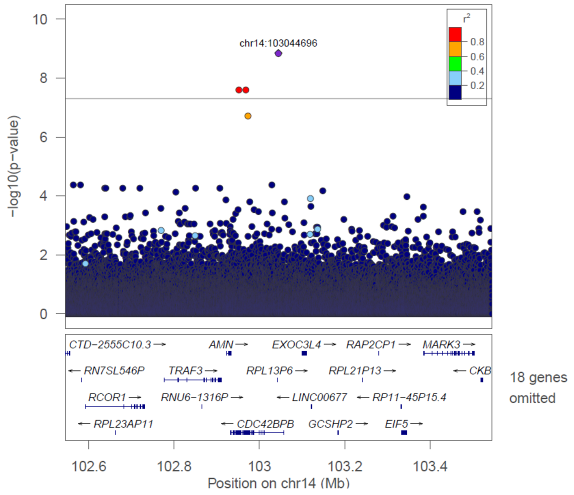
chr11:5299424 − LD: TOPMed − MAF: 0.00889 − MAC: 822



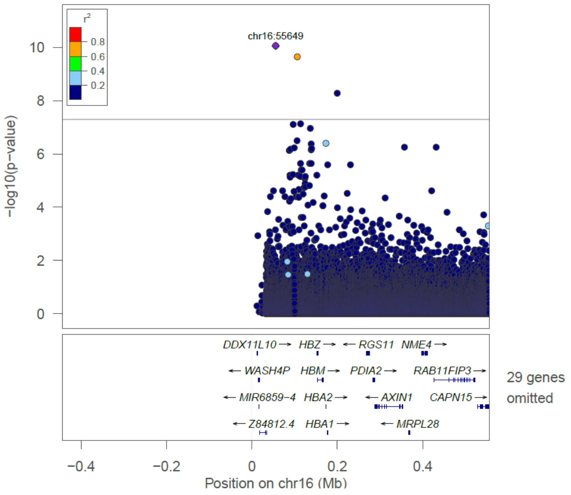
chr11:70462791 − LD: TOPMed − MAF: 7.57e−05 − MAC: 7



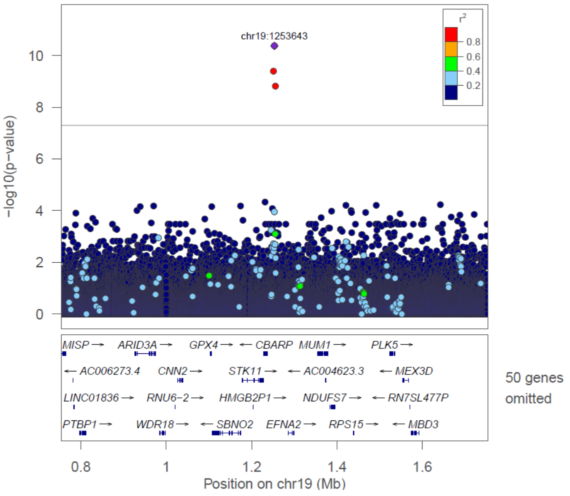
chr14:103044696 − LD: TOPMed − MAF: 0.000108 − MAC: 10



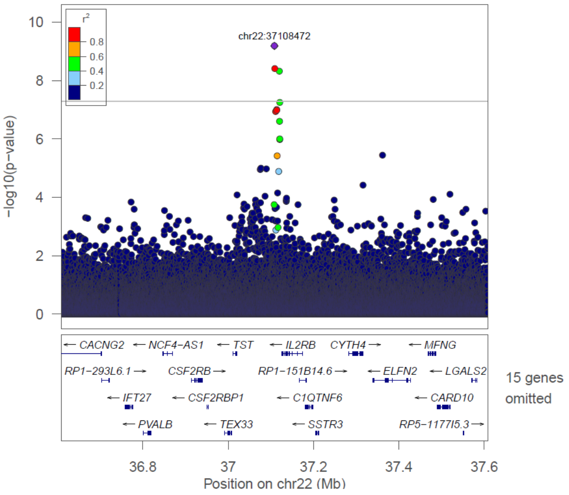
chr16:55649 − LD: TOPMed − MAF: 0.000224 − MAC: 17



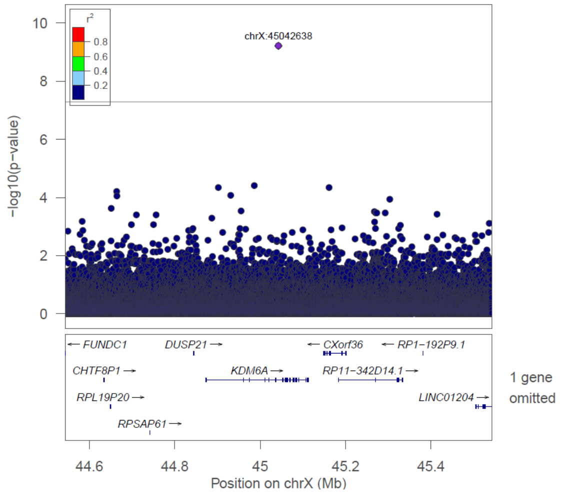
chr19:1253643 − LD: TOPMed − MAF: 0.165 − MAC: 15249



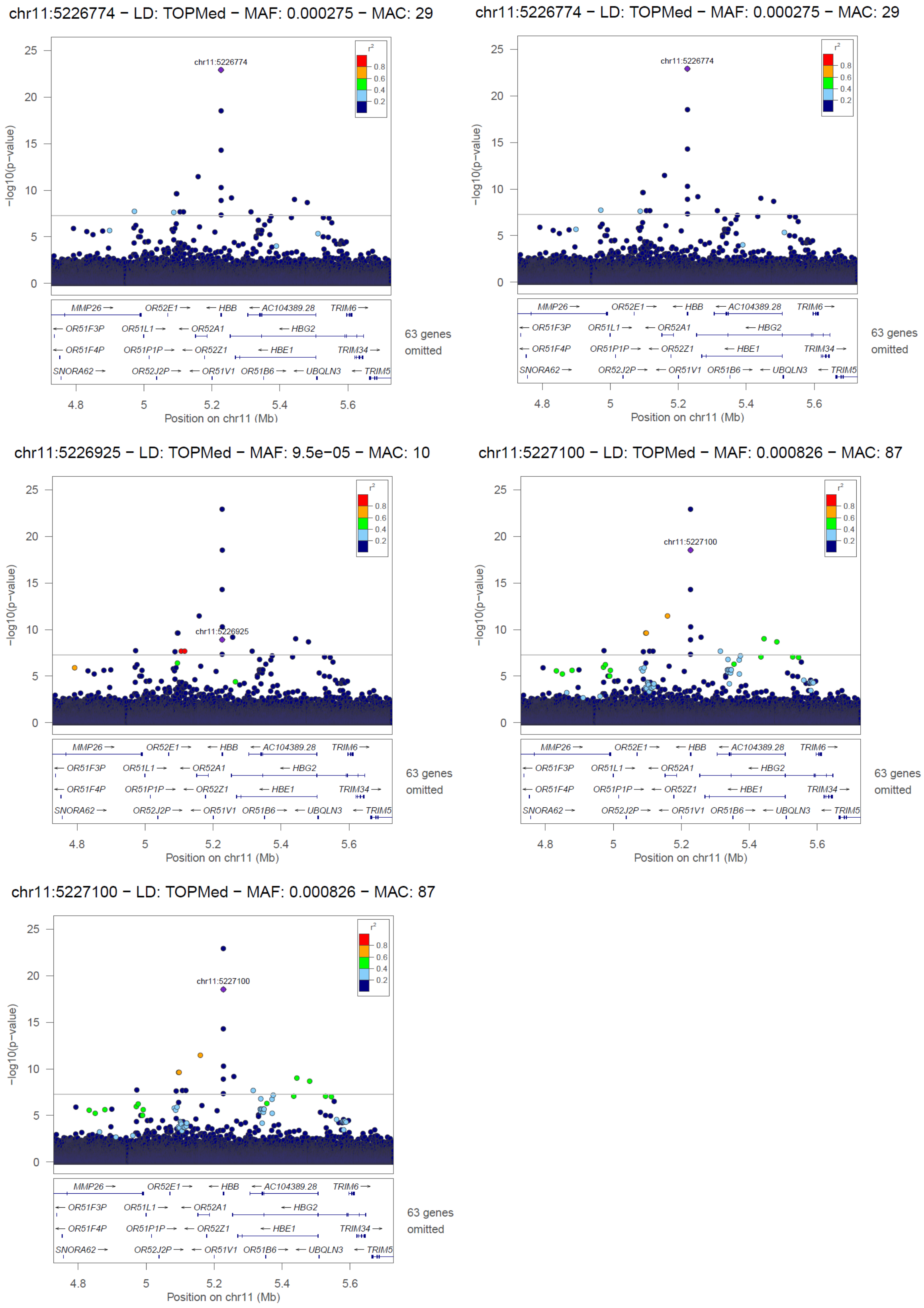
chr22:37108472 − LD: TOPMed − MAF: 0.11 − MAC: 10189



chrX:45042638 − LD: TOPMed − MAF: 0.382 − MAC: 27786



(D)



chr11:5226774 − LD: TOPMed − MAF: 0.000275 − MAC: 29

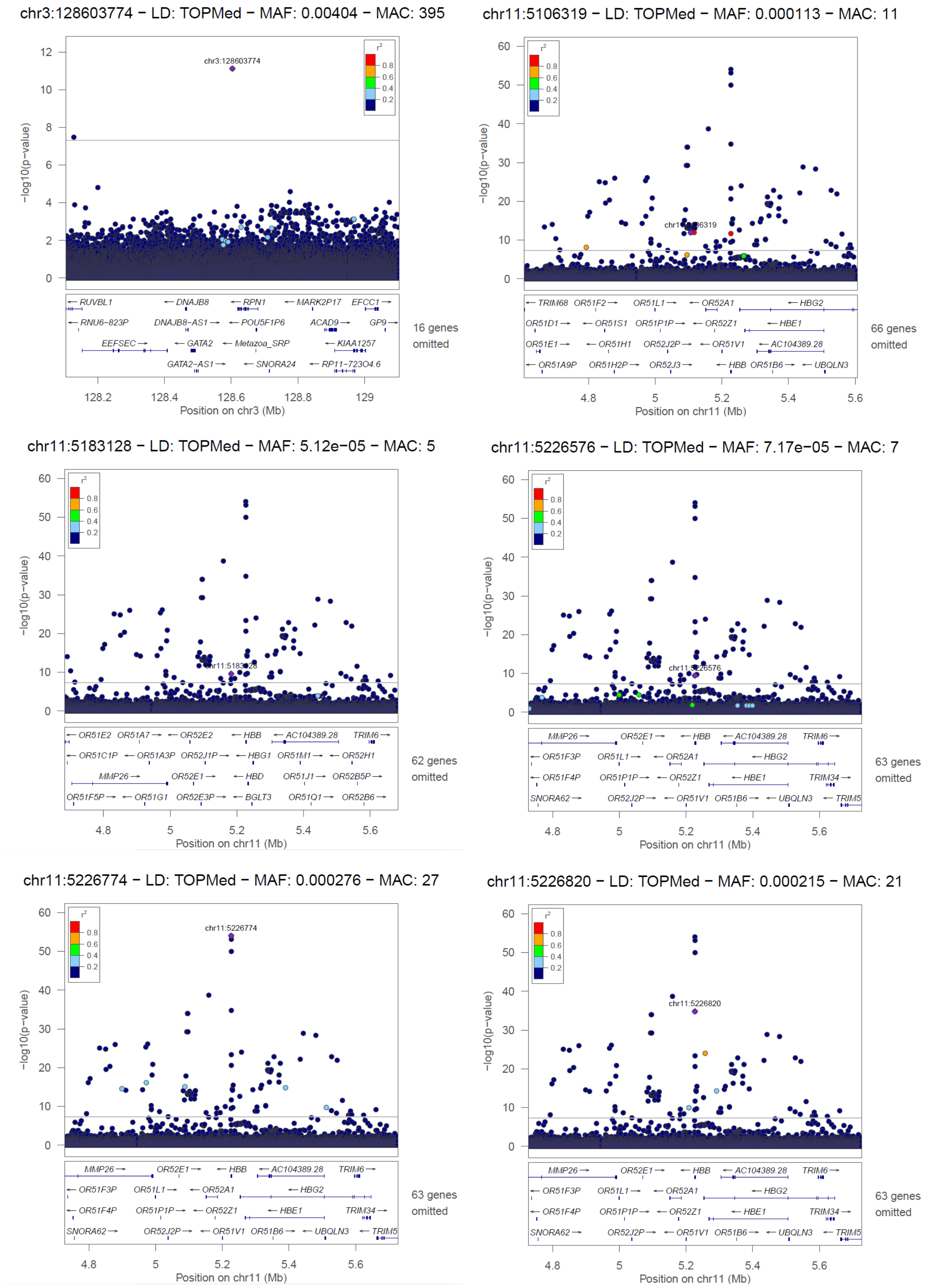
chr11:5226774 − LD: TOPMed − MAF: 0.000275 − MAC: 29

chr11:5226925 − LD: TOPMed − MAF: 9.5e−05 − MAC: 10

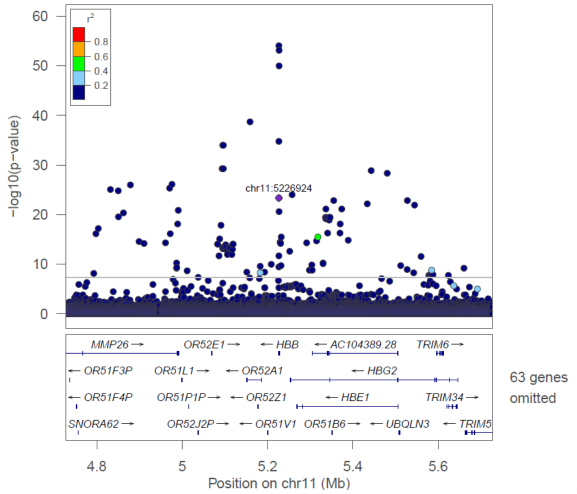
chr11:5227100 − LD: TOPMed − MAF: 0.000826 − MAC: 87

chr11:5227100 − LD: TOPMed − MAF: 0.000826 − MAC: 87

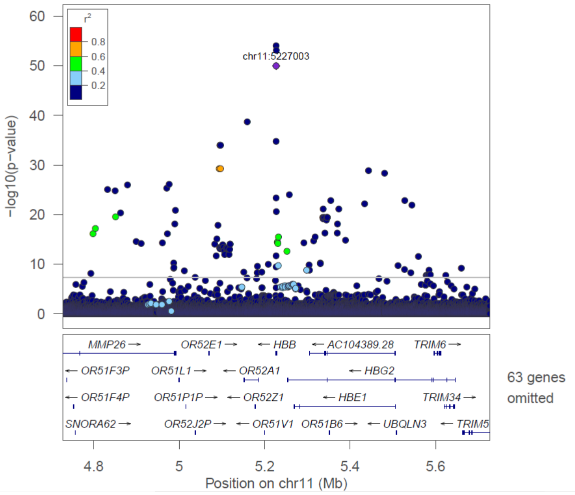
(E)


chr3:128603774 − LD: TOPMed − MAF: 0.00404 − MAC: 395
chr3:128603774
−log10(p−value)
r²
0.8
0.6
0.4
0.2
RUVBL1
DNAJB8
RPN1
MARK2P17
EFCC1
RNU6−823P
DNAJB8−AS1
POU5F1P6
ACAD9
GP9
EEFSEC
GATA2
Metazoa_SRP
KIAA1257
GATA2−AS1
SNORA24
RP11−723O4.6
16 genes omitted
128.2
128.4
128.6
128.8
129
Position on chr3 (Mb)
chr11:5106319 − LD: TOPMed − MAF: 0.000113 − MAC: 11
chr11:5106319
TRIM68
OR51F2
OR51L1
OR52A1
HBG2
OR51D1
OR51S1
OR51P1P
OR52Z1
HBE1
OR51E1
OR51H1
OR52J2P
OR51V1
AC104389.28
OR51A9P
OR51H2P
OR52J3
HBB
OR51B6
UBQLN3
66 genes omitted
4.8
5
5.2
5.4
5.6
Position on chr11 (Mb)
chr11:5183128 − LD: TOPMed − MAF: 5.12e−05 − MAC: 5
chr11:5183128
OR51E2
OR51A7
OR52E2
HBB
AC104389.28
TRIM6
OR51C1P
OR51A3P
OR52J1P
HBG1
OR51M1
OR52H1
MMP26
OR52E1
HBD
OR51J1
OR52B5P
OR51F5P
OR51G1
OR52E3P
BGLT3
OR51Q1
OR52B6
62 genes omitted
chr11:5226576 − LD: TOPMed − MAF: 7.17e−05 − MAC: 7
chr11:5226576
MMP26
OR52E1
HBB
AC104389.28
TRIM6
OR51F3P
OR51L1
OR52A1
HBG2
OR51F4P
OR51P1P
OR52Z1
HBE1
TRIM34
SNORA62
OR52J2P
OR51V1
OR51B6
UBQLN3
TRIM5
63 genes omitted
chr11:5226774 − LD: TOPMed − MAF: 0.000276 − MAC: 27
chr11:5226774
chr11:5226820 − LD: TOPMed − MAF: 0.000215 − MAC: 21
chr11:5226820
0
10
20
30
40
50
60

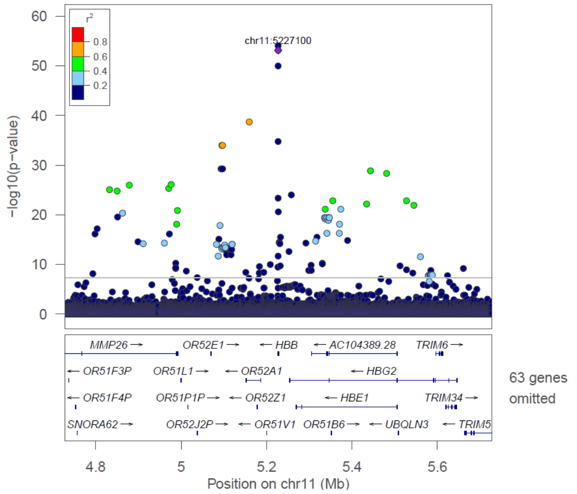
chr11:5226924 − LD: TOPMed − MAF: 0.000143 − MAC: 14



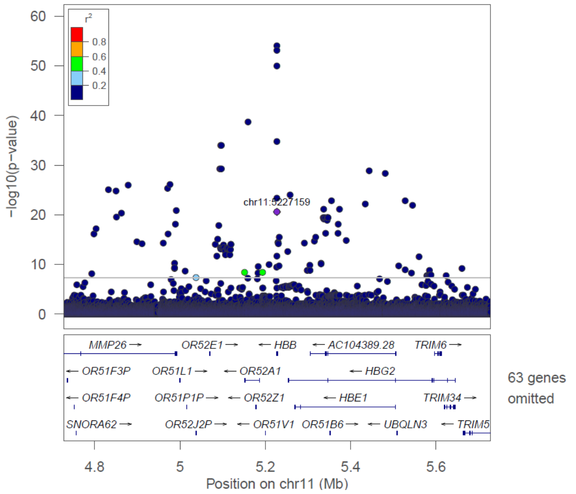
chr11:5227003 − LD: TOPMed − MAF: 0.00412 − MAC: 402



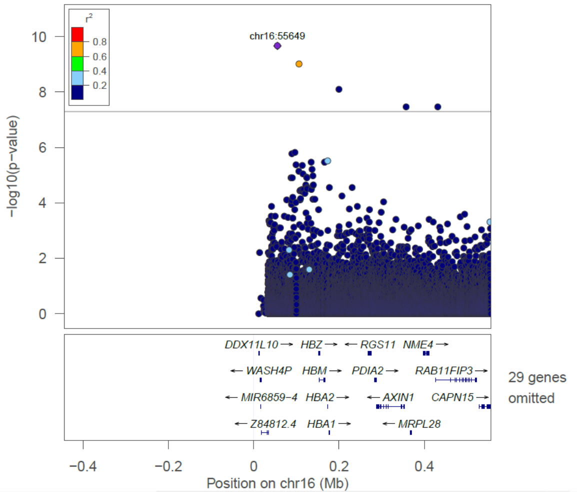
chr11:5227100 − LD: TOPMed − MAF: 0.00085 − MAC: 83



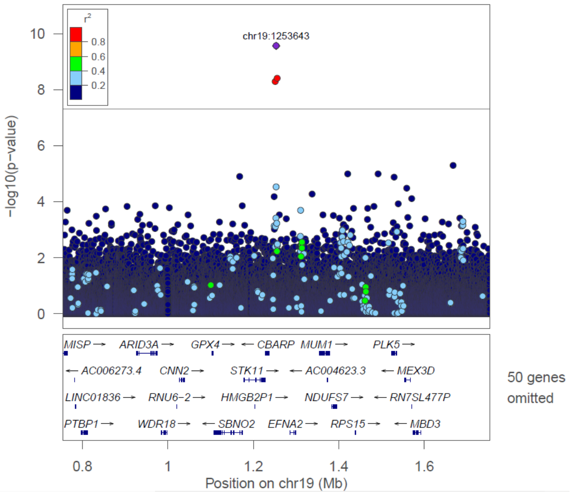
chr11:5227159 − LD: TOPMed − MAF: 0.000256 − MAC: 25

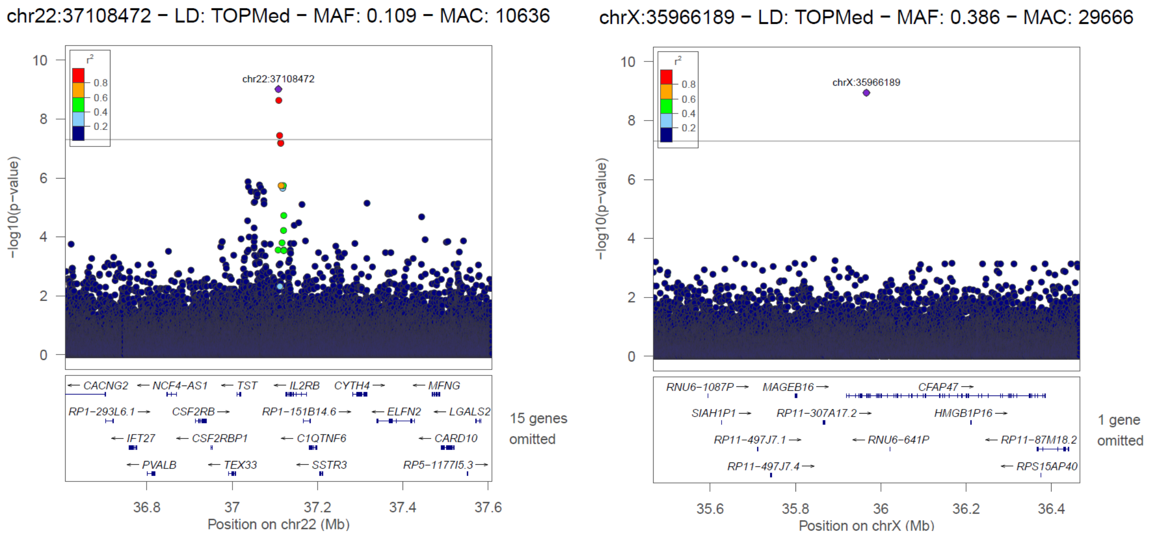


chr16:55649 − LD: TOPMed − MAF: 0.000217 − MAC: 17

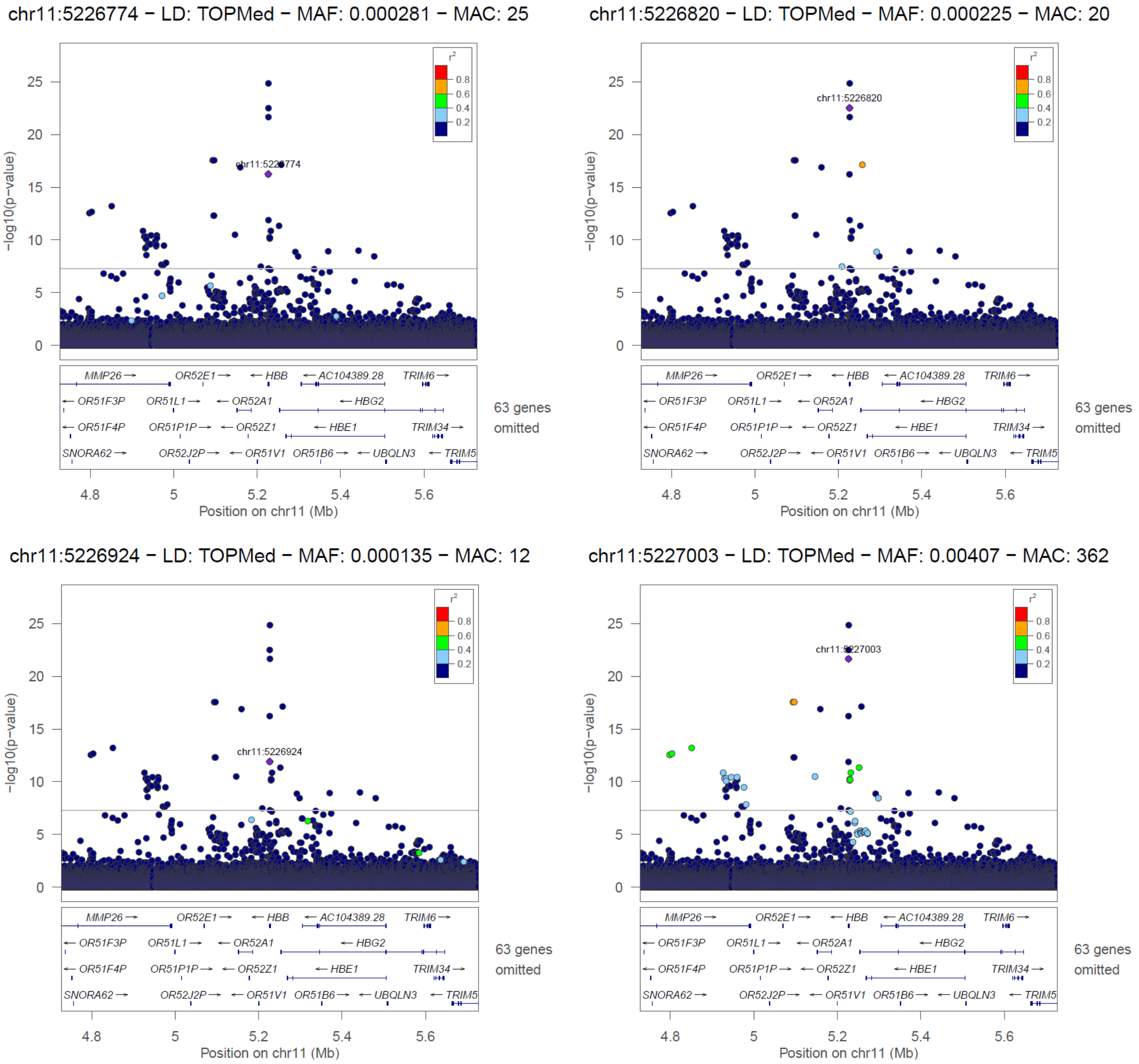


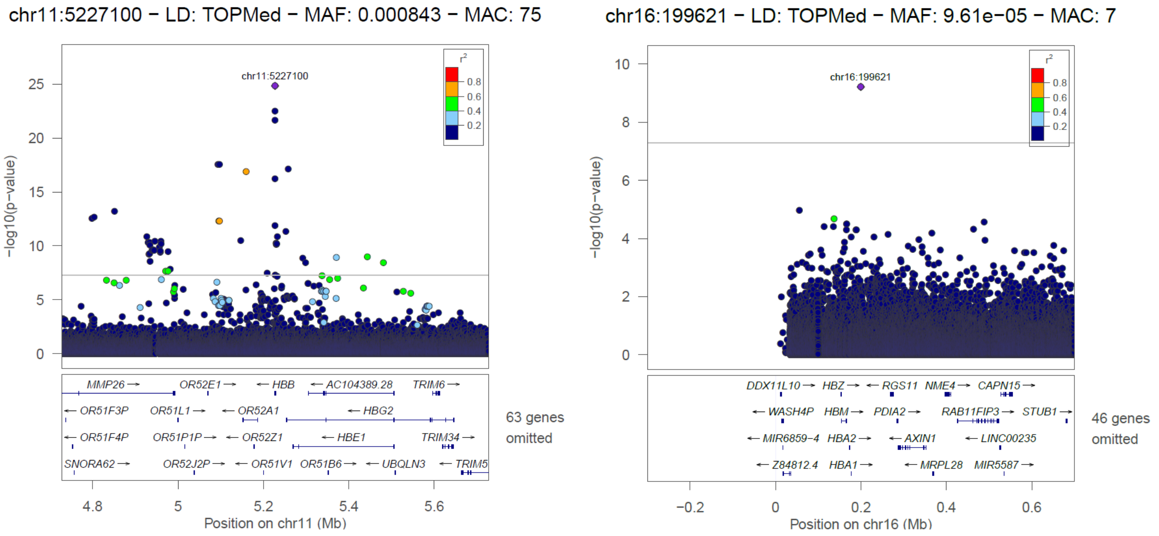
chr19:1253643 − LD: TOPMed − MAF: 0.167 − MAC: 16328




chr22:37108472 − LD: TOPMed − MAF: 0.109 − MAC: 10636
−log10(p−value)
r²
0.8
0.6
0.4
0.2
chr22:37108472
CACNG2
NCF4−AS1
TST
IL2RB
CYTH4
MFNG
RP1−293L6.1
CSF2RB
RP1−151B14.6
ELFN2
LGALS2
IFT27
CSF2RBP1
C1QTNF6
CARD10
PVALB
TEX33
SSTR3
RP5−1177I5.3
15 genes omitted
36.8
37
37.2
37.4
37.6
Position on chr22 (Mb)
chrX:35966189 − LD: TOPMed − MAF: 0.386 − MAC: 29666
−log10(p−value)
chrX:35966189
RNU6−1087P
MAGEB16
CFAP47
SIAH1P1
RP11−307A17.2
HMGB1P16
RP11−497J7.1
RNU6−641P
RP11−87M18.2
RP11−497J7.4
RPS15AP40
1 gene omitted
35.6
35.8
36
36.2
36.4
Position on chrX (Mb)

(F)


chr11:5226774 − LD: TOPMed − MAF: 0.000281 − MAC: 25
−log10(p−value)
r²
0.8
0.6
0.4
0.2
chr11:5226774
chr11:5226820 − LD: TOPMed − MAF: 0.000225 − MAC: 20
chr11:5226820
chr11:5226924 − LD: TOPMed − MAF: 0.000135 − MAC: 12
chr11:5226924
chr11:5227003 − LD: TOPMed − MAF: 0.00407 − MAC: 362
chr11:5227003
MMP26
OR52E1
HBB
AC104389.28
TRIM6
OR51F3P
OR51L1
OR52A1
HBG2
OR51F4P
OR51P1P
OR52Z1
HBE1
TRIM34
SNORA62
OR52J2P
OR51V1
OR51B6
UBQLN3
TRIM5
63 genes omitted
4.8
5
5.2
5.4
5.6
Position on chr11 (Mb)


chr11:5227100 − LD: TOPMed − MAF: 0.000843 − MAC: 75
−log10(p−value)
chr11:5227100
r²
0.8
0.6
0.4
0.2
MMP26 OR52E1 HBB AC104389.28 TRIM6
OR51F3P OR51L1 OR52A1 HBG2
OR51F4P OR51P1P OR52Z1 HBE1 TRIM34
SNORA62 OR52J2P OR51V1 OR51B6 UBQLN3 TRIM5
63 genes omitted
4.8 5 5.2 5.4 5.6
Position on chr11 (Mb)
chr16:199621 − LD: TOPMed − MAF: 9.61e−05 − MAC: 7
−log10(p−value)
chr16:199621
DDX11L10 HBZ RGS11 NME4 CAPN15
WASH4P HBM PDIA2 RAB11FIP3 STUB1
MIR6859−4 HBA2 AXIN1 LINC00235
Z84812.4 HBA1 MRPL28 MIR5587
46 genes omitted
−0.2 0 0.2 0.4 0.6
Position on chr16 (Mb)

(G)

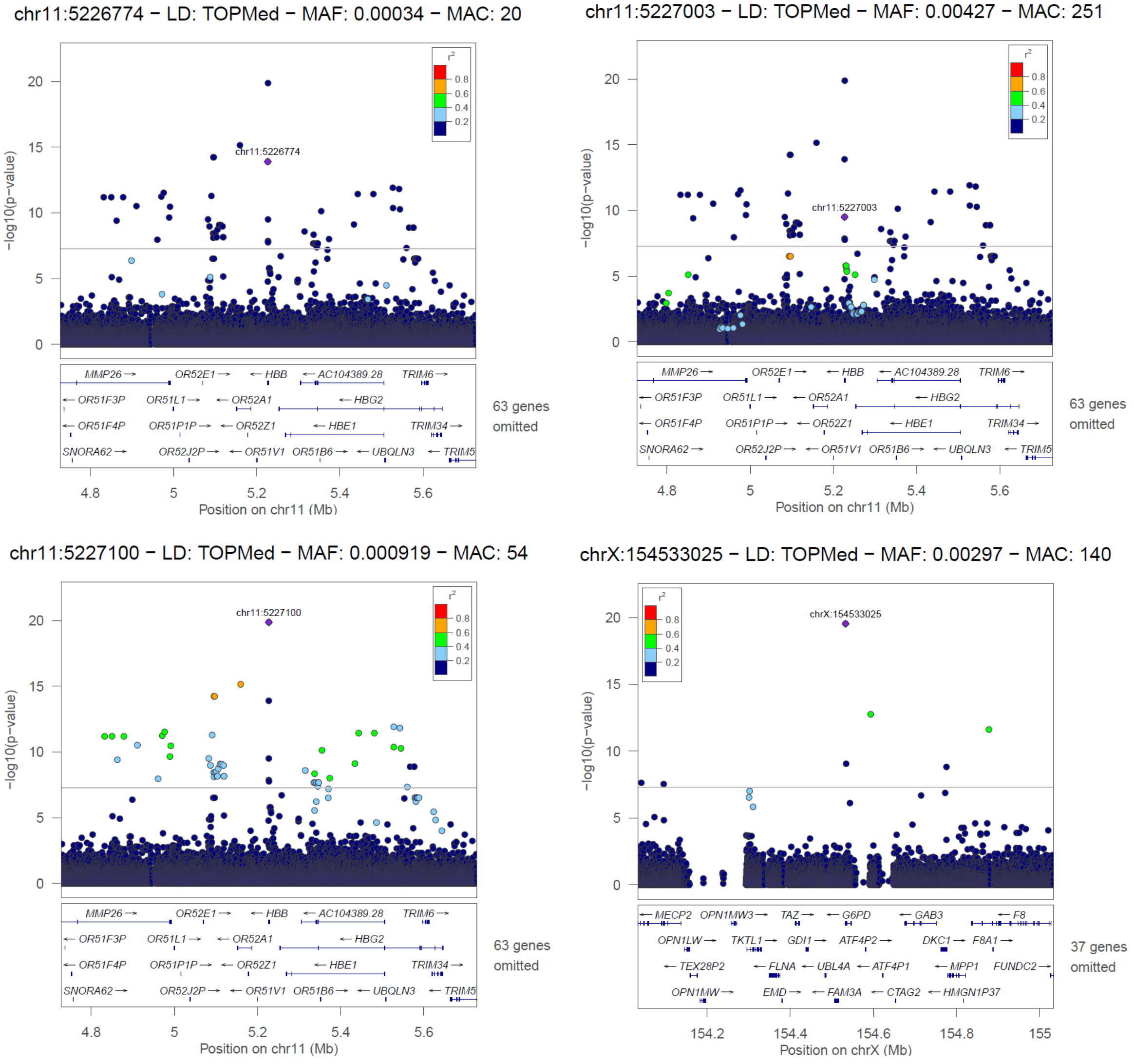
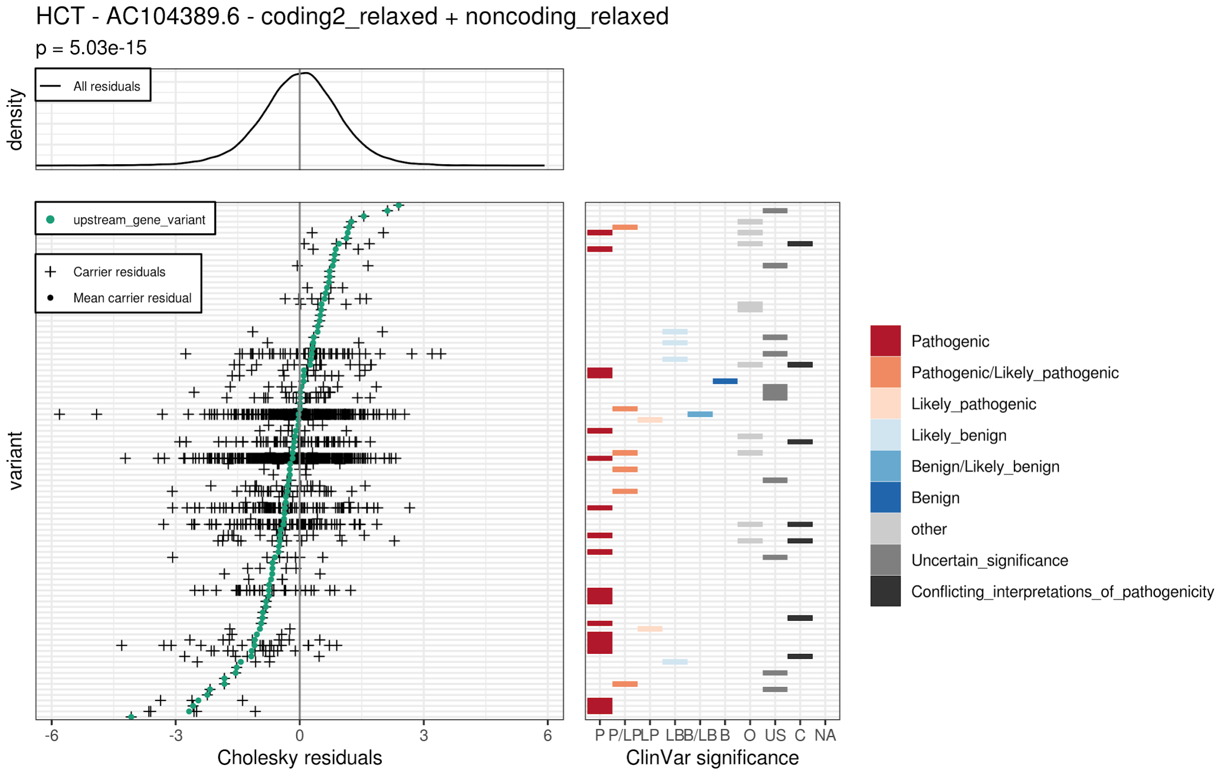

chr11:5226774 − LD: TOPMed − MAF: 0.00034 − MAC: 20
−log10(p−value)
chr11:5226774
63 genes omitted
4.8 5 5.2 5.4 5.6
Position on chr11 (Mb)
chr11:5227003 − LD: TOPMed − MAF: 0.00427 − MAC: 251
−log10(p−value)
chr11:5227003
63 genes omitted
4.8 5 5.2 5.4 5.6
Position on chr11 (Mb)
chr11:5227100 − LD: TOPMed − MAF: 0.000919 − MAC: 54
−log10(p−value)
chr11:5227100
63 genes omitted
4.8 5 5.2 5.4 5.6
Position on chr11 (Mb)
MMP26 OR52E1 HBB AC104389.28 TRIM6
OR51F3P OR51L1 OR52A1 HBG2
OR51F4P OR51P1P OR52Z1 HBE1 TRIM34
SNORA62 OR52J2P OR51V1 OR51B6 UBQLN3 TRIM5
chrX:154533025 − LD: TOPMed − MAF: 0.00297 − MAC: 140
−log10(p−value)
chrX:154533025
MECP2 OPN1MW3 TAZ G6PD GAB3 F8
OPN1LW TKTL1 GDI1 ATF4P2 DKC1 F8A1
TEX28P2 FLNA UBL4A ATF4P1 MPP1 FUNDC2
OPN1MW EMD FAM3A CTAG2 HMGN1P37
37 genes omitted
154.2 154.4 154.6 154.8 155
Position on chrX (Mb)
r²
0.8
0.6
0.4
0.2

