

The American Journal of Human Genetics, Volume 100

Supplemental Data

Practical Approaches for Whole-Genome Sequence

Analysis of Heart- and Blood-Related Traits

Alanna C. Morrison, Zhuoyi Huang, Bing Yu, Ginger Metcalf, Xiaoming Liu, Christie Ballantyne, Josef Coresh, Fuli Yu, Donna Muzny, Elena Feofanova, Navin Rustagi, Richard Gibbs, and Eric Boerwinkle

Supplemental Note: ARIC study

This study investigates 10 heart and blood-related factors measured in Atherosclerosis Risk in Communities (ARIC) study participants. Fasting blood was drawn at baseline and processed according to a standard protocol.¹ Platelet count was obtained within 24 hours after blood collection. Blood was stored at 4 °C between the venipuncture and the platelet count measurement, conducted using automated particle counters (Coulter Diagnostics, Hialeah, Florida, in three field centers; Technicon H-6000, Technicon Corporation, Tarrytown, New York, in one field center).² Total white blood cell count (WBC) and relative proportion of neutrophils were obtained from EDTA-anticoagulated venous blood cells using a Beckman-Coulter Counter (Beckman Coulter, Inc., Fullerton, CA). Absolute neutrophil counts were received by multiplying the neutrophils percentage by overall WBC.³ For the phosphorus (P) and magnesium (Mg) measurements, serum was frozen at -70 °C after the blood collection, until analyzed.⁴ Serum Mg was measured by the Gindler and Heth procedure using metallochromic dye calmagite (1-[1-hydroxy-4-methyl-2-phenylazo]-2-naphthol-4-sulfonic acid).⁵ Serum P was measured using an automated platform (Beckman-Coulter) in which inorganic phosphorus reacts with ammonium molybdate in an acidic solution to form a colored phosphomolybdate complex. The reported intra-assay coefficient of variation was 5.8%.⁶ Phosphorus values were set to missing if estimated Glomerular Filtration Rate (eGFR) as less than 45. The total protein component of lipoprotein (a) [Lp(a), (apolipoprotein(a) + apolipoprotein B)] was measured at the first visit with the 'sandwich' ELISA assay, with the reliability of 0.90.⁷ The protein portion constitutes around one-third of the total protein component of Lp(a). Natural log-transformation was applied to Lp(a) values to normalize their distribution.⁸ Hemoglobin was measured using automated hematology analyzers: Coulter S + IV (calibration S - Cal, Beckman Coulter, Inc, Fullerton, CA) at two sites, Coulter S + III and Coulter S + IV (calibration S-Cal) at one site, and Technicon H-6000 (calibration Fisher, Technicon Corporation, Tarrytown, NY) at one site.² Cardiac troponin T (cTnT) levels were measured in plasma samples collected on the fourth examination

between years 1996–1998. Plasma samples were stored at –70 to –80 °C and thawed before testing. cTnT levels were measured with Elecsys Troponin T high sensitive assay (Roche Diagnostics, Indianapolis, IN, range of detection: 0.003 – 10 µg/L.⁹ For participants with cTnT levels below the lowest detectable limit, the value of cTnT was set to 0.003 ug/L. Participants with prevalent coronary heart disease or prevalent heart failure at the fourth examination were excluded from the analysis.¹⁰ Natural log-transformation was applied to cTnT values before the analysis. N-terminal pro-B-type natriuretic peptide (NT-proBNP) was measured by an electrochemiluminescent immunoassay on an automated Cobas e411 analyzer (Roche Diagnostics, Indianapolis, IN) with lower limit of detection ≤5 pg/mL and coefficient of variation 3.5 – 4.7%.¹¹ Individuals with NT-proBNP below the detection limit were assigned an NT-proBNP value of 2.5 pg/ml. Participants with prevalent heart failure at the fourth visit were excluded from the analysis. NT-proBNP values were natural log-transformed before the analysis. C-reactive protein (CRP) levels were also measured at the fourth examination by immunoturbidimetric CRP-Latex (II) high-sensitivity assay (Denka Seiken, Tokyo, Japan) on a Hitachi 911 analyzer (Roche Diagnostics, Indianapolis, Indiana).¹² CRP values were natural log-transformed before the analysis. A homogeneous assay method was used for the direct measurement of small dense low-density lipoprotein cholesterol (sdLDL-C) in plasma (sd-LDL-EX “Seiken”, Denka Seiken, Tokyo, Japan) on a Hitachi 917 automated chemistry analyzer, with the reliability coefficient of 0.92 (based on 435 blinded quality control replicates).¹³

Supplemental Figure Legends

Figure S1. Proportion of single nucleotide variants within frequency bins in African Americans. MAF = minor allele frequency.

Figure S2. Distribution of single nucleotide variants (SNVs) per sliding window, regulatory domain, and the first intron motifs in African Americans. SNVs included here have $MAF \leq 5\%$ and $MAC \geq 3$

Figure S3. Common variants ($MAF > 5\%$), with the significance threshold line at p-value equal to 5×10^{-8}

Figure S4. Histogram contrasting individuals carrying risk alleles that increase $Lp(a)$ levels versus non-carriers. Individuals in red are carriers of 2 or more alleles that increase $Lp(a)$ levels. Individuals in grey do not carry any alleles associated with increased $Lp(a)$ levels.

Figure S5. Histogram contrasting individuals carrying risk alleles that decrease $Lp(a)$ levels versus non-carriers. Individuals in blue are carriers of 2 or more alleles that decrease $Lp(a)$ levels. Individuals in grey do not carry any alleles associated with decreased $Lp(a)$ levels.

Figure S6. QQ plot for unweighted tests involving sliding windows compared to weighted tests incorporating CADD scores. Panel A. T5 test for $Lp(a)$ levels. Panel B. SKAT for $Lp(a)$ levels. Panel C. T5 test for neutrophil count. Panel D. SKAT for neutrophil count. Blue line is weighted. Red line is unweighted.

Figure S7. QQ plot for unweighted tests involving sliding windows compared to weighted tests incorporating Eigen scores. Panel A. T5 test for $Lp(a)$ levels. Panel B. SKAT for $Lp(a)$ levels. Panel C. T5 test for neutrophil count. Panel D. SKAT for neutrophil count. Blue line is weighted. Red line is unweighted.

Figure S8. QQ plot for unweighted tests involving sliding windows compared to weighted tests incorporating quartic-scaled CADD scores. Panel A. T5 test for $Lp(a)$ levels. Panel B. SKAT for $Lp(a)$ levels. Panel C. T5 test for neutrophil count. Panel D. SKAT for neutrophil count. Blue line is weighted. Red line is unweighted.

Figure S9. Average CADD for significant windows compared to all sliding windows. **A.** Average CADD for significant T5 windows. The histogram represents the average CADD for all sliding windows. The vertical red lines plot the average CADD for the significant sliding windows in Table S2. **B.** Average CADD for significant SKAT windows. The histogram represents the average CADD for all sliding windows. The vertical red lines plot the average CADD for the significant sliding windows in Table S3.

Figure S10. Average quartic CADD for significant windows compared to all sliding windows. **A.** Average quartic CADD for significant T5 windows. The histogram represents the average quartic CADD for all sliding windows. The vertical red lines plot the average quartic CADD for the significant sliding windows in Table S2. **B.** Average quartic CADD for significant SKAT windows. The histogram represents the average quartic CADD for all sliding windows. The vertical red lines plot the average quartic CADD for the significant sliding windows in Table S3.

Figure S11. Average Eigen score for significant non-coding windows compared to all non-coding sliding windows. **A.** Average Eigen score for significant T5 non-coding windows. The

histogram represents the average Eigen score for all non-coding sliding windows. The vertical red lines plot the average Eigen score for the significant non-coding sliding windows in Table S2. **B.** Average Eigen score for significant SKAT non-coding windows. The histogram represents the average Eigen score for all non-coding sliding windows. The vertical red lines plot the average Eigen score for the significant non-coding sliding windows in Table S3.

Figure S12. Quantile-quantile (QQ) plots related to the results in Tables 1-5. **A.** QQ plots corresponding to Table 1 (T5 sliding window analyses) and also Supplemental Table S2. **B.** QQ plots corresponding to Table 2 (SKAT sliding window analyses) and also Supplemental Table S3. **C.** QQ plots corresponding to Table 3 (T5 regulatory domain analyses). **D.** QQ plots corresponding to Table 4 (SKAT regulatory domain analyses). **E.** QQ plots corresponding to Table 5 (SKAT first intron analyses).

Figure S1.