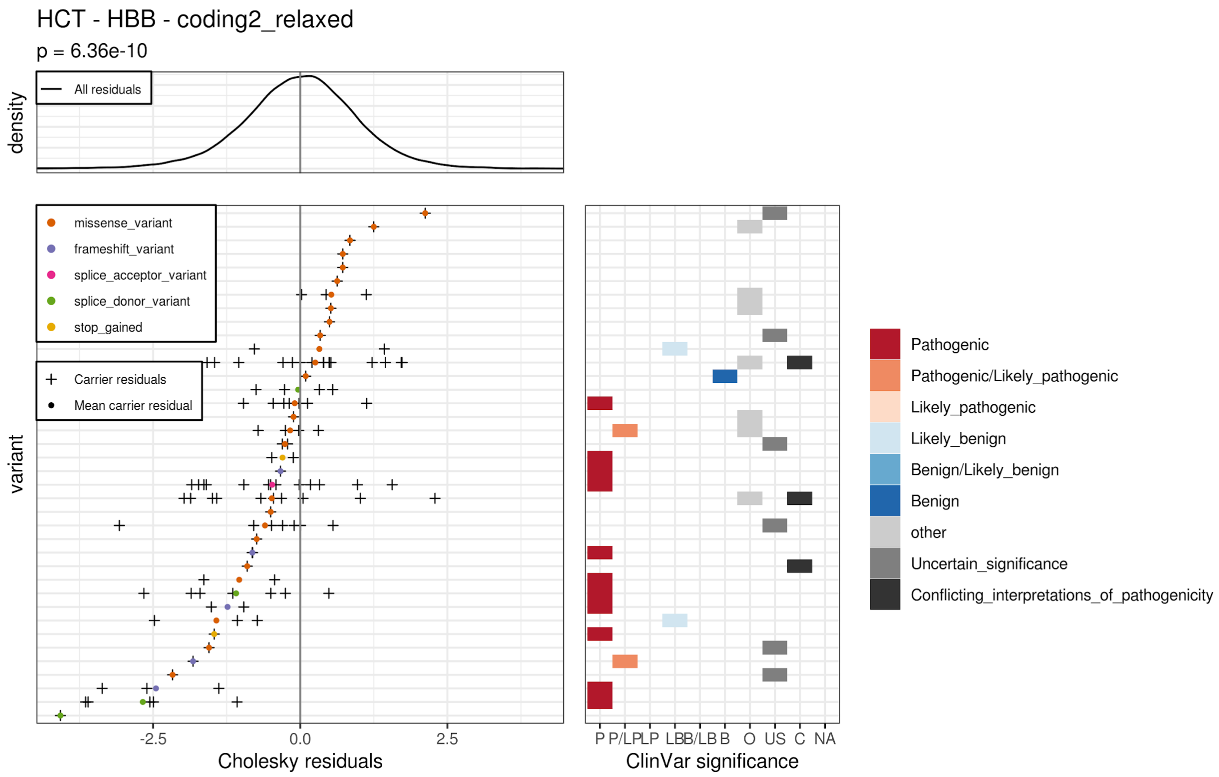
Figure S5. Rare variants identified in the aggregated analysis in TOPMed. (A) HCT; (B) HGB; (C) MCH; (D) MCHC; (E) MCV; (F) RBC; (G) RDW.

(A)

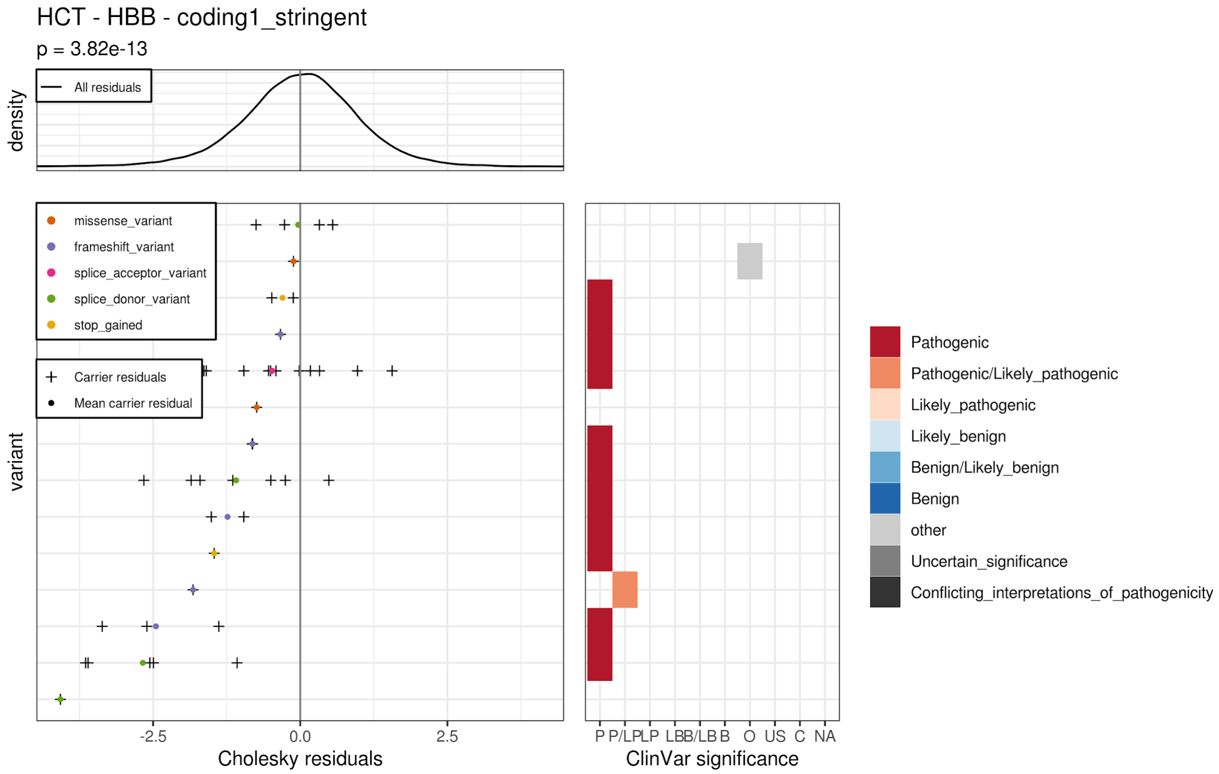
(A1)



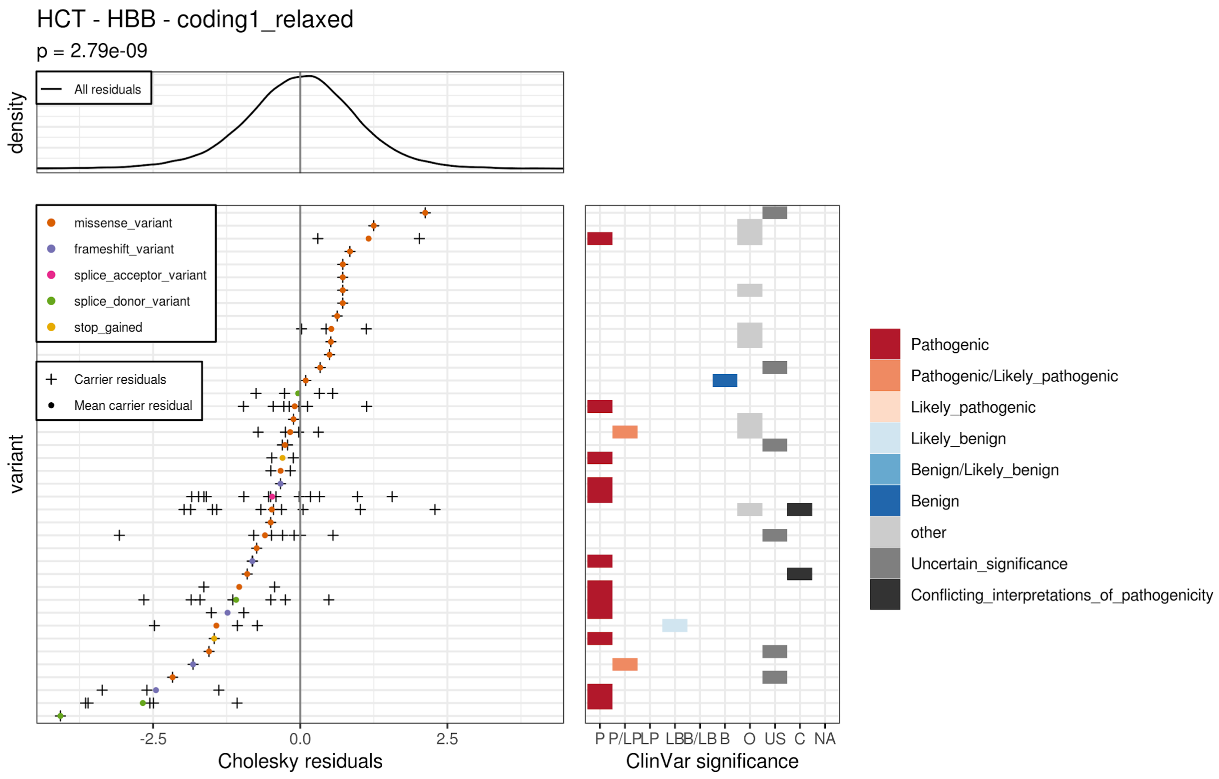
(A2)


HCT - HBB - coding2_relaxed
p = 6.36e-10
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

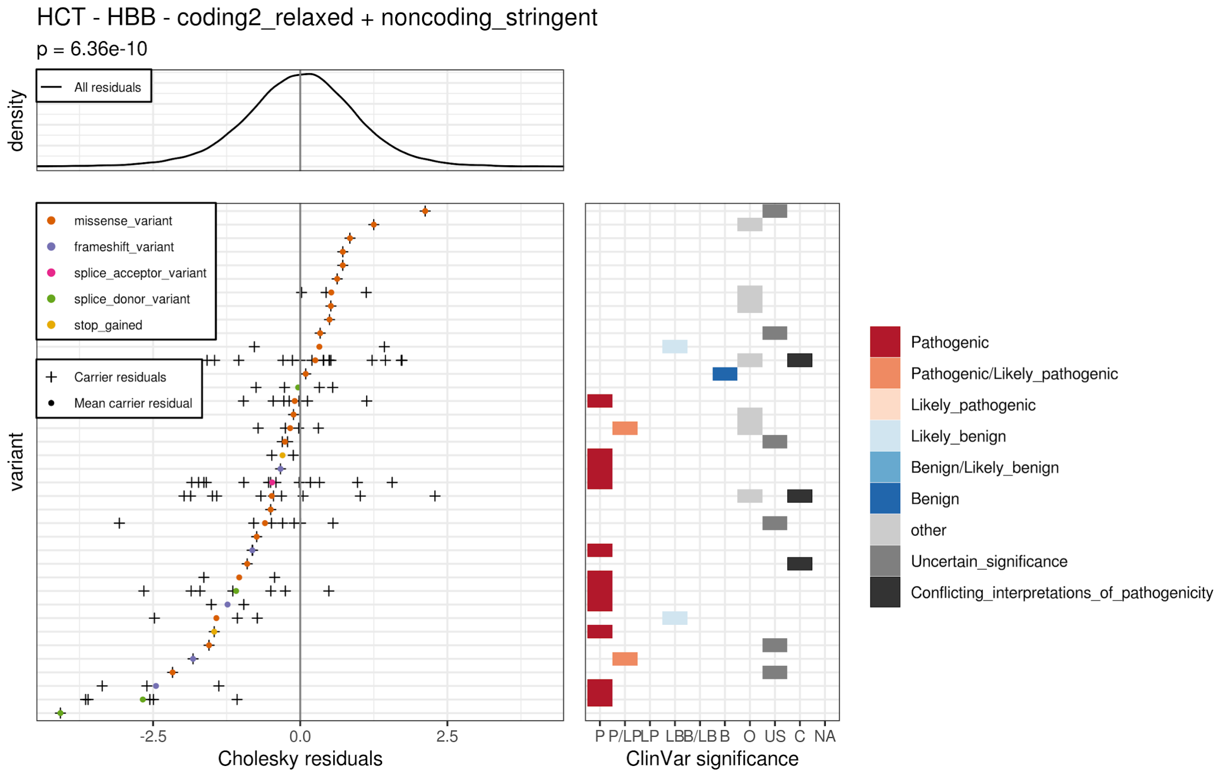
(A3)


HCT - HBB - coding1_stringent
p = 3.82e-13
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(A4)

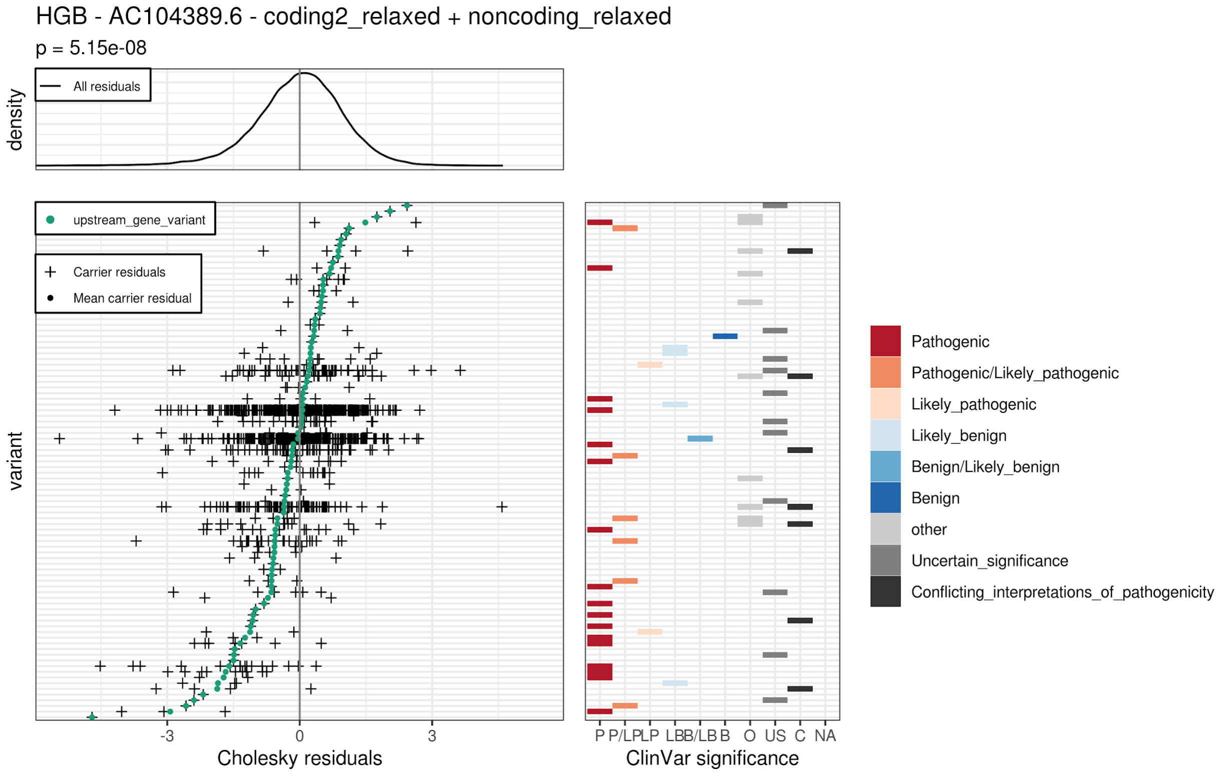

HCT - HBB - coding1_relaxed
p = 2.79e-09
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(A5)

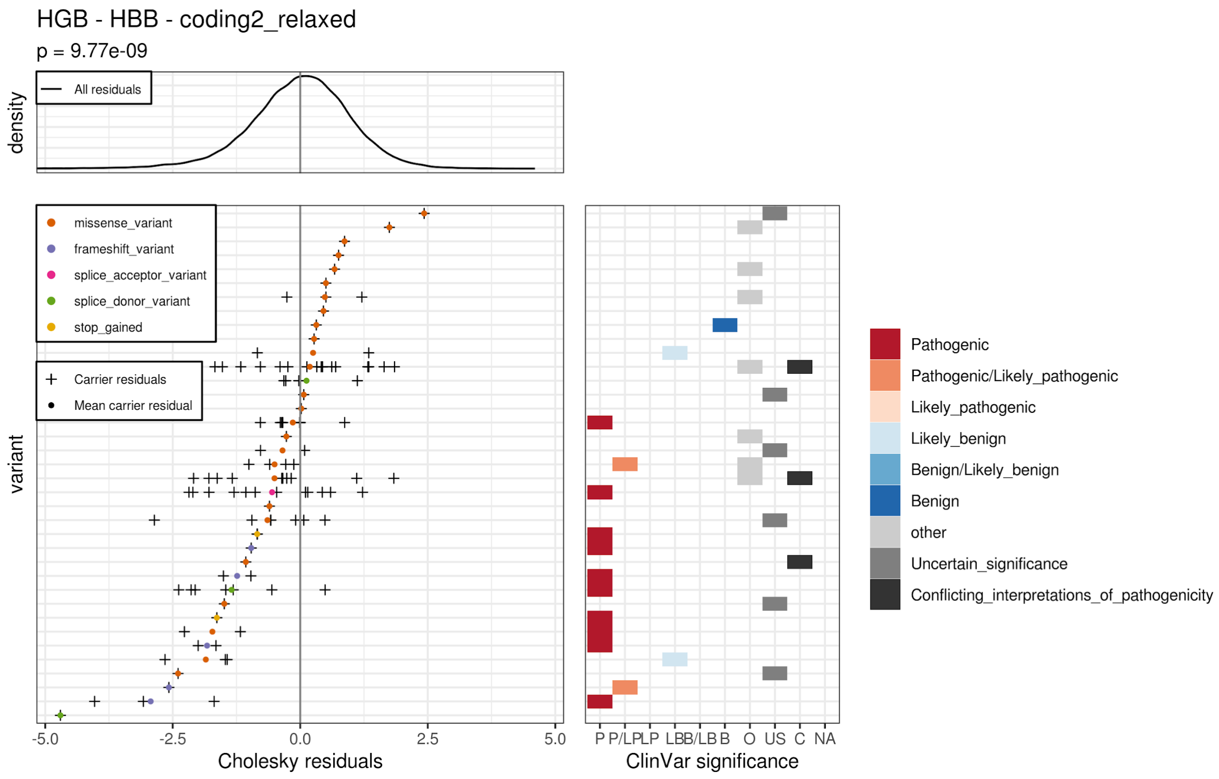

HCT - HBB - coding2_relaxed + noncoding_stringent
p = 6.36e-10
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(B)

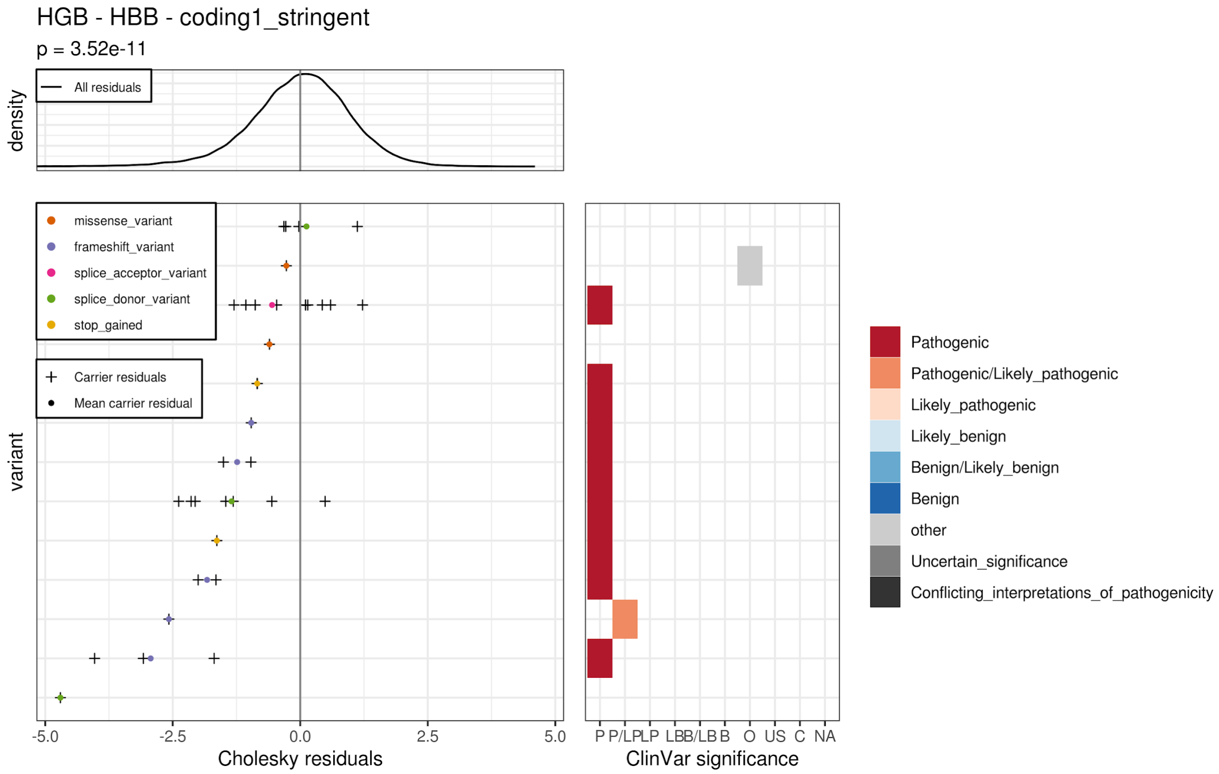
(B1)


HGB - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 5.15e-08
density
All residuals
upstream_gene_variant
Carrier residuals
Mean carrier residual
variant
-3
0
3
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

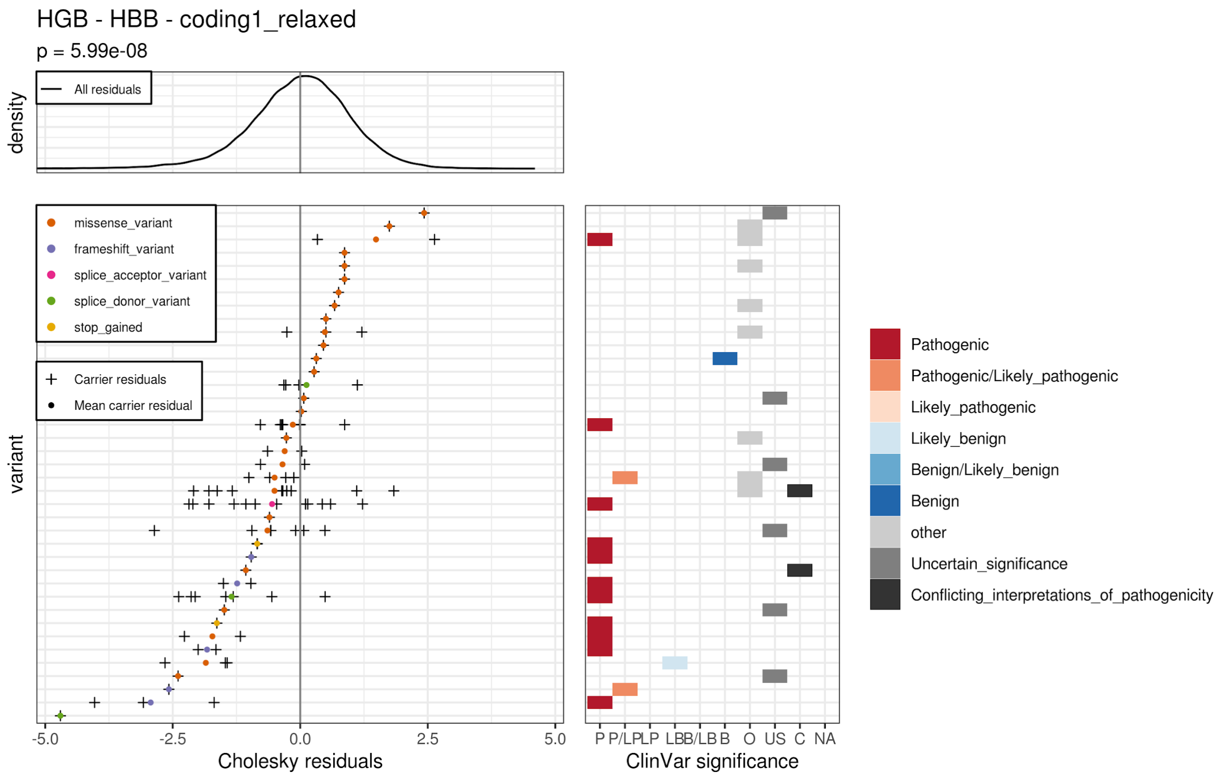
(B2)


HGB - HBB - coding2_relaxed
p = 9.77e-09
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

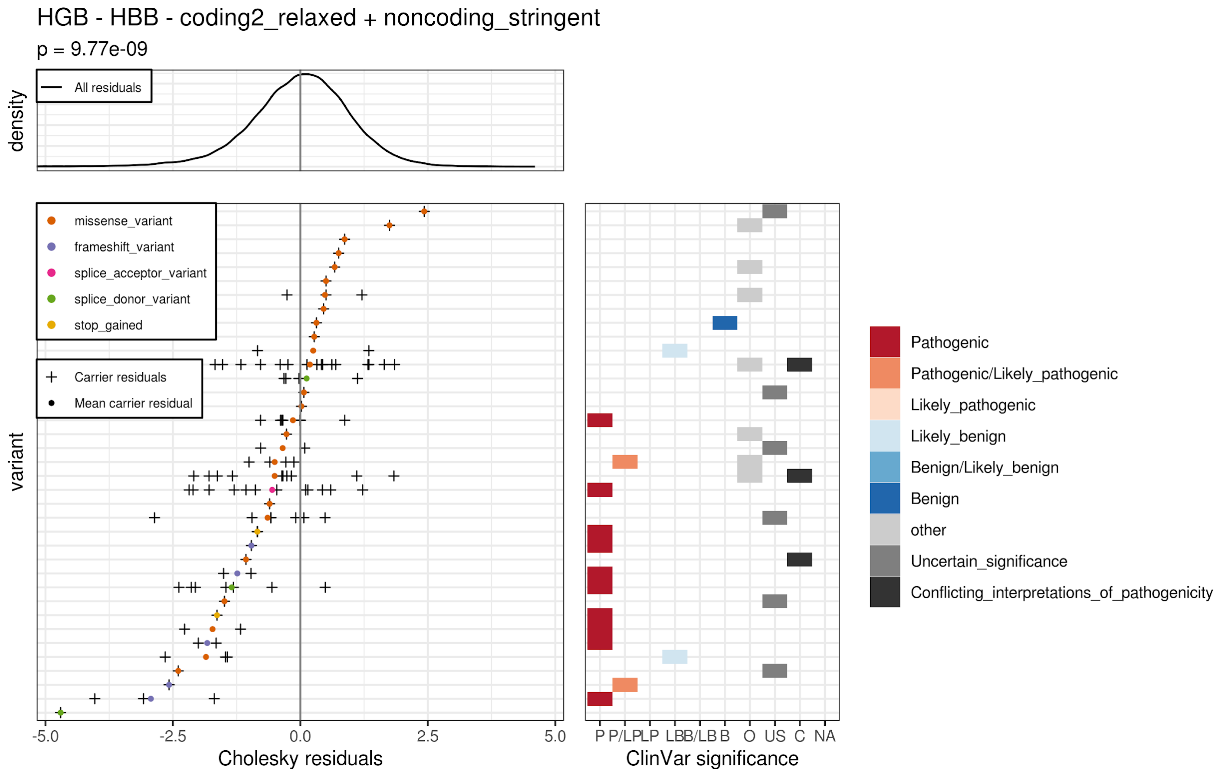
(B3)


HGB - HBB - coding1_stringent
p = 3.52e-11
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(B4)

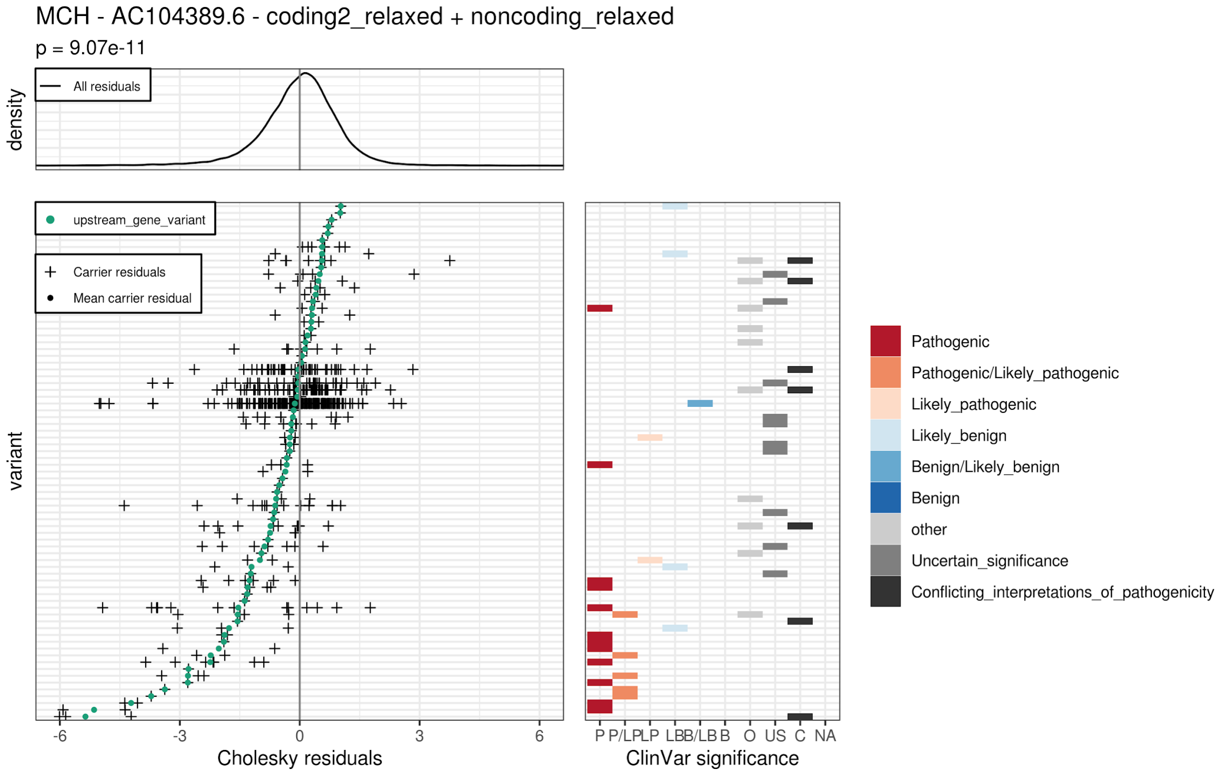

HGB - HBB - coding1_relaxed
p = 5.99e-08
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(B5)

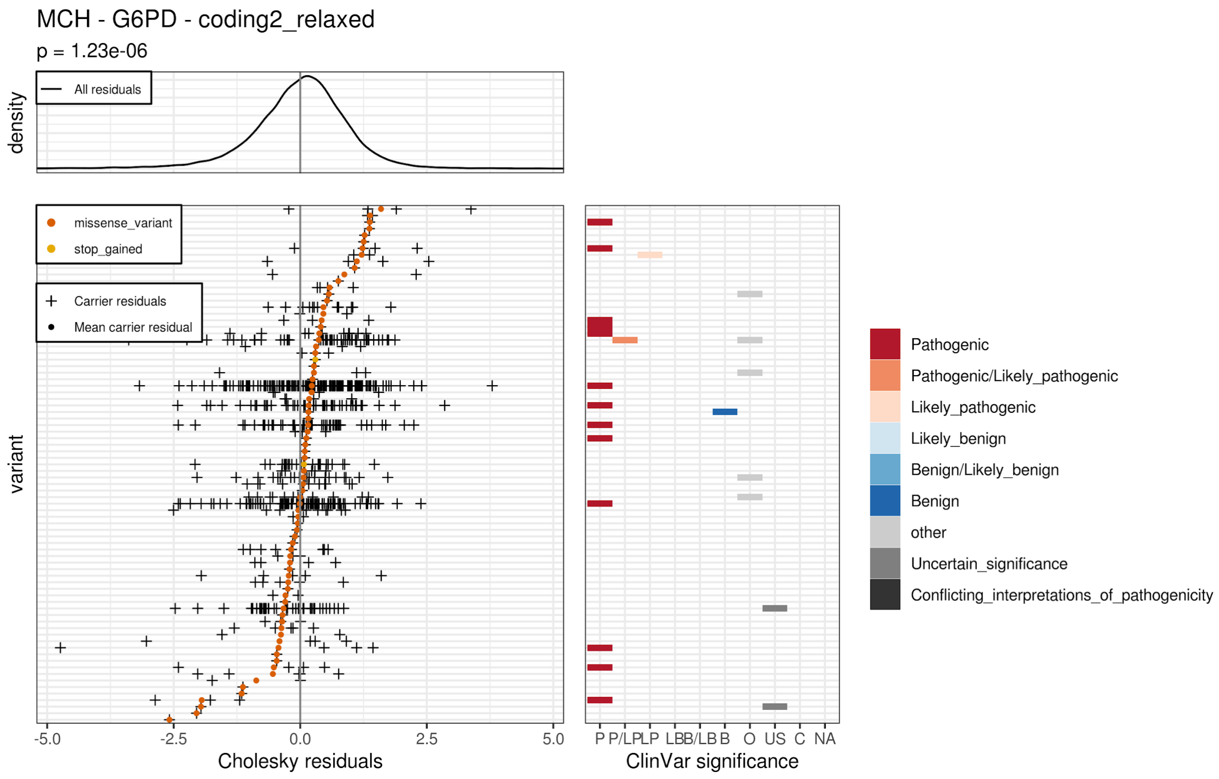

HGB - HBB - coding2_relaxed + noncoding_stringent
p = 9.77e-09
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(C)

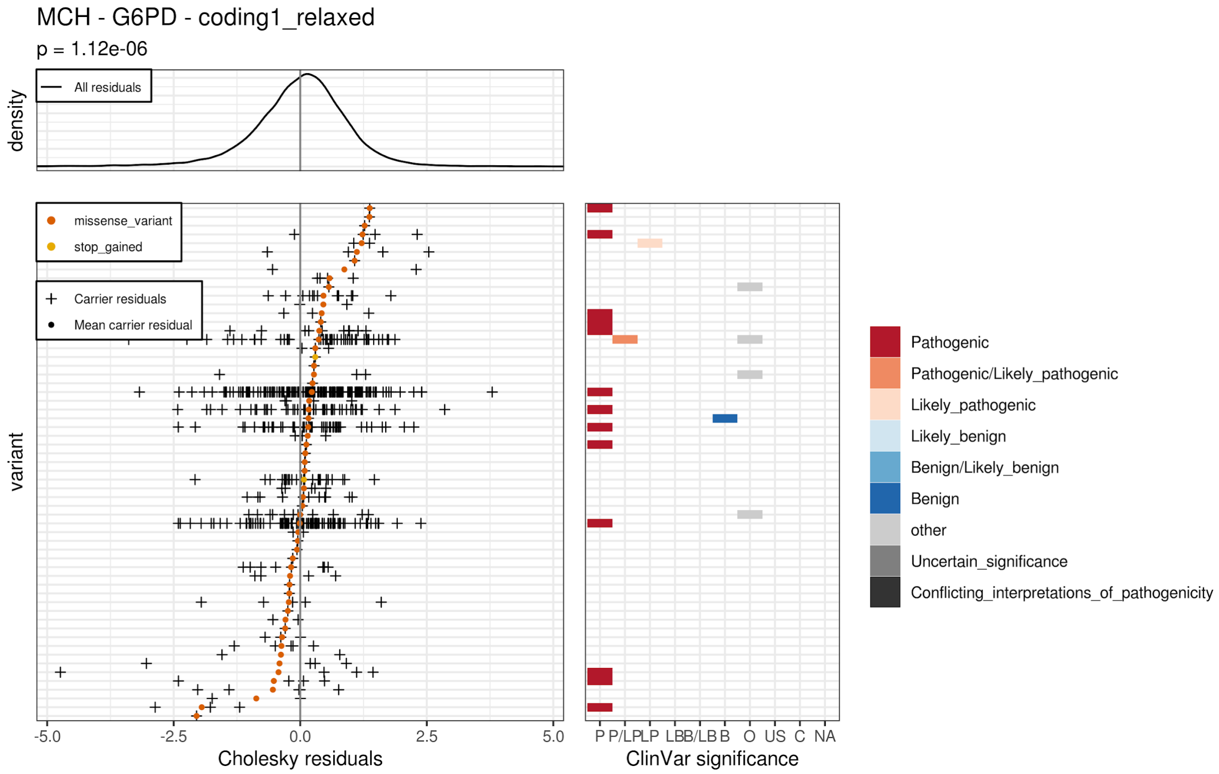
(C1)


MCH - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 9.07e-11
density
All residuals
variant
upstream_gene_variant
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

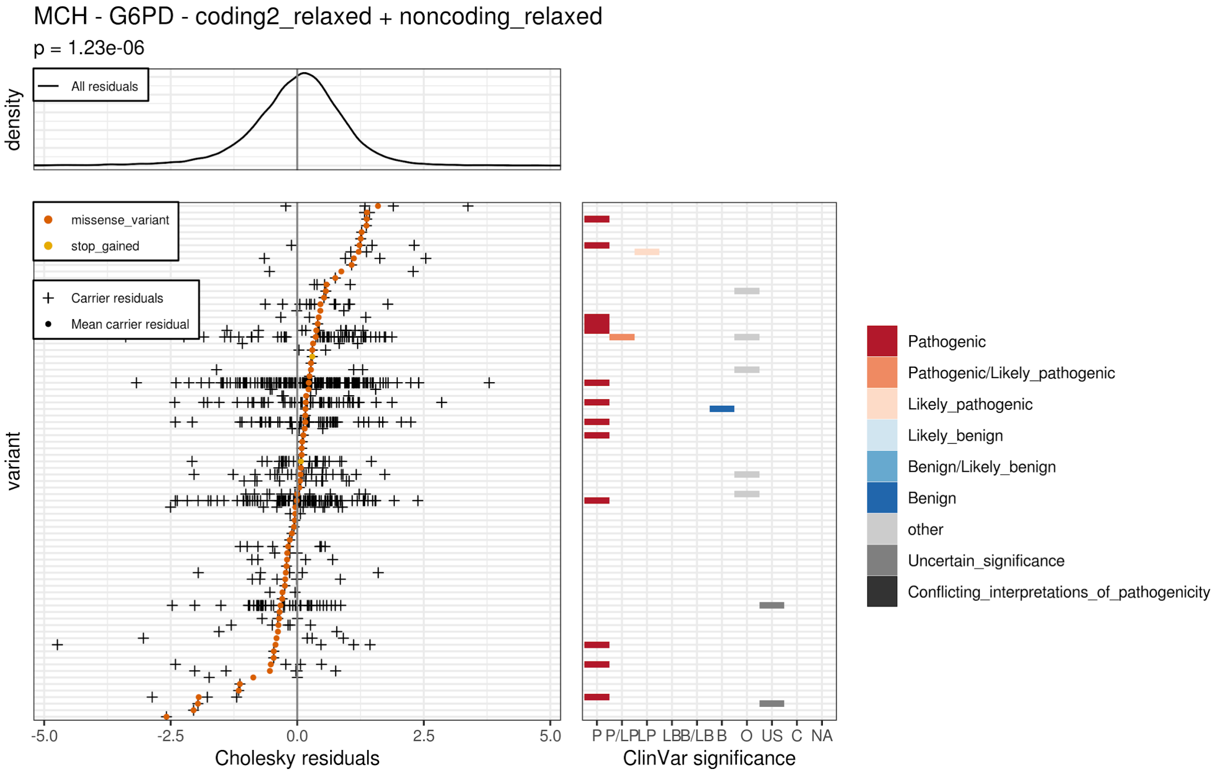
(C2)


MCH - G6PD - coding2_relaxed
p = 1.23e-06
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

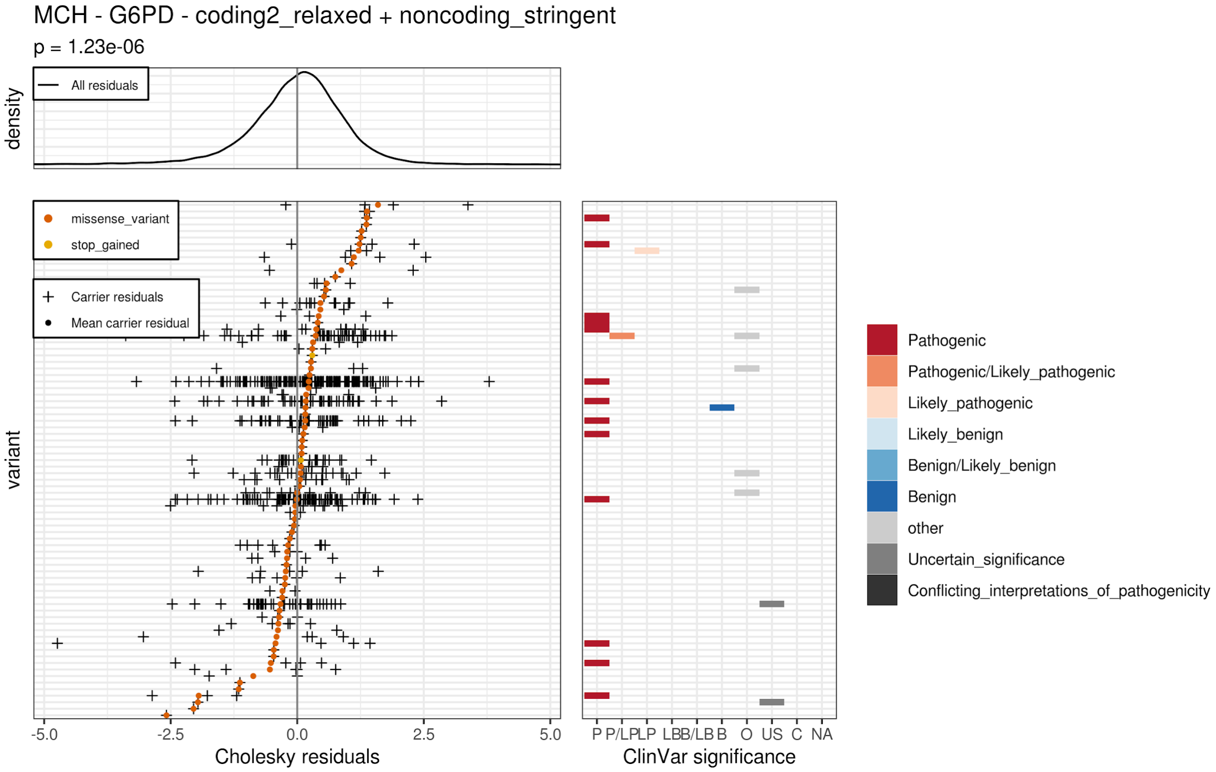
(C3)


MCH - G6PD - coding1_relaxed
p = 1.12e-06
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

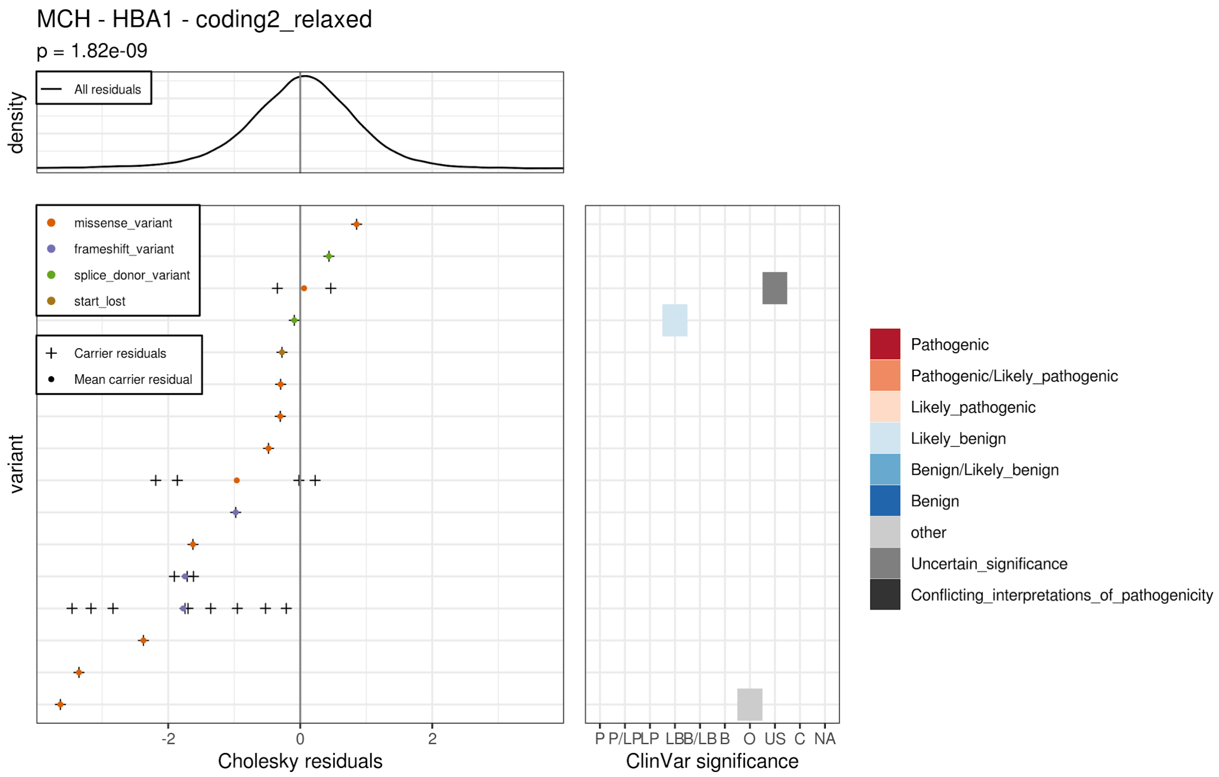
(C4)


MCH - G6PD - coding2_relaxed + noncoding_relaxed
p = 1.23e-06
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

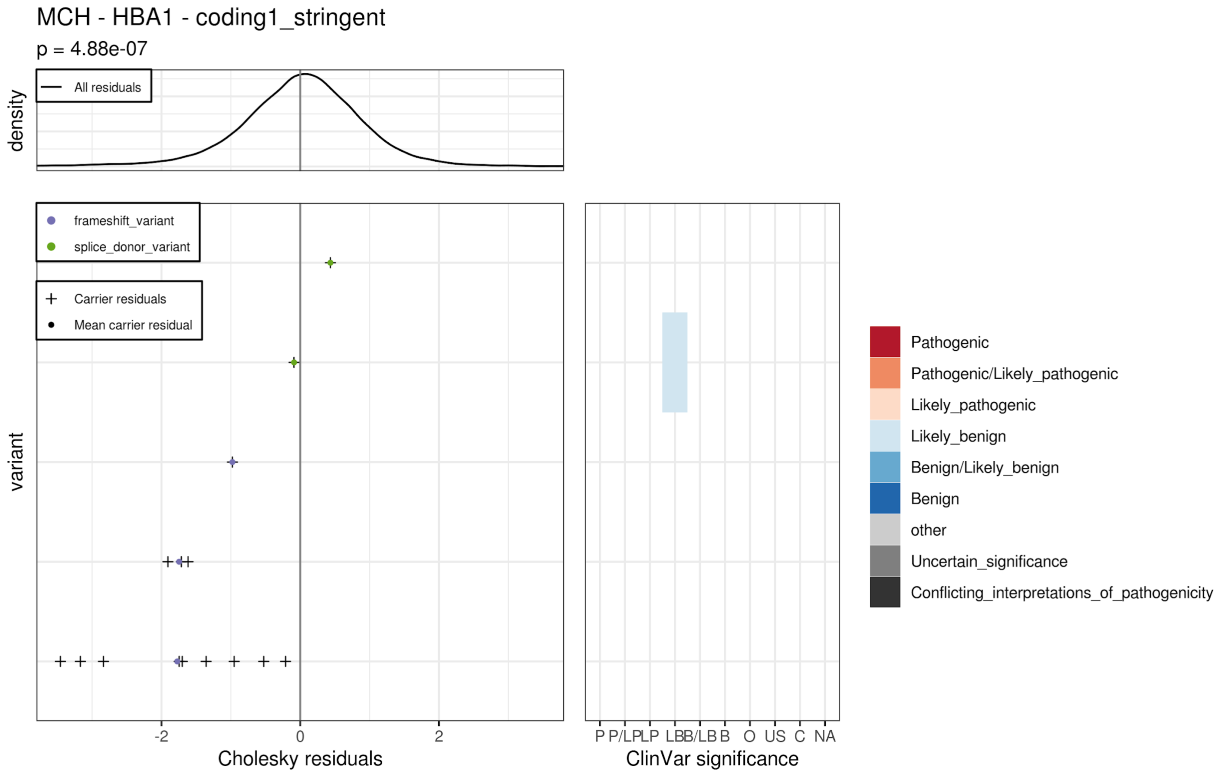
(C5)


MCH - G6PD - coding2_relaxed + noncoding_stringent
p = 1.23e-06
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

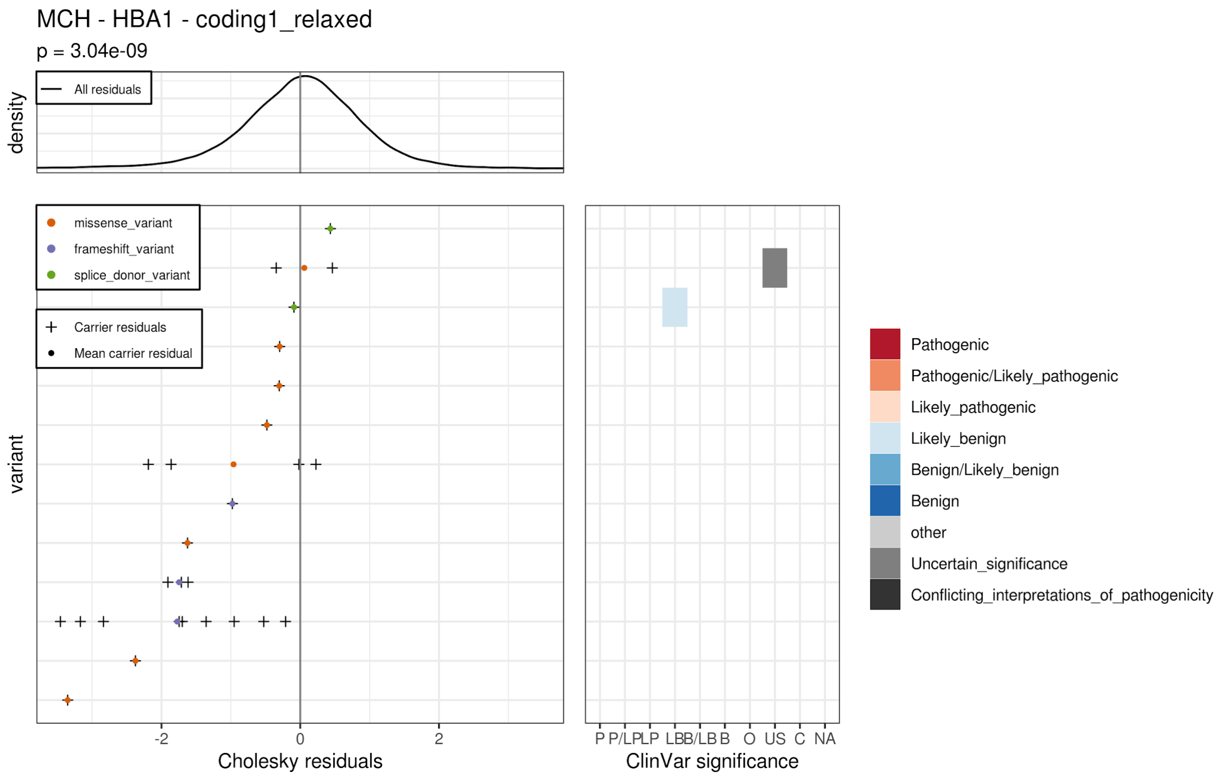
(C6)


MCH - HBA1 - coding2_relaxed
p = 1.82e-09
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
start_lost
Carrier residuals
Mean carrier residual
-2
0
2
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

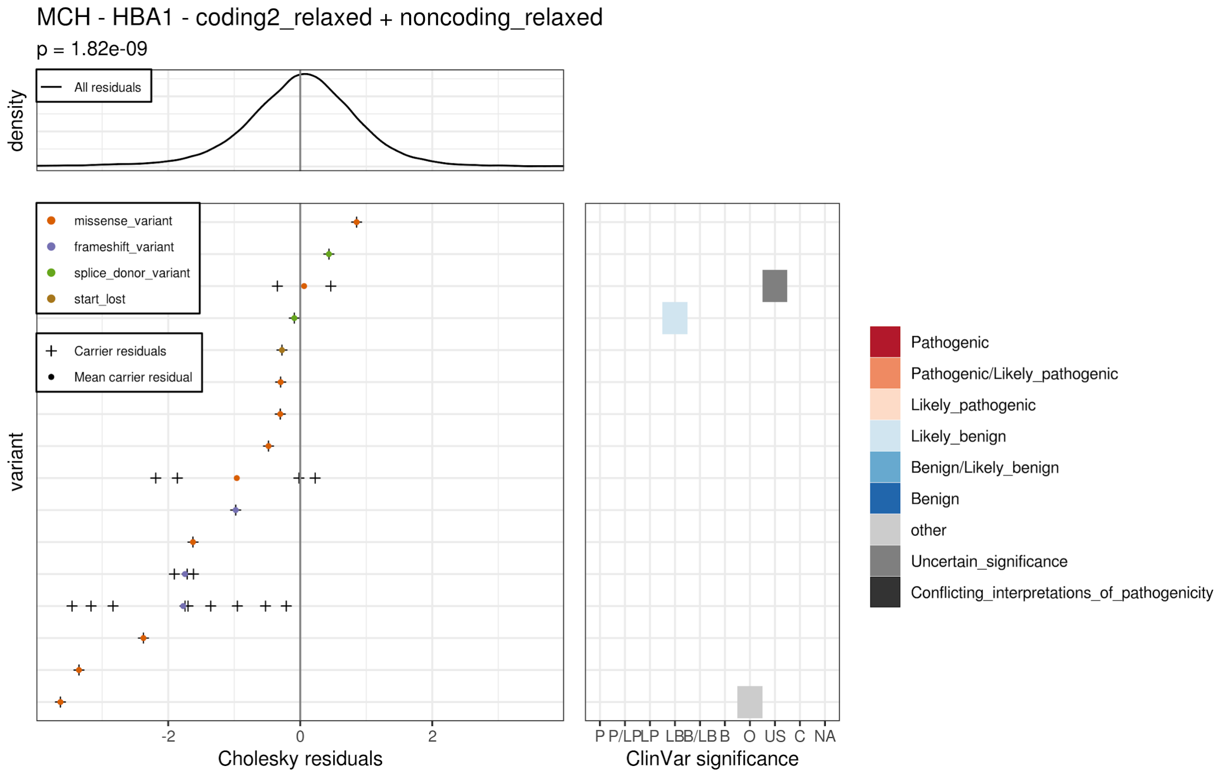
(C7)


MCH - HBA1 - coding1_stringent
p = 4.88e-07
density
All residuals
variant
frameshift_variant
splice_donor_variant
Carrier residuals
Mean carrier residual
-2
0
2
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

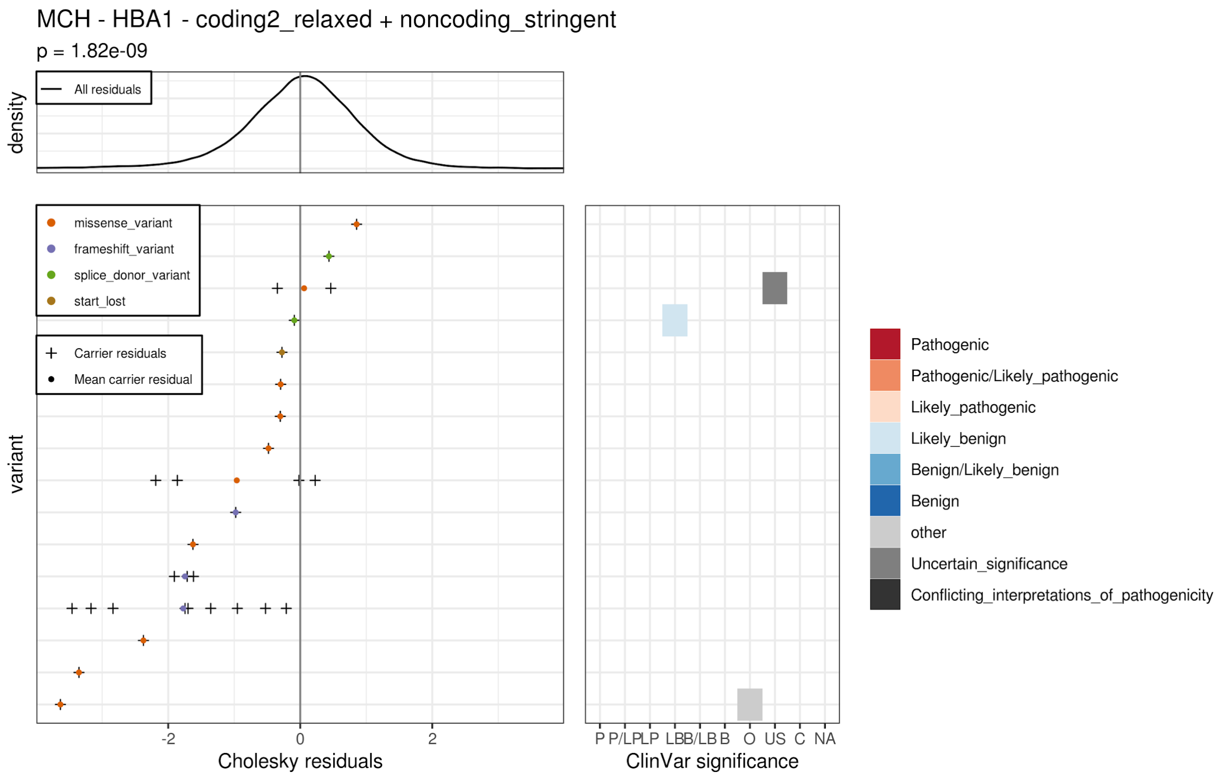
(C8)


MCH - HBA1 - coding1_relaxed
p = 3.04e-09
density
All residuals
missense_variant
frameshift_variant
splice_donor_variant
Carrier residuals
Mean carrier residual
variant
-2
0
2
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

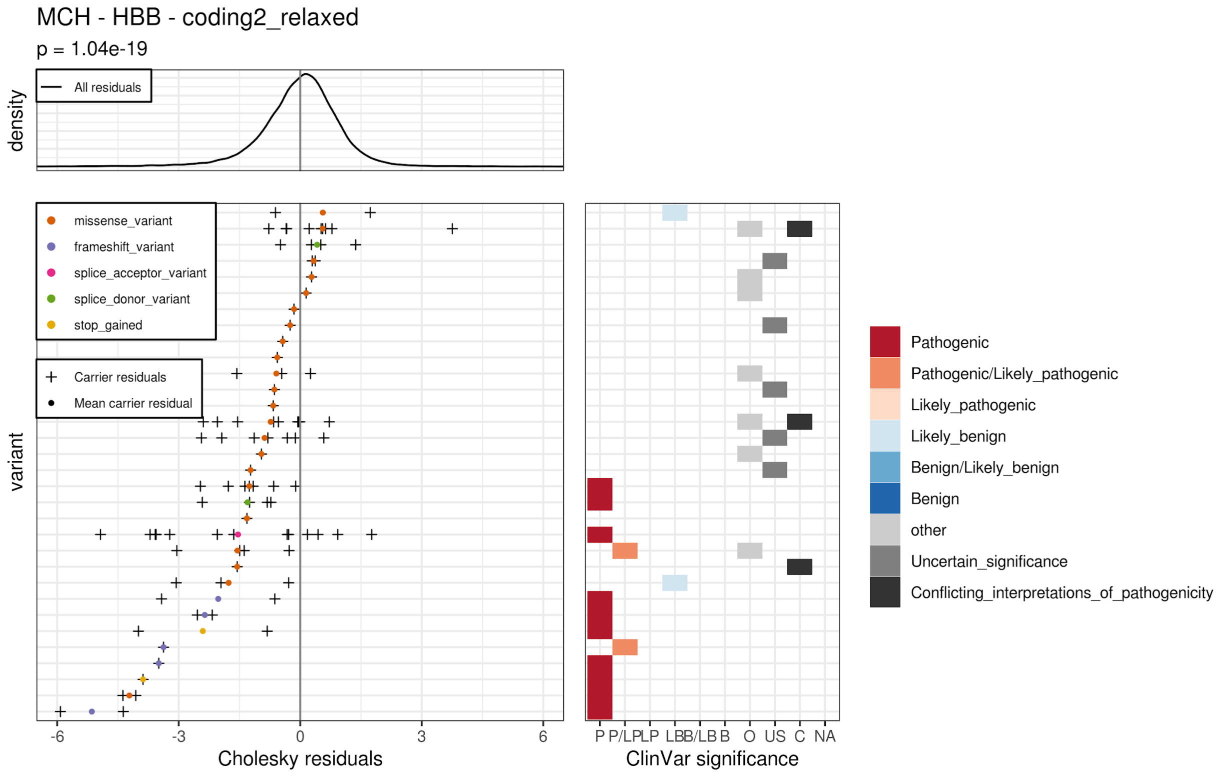
(C9)


MCH - HBA1 - coding2_relaxed + noncoding_relaxed
p = 1.82e-09
density
All residuals
missense_variant
frameshift_variant
splice_donor_variant
start_lost
Carrier residuals
Mean carrier residual
variant
-2
0
2
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

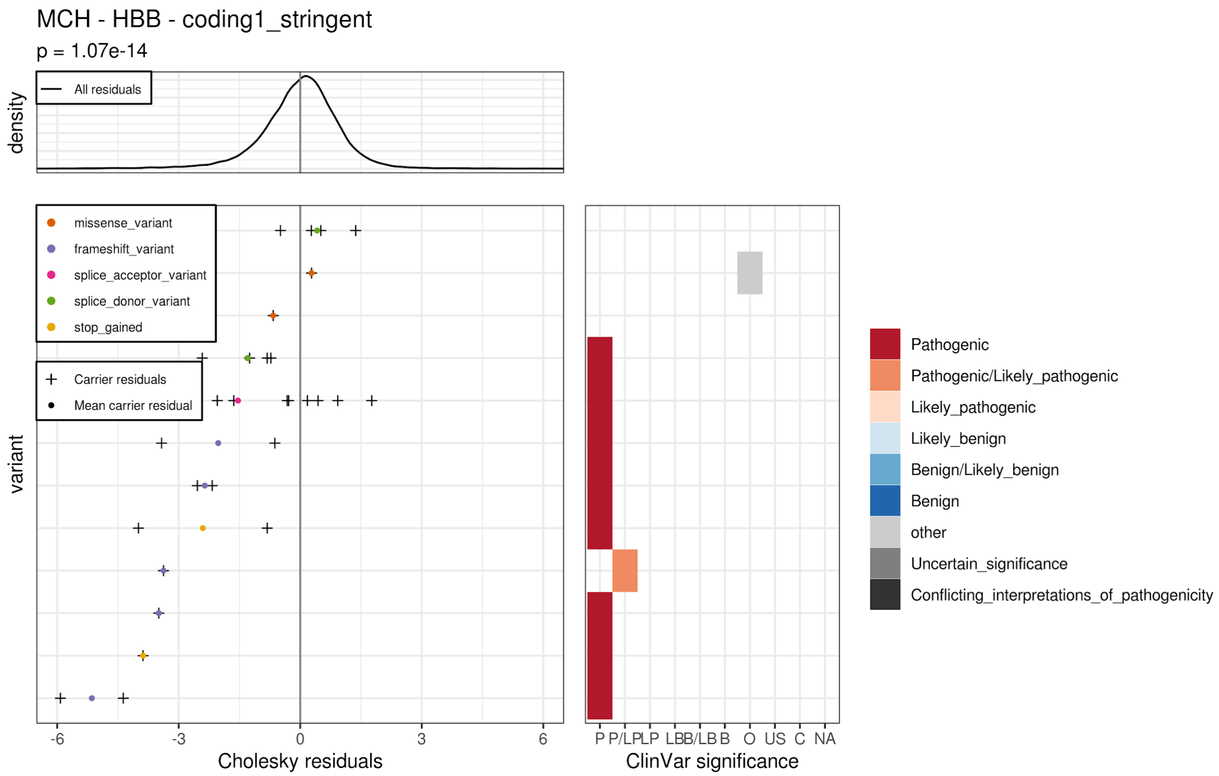
(C10)


MCH - HBA1 - coding2_relaxed + noncoding_stringent
p = 1.82e-09
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
start_lost
Carrier residuals
Mean carrier residual
-2
0
2
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

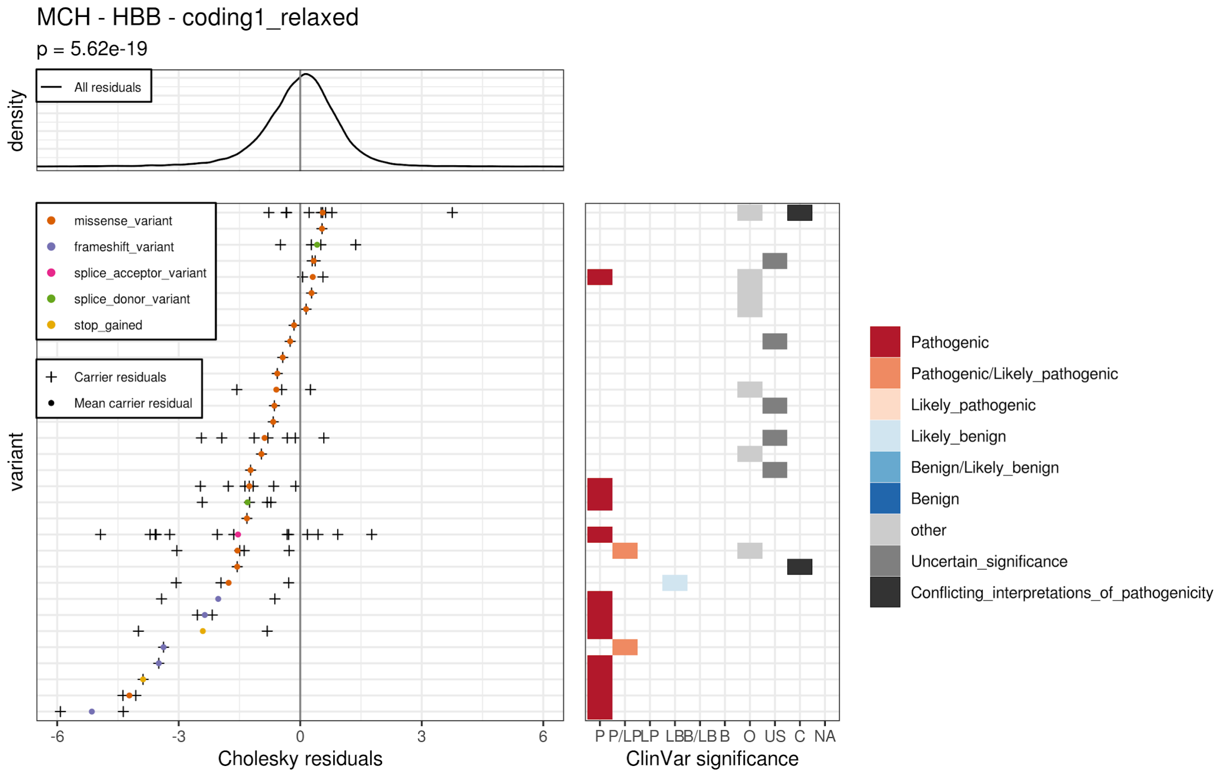
(C11)


MCH - HBB - coding2_relaxed
p = 1.04e-19
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

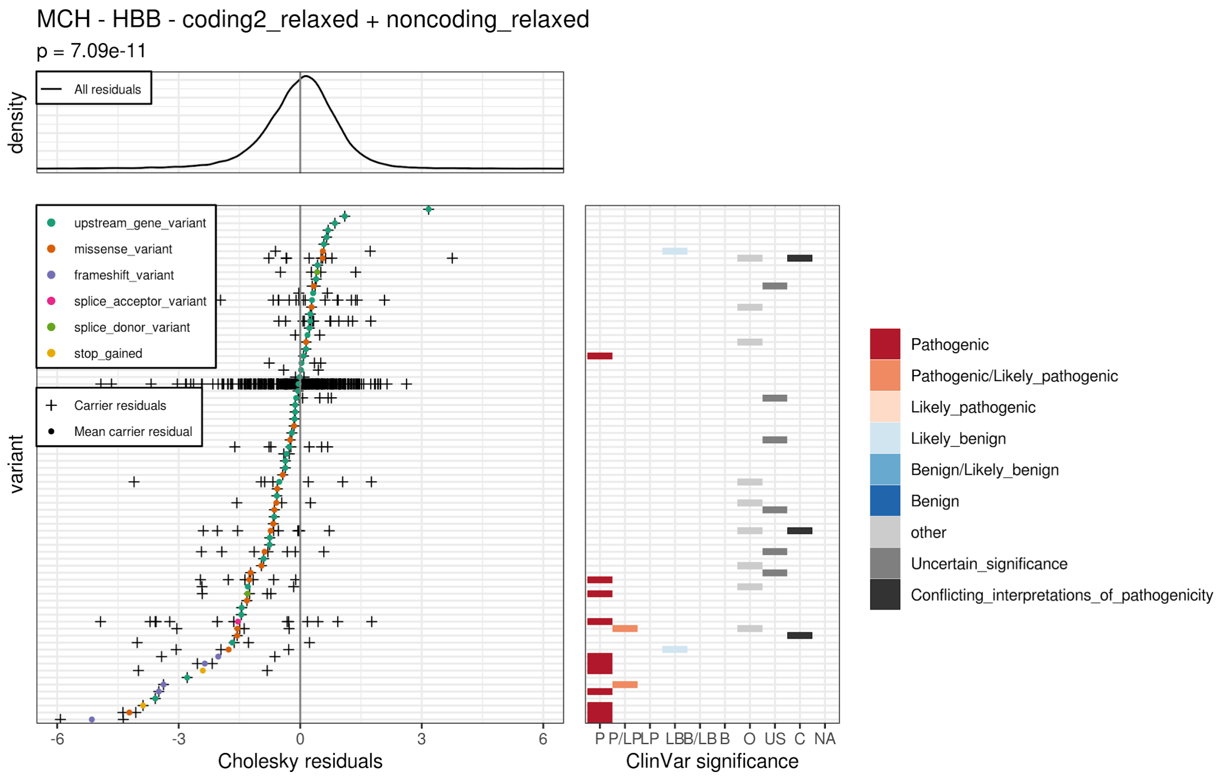
(C12)


MCH - HBB - coding1_stringent
p = 1.07e-14
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