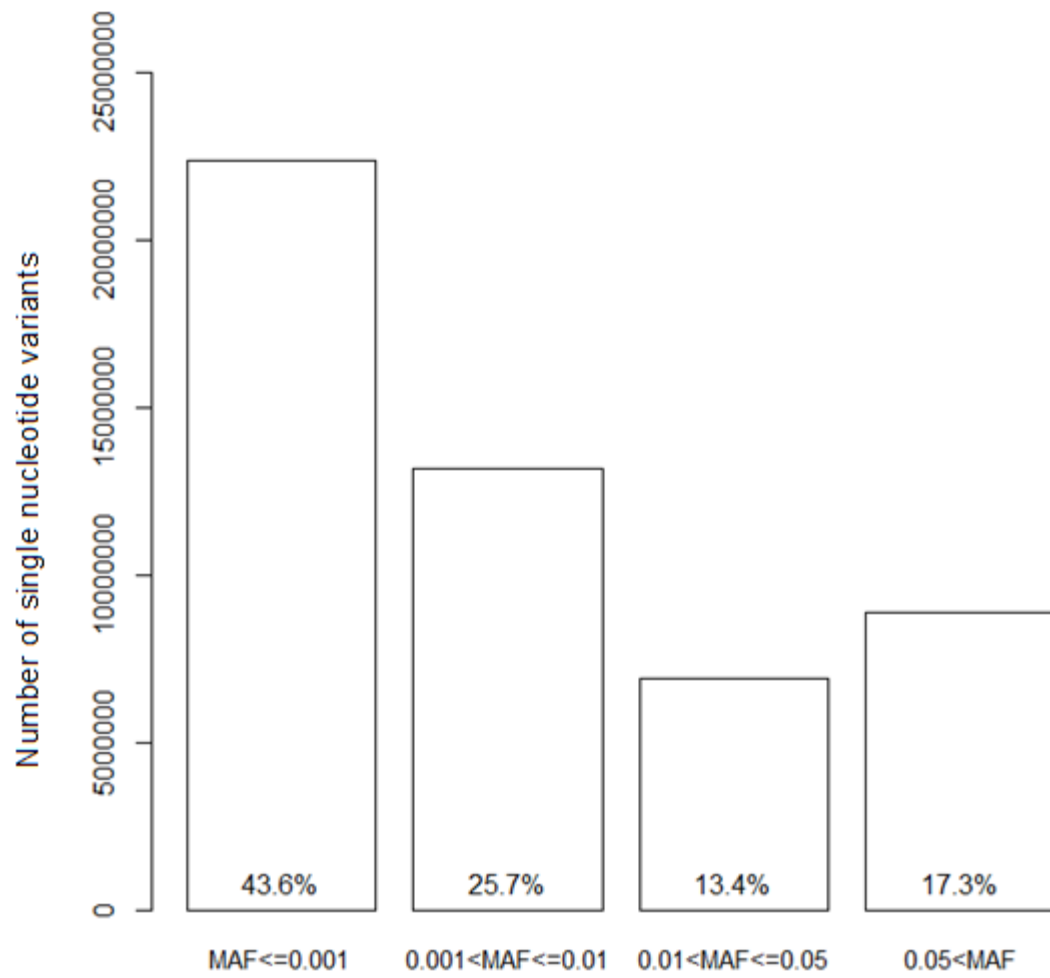


Figure S2.

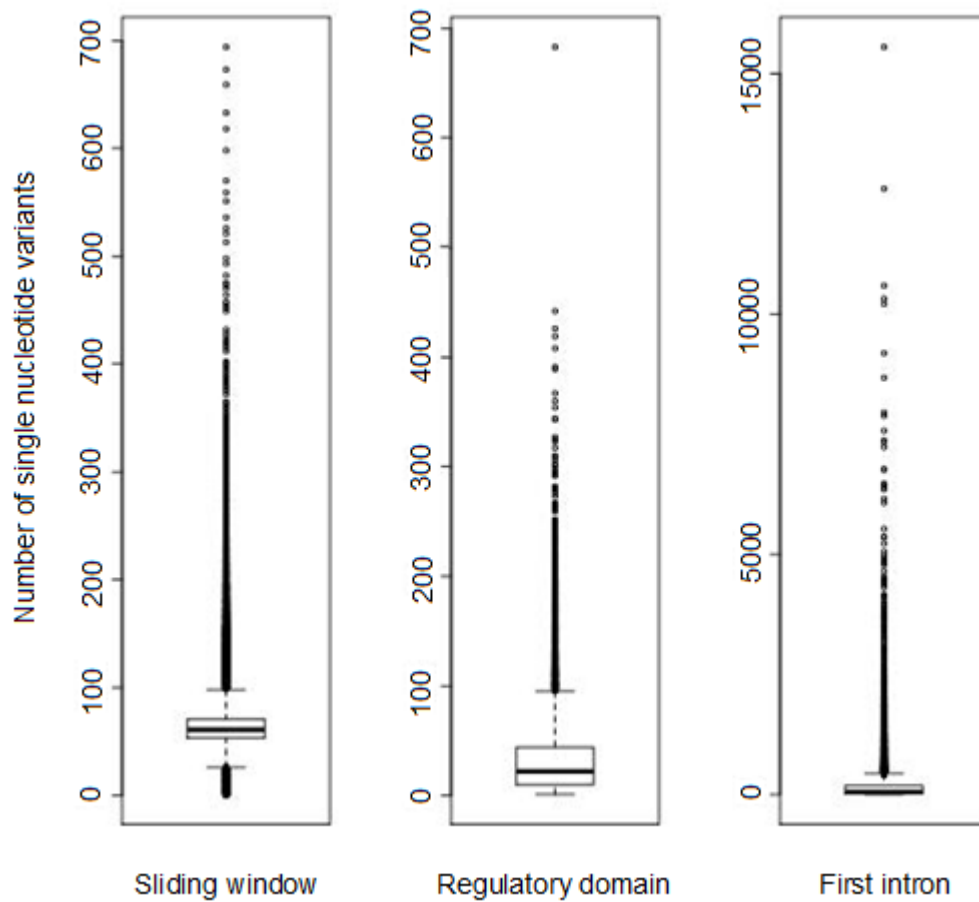


Figure S3.

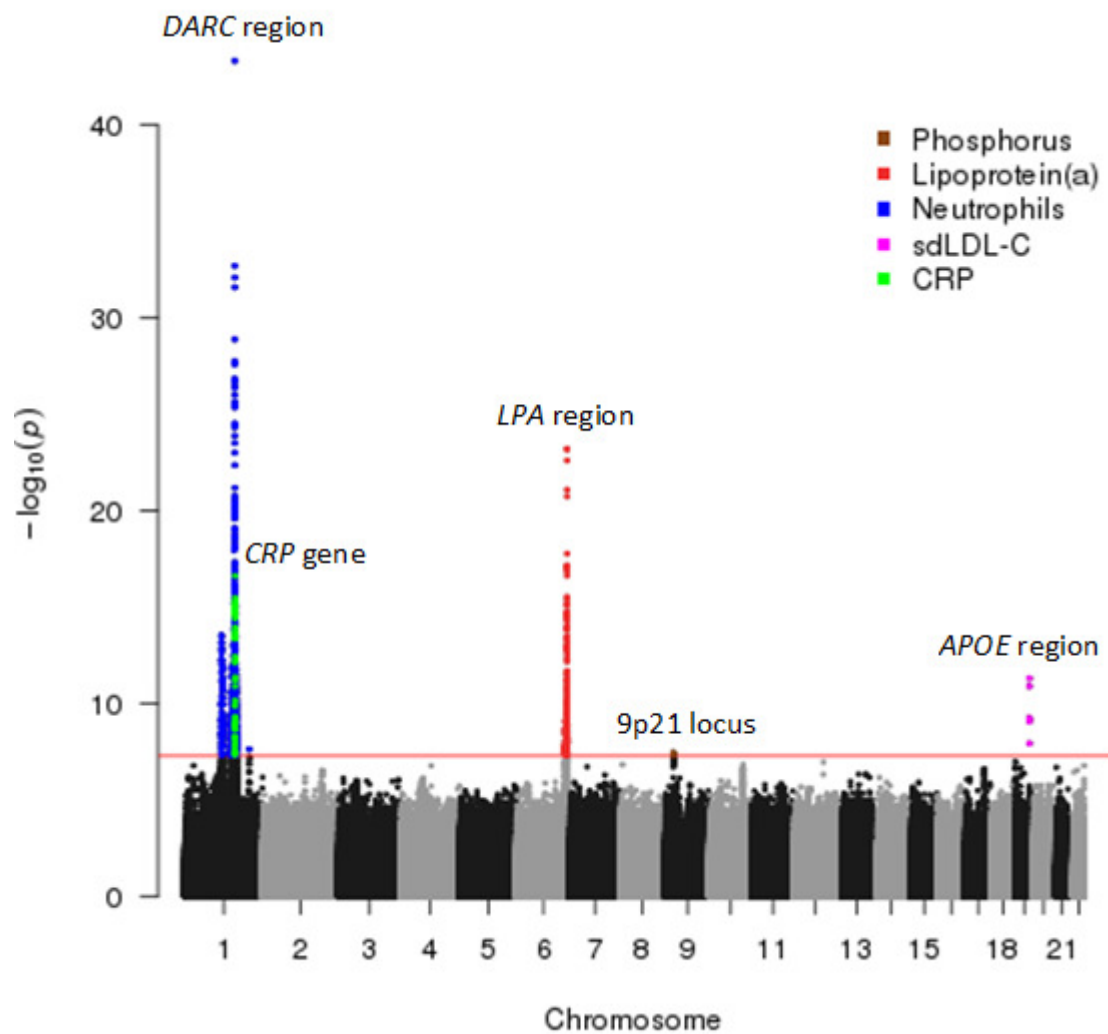


Figure S4.