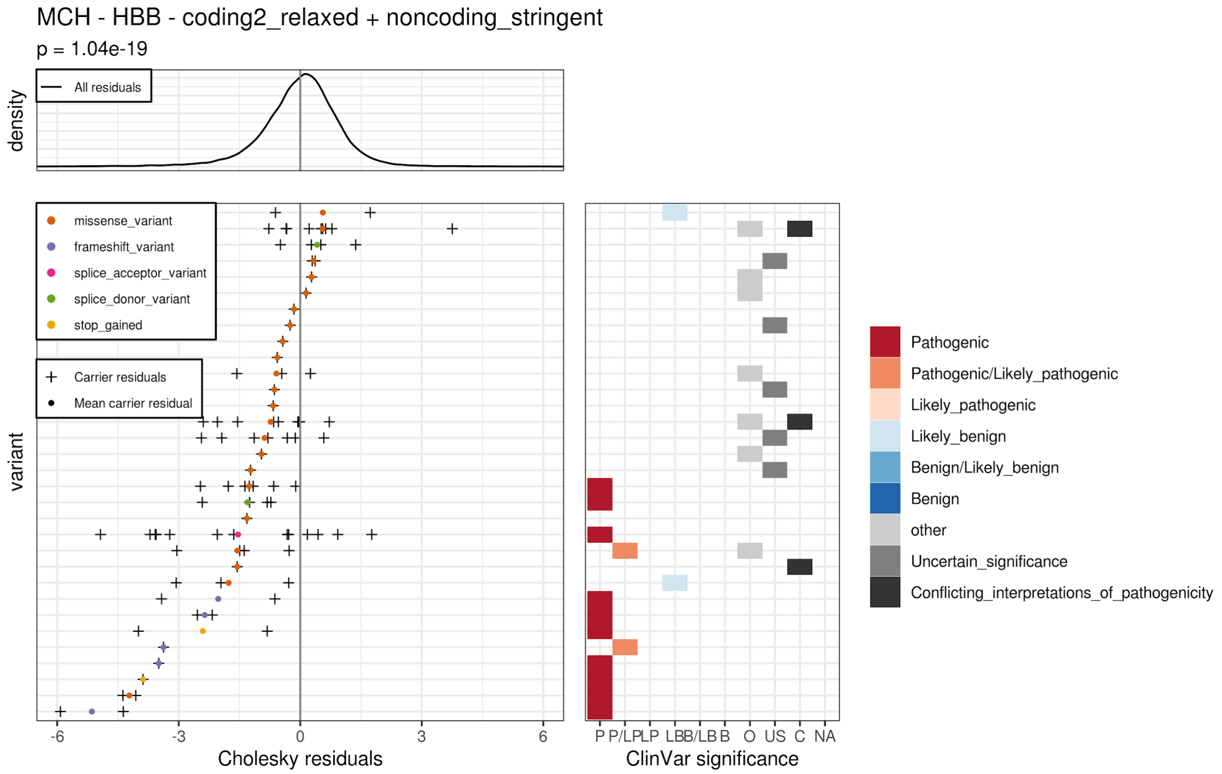
(C13)


MCH - HBB - coding1_relaxed
p = 5.62e-19
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

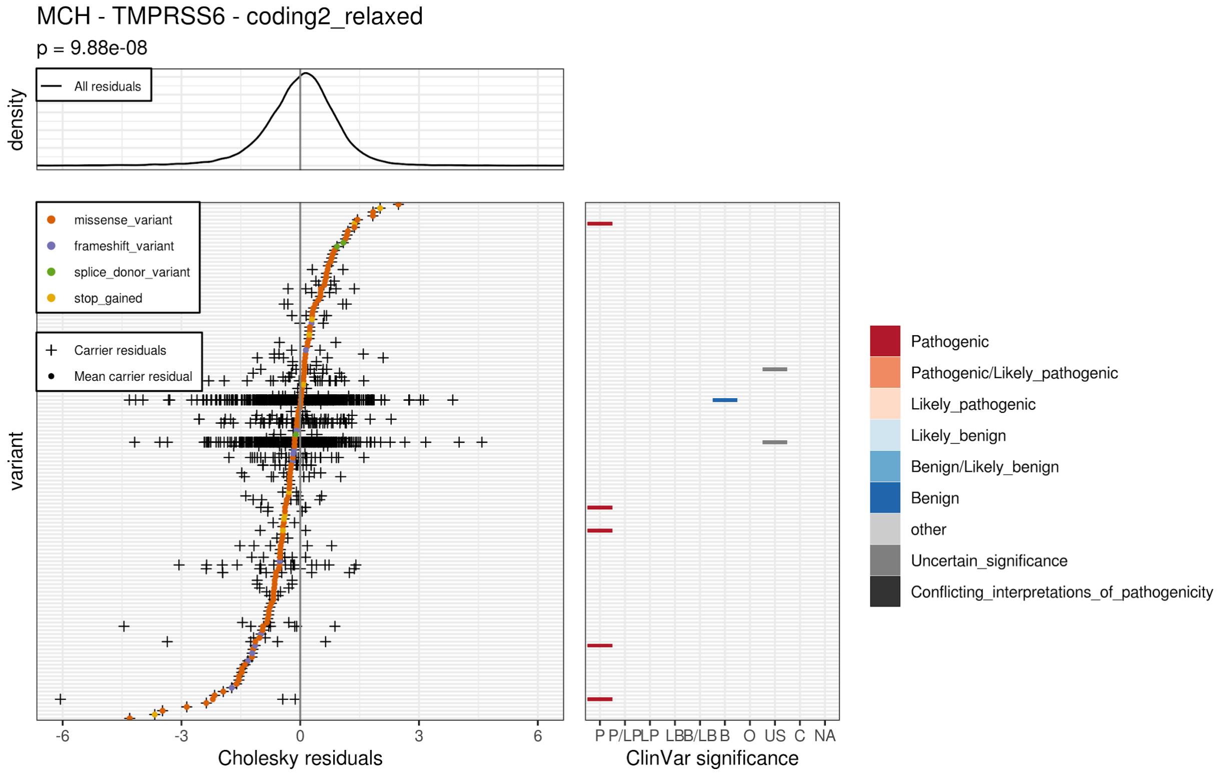
(C14)


MCH - HBB - coding2_relaxed + noncoding_relaxed
p = 7.09e-11
All residuals
density
upstream_gene_variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

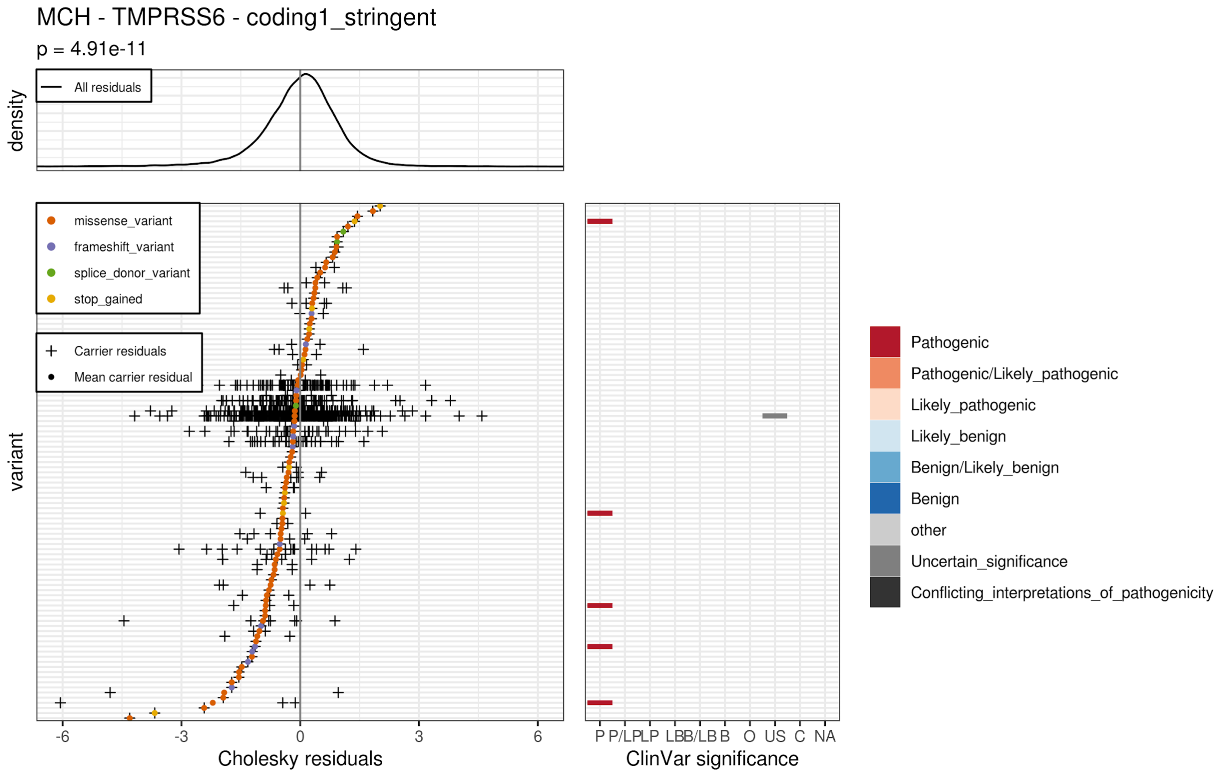
(C15)


MCH - HBB - coding2_relaxed + noncoding_stringent
p = 1.04e-19
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

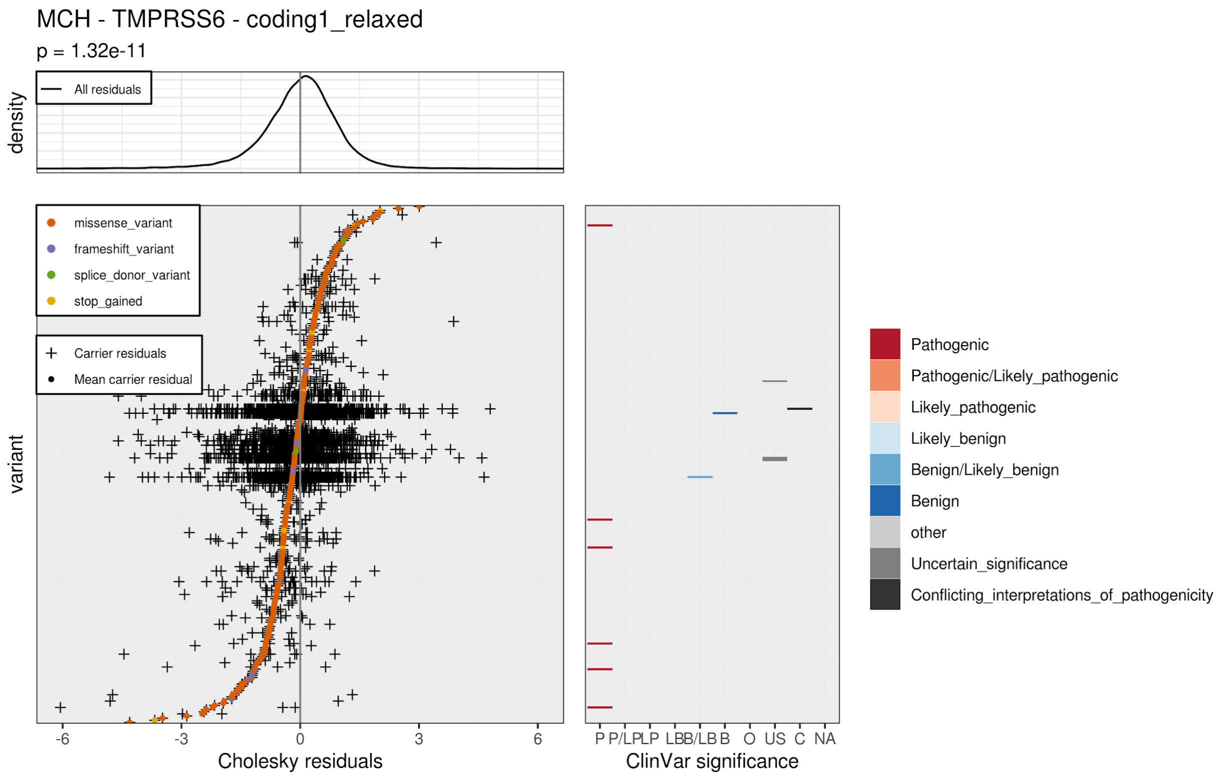
(C16)


MCH - TMPRSS6 - coding2_relaxed
p = 9.88e-08
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

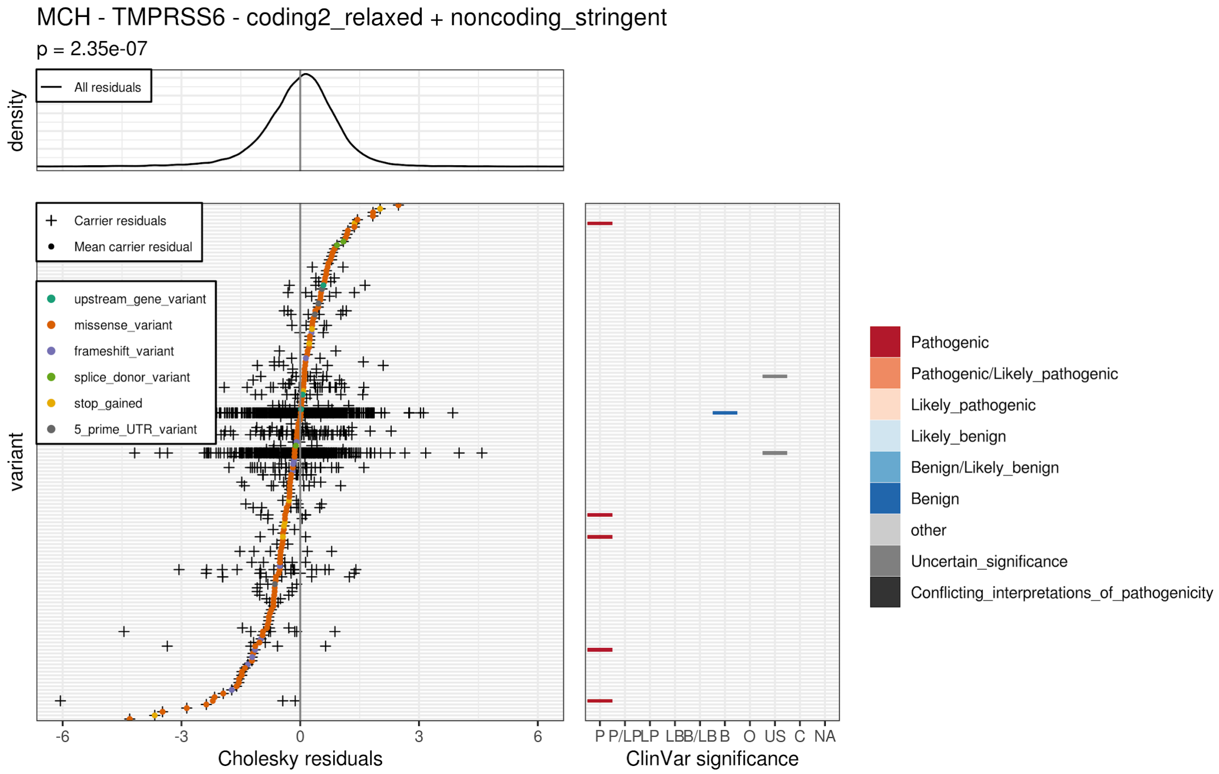
(C17)


MCH - TMPRSS6 - coding1_stringent
p = 4.91e-11
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(C18)

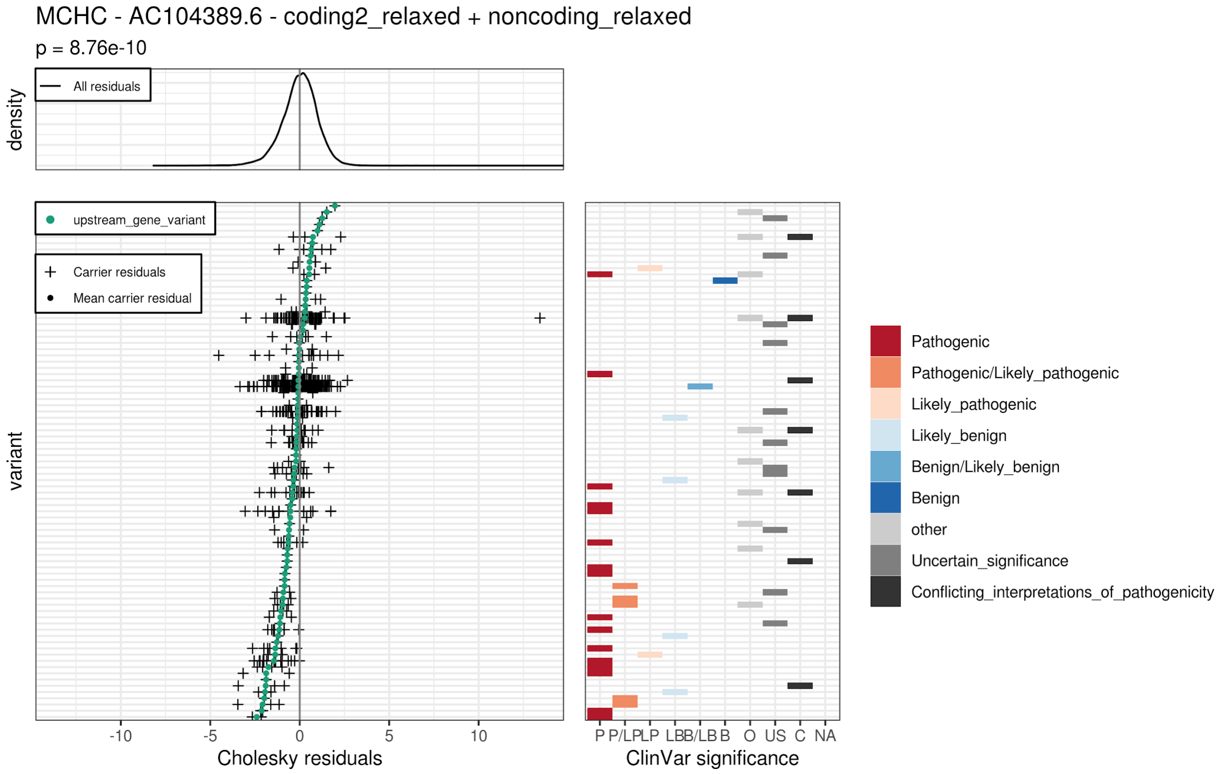

MCH - TMPRSS6 - coding1_relaxed
p = 1.32e-11
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(C19)

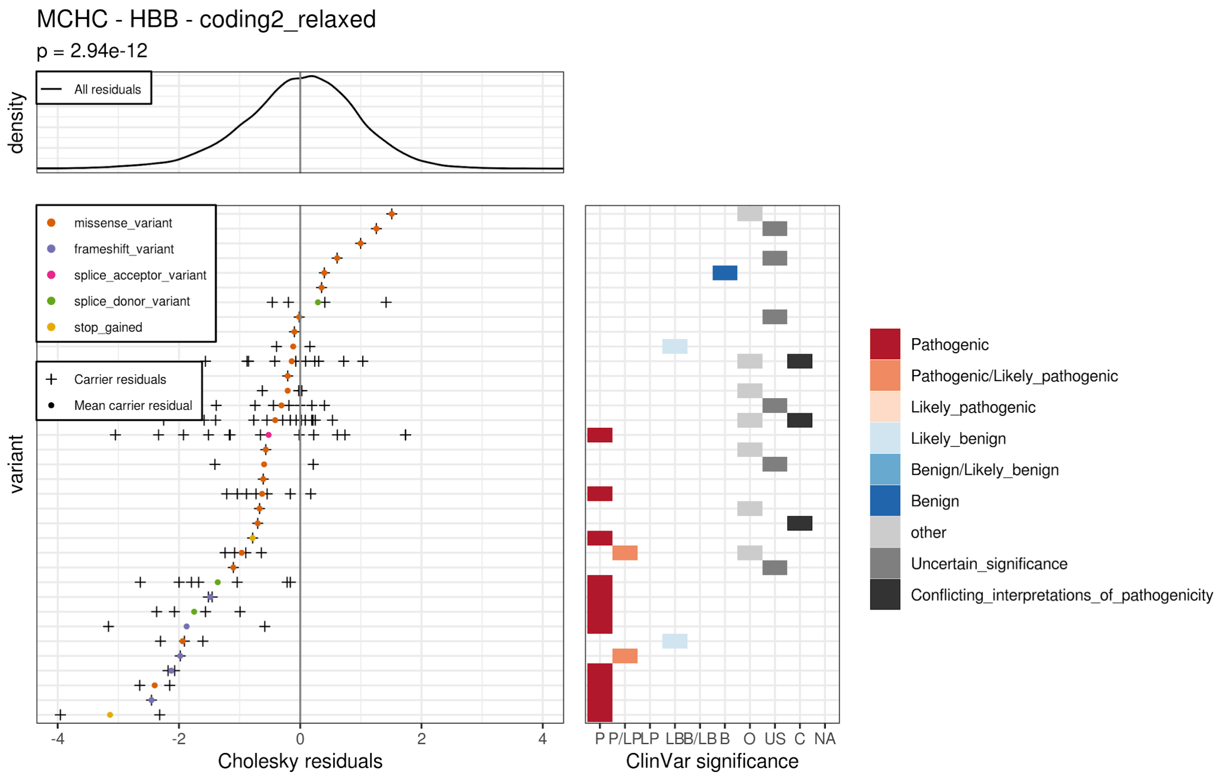

MCH - TMPRSS6 - coding2_relaxed + noncoding_stringent
p = 2.35e-07
density
All residuals
variant
Carrier residuals
Mean carrier residual
upstream_gene_variant
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
5_prime_UTR_variant
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(D)

(D1)

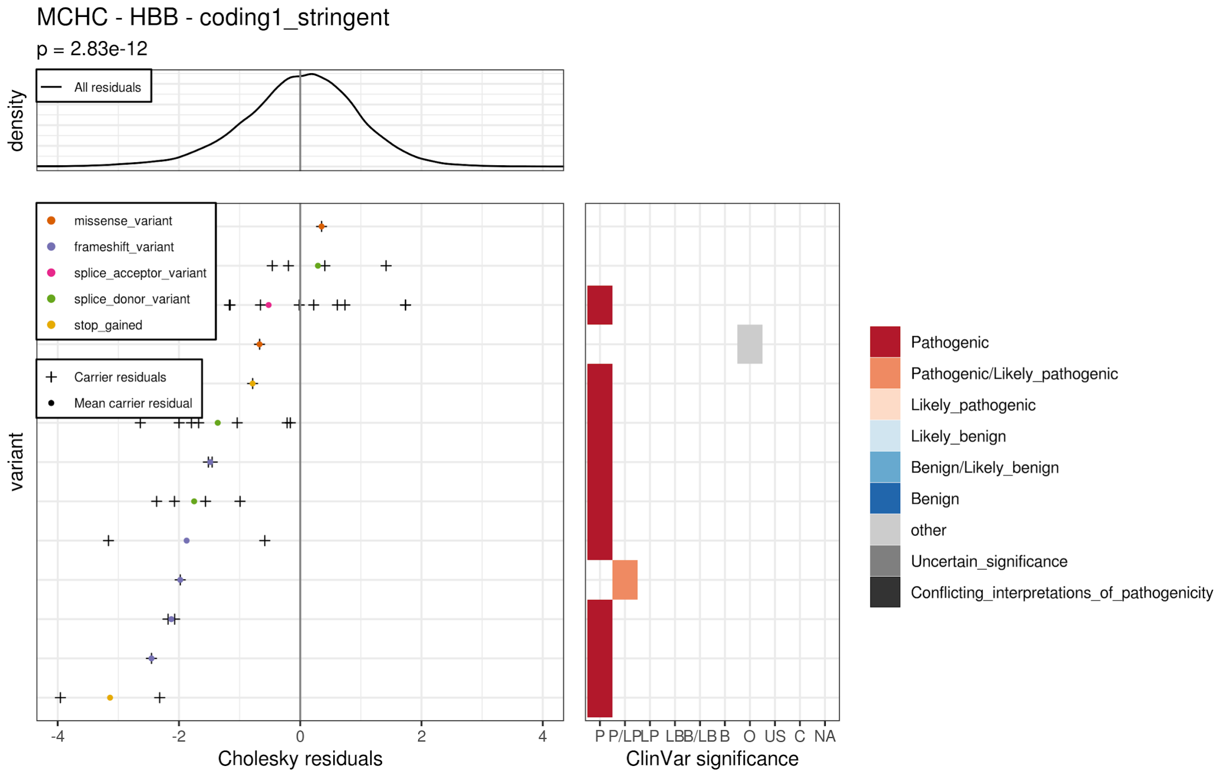

MCHC - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 8.76e-10
density
All residuals
upstream_gene_variant
Carrier residuals
Mean carrier residual
variant
-10
-5
0
5
10
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(D2)



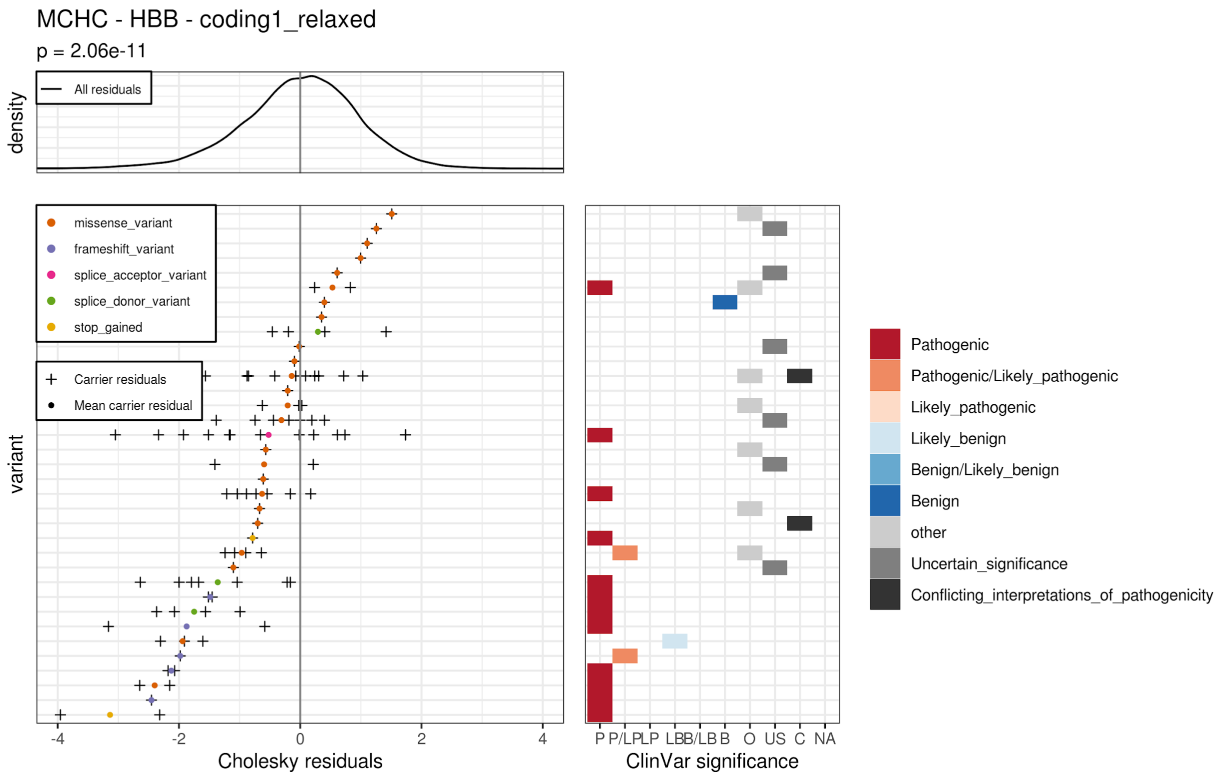
MCHC - HBB - coding2_relaxed
p = 2.94e-12
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(D3)

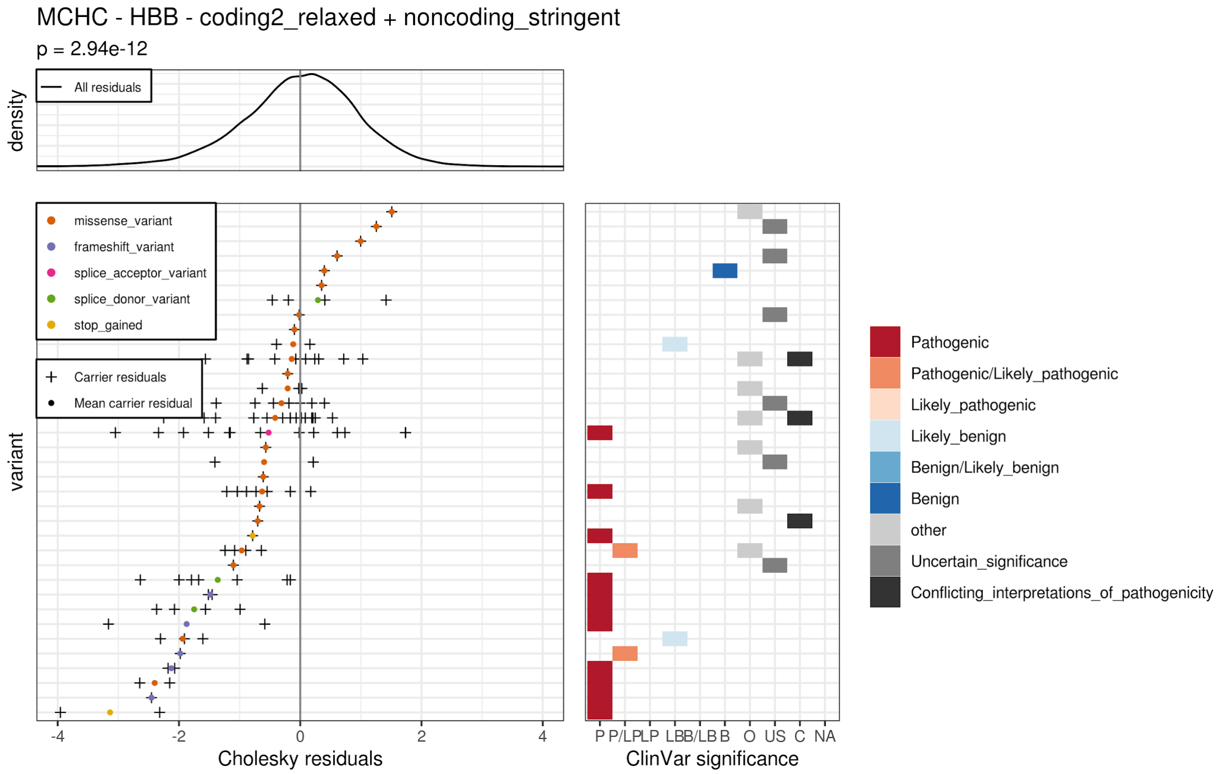


MCHC - HBB - coding1_stringent
p = 2.83e-12
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(D4)

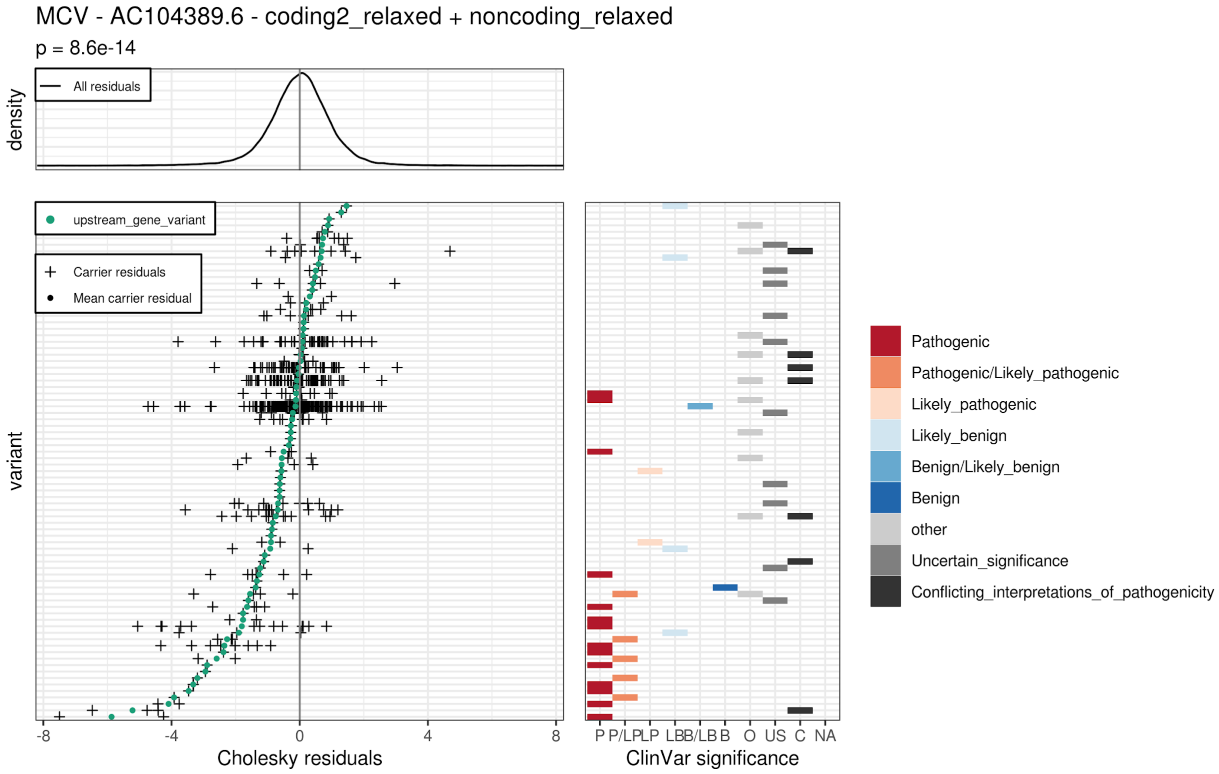

MCHC - HBB - coding1_relaxed
p = 2.06e-11
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(D5)

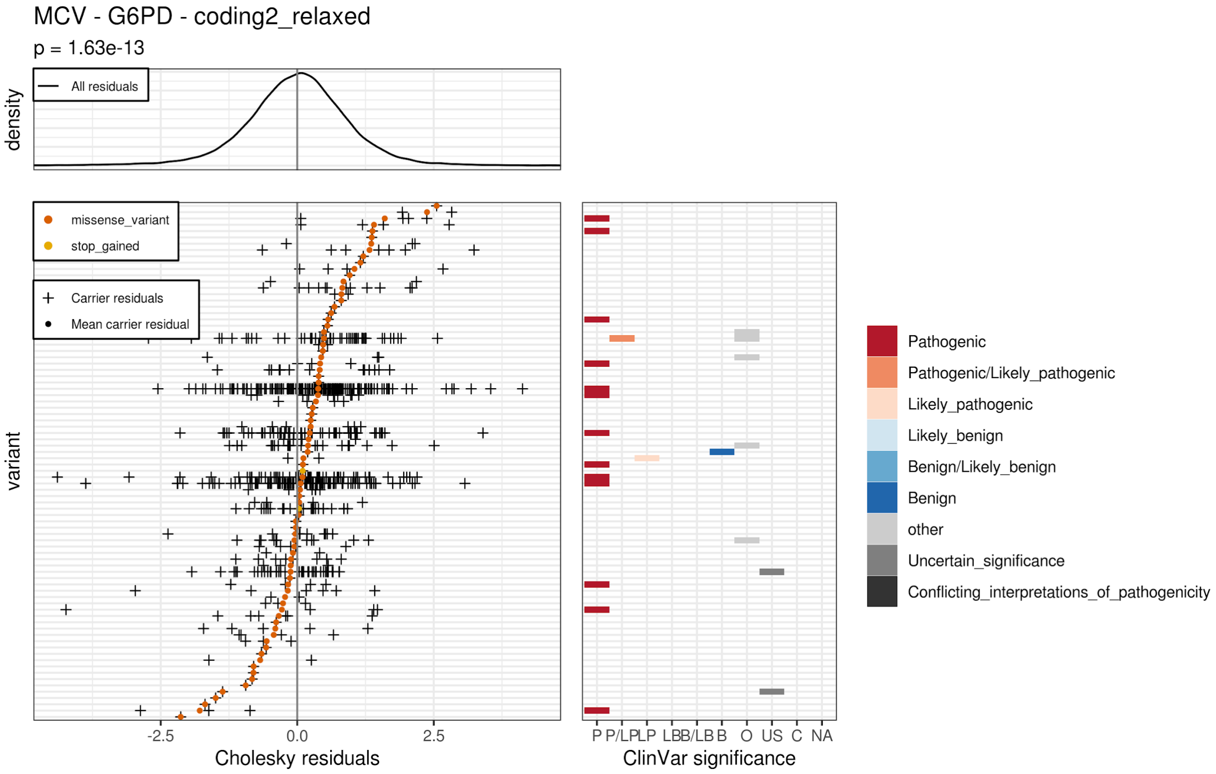

MCHC - HBB - coding2_relaxed + noncoding_stringent
p = 2.94e-12
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(E)

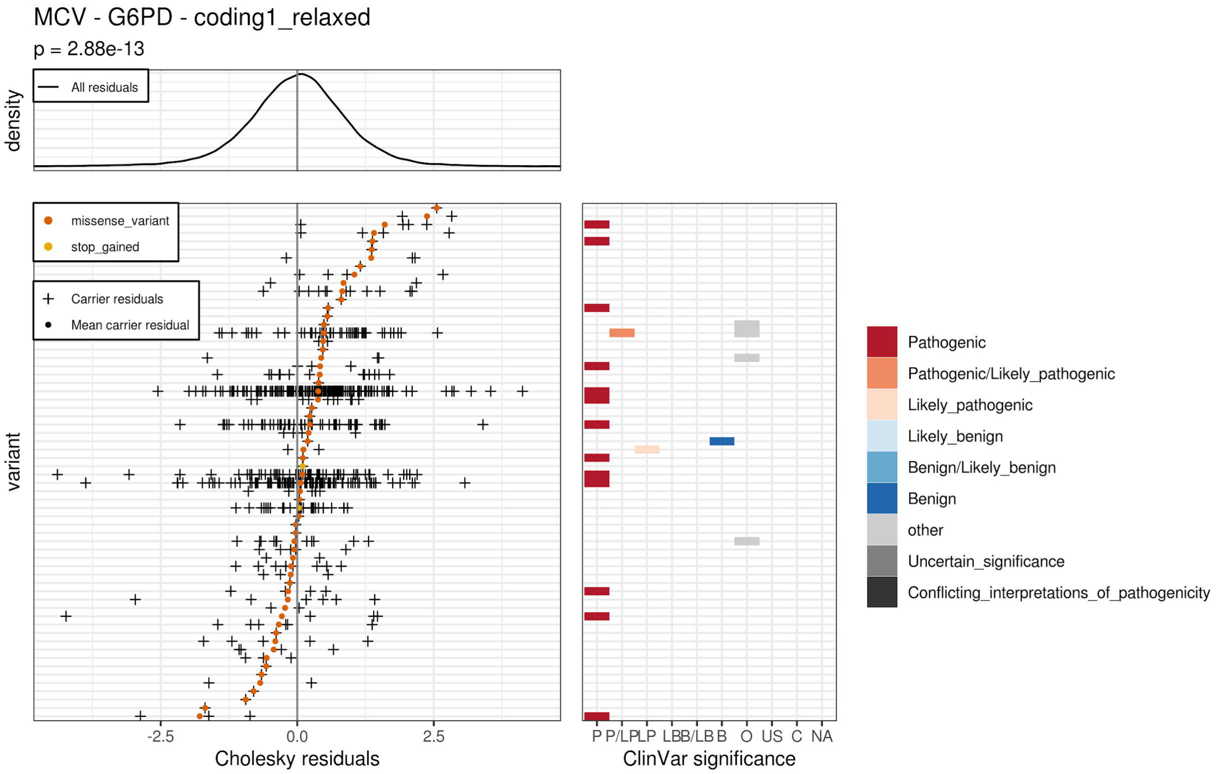
(E1)


MCV - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 8.6e-14
density
All residuals
upstream_gene_variant
Carrier residuals
Mean carrier residual
variant
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

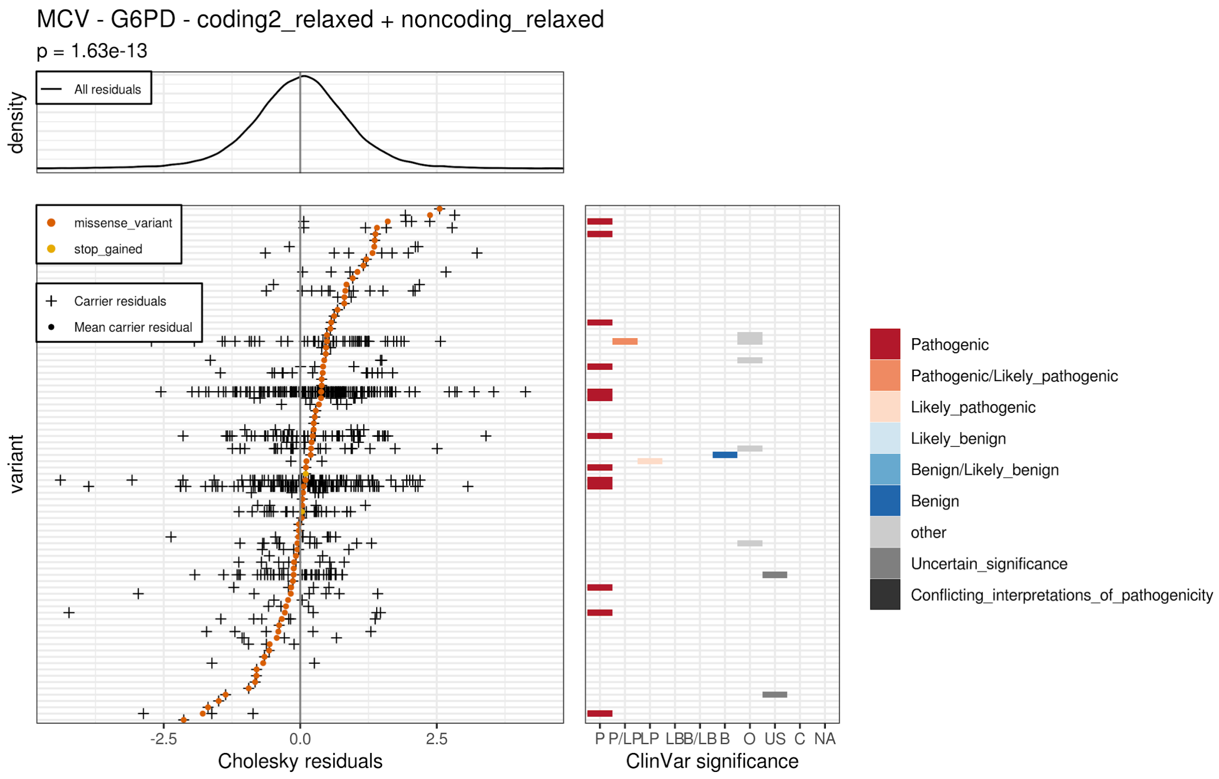
(E2)


MCV - G6PD - coding2_relaxed
p = 1.63e-13
density
All residuals
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(E3)


MCV - G6PD - coding1_relaxed
p = 2.88e-13
density
All residuals
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2.5
0.0
2.5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

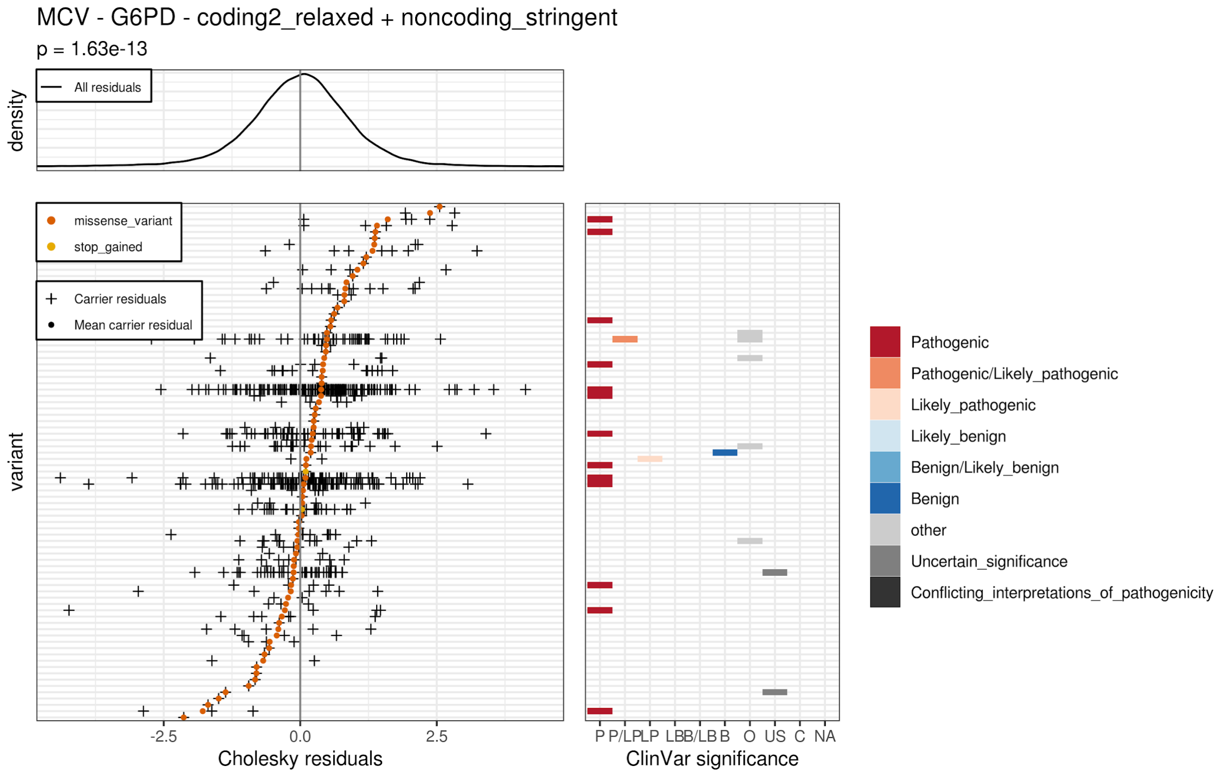
(E4)



MCV - G6PD - coding2_relaxed + noncoding_relaxed

p = 1.63e-13

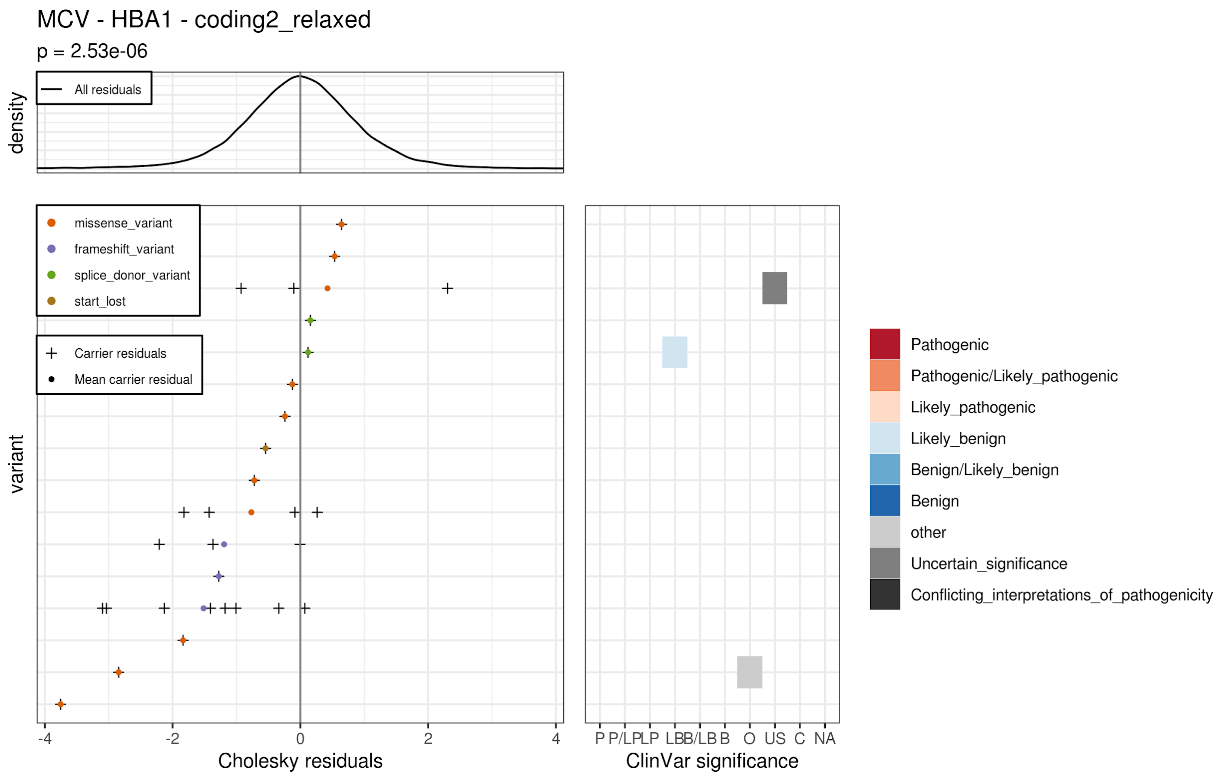
(E5)



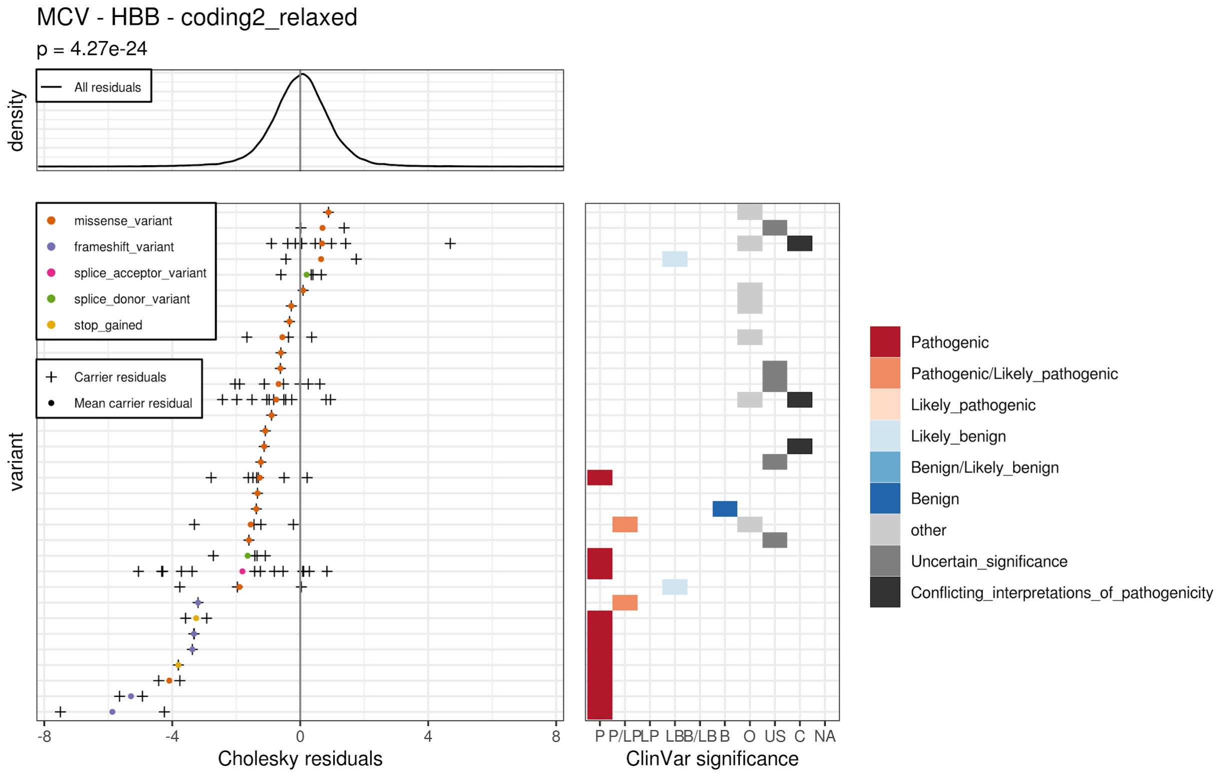
MCV - G6PD - coding2_relaxed + noncoding_stringent

p = 1.63e-13

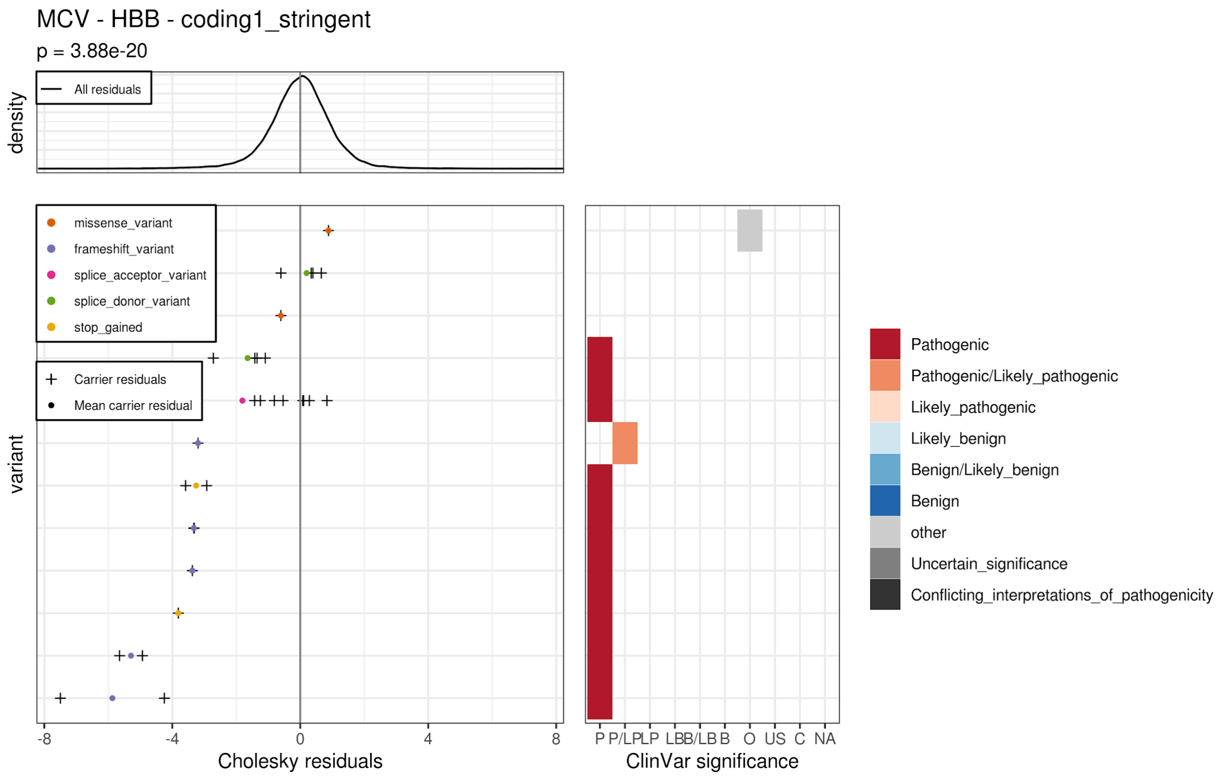
(E6)


MCV - HBA1 - coding2_relaxed
p = 2.53e-06
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
start_lost
Carrier residuals
Mean carrier residual
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

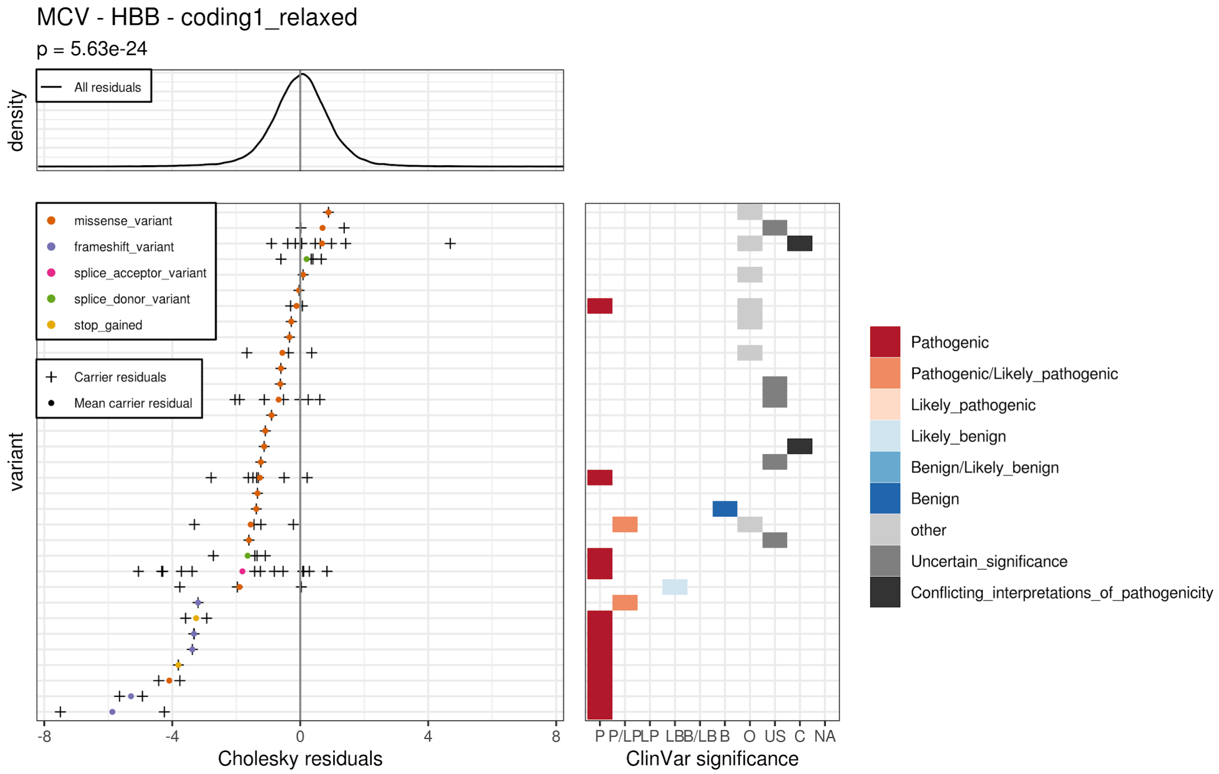
(E7)


MCV - HBB - coding2_relaxed
p = 4.27e-24
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

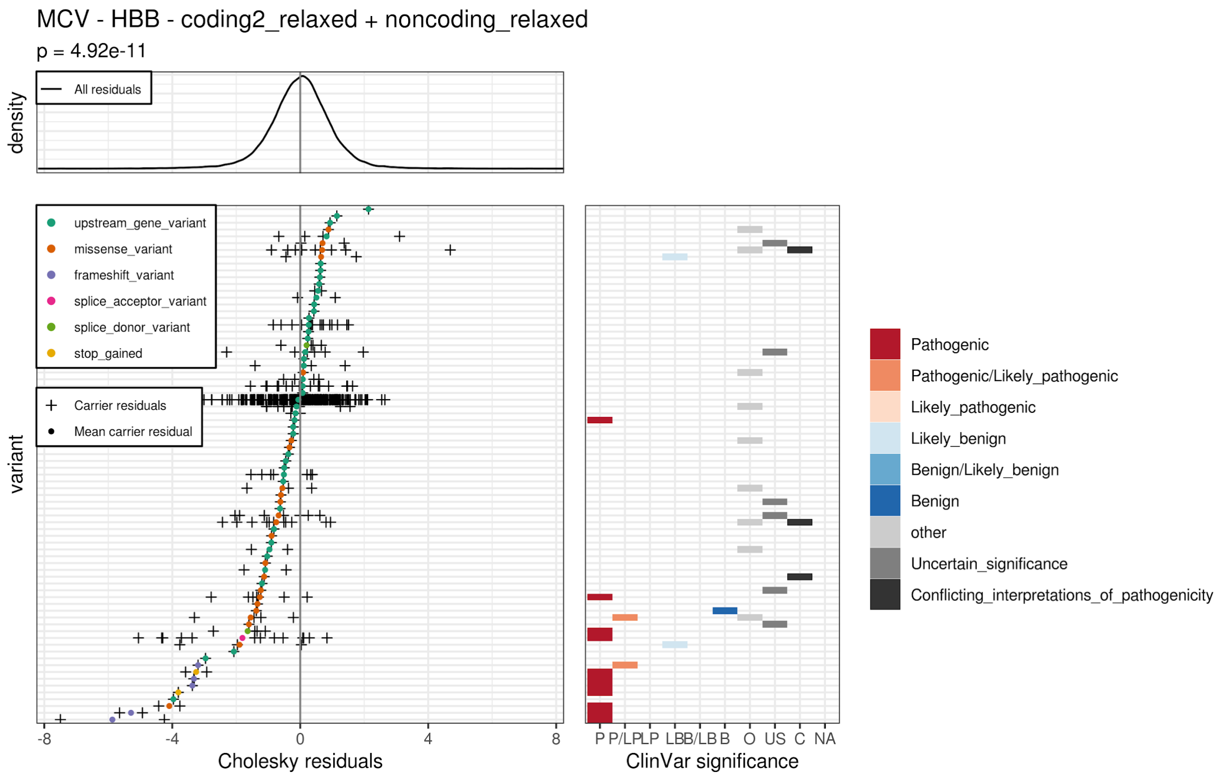
(E8)


MCV - HBB - coding1_stringent
p = 3.88e-20
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

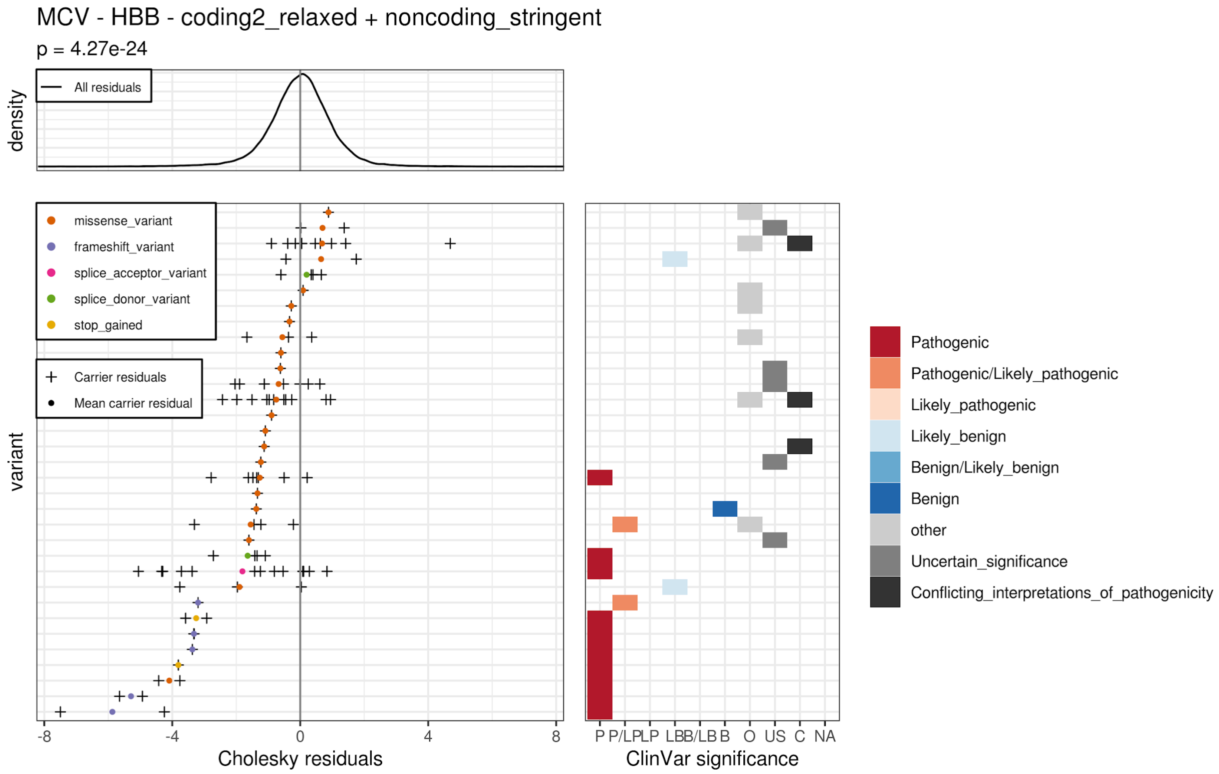
(E9)


MCV - HBB - coding1_relaxed
p = 5.63e-24
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

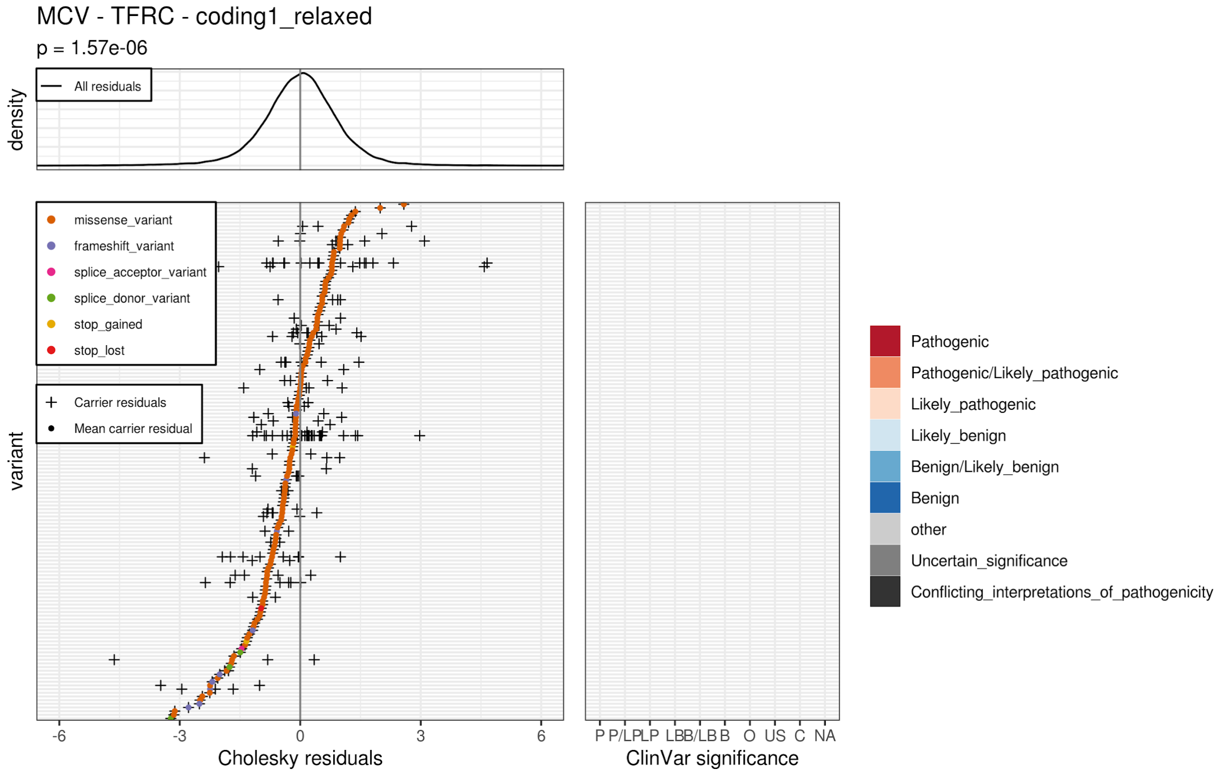
(E10)


MCV - HBB - coding2_relaxed + noncoding_relaxed
p = 4.92e-11
density
All residuals
variant
upstream_gene_variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(E11)


MCV - HBB - coding2_relaxed + noncoding_stringent
p = 4.27e-24
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

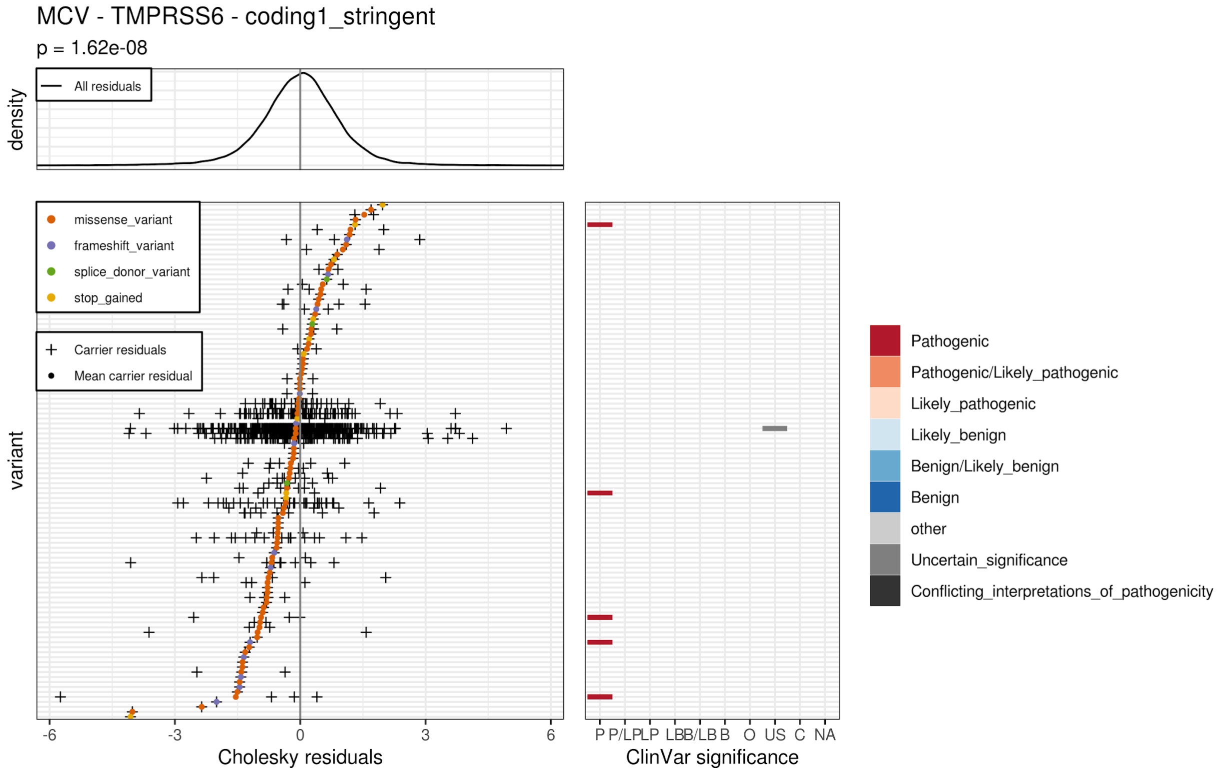
(E12)



MCV - TFRC - coding1_relaxed

p = 1.57e-06

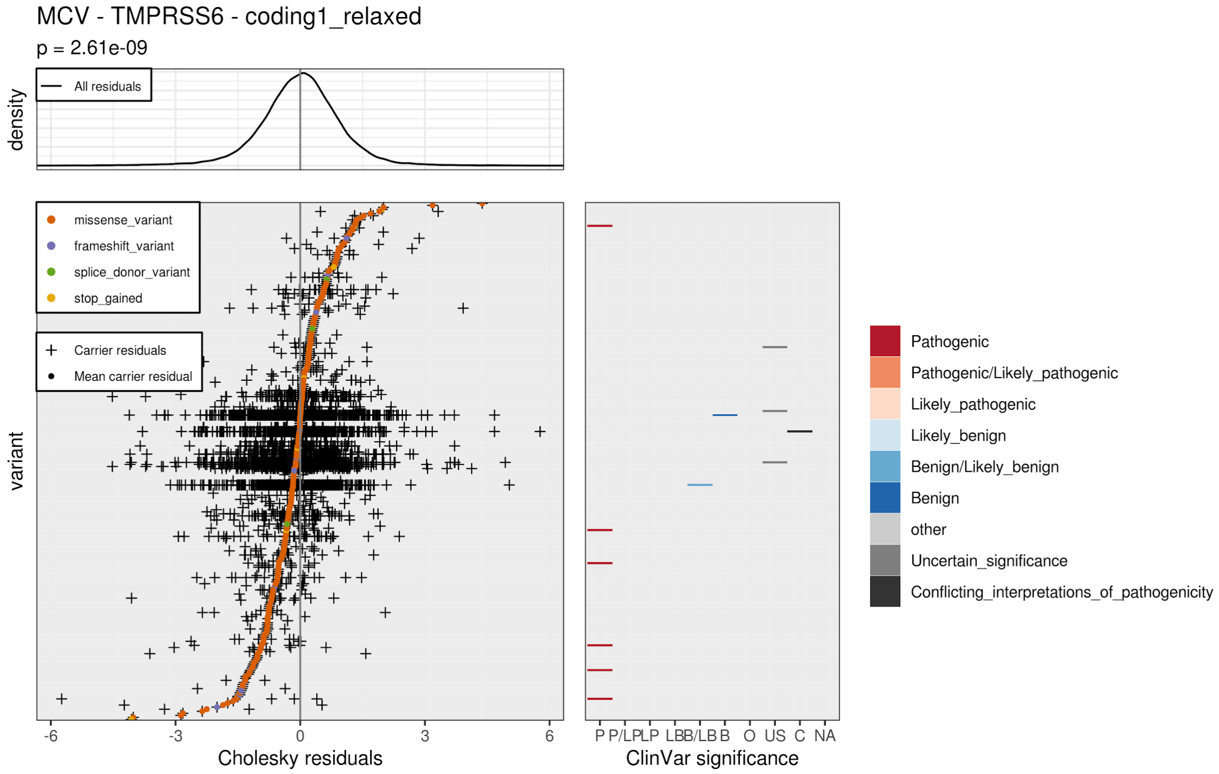
(E13)