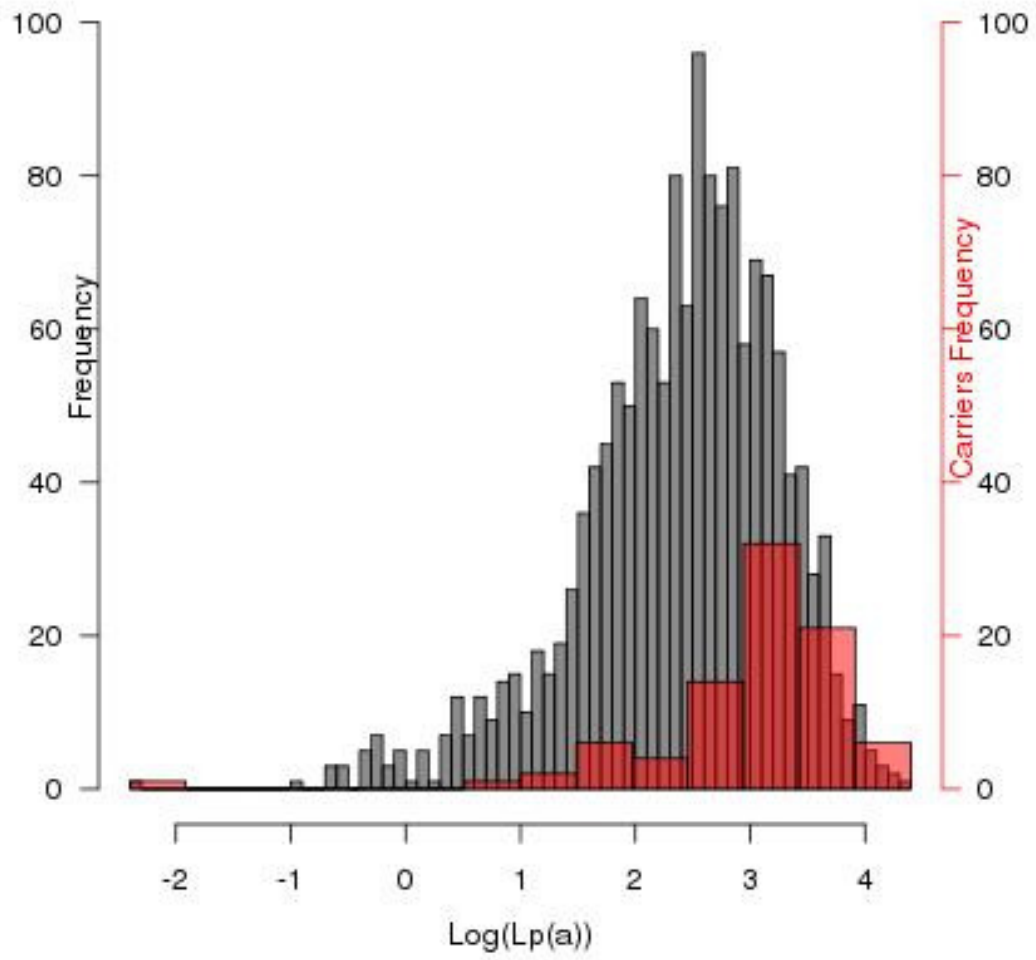


Figure S5.

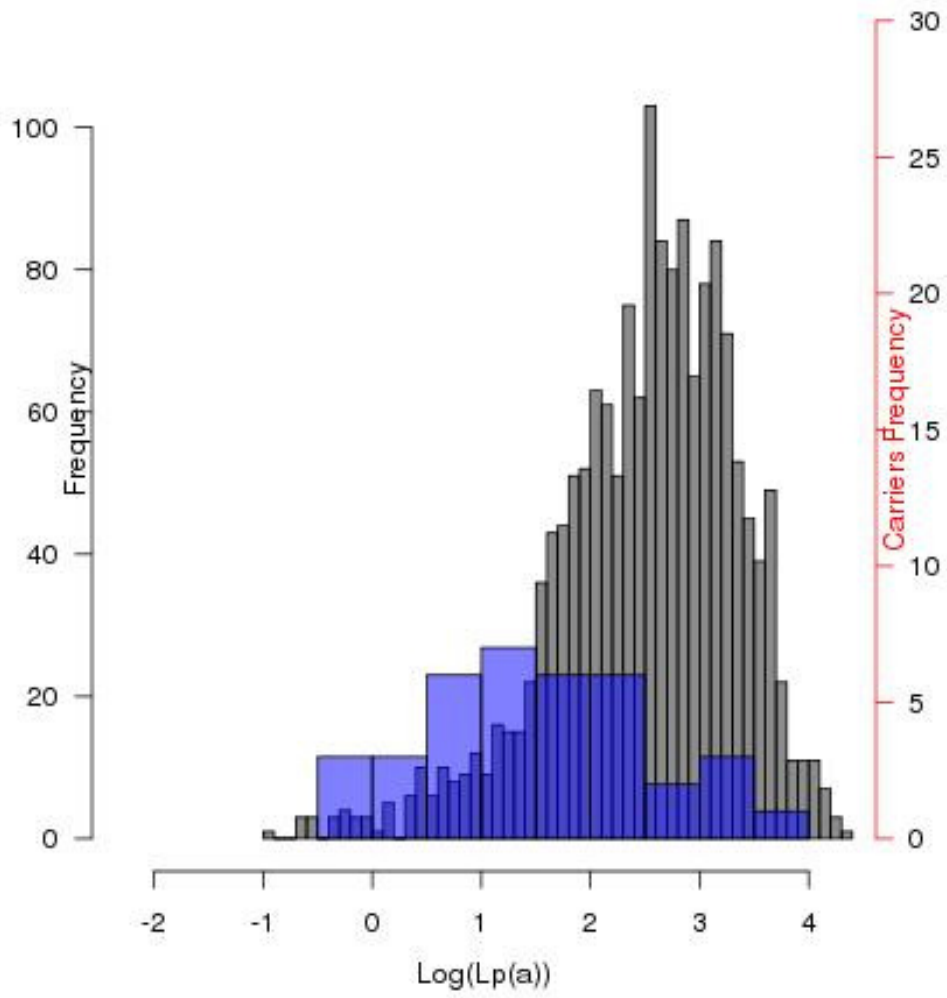


Figure S6.