MCV - TMPRSS6 - coding1_stringent

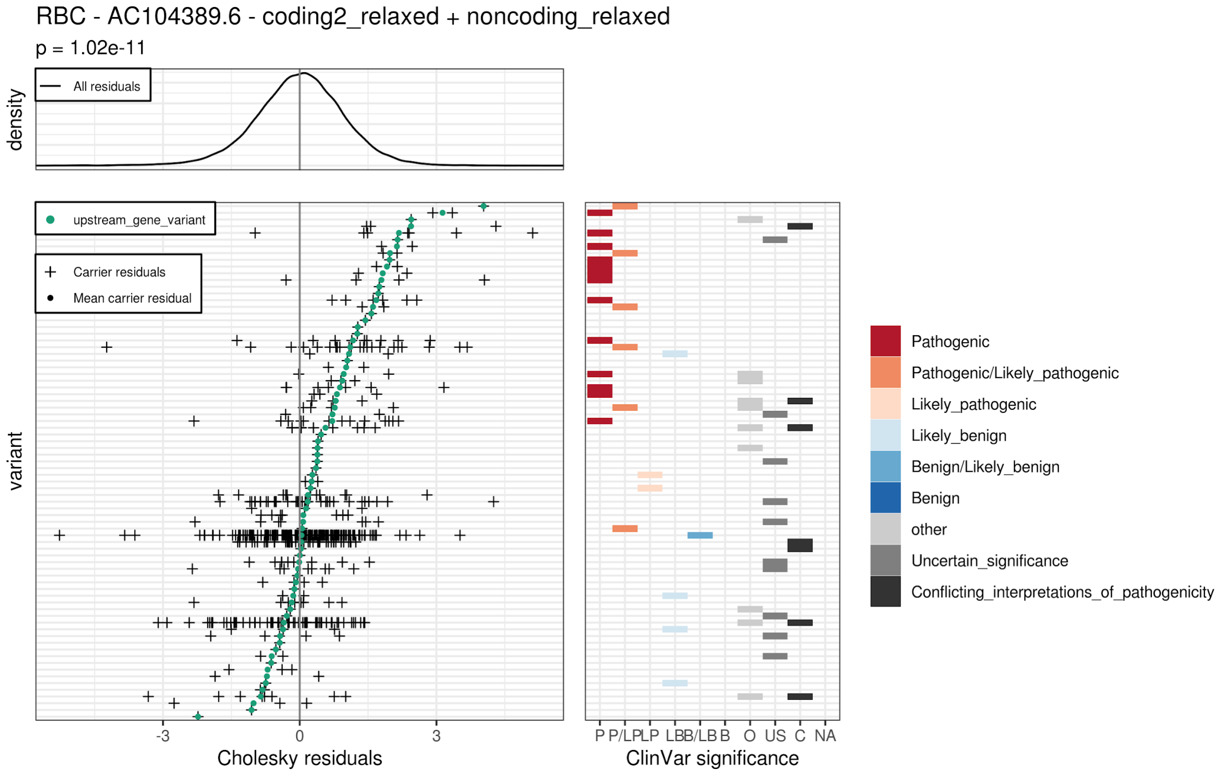
p = 1.62e-08

(E14)

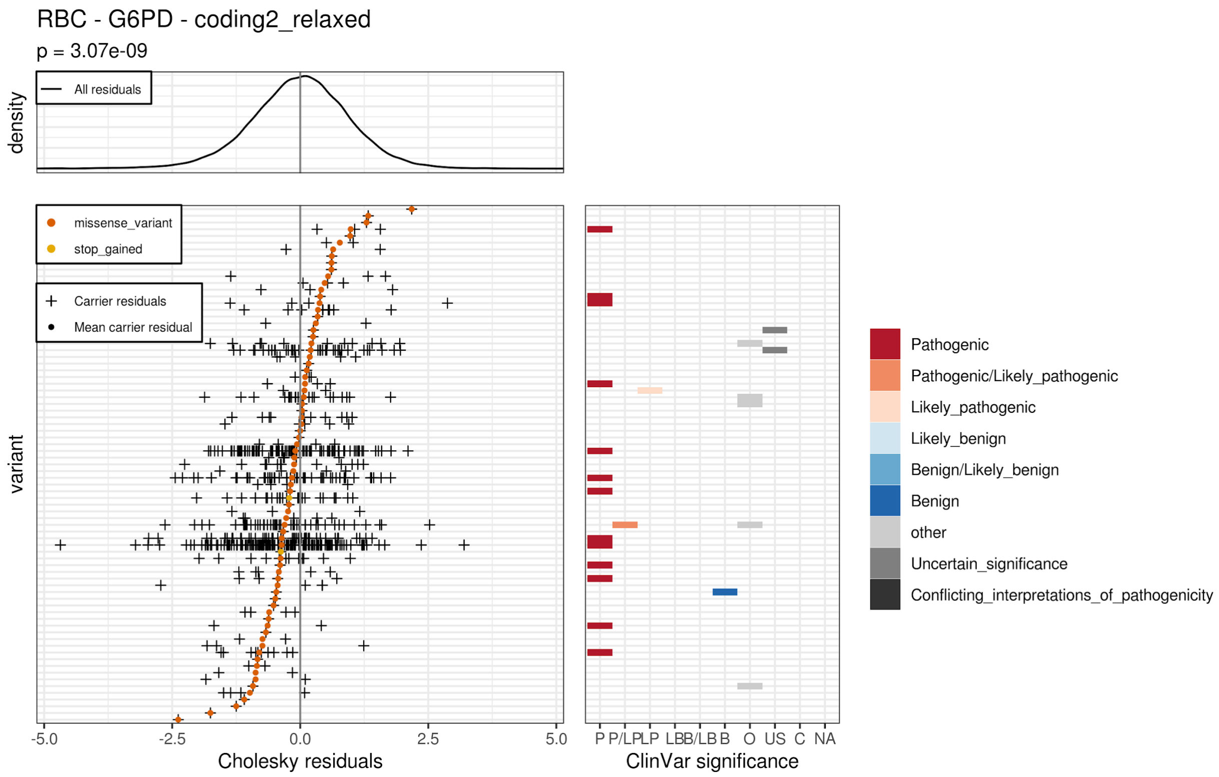

MCV - TMPRSS6 - coding1_relaxed
p = 2.61e-09
density
All residuals
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-6
-3
0
3
6
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(F)

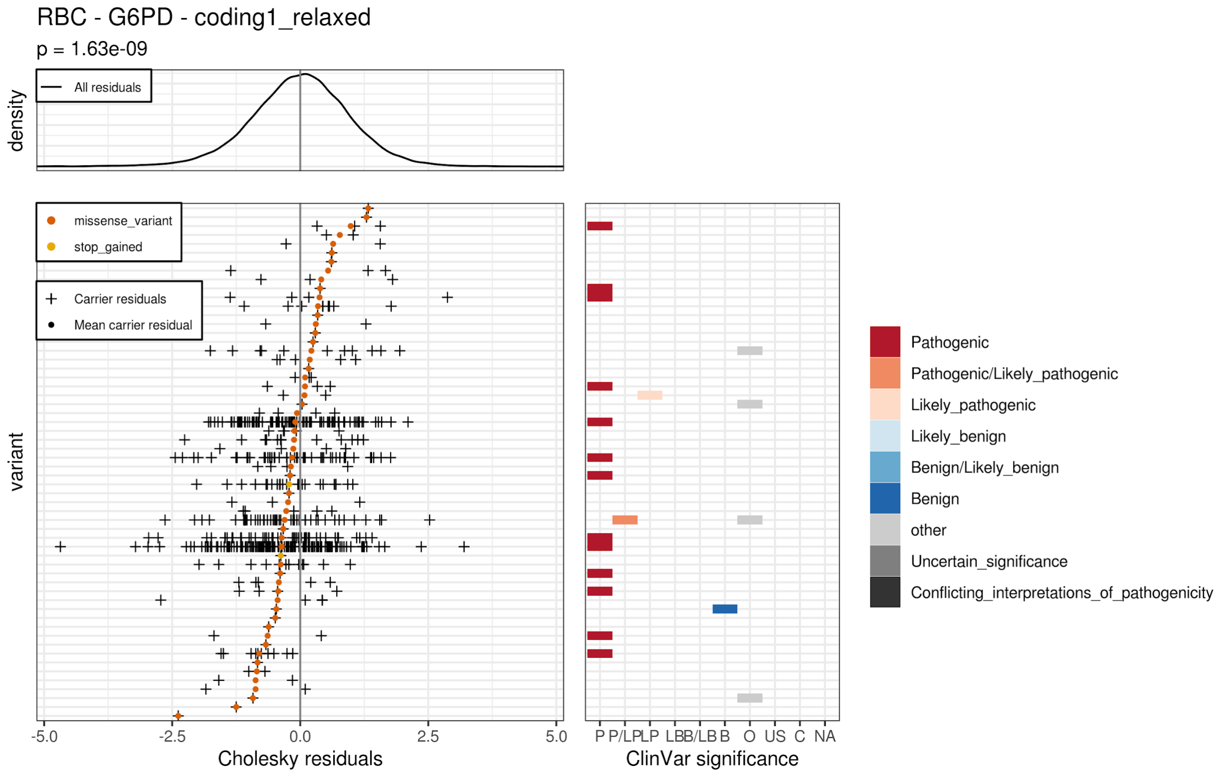
(F1)


RBC - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 1.02e-11
density
All residuals
upstream_gene_variant
Carrier residuals
Mean carrier residual
variant
-3
0
3
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

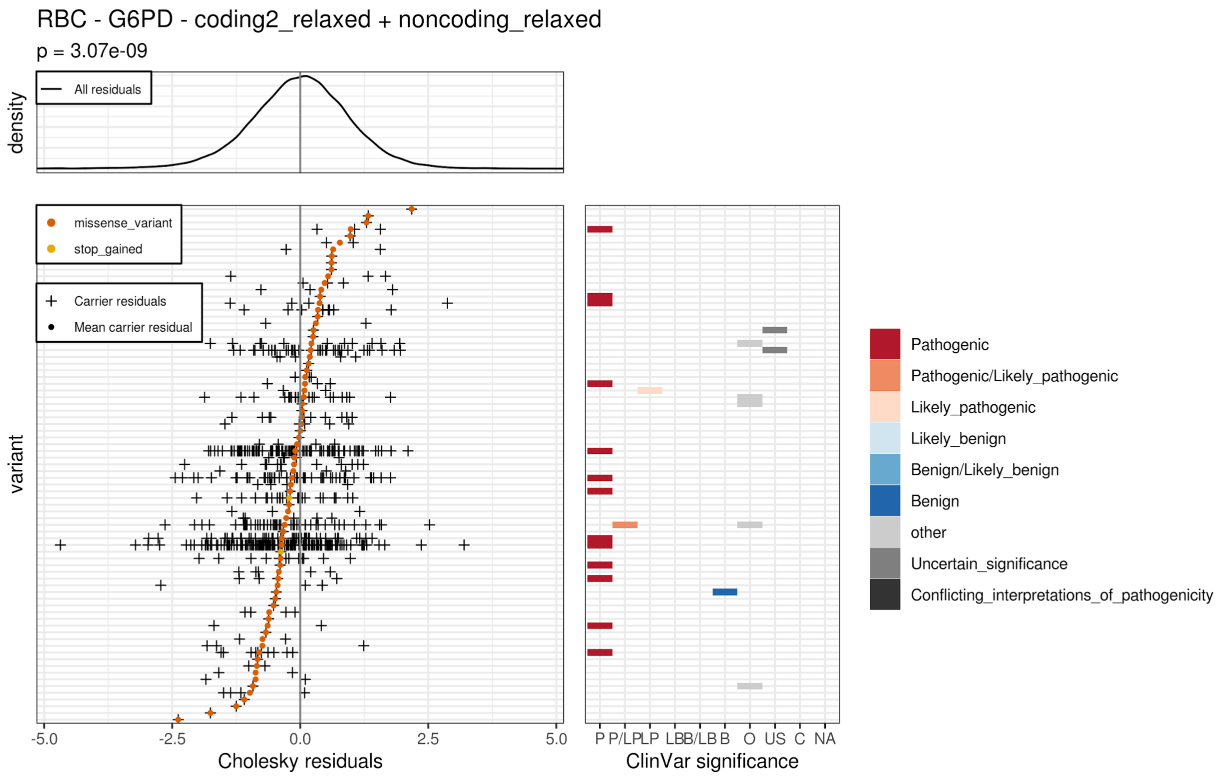
(F2)


RBC - G6PD - coding2_relaxed
p = 3.07e-09
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

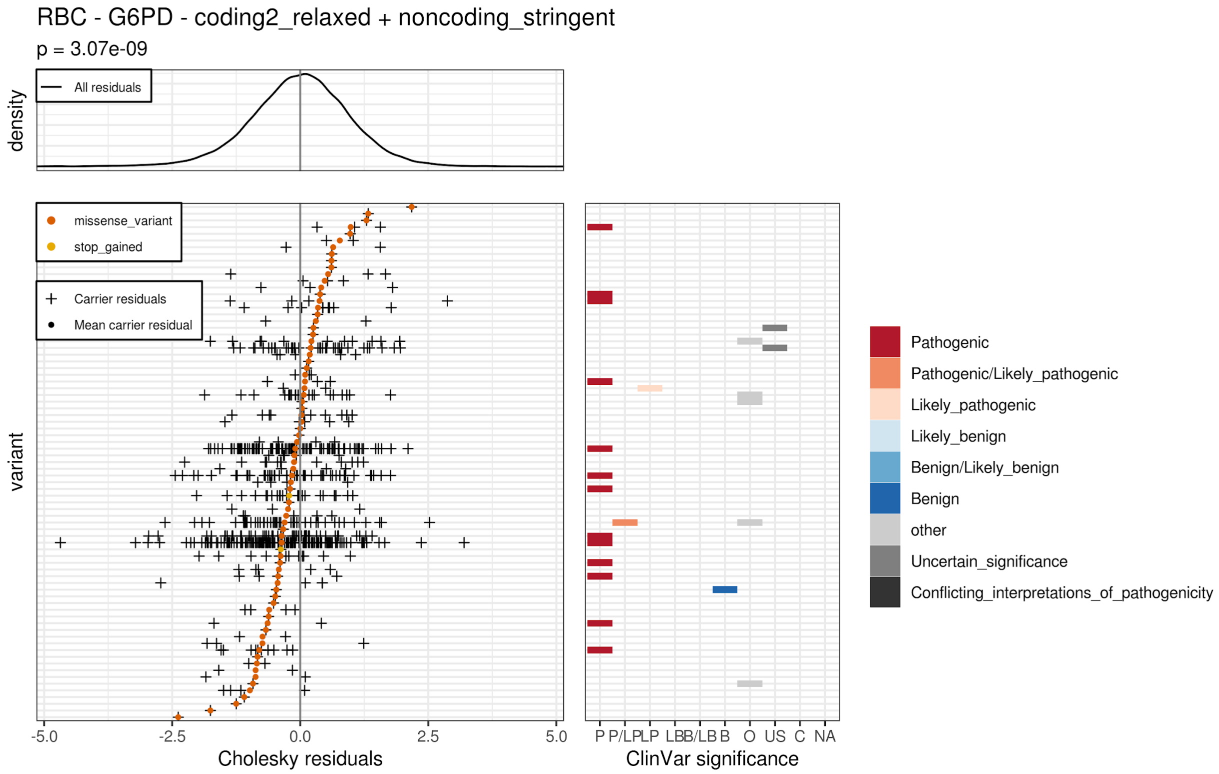
(F3)


RBC - G6PD - coding1_relaxed
p = 1.63e-09
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

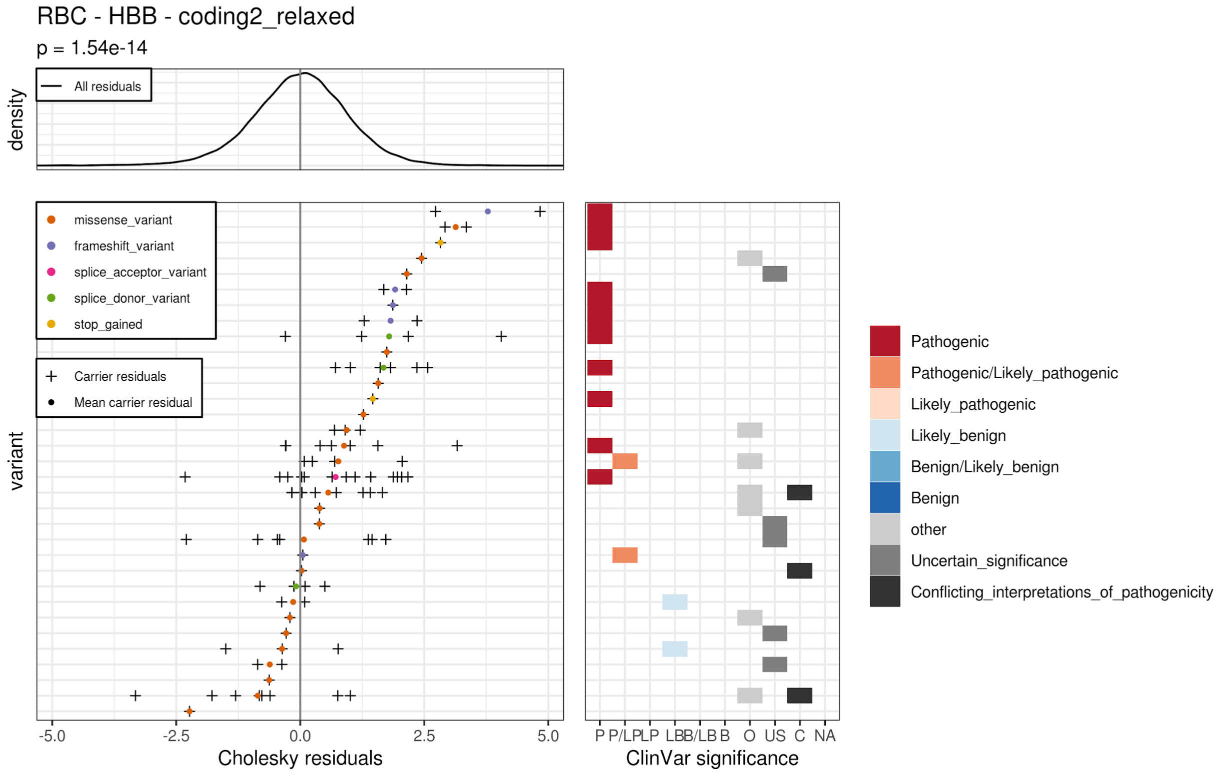
(F4)


RBC - G6PD - coding2_relaxed + noncoding_relaxed
p = 3.07e-09
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

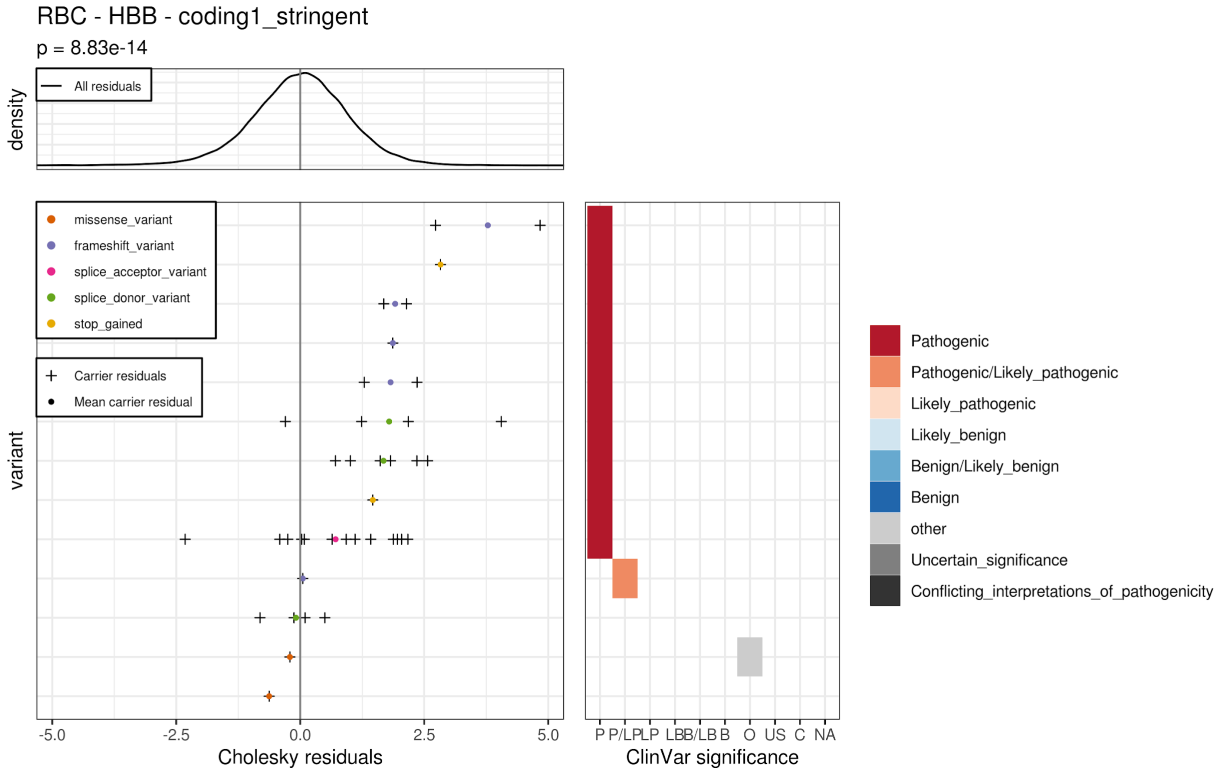
(F5)


RBC - G6PD - coding2_relaxed + noncoding_stringent
p = 3.07e-09
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

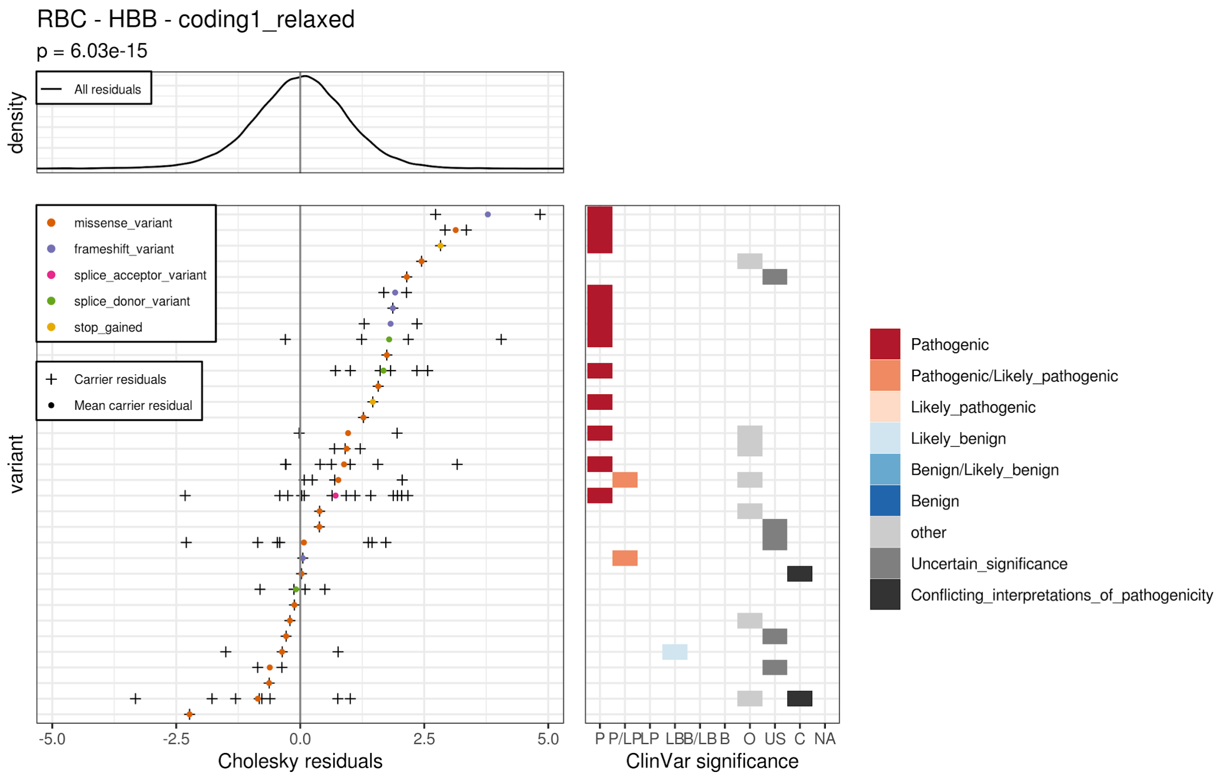
(F6)


RBC - HBB - coding2_relaxed
p = 1.54e-14
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

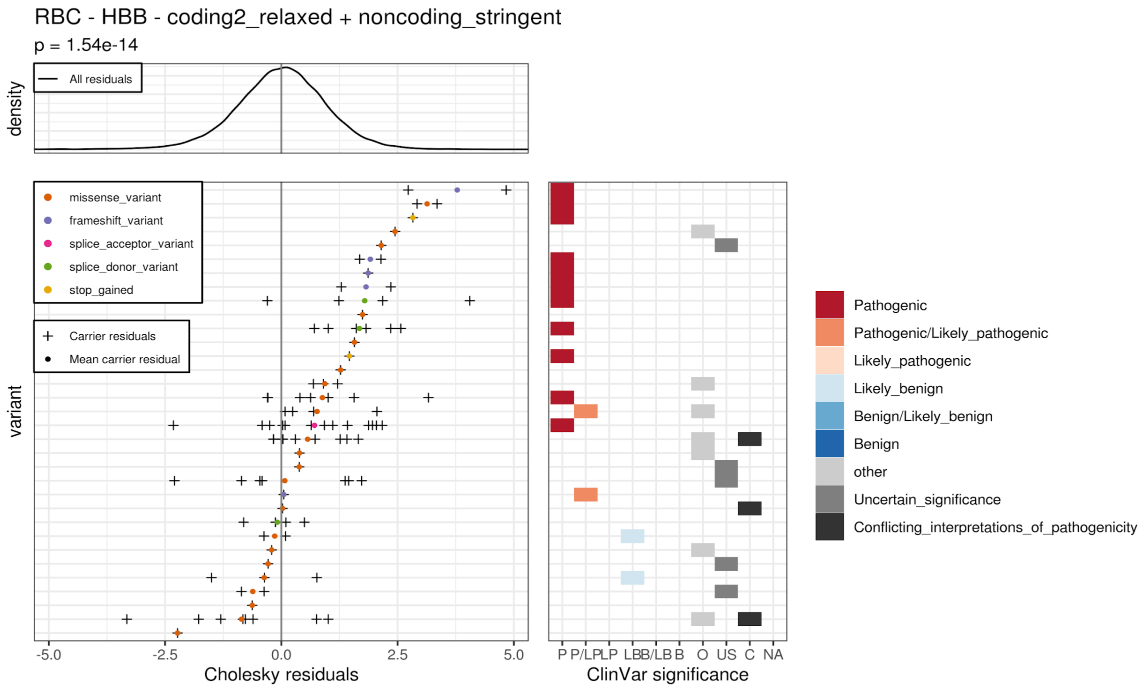
(F7)


RBC - HBB - coding1_stringent
p = 8.83e-14
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(F8)

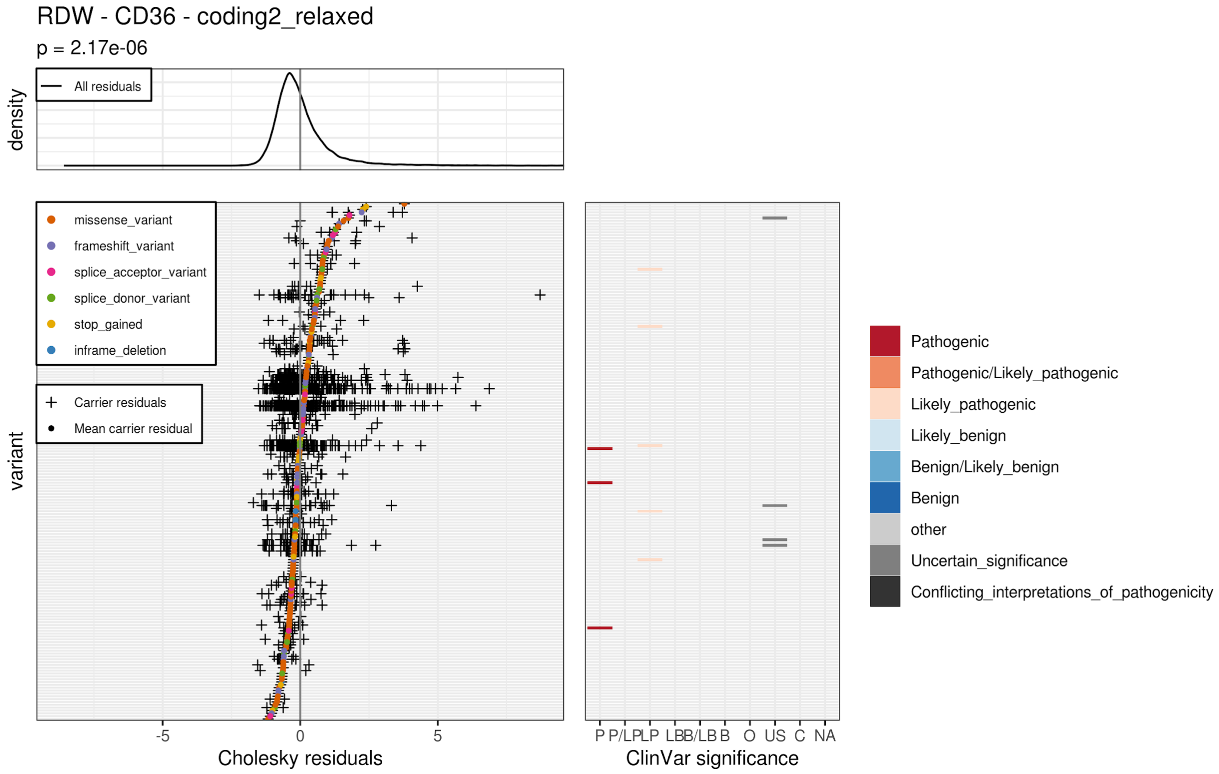

RBC - HBB - coding1_relaxed
p = 6.03e-15
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(F9)


RBC - HBB - coding2_relaxed + noncoding_stringent
p = 1.54e-14
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-5.0
-2.5
0.0
2.5
5.0
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

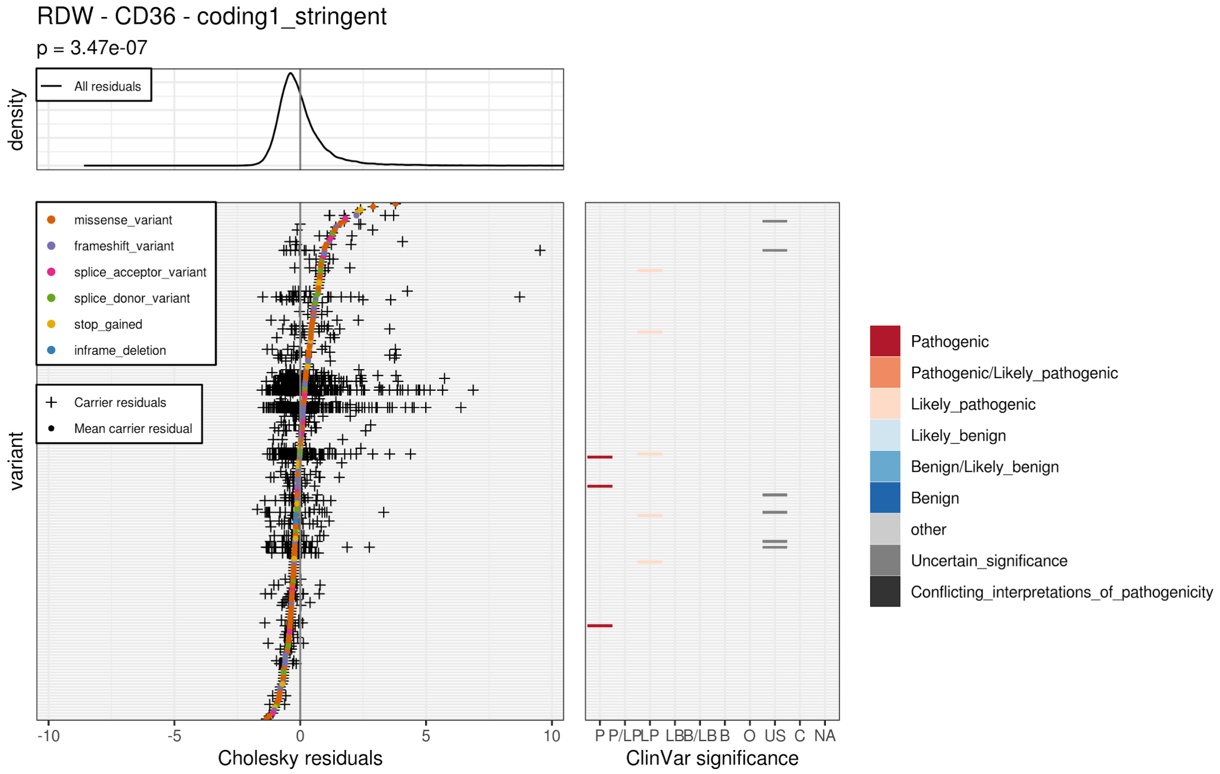
(G)

(G1)

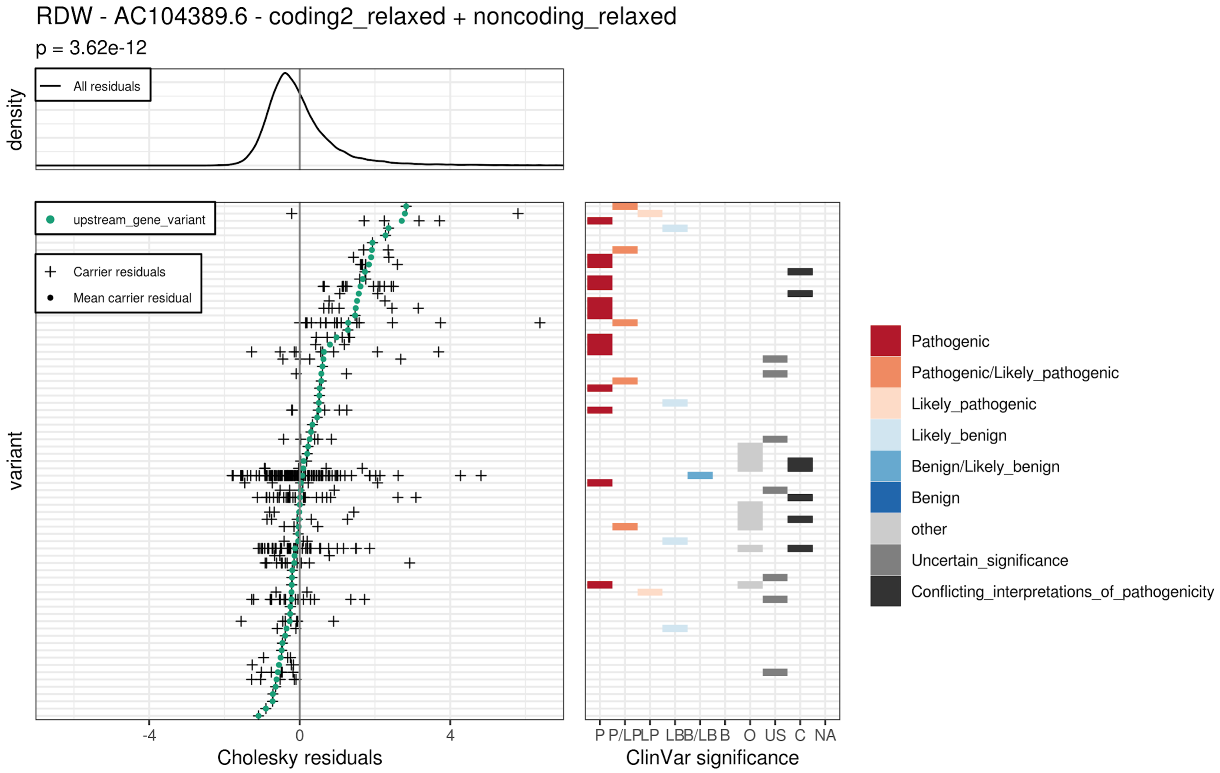


RDW - CD36 - coding2_relaxed
p = 2.17e-06
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
inframe_deletion
Carrier residuals
Mean carrier residual
variant
-5
0
5
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