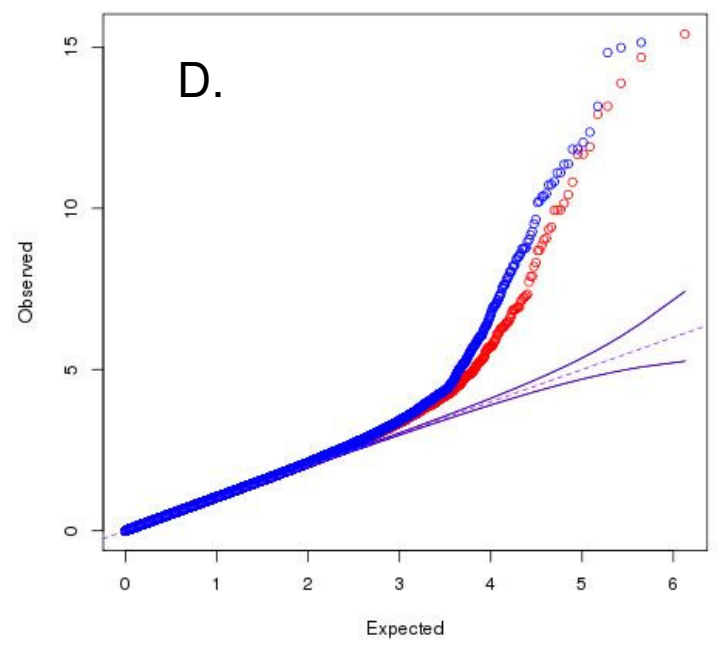
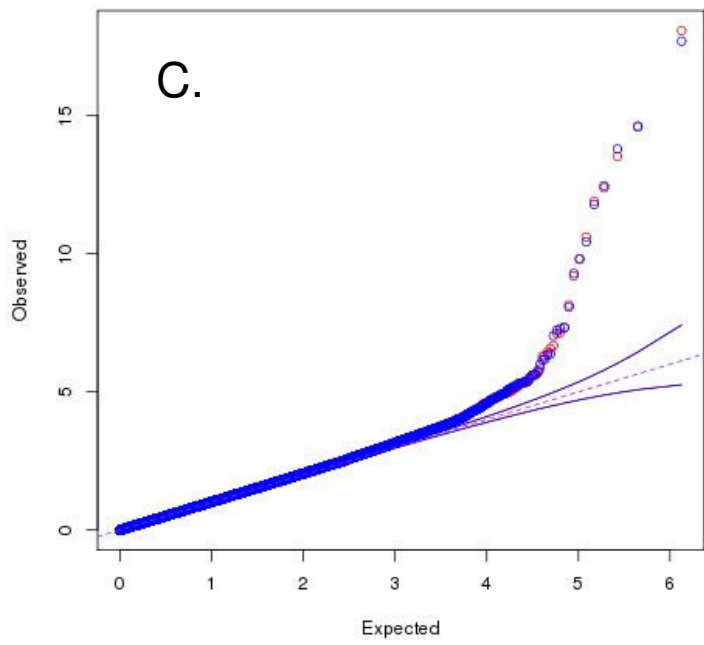
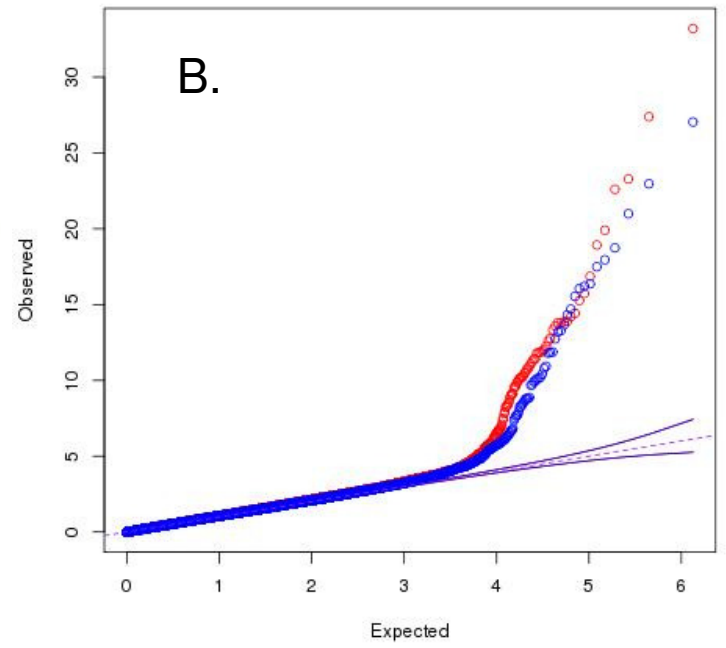
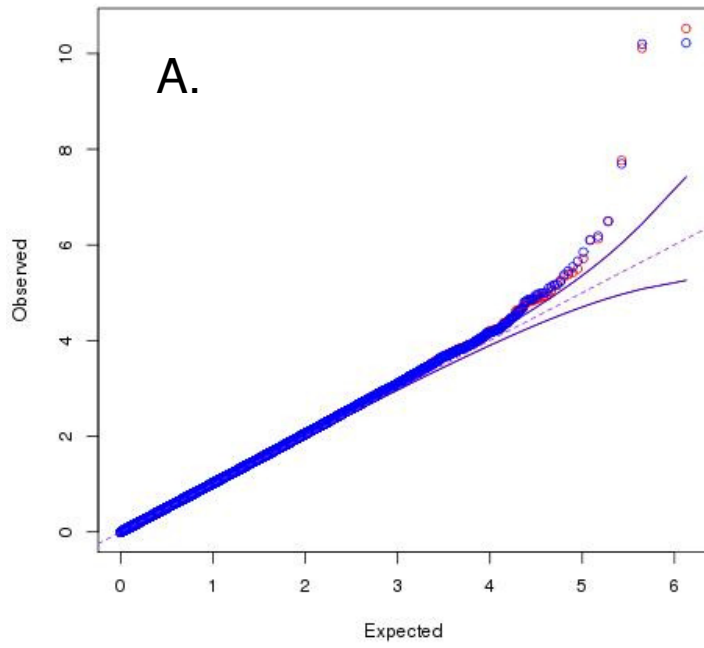


Figure S7.

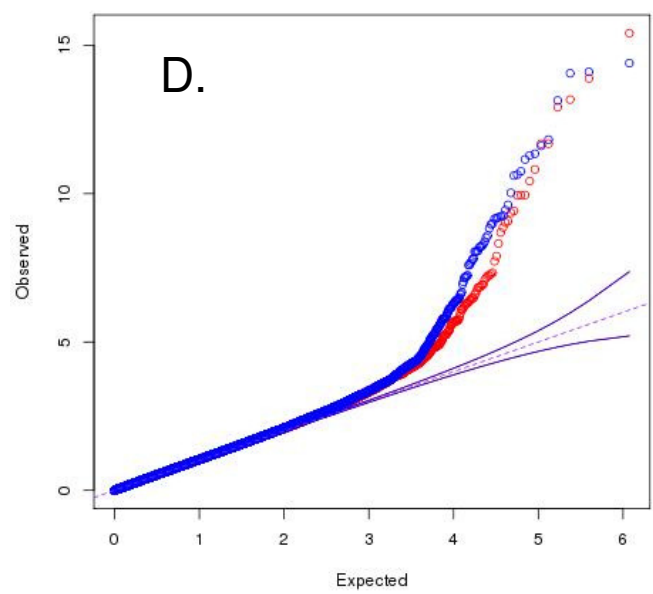
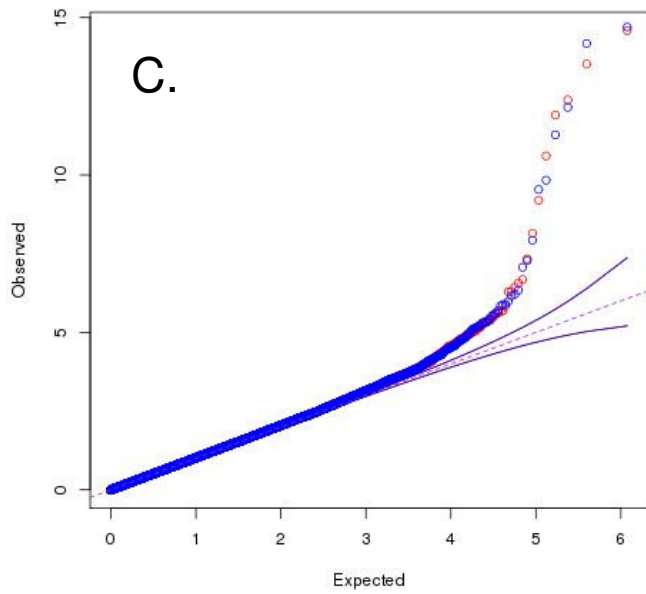
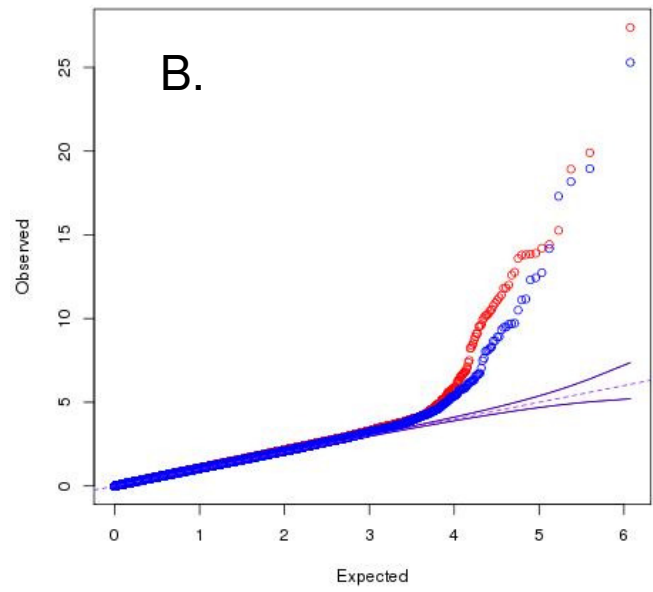
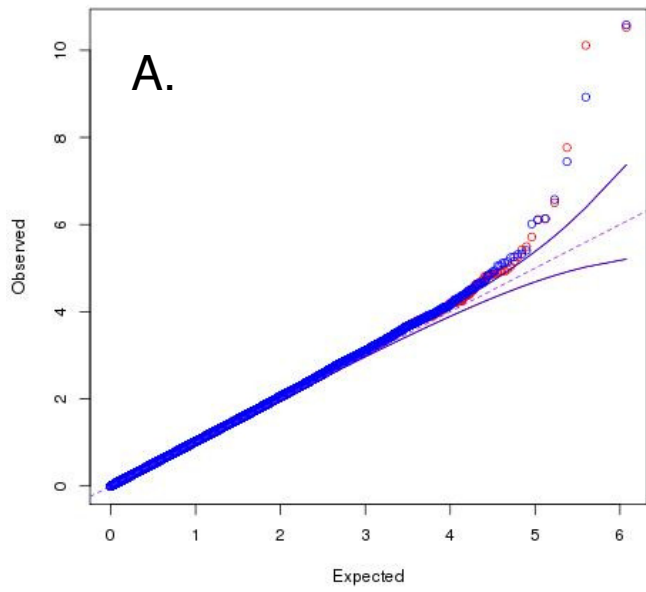


Figure S8.