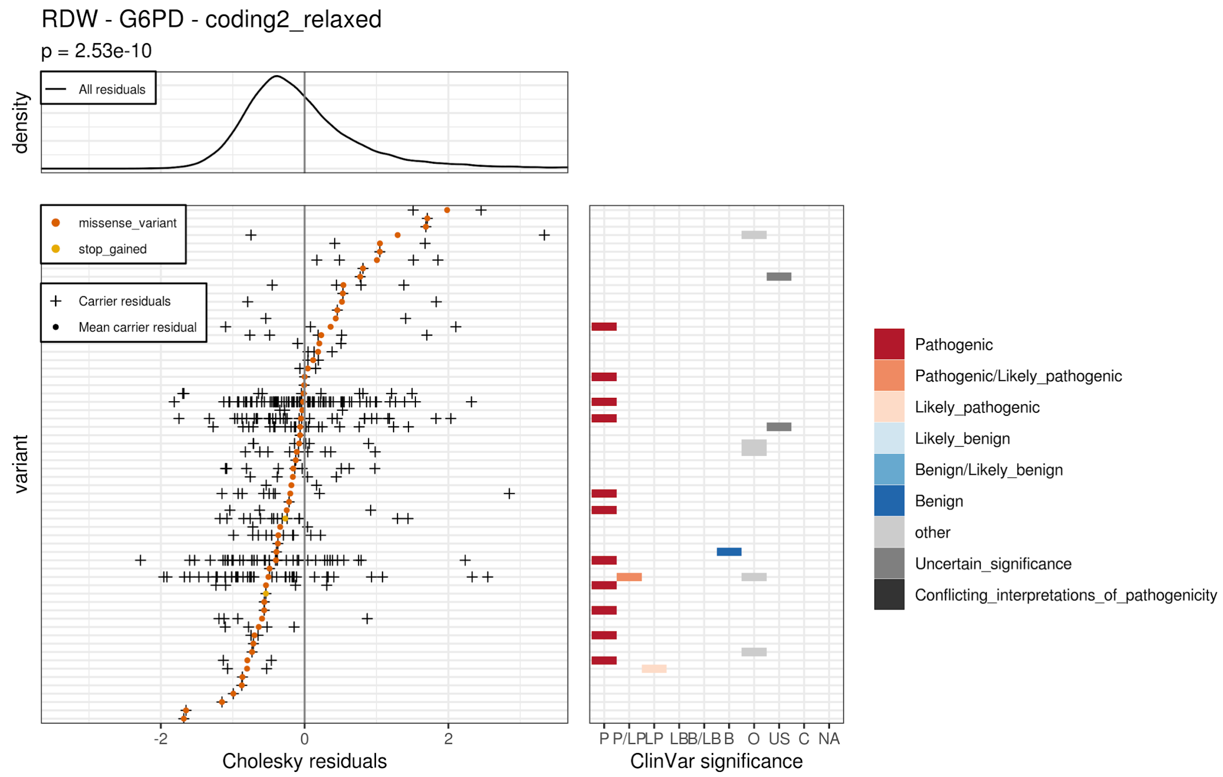
(G2)


RDW - CD36 - coding1_stringent
p = 3.47e-07
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
inframe_deletion
Carrier residuals
Mean carrier residual
-10
-5
0
5
10
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

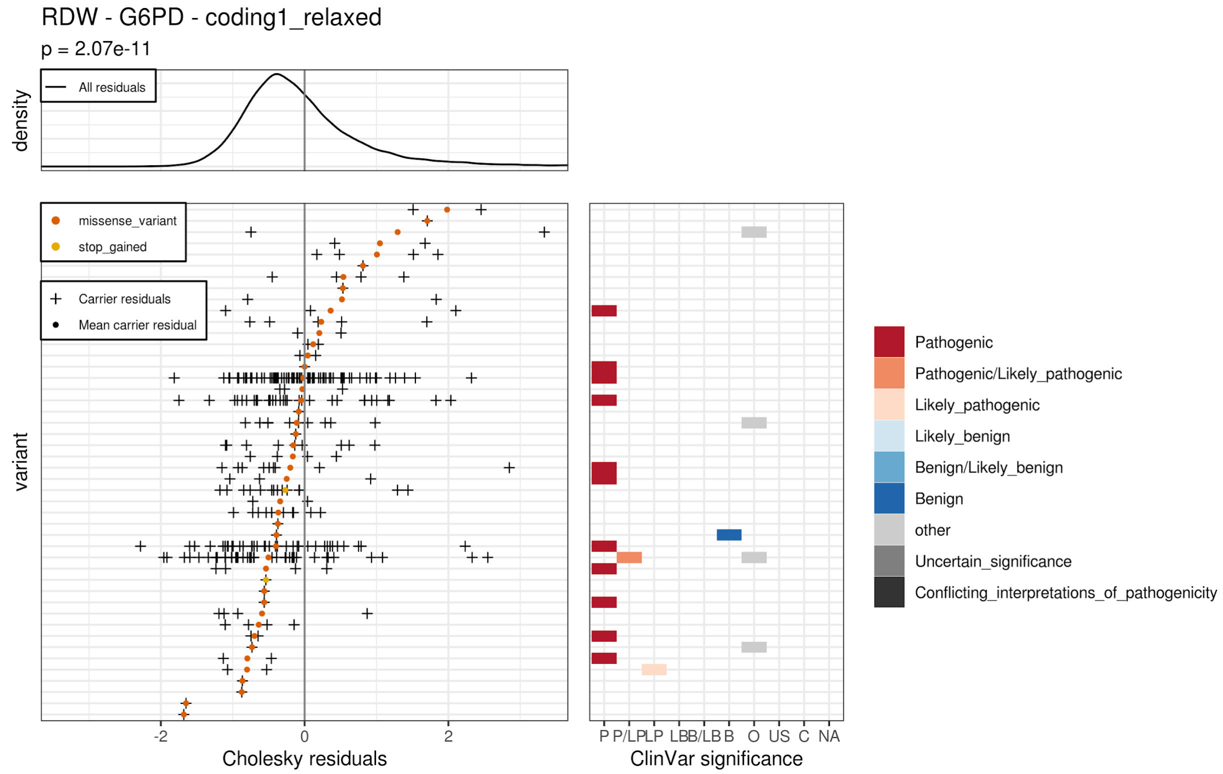
(G3)


RDW - AC104389.6 - coding2_relaxed + noncoding_relaxed
p = 3.62e-12
density
All residuals
variant
upstream_gene_variant
Carrier residuals
Mean carrier residual
-4
0
4
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

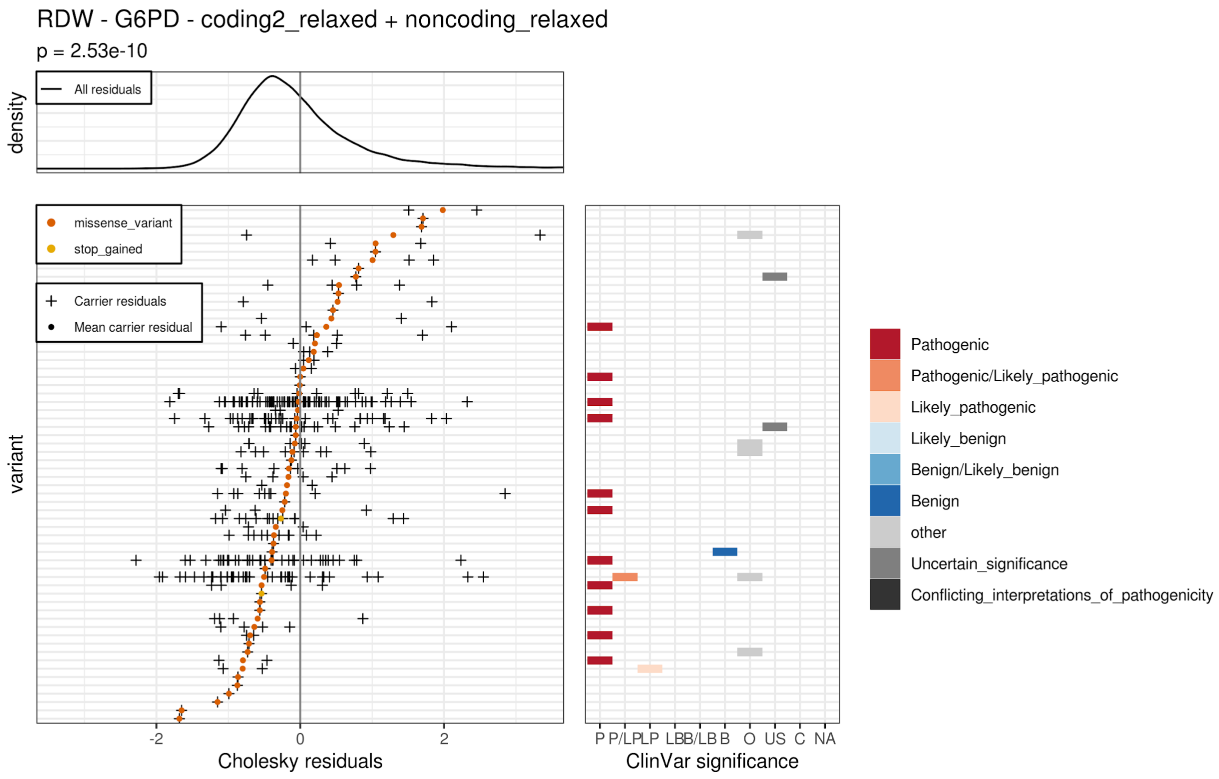
(G4)


RDW - G6PD - coding2_relaxed
p = 2.53e-10
All residuals
density
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2
0
2
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

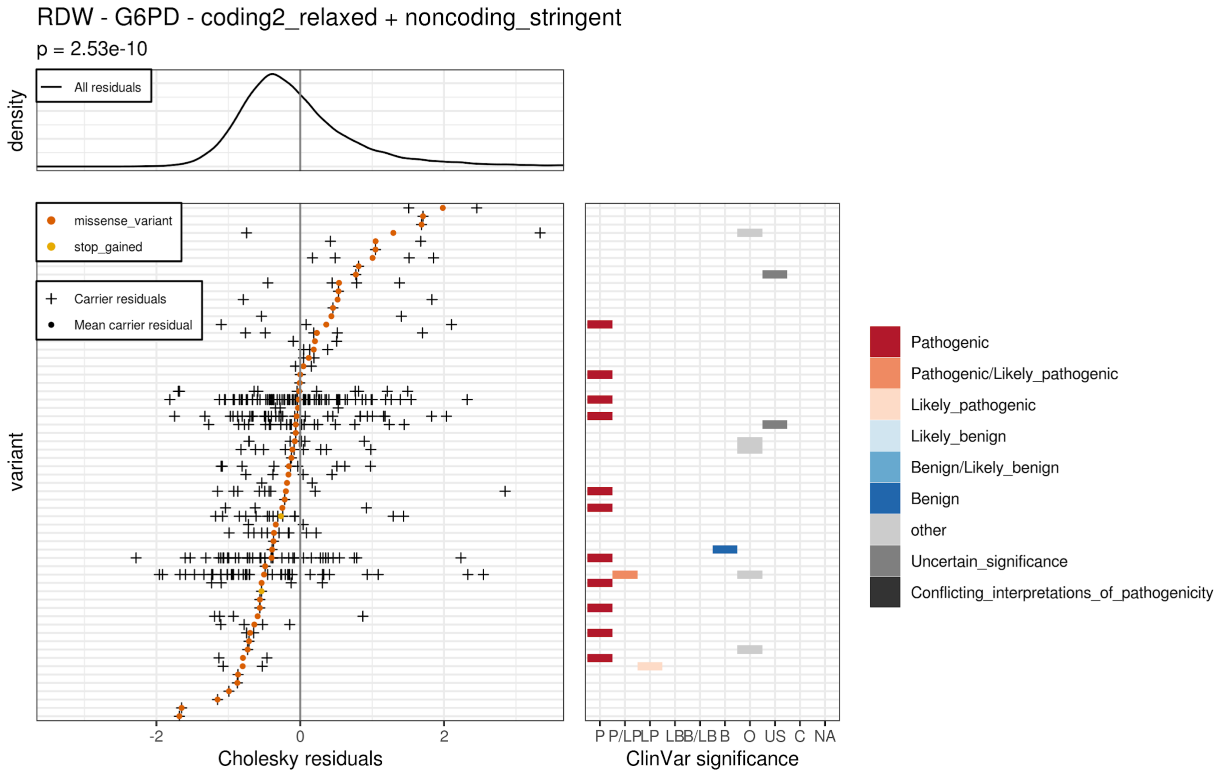
(G5)


RDW - G6PD - coding1_relaxed
p = 2.07e-11
All residuals
density
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-2
0
2
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

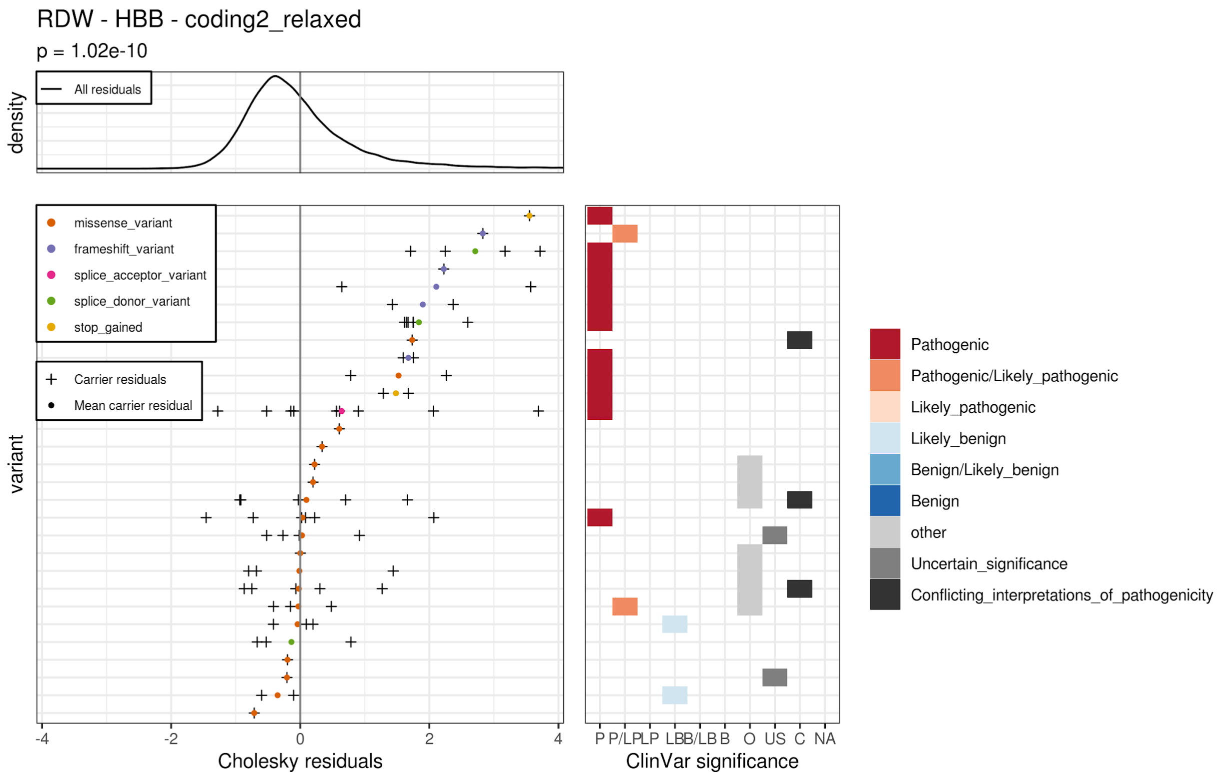
(G6)


RDW - G6PD - coding2_relaxed + noncoding_relaxed
p = 2.53e-10
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-2
0
2
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

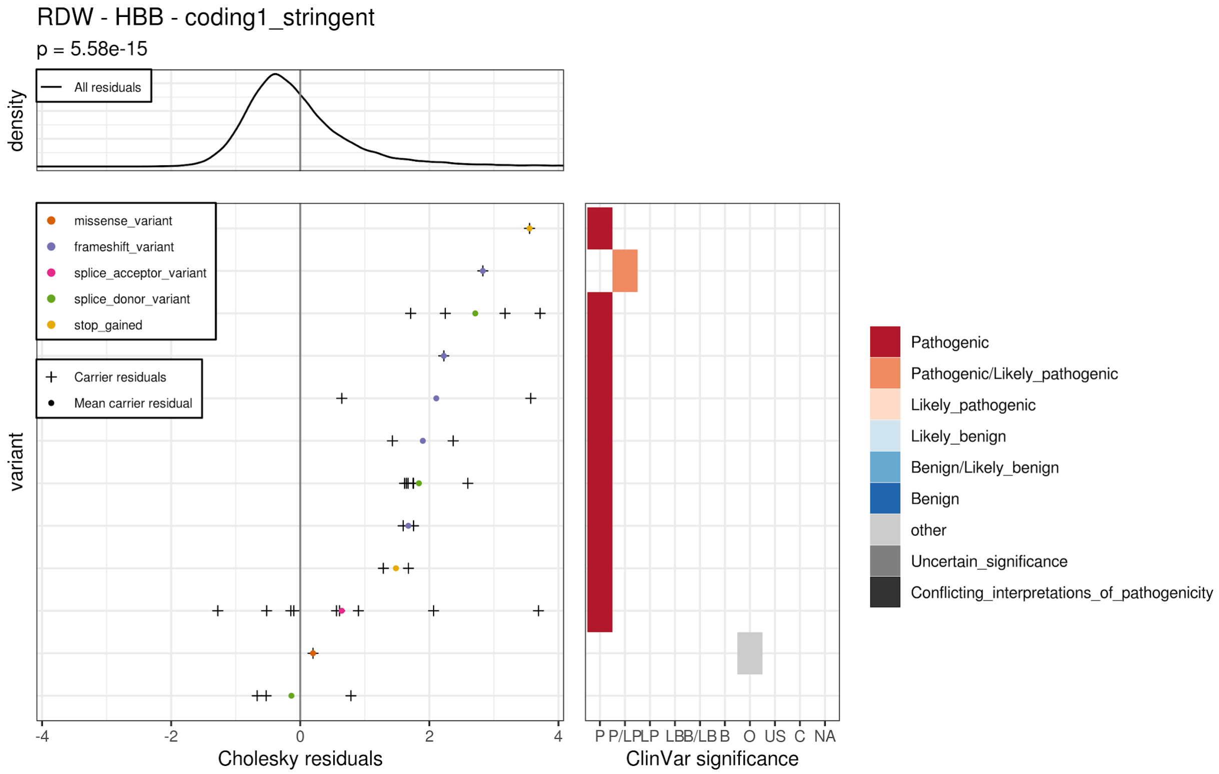
(G7)


RDW - G6PD - coding2_relaxed + noncoding_stringent
p = 2.53e-10
density
All residuals
variant
missense_variant
stop_gained
Carrier residuals
Mean carrier residual
-2
0
2
Cholesky residuals
P P/LPLP LBB/LB B O US C NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

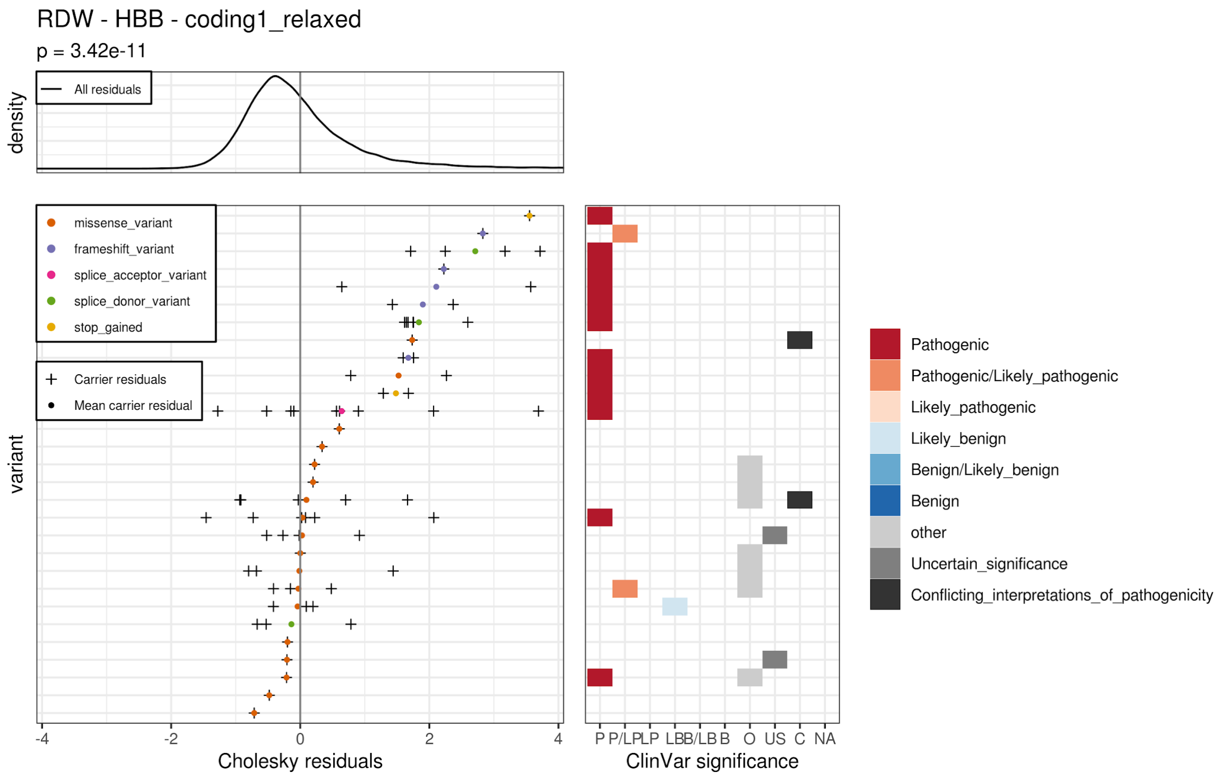
(G8)


RDW - HBB - coding2_relaxed
p = 1.02e-10
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(G9)

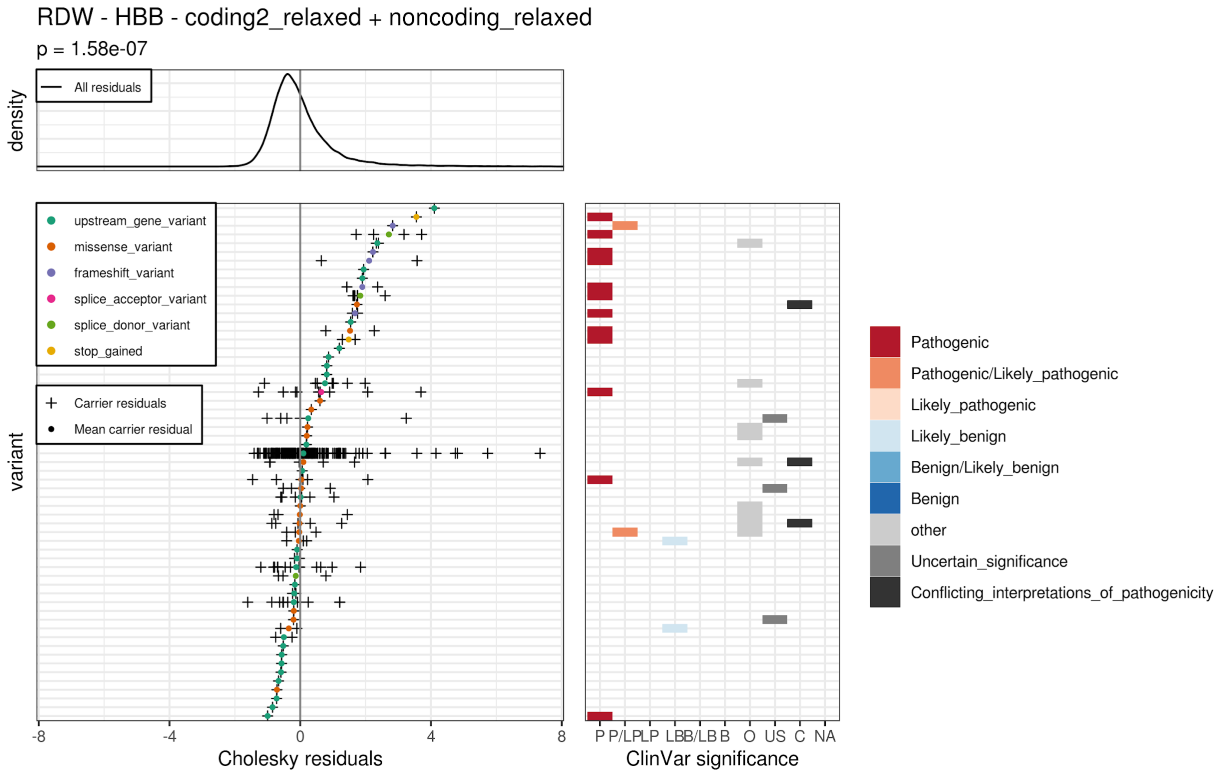

RDW - HBB - coding1_stringent
p = 5.58e-15
density
All residuals
variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(G10)



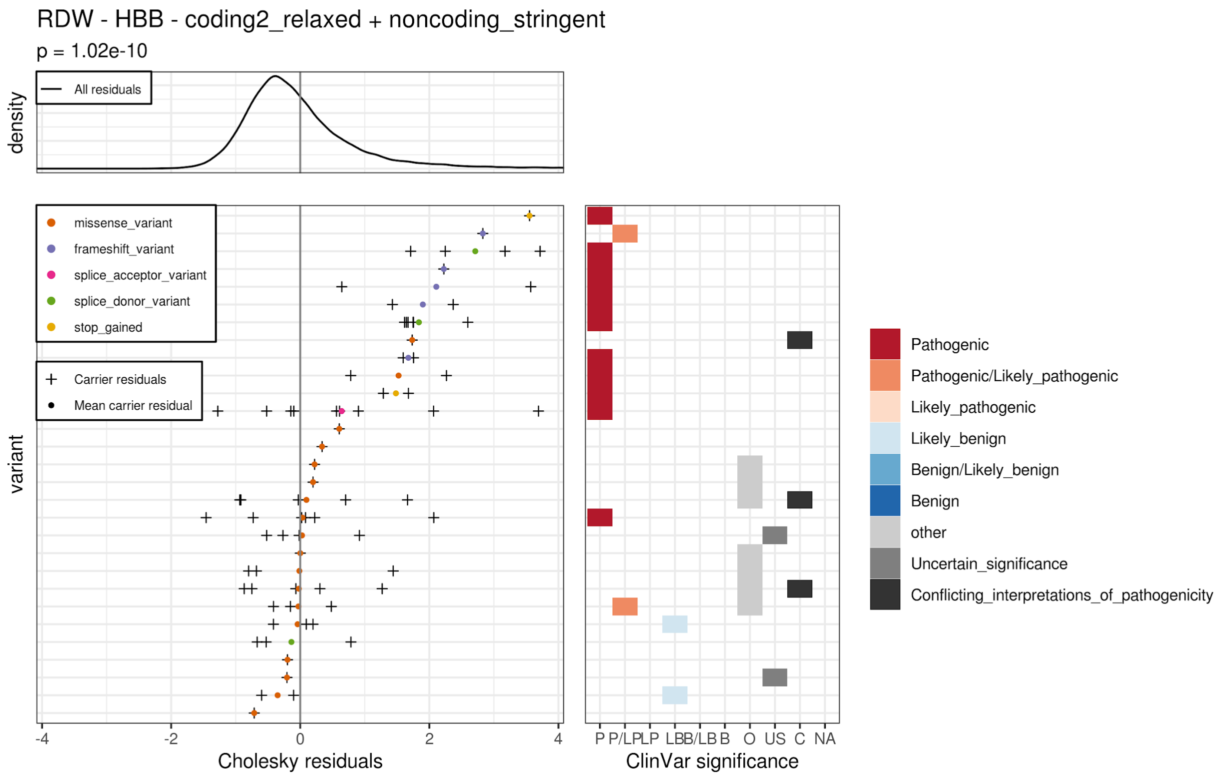
RDW - HBB - coding1_relaxed
p = 3.42e-11
density
All residuals
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(G11)

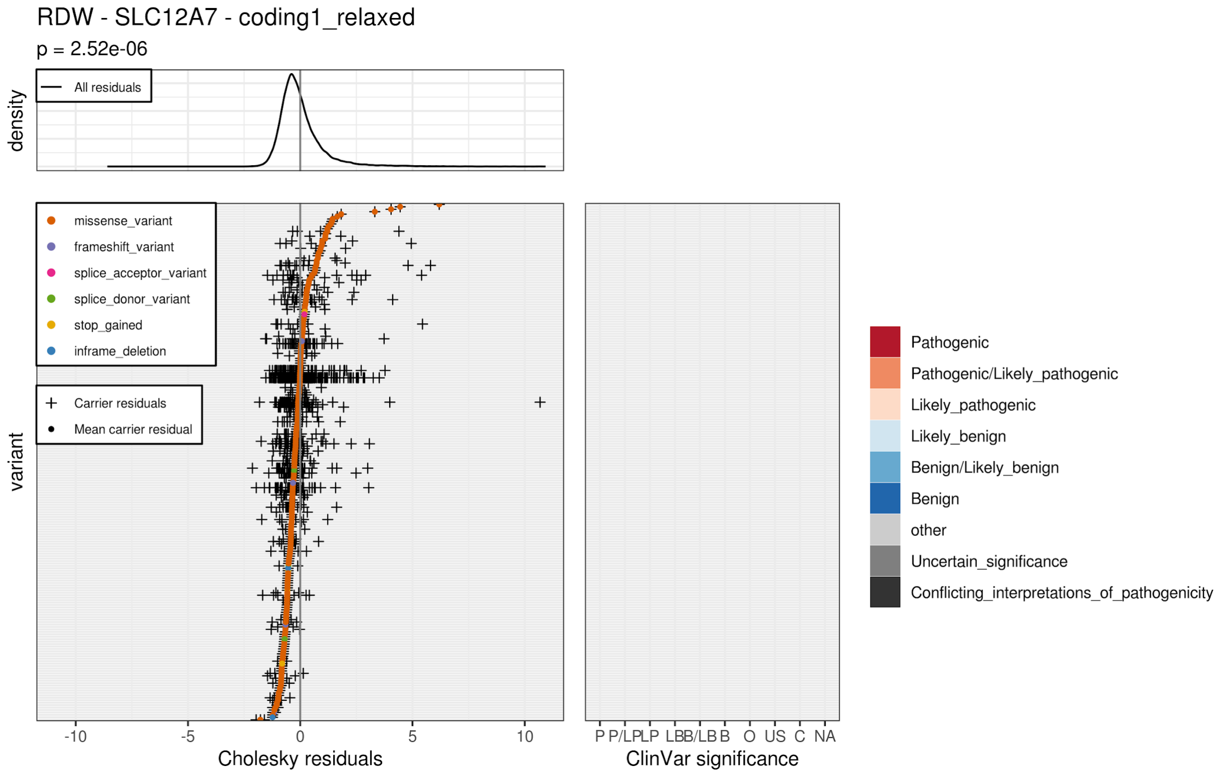


RDW - HBB - coding2_relaxed + noncoding_relaxed
p = 1.58e-07
density
All residuals
upstream_gene_variant
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-8
-4
0
4
8
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

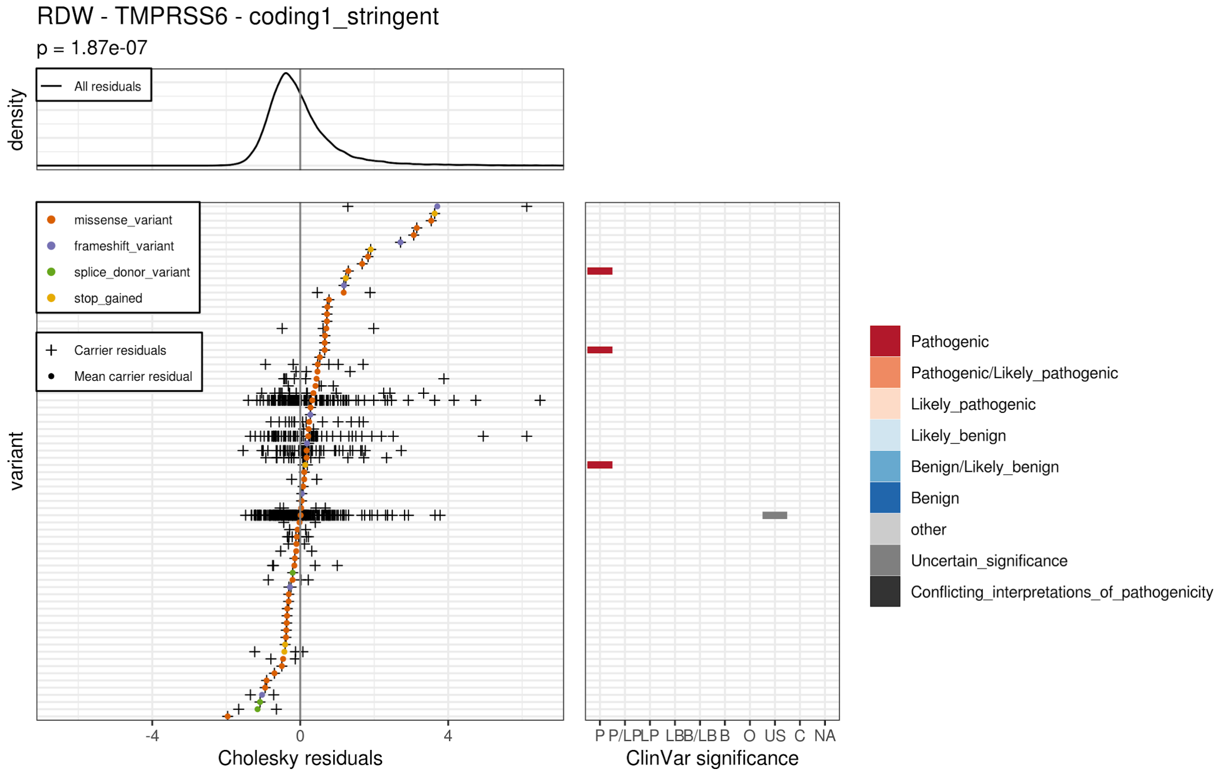
(G12)


RDW - HBB - coding2_relaxed + noncoding_stringent
p = 1.02e-10
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
variant
-4
-2
0
2
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(G13)


RDW - SLC12A7 - coding1_relaxed
p = 2.52e-06
All residuals
density
missense_variant
frameshift_variant
splice_acceptor_variant
splice_donor_variant
stop_gained
inframe_deletion
Carrier residuals
Mean carrier residual
variant
-10
-5
0
5
10
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

(G14)


RDW - TMPRSS6 - coding1_stringent
p = 1.87e-07
density
All residuals
variant
missense_variant
frameshift_variant
splice_donor_variant
stop_gained
Carrier residuals
Mean carrier residual
-4
0
4
Cholesky residuals
P
P/LP
LP
LB
B/LB
B
O
US
C
NA
ClinVar significance
Pathogenic
Pathogenic/Likely_pathogenic
Likely_pathogenic
Likely_benign
Benign/Likely_benign
Benign
other
Uncertain_significance
Conflicting_interpretations_of_pathogenicity

Supplemental Tables

Table S1. Counts of participants by HARE group for each RBC phenotype

	Amish	Asian	Black	Central American	Cuban	Dominican	Mexican	Puerto Rican	South American	White
HCT	1102	654	14474	708	2037	2049	3556	4977	708	32222
HGB	1102	653	14454	708	2037	2048	3556	4974	706	32223
MCH	1102	447	11246	708	2002	2049	3435	4934	706	19612
MCHC	1102	447	13112	708	2002	2049	3434	4934	706	24154
MCV	1102	447	12285	708	2002	2049	3432	4934	706	21165
RBC	1102	384	10747	682	1984	1938	3413	4448	654	19118
RDW	0	447	6776	662	2002	1936	1898	4833	647	10184

HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width.

Table S2. Pairwise trait correlation (upper triangle) and the number of samples used to calculate the correlations (lower triangle)

	HCT	**HGB**	**MCH**	**MCHC**	**MCV**	**RBC**	**RDW**
HCT		0.93211	0.21225	0.03469	0.23772	0.74892	-0.27781
HGB	62447		0.35214	0.31502	0.27268	0.70519	-0.38713
MCH	46099	46083		0.52655	0.87826	-0.33338	-0.44973
MCHC	52628	52612	46109		0.16304	-0.05466	-0.37418
MCV	48807	48791	46116	48816		-0.3458	-0.3531
RBC	44326	44309	44430	44334	44340		-0.0827
RDW	29244	29235	29350	29254	29261	27572	

HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width.

Table S3. Basic characteristics of each participating study in TOPMed stratified by race/ethnicity
See Excel file.

Table S4. Number of SNVs and indels tested for each RBC trait in TOPMed

Trait	Indel	SNV	Total
RDW	5356149	70775433	76131582
RBC	6497451	86154571	92652022
MCH	6637757	88043681	94681438
MCV	6834916	90671951	97506867
MCHC	7089902	94110688	101200590
HGB	7719013	102632640	110351653
HCT	7722116	102674666	110396782

HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width.

Table S5. Previously reported variants and indication of inclusion in the conditional analysis in TOPMed

See Excel file.

Table S6. Guide sequence used in the present study

Guide	**Sequence (5'-3')**
Guides for CRISPR/Cas9 editing	
RUVBL1	ACTACTTACCAATGGCCCTG
Neutral locus	GTAAGCTTAAAACATTAGTA
Guide for C base editing	
rs112097551_C9	GCAAGTAACGGATGCAGGGA

Table S7. Summary of PCR primers used in the present study

Gene symbol	**Direction**	**Sequence (5'-3')**
PCR primers for Sanger sequencing		
RUVBL1	Forward	ACTACTTACCAATGGCCCTG
	Reverse	GAGACAGAGAATCCCATGGG
RPN1	Forward	GTAGGTCCTCAGAGCGCGTG
	Reverse	CAGAGTCATCCAAAATAAGG
rs112097551	Forward	TCCTCTGTCCTTCCTTTCC
	Reverse	CATCTTGCCGATCTCTGAAC
Neutral locus	Forward	CCATGAGACAAGGAAGTAGTG
	Reverse	AGCAGTGGTGAGGAGAATA
Real-time qPCR primers		
EEFSEC	Forward	GAGCGGCAAGTTCAAGAT
	Reverse	GTGGGTGTCGAAGACATAAC
GATA2	Forward	TACAGCAGCGGACTCTT
	Reverse	GGTTCTGCCCATTCATCTT
RPN1	Forward	ACCAGCCACCTCCTTATT
	Reverse	GGTCCACAAACCTCATCTTC
RAB7A	Forward	CCTAGATAGCTGGAGAGATGAG
	Reverse	CTGGTCTCAAAGTAGGGAATG
RUVBL1	Forward	AAGGAGACCAAGGAAGTTTATG
	Reverse	CAGCTTCTACTCGCTCTTTC
GAPDH	Forward	ACCCAGAAGACTGTGGATGG
	Reverse	TTCAGCTCAGGGATGACCTT

Table S8. Lambda values in the single-variant association analyses in TOPMed

Trait	Lambda values		
	Unconditional analysis	Trait-specific conditional analysis	Trait-agnostic conditional analysis
HCT	1.021	1.020	1.015
HGB	1.019	1.019	1.015
MCH	1.036	1.034	1.029
MCHC	1.024	1.022	1.017
MCV	1.038	1.036	1.030
RBC	1.025	1.021	1.018
RDW	1.033	1.024	1.019

HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width.