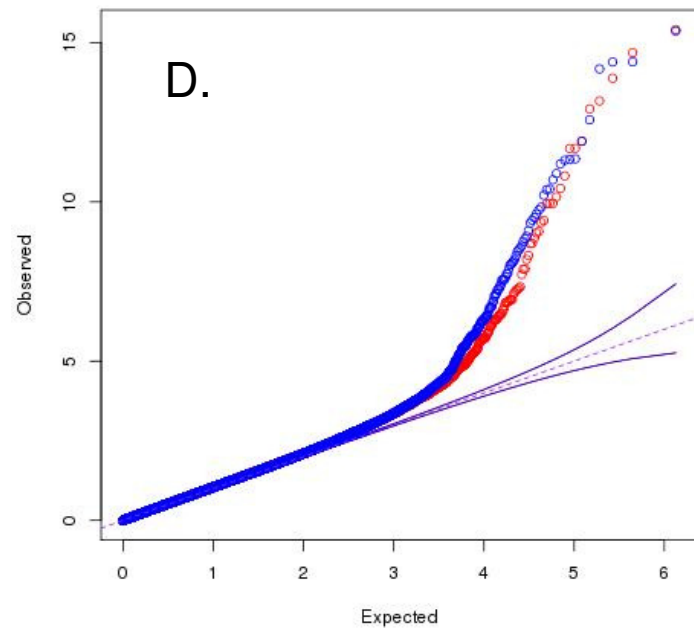
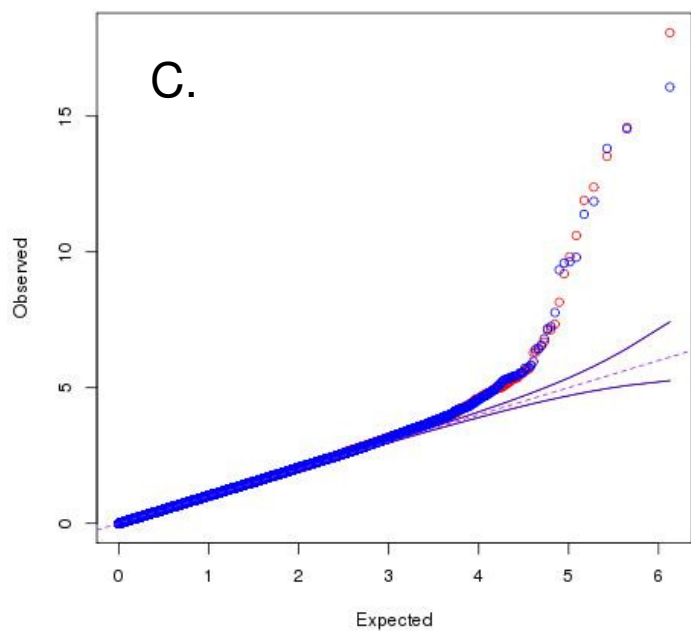
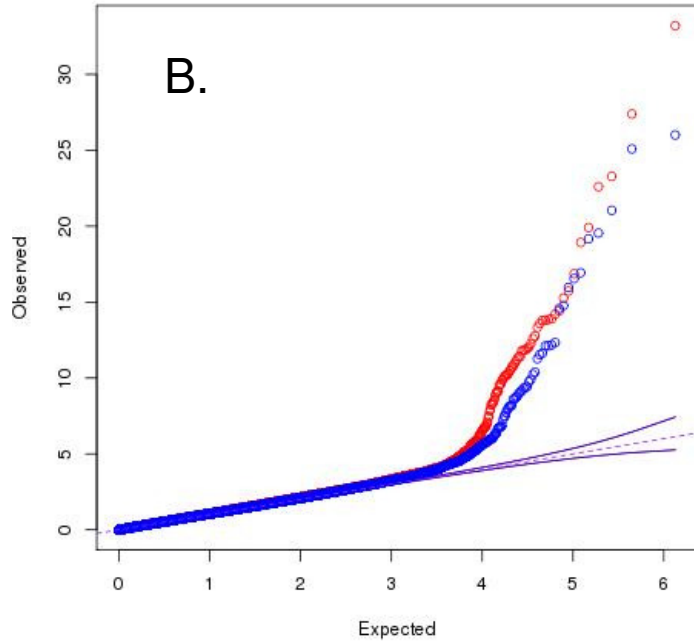
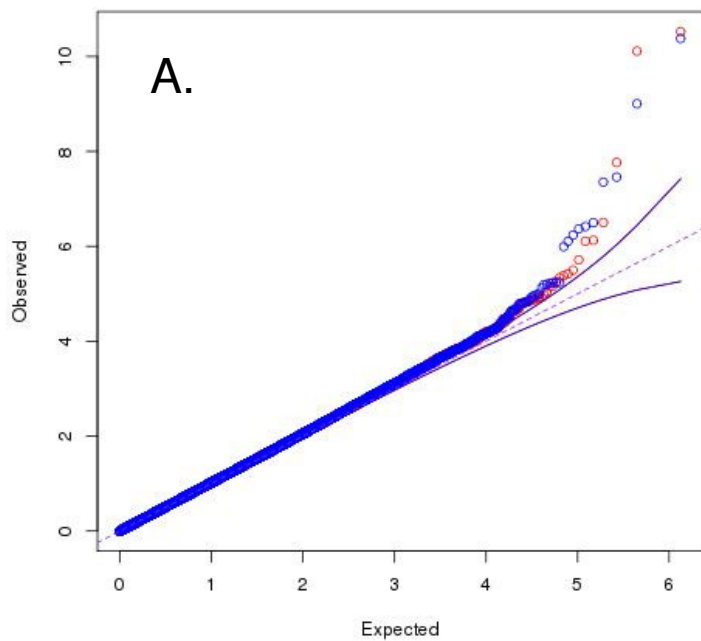
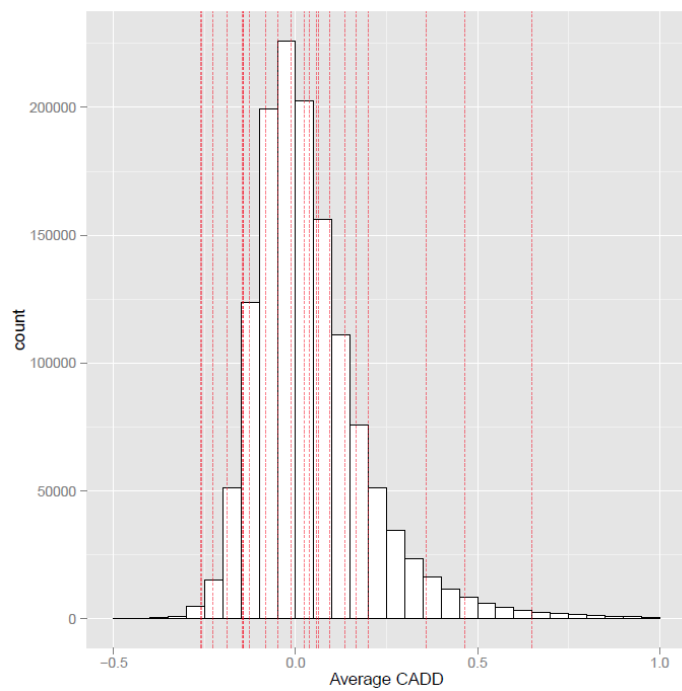


Figure S9.

A.



B.

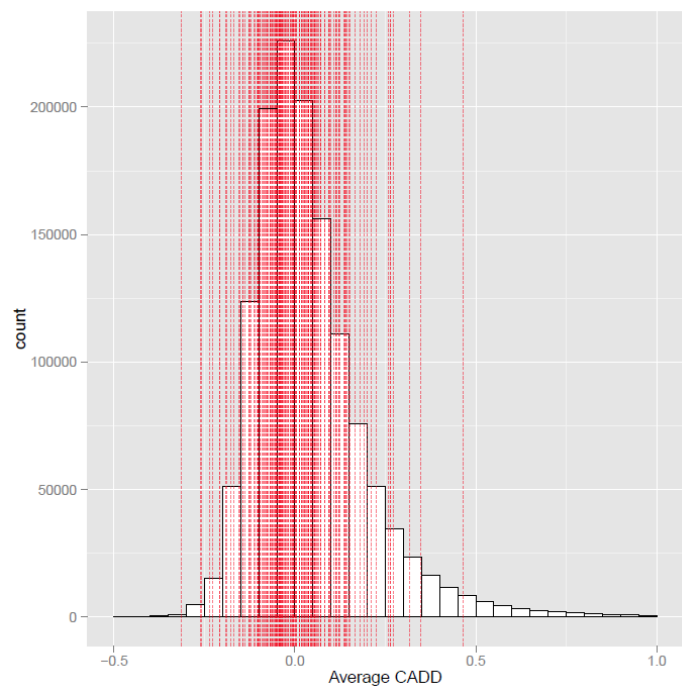
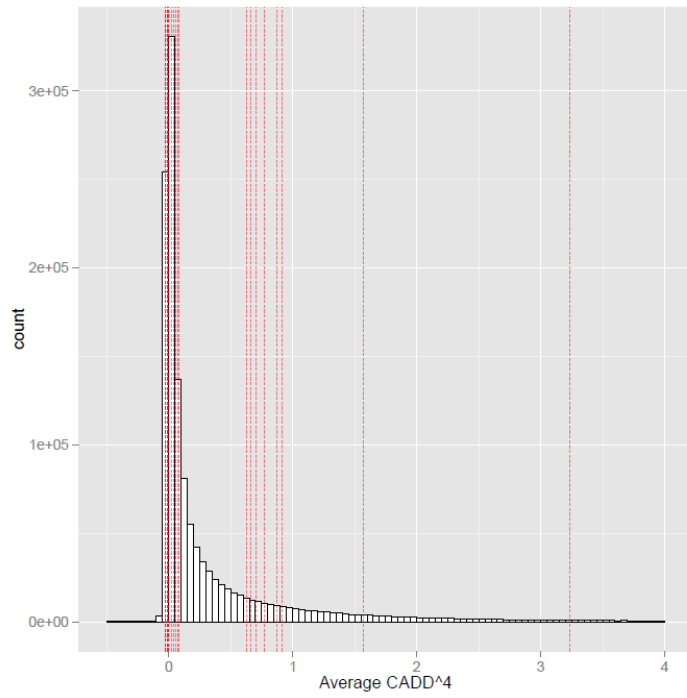


Figure S10.

A.



B.

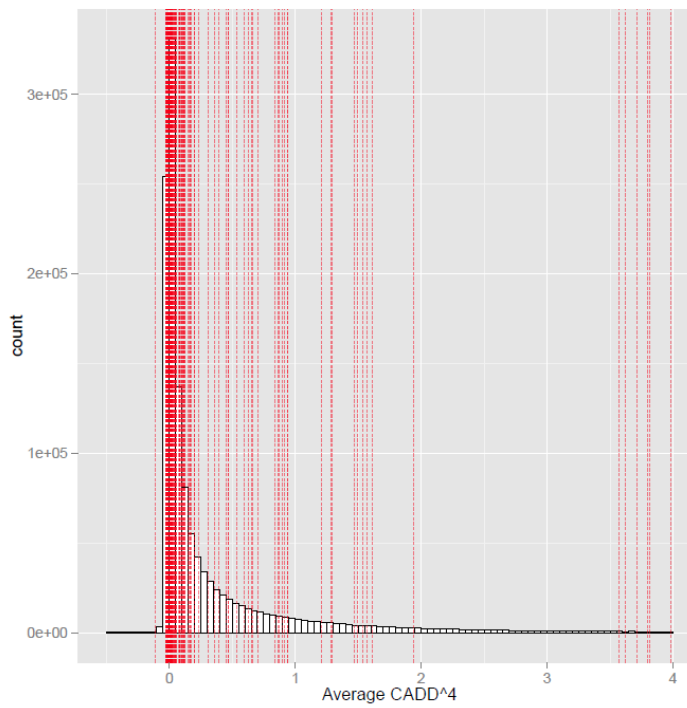
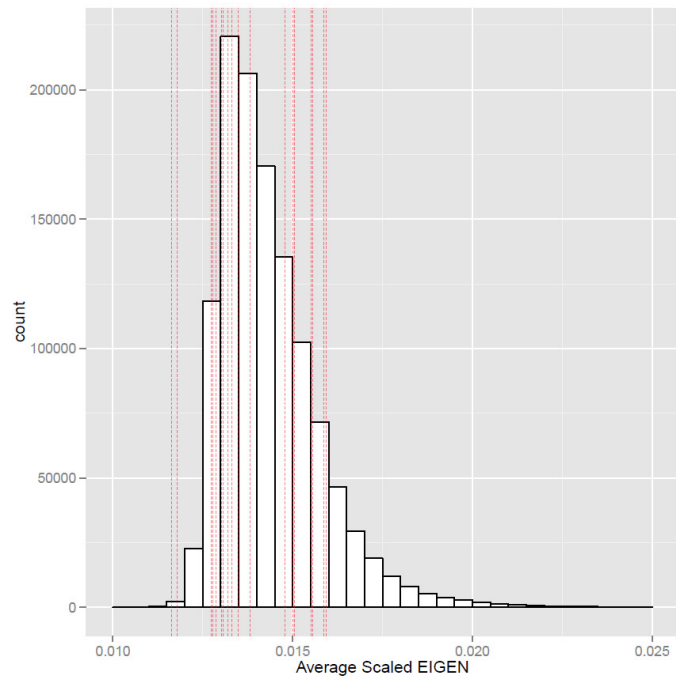


Figure S11.

A.



B.

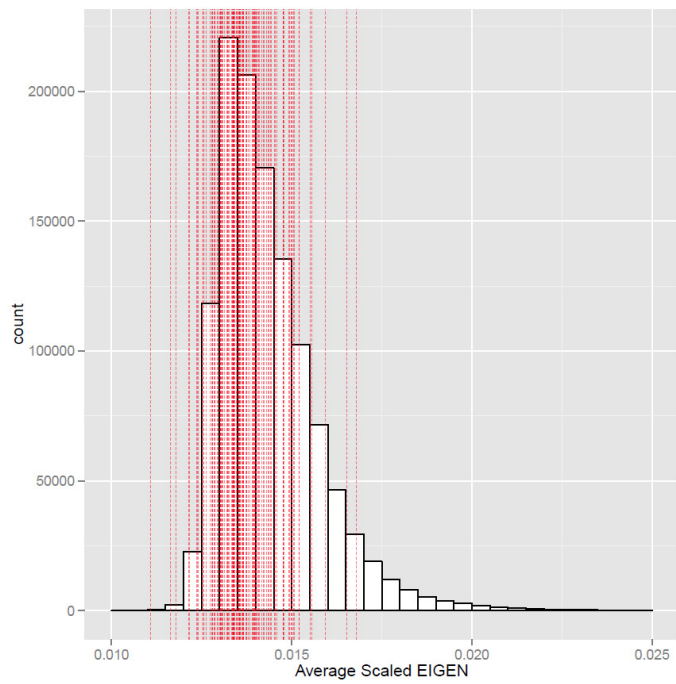
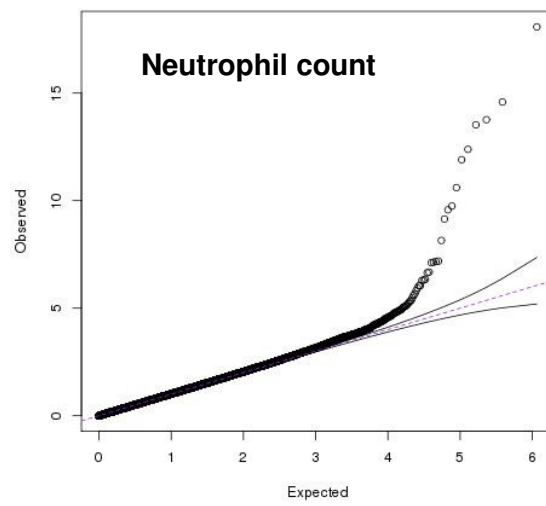
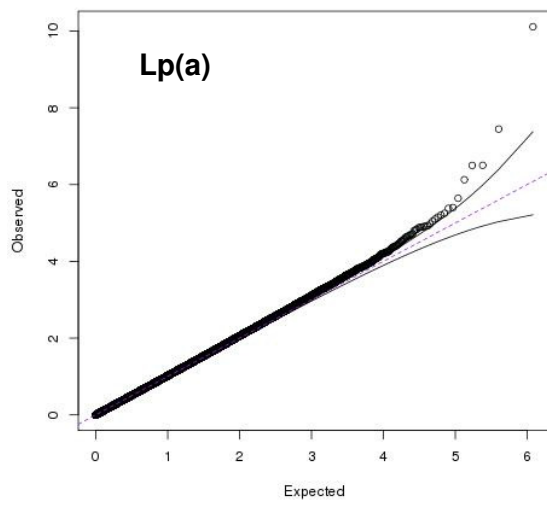
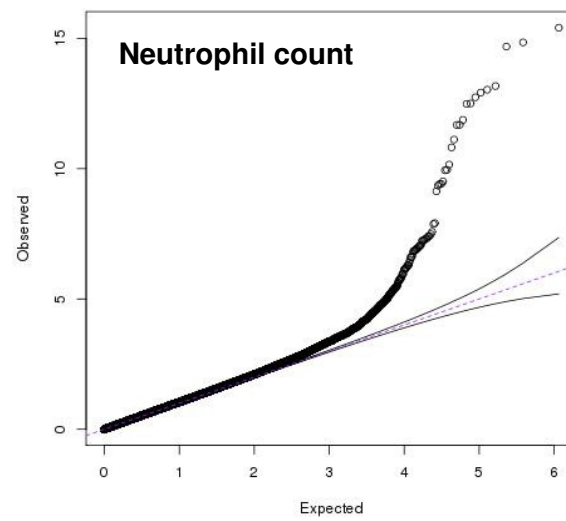
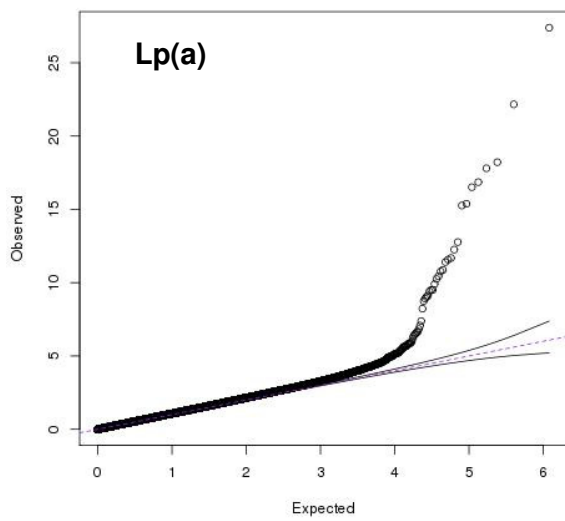


Figure S12.