Table S9. Lead variants at the genome-wide significant loci of the marginal tests in TOPMed
See Excel file.

Table S10. Genome-wide significant variants at the 12 novel loci in the trait-specific conditional analysis in TOPMed

See Excel file.

Table S11. Ancestry-specific allele frequencies of the 14 novel lead variants at the 12 loci

Variant	Chr	Pos	Gene	Alternative allele	Reference allele	Alternative allele frequencies (%)			
						European	African	Hispanic Latino	East Asian
rs112097551	3	1.3E+08	*RNP1*	A	G	0.069	0.940	0.400	0
rs116635225	5	9.6E+07	*ELL2*	A	G	0.074	3.900	0.700	0
rs986415672	10	1.3E+08	*10q26*	T	C	0.011	0	0	0
rs11549407	11	5226774	*HBB*	A	G	0.016	0	0	0
rs34598529	11	5227100	*HBB*	C	T	0	0.320	0	0
rs535577177	11	7E+07	*SHANK2*	A	G	0	0	0.100	0
rs370308370	14	1E+08	*EIF5/MARK3*	A	G	0	0	0	0.910
rs868351380	16	55649	*HBA1/HBA2*	C	G	0.005	0	0.400	0
rs372755452	16	199621	*HBA1/HBA2*	A	AG	0	0	0	1.100
rs763477215	16	8.9E+07	*PIEZO1*	A	ATCT	0.355	0	0	0.050
rs73494666	19	1253643	*MIDN*	T	C	0.614	51.7	4.700	0
rs1368500441	19	2.9E+07	*19q12*	A	G	0.005	0	0	0
rs228914	22	3.7E+07	*TMPRSS6*	A	C	88.7	96.6	79.2	99.8
rs76723693	X	1.5E+08	*G6PD*	G	A	0	0.563	0.077	0

Chr, chromosome; Pos, position.

Table S12. Replication results of the novel findings and the lead independent signals
See Excel file.

Table S13. Independent signals in the step-wise conditional analysis

See Excel file.

Table S14. Phenotypic variance explained by variants identified in the single variant association analysis

Trait	All	Known	Novel
HCT	0.034	0.033	0.001
HGB	0.043	0.040	0.003
MCH	0.213	0.184	0.030
MCHC	0.047	0.041	0.006
MCV	0.179	0.153	0.028
RBC	0.126	0.117	0.010
RDW	0.118	0.109	0.009

HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width.

Table S15. Summary of significant genes in the aggregated association analysis in TOPMed
See Excel file.

Table S16. Summary of significant genes in the aggregated association analysis adjusting for known and novel findings in TOPMed

See Excel file.

Table S17. Annotation of the rare variants identified in the aggregated analysis in TOPMed
See Excel file.

Table S18. Summary of pLoF and pKO variants in TOPMed freeze8 data [1]

Population	N [2]	No. of pLoF variants	No. of genes with at least one individual who is a pKO
African	9,870	55,750	1,617
Asian	231	4,377	395
European	25,569	114,401	1,634
Hispanic	9,757	53,105	1,557

pLoF, predicted loss-of-function; pKO, predicted gene knockout; N, sample size.

[1] No minor allele frequency filter was applied.

[2] Sample sizes represented the number of individuals with blood-cell traits and genotype data available.

Table S19. pLoF variants associated with RBC traits at P<1E-4 in TOPMed [1]

Population	Trait	Chr	Gene	rsID	Variant	MAF (%)	Type	Beta	SE	*P*
African	MCV	2	*WDSUB*	rs377262700	chr2:159236041_G_A	0.021	stopgain	-2.682	0.592	6.09E-06
African	RDW	7	*CD36*	rs3211938	chr7:80671133_T_G[2]	9.340	stopgain	0.244	0.043	**1.24E-08**
Hispanic	MCV	4	*SNX25*	rs1200775460	chr4:185339389_AG_A	0.022	frameshift	2.441	0.543	7.08E-06
Hispanic	MCH	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.022	stopgain	-2.611	0.505	**2.36E-07**
Hispanic	MCV	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.022	stopgain	-2.970	0.506	**4.58E-09**
Hispanic	RBC	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.022	stopgain	2.272	0.506	7.18E-06
European	HCT	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.018	stopgain	-1.545	0.333	3.39E-06
European	HGB	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.018	stopgain	-2.052	0.332	**6.65E-10**
European	MCH	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.017	stopgain	-2.974	0.494	**1.73E-09**
European	MCV	11	*HBB*	rs11549407	chr11:5226774_G_A[3]	0.015	stopgain	-2.981	0.494	**1.66E-09**
European	HCT	11	*CD6*	rs759187282	chr11:61017803_G_T	0.006	stopgain	-2.573	0.575	7.73E-06
European	HGB	11	*CD6*	rs759187282	chr11:61017803_G_T	0.006	stopgain	-2.575	0.574	7.39E-06
Meta-analysis	RDW	1	*SMIM1*	rs566629828	chr1:3775433_AGTCAGCCTAGGGGCTGT_A[4]	1.610	frameshift	0.303	0.068	8.22E-06
Meta-analysis	MCV	18	*SERPINB11*	rs760239610	chr18:63712688_C_CATCAGGTA	0.150	frameshift	-0.681	0.150	5.60E-06

pLoF, predicted loss-of-function; HCT, hematocrit; HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; RBC, red blood cell count; RDW, red blood cell width; Chr, chromosome; MAF, minor allele frequency.

[1] The P values of pLoF variants that reached genome-wide significance were in bold (African: P<8.97E-7; Hispanic: P<9.42E-7; European: P<4.37E-7).

[2] Well known CD36 null allele.

[3] Well known beta-thalassemia allele.

[4] This frameshift indel is responsible for the Vel blood group.

Table S20. pKO variants associated with RBC traits at P<1E-4 in TOPMed[1]

Population	Trait	Chr	Gene	rsID	Variants	MAF (%)	Type	N		Beta	SE	*P*
								Total	KO			
African	MCH	7	*ZNF3*	rs777843966	chr7:100064797_GT_G	0.008	frameshift	6042	200	0.277	0.070	8.53E-05
				rs987730433	chr7:100064875_C_CA	0.008	frameshift					
				rs71689664	chr7:100064888_GTAGT_G	18.3	frameshift					
				rs745468385	chr7:100071151_ACT_A	0.008	frameshift					
				rs774923137	chr7:100071181_TG_T	0.008	frameshift					
				rs988854061	chr7:100079535_C_T	0.008	splicing					
African	MCV	7	*ZNF3*	rs777843966	chr7:100064797_GT_G	0.007	frameshift	7198	239	0.267	0.064	3.60E-05
				rs987730433	chr7:100064875_C_CA	0.007	frameshift					
				rs71689664	chr7:100064888_GTAGT_G	18.4	frameshift					
				rs745468385	chr7:100071151_ACT_A	0.007	frameshift					
				rs774923137	chr7:100071181_TG_T	0.007	frameshift					
				rs988854061	chr7:100079535_C_T	0.007	splicing					

pKO, predicted gene knockout; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; Chr, chromosome; MAF, minor allele frequency.

[1] No pKO variant reached genome-wide significance (African: P<3.09E-5; Hispanic: P<3.21E-5; European: P<3.06E-5).

Supplemental Methods

Participating studies

Amish

The Amish Complex Disease Research Program includes a set of large community-based studies focused largely on cardiometabolic health carried out in the Old Order Amish (OOA) community of Lancaster, Pennsylvania (http://medschool.umaryland.edu/endocrinology/amish/research-program.asp). The OOA population of Lancaster County, PA immigrated to the Colonies from Western Europe in the early 1700's. There are now over 30,000 OOA individuals in the Lancaster area, nearly all of whom can trace their ancestry back 12-14 generations to approximately 700 founders. Investigators at the University of Maryland School of Medicine have been studying the genetic determinants of cardiometabolic health in this population since 1993. To date, over 7,000 Amish adults have participated in one or more of our studies.

Due to their ancestral history, the OOA are enriched for rare exonic variants that arose in the population from a single founder (or small number of founders) and propagated through genetic drift. Many of these variants have large effect sizes and identifying them can lead to new biological insights about health and disease. The parent study for this WGS project provides one (of multiple) examples. In our parent study, we identified through a genome-wide association analysis a haplotype that was highly enriched in the OOA that is associated with very high LDL-cholesterol levels. At the present time, the identity of the causative SNP – and even the implicated gene – is not known because the associated haplotype contains numerous genes, none of which are obvious lipid candidate genes. A major goal of the WGS that will be obtained through the NHLBI TOPMed Consortium will be to identify functional variants that underlie some of the large effect associations observed in this unique population.

ARIC

The ARIC study is a population-based cohort study consisting of 15,792 men and women that were drawn from four U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi)[1]. It was designed to investigate the causes of atherosclerosis and its clinical outcomes, and variation in cardiovascular risk factors, medical care, and disease by race, sex, location, and date. For TOPMed WGS, the study over-sampled participants with incident VTE. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing.

BioMe

The Charles Bronfman Institute for Personalized Medicine at Mount Sinai Medical Center (MSMC), BioMe Biobank, founded in September 2007, is an ongoing, broadly-consented electronic health record-linked clinical care biobank that enrolls participants non-selectively from the Mount Sinai Medical Center patient population. The MSMC serves diverse local communities of upper Manhattan, including Central Harlem (86% African American), East Harlem (88% Hispanic/Latino), and Upper East Side (88% Caucasian/White) with broad health disparities.

CARDIA

The Coronary Artery Risk Development in Young Adults (CARDIA) Study is a study examining the development and determinants of clinical and subclinical cardiovascular disease and their risk factors. It began in 1985-1986 with a group of 5,115 black and white men and women aged 18-30 years. The participants were selected so that there would be approximately the same number of people in subgroups of race, gender, education (high school or less and more than high school) and age (18-24 and 25-30) in each of 4 centers: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA.

CHS

The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults 65 years and older conducted across four field centers [2]. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of people on Medicare eligibility lists from four US communities. Subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. Institutional review committees at each field center approved the CHS, and participants gave informed consent. Blood samples were drawn from all participants at their baseline examination, and DNA was subsequently extracted from available samples. These analyses were limited to participants with available DNA who also consented to genetic studies. Participants were examined annually from enrollment to 1999 and continued to be under surveillance for stroke following 1999.

COPDGene

COPDGene (also known as the Genetic Epidemiology of COPD Study) is an NIH-funded, multicenter study. A study population of more than 10,000 smokers (1/3 African American and 2/3 non-Hispanic White) has been characterized with a study protocol including pulmonary function tests, chest CT scans, six minute walk testing, and multiple questionnaires. Five years after this initial visit, all available study participants are being brought back for a follow-up visit with a similar study protocol. This study has been used for epidemiologic and genetic studies. Previous genetic analysis in this study has been based on genome-wide SNP genotyping data. Approximately 1,900 subjects underwent whole genome sequencing in this NHLBI WGS project, including severe COPD subjects and non-COPD smoking controls. The COPDGene Study web site is: http://www.copdgene.org/.

FHS

FHS is a three-generation, single-site, community-based, ongoing cohort study that was initiated in 1948 to investigate prospectively the risk factors for CVD including stroke. It now comprises 3 generations of participants: the Original cohort followed since 1948 [3]; their Offspring and spouses of the Offspring, followed since 1971 [4]; and children from the largest Offspring families enrolled in 2002 (Gen 3) [5]. The Original cohort enrolled 5,209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5,124 persons (including 3,514 biological offspring) who have been examined approximately once every 4 years. The Gen 3 cohort contains 4,095 participants.

GeneSTAR

In 1982 The Johns Hopkins Sibling and Family Heart Study was created to study patterns of coronary heart disease and related risk factors in families with early-onset coronary disease, identified from10 Baltimore area Hospitals. GeneSTAR continues to study mechanisms of coronary heart disease and stroke in families using novel models and exciting new methods. GeneSTAR is a family-based study in initially healthy brothers and sisters, and offspring of people with early-onset coronary disease, The goal is to discover and amplify mechanisms of stroke and coronary heart disease. Our African American and European American family cohort has undergone extensive screening, genetic testing, and follow-up for new cardiovascular disease, stroke, and other clinical events for 5 to 32 years.

HCHS/SOL

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a multi-center study of Hispanic/Latino populations with the goal of determining the role of acculturation in the prevalence and development of diseases, and to identify other traits that impact Hispanic/Latino health [6] . The study is sponsored by the National Heart, Lung, and Blood Institute (NHLBI) and other institutes, centers, and offices of the National Institutes of Health (NIH). Recruitment began in 2006 with a target population of 16,000 persons of Cuban, Puerto Rican, Dominican, Mexican or Central/South American origin. Participants were recruited through four sites affiliated with San Diego State University, Northwestern University in Chicago, Albert Einstein College of Medicine in Bronx, New York, and the University of Miami. Recruitment was implemented through a two-stage area household probability design [6]. The study enrolled 16,415 participants who were self-identified Hispanic/Latino and aged 18-74 years and the extensive psycho-social and clinical assessments were conducted during 2008-2011. Annual telephone follow-up interviews are ongoing since study inception. During the 2014-2017 second visit, the participants were re-examined again of various health outcomes of interest.

JHS

The Jackson Heart Study (JHS, https://www.jacksonheartstudy.org/jhsinfo/) is a large, community-based, observational study whose participants were recruited from urban and rural areas of the three counties (Hinds, Madison and Rankin) that make up the Jackson, MS metropolitan statistical area (MSA). Participants were enrolled from each of 4 recruitment pools: random, 17%; volunteer, 30%; currently enrolled in the Atherosclerosis Risk in Communities (ARIC) Study, 31% and secondary family members, 22%. Recruitment was limited to non-institutionalized adult African Americans 35-84 years old, except in a nested family cohort where those 21 to 34 years of age were also eligible. The final cohort of 5,301 participants included 6.59% of all African American Jackson MSA residents aged 35-84 during the baseline exam (N-76,426, US Census 2000). Among these, approximately 3,700 gave consent that allows genetic research and deposition of data into dbGaP. Major components of three clinic examinations (Exam 1 – 2000-2004; Exam 2 – 2005-2008; Exam 3 – 2009-2013) include medical history, physical examination, blood/urine analytes and interview questions on areas such as: physical activity; stress, coping and spirituality; racism and discrimination; socioeconomic position; and access to health care. Extensive clinical phenotyping includes anthropometrics, electrocardiography, carotid ultrasound, ankle-brachial blood pressure index, echocardiography, CT chest and abdomen for coronary and aortic calcification, liver fat, and subcutaneous and visceral fat measurement, and cardiac MRI. At 12-month intervals after the baseline clinic visit (Exam 1), participants have been contacted by telephone to: update information; confirm vital statistics; document interim medical events, hospitalizations, and functional status; and obtain additional sociocultural information. Questions about medical events, symptoms of cardiovascular disease and functional status are repeated annually. Ongoing cohort surveillance includes abstraction of medical records and death certificates for relevant International Classification of Diseases (ICD) codes and adjudication of nonfatal events and deaths. CMS data are currently being incorporated into the dataset.

MESA

The MESA study is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease [7]. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Thirty-eight percent of the recruited participants are white, 28 percent African-American, 22 percent

Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and the University of California - Los Angeles.

SAFS

The San Antonio Family Study (SAFS) is a complex pedigree-based mixed longitudinal study designed to identify low frequency or rare variants influencing susceptibility to cardiovascular disease, using WGS information from 2,590 individuals in large Mexican American pedigrees from San Antonio, Texas. The major objectives of this study are to identify low frequency or rare variants in and around known common variant signals for CVD, as well as to find novel low frequency or rare variants influencing susceptibility to CVD.

WHI

The Women's Health Initiative (WHI) is a long-term, prospective, multi-center cohort study that investigates post-menopausal women's health [8]. WHI was funded by the National Institutes of Health and the National Heart, Lung, and Blood Institute to study strategies to prevent heart disease, breast cancer, colon cancer, and osteoporotic fractures in women 50-79 years of age. WHI involves 161,808 women recruited between 1993 and 1998 at 40 centers across the US. The study consists of two parts: the WHI Clinical Trial which was a randomized clinical trial of hormone therapy, dietary modification, and calcium/Vitamin D supplementation, and the WHI Observational Study, which focused on many of the inequities in women's health research and provided practical information about the incidence, risk factors, and interventions related to heart disease, cancer, and osteoporotic fractures. For TOPMed WGS, the study over-sampled participants with incident stroke and VTE. The remaining samples were age- and ethnicity-matched controls without stroke or VTE.

Phenotype harmonization

Because multiple studies contributed to the analysis, RBC phenotypes were harmonized across studies such that they could be analyzed together (https://www.biorxiv.org/content/10.1101/2020.06.18.146423v1). For most studies, variables were obtained from dbGaP. Data for the BioME study, the COPDGene study, and some participants in the WHI study were transferred directly to investigators. The study variables were QC'd to identify recording errors or other problematic data. After QC, variables were converted to a consistent measurement unit across studies. When possible, harmonized variables were calculated using study-derived variables. If QC uncovered data quality issues with the study-derived variable or if the study did not provide the specific variable of interest, the harmonized variable was instead calculated from the components (e.g., MCH = 10 * hemoglobin / red cell count). Finally, QC of the harmonized variable was performed to check that no large differences between studies remained.

Both FHS and ARIC measured phenotypes at multiple times in a given participant. For these studies, a single measurement for each participant was selected for each trait. To maximize the sample size and minimize batch effects within studies, harmonized data for different RBC traits for a single subject may have been measured at different visits. Within FHS, only the Offspring cohort had measurements at multiple exams. For these participants, the measurement at the most recent exam was chosen. For the ARIC study, the visit with the most non-missing phenotype values across all participants was chosen first. For subjects without measurements at this visit, the visit with the next most non-missing values was chosen, and so forth. As a consequence, values for the same participant for different RBC phenotypes were sometimes measured at different visits."

Trait-specific QC procedures were also performed. We excluded participants with HCT values >80% and those with HCT <5% (n=14). Similarly for HGB, we excluded participants with HGB measurements >30 g/dL and <5 g/dL (n=31). For participants from WHI, for both the HCT and HGB analyses, we excluded those with an HCT/HGB ratio >7 (n=11). Participants with MCH values >75 pg were excluded from analysis (n=2). In MCHC, we excluded participants with measurements ≥60 g/dL (n=1). For the MCV analysis, outliers with values >150 fL were excluded (n=2). In the RBC and RDW analyses, no participants were excluded based upon measured trait values.

Statistical analyses

Genetic Ancestry and Relatedness

Principal components (PCs) of genetic ancestry and pairwise relatedness measures were estimated for all 140,062 samples included in the TOPMed 'freeze 8' genotype release. Autosomal genetic variants passing the quality filter with a MAF > 0.01 and missing call rate < 0.01 were LD-pruned with an r^2 threshold of 0.1 to obtain a set of 638,486 effectively independent variants for genetic ancestry and relatedness estimation. PC-AiR [9] was used to obtain ancestry informative PCs robust to familial relatedness; the first 11 PCs showed evidence of population structure. PC-Relate [10] was then used to estimate pairwise kinship coefficients (KCs) for all pairs of samples, conditional on the genetic ancestry captured by PC-AiR PCs 1-11; these KC estimates reflect only recent genetic relatedness, e.g. due to pedigree structure. The PC-Relate KC estimates were used to construct a 4th degree sparse, block-diagonal, empirical kinship matrix (KM) for association testing, using the procedure recommended in Gogarten et al [11].: any pair of samples with estimated KC > $2^{(-11/2)}$ ~ 0.022 were clustered in the same block; all KC estimates within a block of samples were kept, regardless of value; and all KC estimates between blocks were set to 0. By using a sparse block-diagonal KM, the association tests are more computationally efficient yet recent genetic relatedness is still accounted for. We subset the freeze-wide PCs and sparse KM to the appropriate set of participants for each analysis.

HARE for Imputation of Race/Population Membership using Genetic Ancestry

Ancestry groups were based on a combination of participants reported race/ethnicity and genetic ancestry represented by PCs from PC-AiR[9]. To infer race/population group membership for participants with missing values, we used the HARE method [12] . HARE is a machine learning algorithm that uses a support vector machine (SVM) to determine stratum assignment, taking as input genetically estimated PC values and reported race/ethnicity for each participant. Strata are defined by the unique reported race/ethnicity values provided, then the HARE SVM uses the input (training) data to learn the probability of stratum membership across the entire PC space. The output of HARE consists of multinomial probability vectors of stratum membership for each participant. HARE was run on a subset of samples included in the TOPMed freeze 8 genotype release; specifically, samples for participants from non-US populations (e.g. Costa Rica) and the Amish participants (because they were very distinct in PC space) were excluded from the HARE analysis. HARE was run using the first 9 PC-AiR PCs generated on this subset of samples to represent genetic ancestry with the following reported race/population groups: Asian, Black, Central American, Cuban, Dominican, Mexican, Puerto Rican, South American, and White. The genetic data from the 31,918 participants with either unreported or non-specific (e.g. 'Multiple' or 'Other') race and population membership was included in the HARE analysis, but they were not used to train the SVM. These participants were assigned to a population stratum based on their highest HARE output probability of membership. All other participants remained in the population stratum corresponding to their reported race/population group. Amish participants were assigned to their own stratum.

Fitting the Linear Mixed Model

The linear mixed model (LMM) can be written as $Y = G\beta + X\alpha + \epsilon$, where Y is the $(n \; x \; 1)$ vector of outcome values; G is an $(n \; x \; m)$ matrix of alternate allele counts for each of the n individuals at the m variants of interest ($m = 1$ for a single variant analysis) with effect sizes given by the $(m \; x \; 1)$ vector β; X is the $(n \; x \; k)$ matrix of fixed effect covariates including an intercept with effect sizes given by the $(k \; x \; 1)$ vector α; and $\epsilon \sim N(0, \Sigma)$ is the $(n \; x \; 1)$ vector of errors with covariance matrix Σ that captures both genetic covariance due to relatedness/kinship and residual variance structure. Given the true Σ, we could estimate β using generalized least squares (GLS). However, we can simplify this GLS problem to an ordinary least squares (OLS) problem by pre-multiplying both sides of the equation by the matrix C, the Cholesky-decomposition of Σ^{-1}, such that $C'C = \Sigma^{-1}$ and $C'\Sigma C = I$, where I is the $(n \; x \; n)$ identity matrix. Further, by the Frisch-Waugh Lovell theorem[13], we can adjust for the covariates in the new OLS model, CX, by pre-multiplying CY and CG by the annihilator matrix $[I - (CX)((CX)'(CX))^{-1}(CX)']$. Ultimately, the original GLS problem can be re-written as the linear regression model $Y^* = G^*\beta + \epsilon^*$, where $Y^* = MY$, $G^* = MG$, and $M = [I - CX(X'C'CX)^{-1}X'C']C$. In practice, we use REML to estimate $\hat{\Sigma}$ under the null hypothesis that $\beta = 0$ (i.e. fit the null model) and calculate the estimate of the matrix $\widehat{M}$.

Score Tests and Approximate Variant Effect Sizes

Given Y^* and G^*, a joint score test for the set of m variants can be performed, where the score is $U = G^{*\prime}Y^*$, the variance of the score is $V = G^{*\prime}G^*$, and the test statistic is $T_G = U'V^{-1}U \sim \chi^2_m$. The score and the Wald tests are approximately asymptotically equivalent when β is small (as is typical for GWAS), so the variant effect sizes can be reasonably approximated from the score test as $\hat{\beta} \approx V^{-1}U = (G^{*\prime}G^*)^{-1}(G^{*\prime}Y^*)$, and their covariance matrix can be reasonably approximated from the score test as $\widehat{var}(\hat{\beta}) \approx V^{-1} = (G^{*\prime}G^*)^{-1}$.[14] Note that these are the score tests used for the single variant association analysis, where each variant genome-wide is tested individually (i.e. $m = 1$).

Proportion of Variance Explained Jointly by a set of variants

To estimate the proportion of phenotypic variance explained (PVE) by the m variants in G, we use the formula $PVE = 1 - RSS_1/RSS_0$, where RSS_0 and RSS_1 are the residual sums of squares computed from the null model, and the model including the m variants of interest, respectively. Under the null model, we have that $RSS_0 = Y^{*\prime}Y^*$, and from the model $Y^* = G^*\beta + \epsilon^*$, we have that $RSS_1 = (Y^* - G^*\hat{\beta})'(Y^* - G^*\hat{\beta})$. Using the approximation for $\hat{\beta}$ given above, we get that $RSS_1 \approx Y^{*\prime}Y^* - Y^{*\prime}G^*(G^{*\prime}G^*)^{-1}G^{*\prime}Y^* = Y^{*\prime}Y^* - T_G$, and the estimate $\widehat{PVE} \approx T_G/(Y^{*\prime}Y^*)$. It's worth noting that using this approach to estimate the PVE for the set of m variants jointly should provide a more accurate estimate than estimating the PVE for each variant separately and summing, as this joint approach accounts for the covariance between the variant effect sizes, as measured by V^{-1} (the separate approach is equivalent to $\widehat{PVE} = [U'diag(V)^{-1}U]/(Y^{*\prime}Y^*)$, where $diag(V)$is an $(m\ x\ m)$ matrix of just the diagonal of the V matrix). This joint PVE calculation is implemented in the GENESIS software [11] with the jointScoreTest function.

Conditional analyses

We performed three types of conditional analysis in the discovery stage. The first conditional analyses adjusted each trait for variants that were previously reported to be associated with the particular RBC trait and that passed the QC filter. These variants were pruned to a set with linkage disequilibrium (LD) r2 < 0.8 such that a variant with a more significant p-value was preferentially retained over those with higher p-values. Known variants failing the QC filter were included if any variants within 1 MB of the known variant remained significant after adjusting for the passing variants only. We refer to this analysis as the "RBC trait-specific conditional analysis". In the second conditional analysis, we included all previously reported variants for *any* of the seven RBC traits as well as any failed variants included in any of the trait-specific conditional analyses. Variants were again pruned to LD $r^2 < 0.8$ with preferential selection based on p-value. We refer to the second conditional analysis as the "RBC trait-agnostic conditional analysis". Finally, we performed iterative conditional analysis by chromosome for each trait to identify an independent set of associated variants. For this third conditional analysis, we started with the association results from the trait-specific conditional analysis. For each chromosome, we identified the most significant variant (if any, using a $5x10^{-9}$ threshold) as the 'peak variant' and then fit a new null model adjusted for both the previous set of conditional variants from the trait-specific conditional analyses as well as this peak variant, and calculated new score test statistics. If any variant was significant at the $5x10^{-9}$ level in the new score tests (regardless of its significance level in the original trait-specific conditional results), we performed a second round of conditional analysis, re-estimating the null model and calculating the score test statistics, adjusting for the new peak variant along with the original trait-specific conditional variants and the first peak variant. We continued this procedure iteratively, adding any new 'peak variants' into the list of variants to condition on, re-fitting the null model, and calculating the updated score statistics, until no additional variants were significant at the $5x10^{-9}$ level. Finally, the variants identified across all chromosomes in this iterative conditional analysis were combined into a set of "conditionally-independent variants" for each trait.

Aggregation Strategies

For aggregate association testing, five distinct methods were used to aggregate rare variants into gene-based groups using GENCODE v29 gene model. Three strategies only included coding variants, and two strategies additionally included non-coding variants. Variants were further filtered using one or more deleterious prediction scores to enrich for likely causal variants. The detail method used for each strategy is provided below

1. **Coding filter 1 - Stringent (C1-S)**: This strategy includes high confidence predicted LoF variants inferred using LOFTEE (https://github.com/konradjk/loftee), missense variants predicted deleterious by all of SIFT4G[26633127]<=0.05, Polyphen2_HDIV>0.5[20354512], Polyphen2_HVAR>0.5[20354512], and variants predicted as "Deleterious" by LRT [19602639] and inframe indels or synonymous variants with Fathmm-XF score[28968714] > 0.5
2. **Coding filter 1 - Relaxed (C1-R)**: This strategy is same as C1-S but the missense filter was relaxed to retain variants predicted deleterious by any of SIFT4G, Polyphen2_HDIV, Polyphen2_HVAR scores

3. **Coding filter 2 - Relaxed (C2-R)**: This strategy is the same as C1-S but missense variants were filtered using MetaSVM score and a relatively relaxed set of missense variants was retained by applying MetaSVM score [25552646] > 0 filter
4. **Coding filter 2 - Relaxed & Non-coding filter-Relaxed (C2-R+NC-R)**: This strategy includes variants included in C2-R and additional regulatory variants. Regulatory variants were included if they overlapped with enhancer(s) or promoters linked to a gene using GeneHancer[28605766], or 5 Kb upstream of the Transcription start site. Within these regions only those variants were retained which had Fathmm-XF score > 0.5 **or** overlap with regions labelled as either"CTCF binding sites," "Transcription factor binding sites" as annotated by the Ensembl regulatory build annotation[25887522]
5. **Coding filter 2 - Relaxed & Non-coding filter-Stringent (C2-R +NC-S)** :This strategy includes variants included in C2-R and additional regulatory variants using a stringent filtering criteria. Even in this method regulatory variants in Genehancer[28605766] linked regulatory regions and 5 Kb upstream of the Transcription start site of gene were included. However within these regions only those variants were retained which had Fathmm-XF score > 0.5 **and** which overlapped with regions labelled as "Promoters," "Promoter flanking regions," "Enhancers," "CTCF binding sites," "Transcription factor binding sites" or "Open chromatin regions" as annotated by the Ensembl regulatory build annotation[25887522].

The annotation based variant filtering and gene based aggregation was performed using TOPMed freeze 8 WGSA Google BigQuery annotation database on the BiodataCatalyst powered by Seven Bridges platform (http://doi.org/10.5281/zenodo.3822858). The annotation database was built using variant annotations generated by Whole genome Sequence annotator version v0.8 [26395054] and formatted by WGSAParsr version 6.3.8 (https://github.com/UW-GAC/wgsaparsr). The GENCODE v29 gene model based varint consequences were obtained from Ensembl Variant effect predictor (VEP)[26683364] incorporated within WGSA. When using a deleteriousness prediction score, respective author recommended cut points were used to retain likely deleterious variants.

References

1. (1989). The atherosclerosis risk in communities (aric) study: design and objectives. Am. J. Epidemiol. *129*, 687–702.

2. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., and Newman, A. (1991). The Cardiovascular Health Study: design and rationale. Ann. Epidemiol. *1*, 263–276.

3. Dawber, T.R., and Kannel, W.B. (1966). The Framingham study. An epidemiological approach to coronary heart disease. Circulation *34*, 553–555.

4. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M., and Castelli, W.P. (1975). The Framingham Offspring Study. Design and preliminary data. Prev. Med. *4*, 518–525.

5. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Fox, C.S., Larson, M.G., Murabito, J.M., et al. (2007). The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am. J. Epidemiol. *165*, 1328–1335.

6. Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Avilés-Santa, L.M., Kaplan, R.C., Barnhart, J., Liu, K., Giachello, A., Lee, D.J., Ryan, J., et al. (2010). Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. *20*, 642–649.

7. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A.V., Folsom, A.R., Greenland, P., Jacob, D.R., Kronmal, R., Liu, K., et al. (2002). Multi-Ethnic Study of Atherosclerosis: objectives and design. Am. J. Epidemiol. *156*, 871–881.

8. (1998). Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control. Clin. Trials *19*, 61–109.

9. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet. Epidemiol. *39*, 276–293.

10. Conomos, M.P., Reiner, A.P., Weir, B.S., and Thornton, T.A. (2016). Model-free Estimation of Recent Genetic Relatedness. Am. J. Hum. Genet. *98*, 127–148.

11. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics *35*, 5346–5348.

12. Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., Vujkovic, M., Damrauer, S.M., Pyarajan, S., Gaziano, J.M., et al. (2019). Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. Am. J. Hum. Genet. *105*, 763–772.

13. Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends.

Econometrica: Journal of the Econometric Society, 387-401.

14. Zhou, B., Shi, J., and Whittemore, A.S. (2011). Optimal methods for meta-analysis of genome-wide association studies. Genet. Epidemiol. *35*, 581–591.

Acknowledgements

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). The table below presents study specific omics support information. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

Paul S. de Vries was supported by American Heart Association grant number 18CDA34110116. H.C. and E.J. were supported by the National Eye Institute (NEI) grant R01 EY027004, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK116738 and by the National Cancer Institute (NCI) R01CA2416323. M.P.C and D.J were supported by NHLBI grant U01HL137162. D.E.B. was supported by NHLBI P01HL032262, DP2HL137300, R01HL130733. B.P.K. was supported by NCI R00 CA218870 and NHLBI P01HL142494.

TOPMed Accession #	TOPMed Project	Parent Study Name	TOPMed Phase	Omics Center	Omics Support
phs000956	Amish	Amish	1	Broad Genomics	3R01HL121007-01S1
phs001211	AFGen	ARIC AFGen	1	Broad Genomics	3R01HL092577-06S1
phs001211	VTE	ARIC	2	Baylor	3U54HG003273-12S2 / HHSN268201500015C
phs001644	AFGen	BioMe AFGen	2.4	MGI	3UM1HG008853-01S2
phs001644	BioMe	BioMe	3	Baylor	HHSN268201600033I
phs001644	BioMe	BioMe	3	MGI	HHSN268201600037I
phs001612	CARDIA	CARDIA	3	Baylor	HHSN268201600033I
phs001368	CHS	CHS	3	Baylor	HHSN268201600033I
phs001368	VTE	CHS VTE	2	Baylor	3U54HG003273-12S2 / HHSN268201500015C
phs000951	COPD	COPDGene	1	NWGC	3R01HL089856-08S1
phs000951	COPD	COPDGene	2	Broad Genomics	HHSN268201500014C
phs000951	COPD	COPDGene	2.5	Broad Genomics	HHSN268201500014C
phs000974	AFGen	FHS AFGen	1	Broad Genomics	3R01HL092577-06S1
phs000974	FHS	FHS	1	Broad Genomics	3U54HG003067-12S2
phs001218	AA_CAC	GeneSTAR AA_CAC	2	Broad Genomics	HHSN268201500014C

phs001218	GeneSTAR	GeneSTAR	legacy	Illumina	R01HL112064
phs001218	GeneSTAR	GeneSTAR	2	Psomagen	3R01HL112064-04S1
phs001395	HCHS_SOL	HCHS_SOL	3	Baylor	HHSN268201600033I
phs000964	JHS	JHS	1	NWGC	HHSN268201100037C
phs001416	AA_CAC	MESA AA_CAC	2	Broad Genomics	HHSN268201500014C
phs001416	MESA	MESA	2	Broad Genomics	3U54HG003067-13S1
phs001215	SAFS	SAFS	1	Illumina	3R01HL113323-03S1
phs001215	SAFS	SAFS	legacy	Illumina	R01HL113322
phs001237	WHI	WHI	2	Broad Genomics	HHSN268201500014C

Amish: The TOPMed component of the Amish Research Program was supported by NIH grants R01 HL121007, U01 HL072515, and R01 AG18728. Email Rhea Cosentino (rcosenti@som.umaryland.edu) for additional input.

ARIC: The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN2682017000011, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions.

BioMe: The Mount Sinai BioMe Biobank has been supported by The Andrea and Charles Bronfman Philanthropies and in part by Federal funds from the NHLBI and NHGRI (U01HG00638001; U01HG007417; X01HL134588). We thank all participants in the Mount Sinai Biobank. We also thank all our recruiters who have assisted and continue to assist in data collection and management and are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

CARDIA: The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005).

CHS: Cardiovascular Health Study: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114

from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

COPDGene: The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: http://www.copdgene.org/directory

FHS: The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195 and HHSN268201500001I from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

GeneSTAR: GeneSTAR was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL112064, HL11006, HL118356) and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. We would like to thank our participants and staff for their valuable contributions.

HCHS/SOL: The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

JHS: The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201300049C and HHSN268201300050C), Tougaloo College

(HHSN268201300048C), and the University of Mississippi Medical Center (HHSN268201300046C and HHSN268201300047C) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS.

MESA: MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

SAFS: Collection of the San Antonio Family Study data was supported in part by National Institutes of Health (NIH) grants R01 HL045522, MH078143, MH078111 and MH083824; and whole genome sequencing of SAFS subjects was supported by U01 DK085524 and R01 HL113323. We are very grateful to the participants of the San Antonio Family Study for their continued involvement in our research programs.

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C.