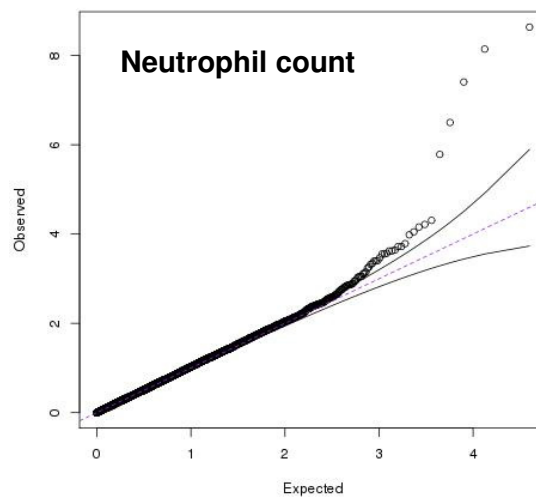
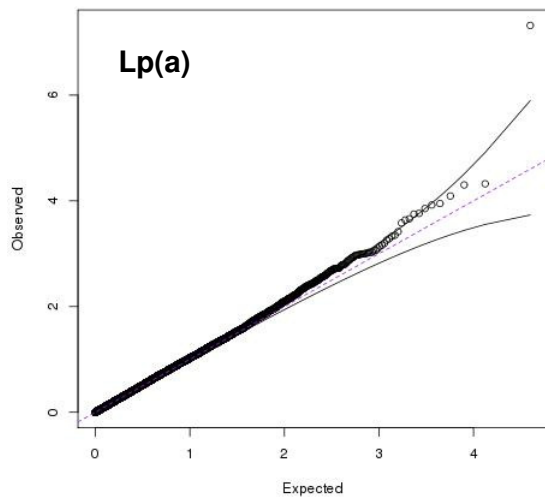
A.



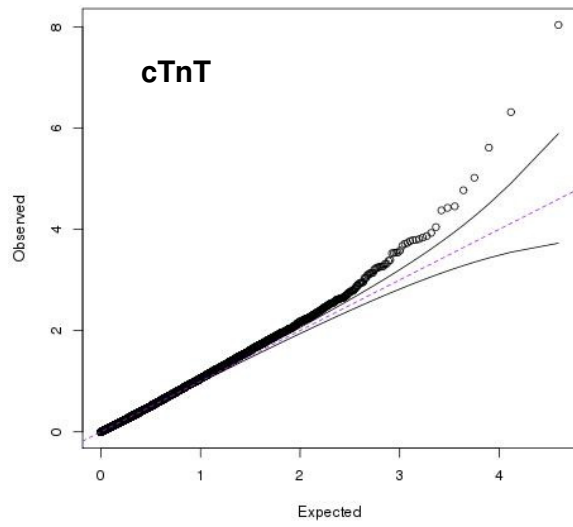
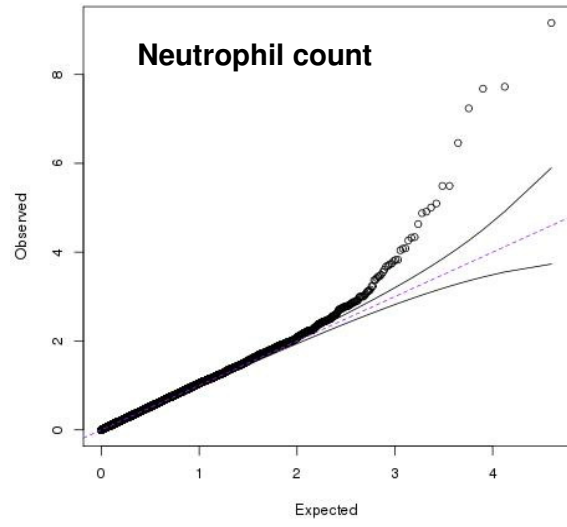
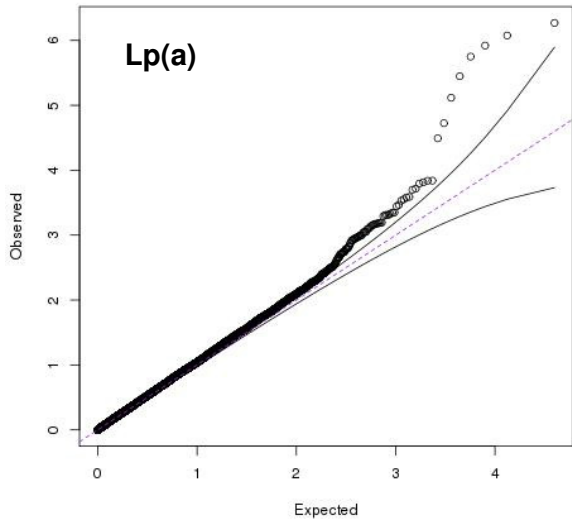
B.



C.



D.



E.

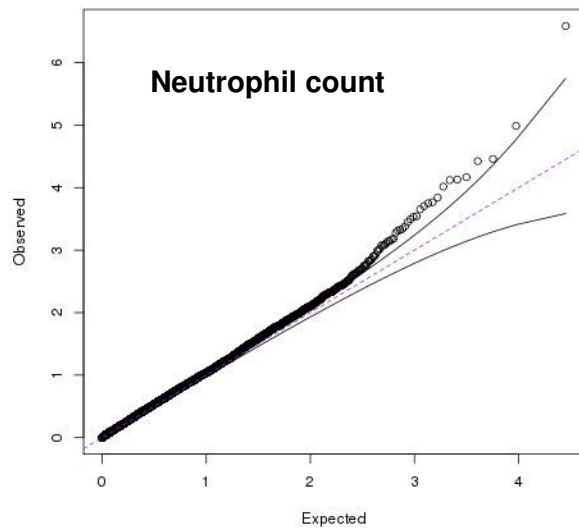
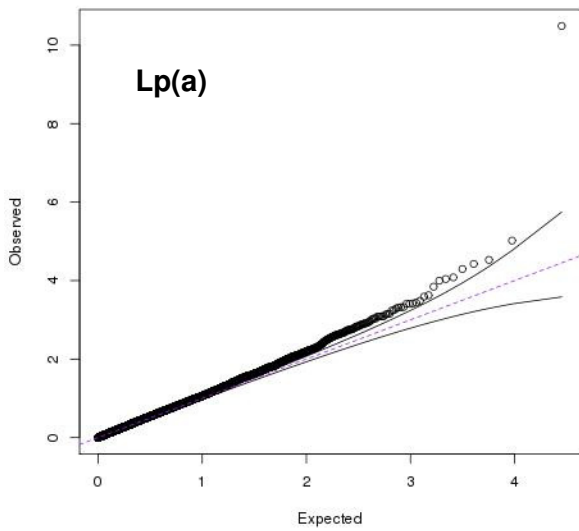


Table S1. Descriptive characteristics of the 10 heart and blood-related traits

Trait	EA (n=1,705)	AA (n=1,860)
Sex (% male)	47%	36%
Age (mean, SD)	54.9 (5.7)	52.8 (5.7)
BMI, kg/m ² (mean, SD)	27.3 (5)	29.7 (6.2)
Neutrophil count (mean, SD)	61.5 (8.2)	46.5 (12.8)
Platelet count (mean, SD)	261.1 (66.4)	257.5 (65.4)
Hemoglobin (mean, SD)	14.1 (1.3)	13.2 (1.5)
Lp(a), mg/dL (mean, SD)	7.9 (8.9)	15.8 (11.9)
sdLDL-c, mg/dL (mean, SD)	46.7 (21.5)	36.6 (17)
CRP, mg/L (mean, SD)	4.7 (6.9)	6.0 (8.4)
cTnT, ug/L (mean, SD)	0.008 (0.01)	0.007 (0.01)
NT-proBNP, pg/ml (mean, SD)	154.1 (268.9)	136.1 (1202.5)
Mg, meq/L (mean, SD)	1.6 (0.2)	1.6 (0.2)
P, mg/dL (mean, SD)	3.4 (0.5)	3.4 (0.5)

Table S4. Significant regulatory domain-based T5 test results in European Americans and conditional results in African Americans

Trait	Gene	European Americans					Conditional analysis				
		cMAF	# SNV	p-value	Beta	SE	cMAF	# SNV	p-value	Beta	SE
Lp(a)	<i>PLG</i>	0.074	27	0.1123	0.10	0.06	0.179	33	3.79E-09	-0.18	0.03
Neutrophil count	<i>DARC</i>	0.036	14	0.2764	-0.88	0.81	0.121	21	0.8635	0.09	0.50

Results shown for all associations with $p < 2.3E-7$ and $MAC_{\geq 3}$

cMAF = cumulative minor allele frequency

Table S5. Significant regulatory domain-based SKAT results in European Americans and conditional results in African Americans

Trait	Gene	European Americans			Conditional Analysis		
		cMAF	# SNV	p-value	cMAF	# SNV	p-value
cTnT	<i>CA9</i>	0.003	2	0.0388	-	-	-
Lp(a)	<i>MFAP5</i>	0.001	3	0.4034	-	-	-
Lp(a)	<i>SLC22A3</i>	0.077	18	0.0044	0.231	35	1.75E-05
Neutrophil count	<i>MNDA</i>	0.132	19	0.1750	0.324	38	0.0231
Neutrophil count	<i>IFI16</i>	0.156	28	0.1957	0.412	54	0.0235
Neutrophil count	<i>HSPA6</i>	0.226	32	0.3427	0.46	65	0.9275
Neutrophil count	<i>DARC</i>	0.036	14	0.3503	0.121	21	0.3927
Neutrophil count	<i>CADM3</i>	0.099	41	0.2484	0.204	54	0.2607
Neutrophil count	<i>PYHIN1</i>	0.108	13	0.3017	0.30	34	0.1276

Results shown for all associations with $p < 2.3E-7$ and $MAC_{\geq 3}$

cMAF = cumulative minor allele frequency

Table S6. Significant first intron-based SKAT results in European Americans and conditional results in African Americans

Trait	Gene	European Americans			Conditional Analysis		
		cMAF	# SNV	p	cMAF	# SNV	p
Lp(a)	<i>SLC22A3</i>	0.697	219	0.0002	4.1943	757	4.03E-10
Neutrophil count	<i>EFNA3</i>	0.127	19	0.7792	0.2499	64	0.0177

Results shown for all associations with $p < 3.5E-7$ and $MAC_{\geq 3}$

cMAF = cumulative minor allele frequency

Table S7. The sentinel SNVs with the lowest p-value for each trait in African Americans

Trait	Chr	Position	rs#	Allele 1 / Allele 2	MAF	Genomic location	Gene	P	Beta	SE
Neutrophil count	1	159174683	rs2814778	T/C	0.17	5' UTR	<i>DARC</i>	4×10^{-44}	-8.35	0.6
CRP	1	159708818	rs112563958	C/T	0.18	intergenic	<i>CRP</i>	2×10^{-17}	0.48	0.06
Lp(a)	6	161019440	rs41271018	A/C	0.05	intronic	<i>LPA</i>	6×10^{-24}	0.67	0.07
P	9	24853029	rs60456827	A/G	0.15	intergenic	9p21	3×10^{-8}	-0.12	0.02
sdLDL-c	19	45413233	rs1065853	G/T	0.12	upstream	<i>APOE</i>	4×10^{-12}	-7.59	1.1

Chr = chromosome

Base pair (bp) position based on hg19

Table S12. Results from sliding window-based test for Lp(a) in African Americans, for the region 6q25.3-6q26 adjusted for rs115848955

Test	Start position (bp)	Stop position (bp)	cMAF	# SNV	p-value*
T5	160990009	160994008	0.159	43	2.72E-12
T5	161030009	161034008	0.473	95	2.23E-10
T5	161070009	161074008	0.499	92	3.13E-09
T5	161074009	161078008	0.281	52	3.39E-09
SKAT	160988009	160992008	0.159	43	7.11E-11
SKAT	160990009	160994008	0.461	84	7.56E-11
SKAT	160992009	160996008	0.540	72	8.00E-12
SKAT	160994009	160998008	0.461	55	1.26E-10
SKAT	161000009	161004008	0.488	73	7.48E-09
SKAT	161002009	161006008	0.546	73	5.50E-13
SKAT	161004009	161008008	0.515	71	2.27E-17
SKAT	161006009	161010008	0.577	79	2.97E-22
SKAT	161008009	161012008	0.457	77	7.70E-25
SKAT	161010009	161014008	0.336	70	7.83E-22
SKAT	161014009	161018008	0.450	71	6.73E-11
SKAT	161016009	161020008	0.320	69	3.58E-12
SKAT	161018009	161022008	0.342	76	6.64E-16
SKAT	161020009	161024008	0.458	64	2.45E-22
SKAT	161022009	161026008	0.573	62	1.87E-17
SKAT	161024009	161028008	0.491	78	5.06E-12
SKAT	161026009	161030008	0.224	85	7.51E-10
SKAT	161028009	161032008	0.343	88	1.56E-11
SKAT	161030009	161034008	0.473	95	1.99E-12
SKAT	161052009	161056008	0.352	49	2.52E-11
SKAT	161054009	161058008	0.314	54	3.15E-11
SKAT	161068009	161072008	0.499	92	8.50E-13
SKAT	161070009	161074008	0.444	54	2.52E-15
SKAT	161072009	161076008	0.383	42	9.16E-12
SKAT	161074009	161078008	0.281	52	3.17E-14
SKAT	161076009	161080008	0.121	32	7.93E-13

*reporting only associations $p < 7.5 \times 10^{-9}$ after conditional analysis

References

1. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., et al. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* 41, 666-676.
2. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labruno, Y., et al. (2011). New gene functions in megakaryopoiesis and platelet formation. *Nature* 480, 201-208.
3. Reich, D., Nalls, M.A., Kao, W.H., Akyzbekova, E.L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.C., Cheng, C.Y., Coresh, J., et al. (2009). Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet* 5, e1000360.
4. The ARIC Investigators. (1987). ARIC: Atherosclerosis Risk in Communities Study: operations manual 10. Clinical chemistry determinations. Bethesda, MD: National Heart, Lung, and Blood Institute.
5. Lutsey, P.L., Alonso, A., Michos, E.D., Loehr, L.R., Astor, B.C., Coresh, J., and Folsom, A.R. (2014). Serum magnesium, phosphorus, and calcium are associated with risk of incident heart failure: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Clin Nutr* 100, 756-764.
6. Kestenbaum, B., Glazer, N.L., Kottgen, A., Felix, J.F., Hwang, S.J., Liu, Y., Lohman, K., Kritchevsky, S.B., Hausman, D.B., Petersen, A.K., et al. (2010). Common genetic variants associate with serum phosphorus concentration. *J Am Soc Nephrol* 21, 1223-1232.
7. Chambless, L.E., McMahon, R.P., Brown, S.A., Patsch, W., Heiss, G., and Shen, Y.L. (1992). Short-term intraindividual variability in lipoprotein measurements: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Epidemiol* 136, 1069-1081.
8. Virani, S.S., Brautbar, A., Davis, B.C., Nambi, V., Hoogeveen, R.C., Sharrett, A.R., Coresh, J., Mosley, T.H., Morrisett, J.D., Catellier, D.J., et al. (2012). Associations between lipoprotein(a) levels and cardiovascular outcomes in black and white subjects: the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* 125, 241-249.
9. Giannitsis, E., Kurz, K., Hallermayer, K., Jarausch, J., Jaffe, A.S., and Katus, H.A. (2010). Analytical validation of a high-sensitivity cardiac troponin T assay. *Clin Chem* 56, 254-261.
10. Yu, B., Barbalic, M., Brautbar, A., Nambi, V., Hoogeveen, R.C., Tang, W., Mosley, T.H., Rotter, J.I., deFilippi, C.R., O'Donnell, C.J., et al. (2013). Association of genome-wide variation with highly sensitive cardiac troponin-T levels in European Americans and Blacks: a meta-analysis from atherosclerosis risk in communities and cardiovascular health studies. *Circ Cardiovasc Genet* 6, 82-88.
11. Olsen, M.H., Hansen, T.W., Christensen, M.K., Gustafsson, F., Rasmussen, S., Wachtell, K., Ibsen, H., Torp-Pedersen, C., and Hildebrandt, P.R. (2007). N-terminal pro-brain natriuretic peptide, but not high sensitivity C-reactive protein, improves cardiovascular risk prediction in the general population. *Eur Heart J* 28, 1374-1381.
12. Schick, U.M., Auer, P.L., Bis, J.C., Lin, H., Wei, P., Pankratz, N., Lange, L.A., Brody, J., Stitzel, N.O., Kim, D.S., et al. (2015). Association of exome sequences with plasma C-reactive protein levels in >9000 participants. *Hum Mol Genet* 24, 559-571.
13. Hoogeveen, R.C., Gaubatz, J.W., Sun, W., Dodge, R.C., Crosby, J.R., Jiang, J., Couper, D., Virani, S.S., Kathiresan, S., Boerwinkle, E., et al. (2014). Small dense low-density lipoprotein-cholesterol concentrations predict risk for coronary heart disease: the Atherosclerosis Risk In Communities (ARIC) study. *Arterioscler Thromb Vasc Biol* 34, 1069-1077.