

RESEARCH ARTICLE

Open Access

Genome-wide association analysis of total cholesterol and high-density lipoprotein cholesterol levels using the Framingham Heart Study data

Li Ma¹, Jing Yang¹, H Birali Runesha², Toshiko Tanaka^{3,4}, Luigi Ferrucci⁴, Stefania Bandinelli⁵ and Yang Da^{*1}

Abstract

Background: Cholesterol concentrations in blood are related to cardiovascular diseases. Recent genome-wide association studies (GWAS) of cholesterol levels identified a number of single-locus effects on total cholesterol (TC) and high-density lipoprotein cholesterol (HDL-C) levels. Here, we report single-locus and epistasis SNP effects on TC and HDL-C using the Framingham Heart Study (FHS) data.

Results: Single-locus effects and pairwise epistasis effects of 432,096 SNP markers were tested for their significance on log-transformed TC and HDL-C levels. Twenty nine additive SNP effects reached single-locus genome-wide significance ($p < 7.2 \times 10^{-8}$) and no dominance effect reached genome-wide significance. Two new gene regions were detected, the *RAB3GAP1-R3HDM1-LCT-MCM6* region of chr02 for TC identified by six new SNPs, and the *OSBPL8-ZDHHC17* region (chr12) for HDL-C identified by one new SNP. The remaining 22 single-locus SNP effects confirmed previously reported genes or gene regions. For TC, three SNPs identified two gene regions that were tightly linked with previously reported genes associated with TC, including rs599839 that was 10 bases downstream *PSRC1* and 3.498 kb downstream *CELSR2*, rs4970834 in *CELSR2*, and rs4245791 in *ABCG8* that slightly overlapped with *ABCG5*. For HDL-C, *LPL* was confirmed by 12 SNPs 8-45 kb downstream, *CETP* by two SNPs 0.5-11 kb upstream, and the *LIPG-ACAA2* region by five SNPs inside this region. Two epistasis effects on TC and thirteen epistasis effects on HDL-C reached the significance of "suggestive linkage". The most significant epistasis effect ($p = 5.72 \times 10^{-13}$) was close to reaching "significant linkage" and was a dominance \times dominance effect of HDL-C between *LMBRD1* (chr06) and the *LRIIG3* region (chr12), and this pair of gene regions had six other D \times D effects with "suggestive linkage".

Conclusions: Genome-wide association analysis of the FHS data detected two new gene regions with genome-wide significance, detected epistatic SNP effects on TC and HDL-C with the significance of suggestive linkage in seven pairs of gene regions, and confirmed some previously reported gene regions associated with TC and HDL-C.

Background

Total cholesterol (TC) is related to coronary diseases and high-density lipoprotein (HDL-C) cholesterol is anti-atherogenic. Genome-wide association studies (GWAS) and human genetic studies have identified a number of genes and gene regions affecting cholesterol phenotypes including TC and HDL-C [1-11]. A meta-analysis of HDL-C levels that include the FHS data has previously

been published [2]. An early report on FHS [12] analyzed TC and HDL-C but used 100 k SNPs and a sample size that was much smaller than the current FHS sample size. Epistasis analysis of TC and HDL-C was unavailable. Here, we apply a quantitative genetics approach to detect additive or dominance single-locus effects and epistasis effects on log-transformed TC and HDL-C using 432,096 SNP markers and over 6000 individuals in FHS. The epistasis effects we tested included additive \times additive (A \times A), additive \times dominance (A \times D) or dominance \times additive (D \times A), and dominance \times dominance (D \times D) effects,

* Correspondence: yda@umn.edu

¹ Department of Animal Science, University of Minnesota, USA
Full list of author information is available at the end of the article

with genetic interpretations of allele × allele, allele × genotype or genotype × allele, and genotype × genotype interactions. The single-locus analysis was intended to detect new targets or confirm existing targets using a method of analysis different from those used in previous reports based on an extended Kempthorne model that allows Hardy-Weinberger disequilibrium and linkage disequilibrium [13] for GWAS analysis of the FHS data while the epistasis analysis of TC and HDL-C was the first such attempt using the FHS data and the 500 k SNP panel.

Results

The single-locus tests detected nine SNPs with additive (or allelic) effects on TC and twenty SNPs with additive effects on HDL-C that reached genome-wide significance (Tables 1-2). No dominance effect reached genome-wide significance. Among the twenty nine SNP effects, twenty were new effects that were not reported in previous studies and nine were previously reported to be associated with various cholesterol phenotypes [1-12]. Seven SNPs identified two new gene regions while the remaining twenty two SNPs confirmed previously reported gene

regions. Two epistasis effects on TC and thirteen epistasis effects on HDL-C representing seven pairs of gene regions reached the significance of "suggestive linkage".

Single-locus effects

For TC, nine SNPs with additive (or allelic) effects reached genome-wide significance with $p < 7.2 \times 10^{-8}$ (Table 1). Six SNPs inside or near four genes identified a new chr02 region containing *RAB3GAP1*, *R3HDM1*, *LCT* and *MCM6* to be associated with TC (Figure 1A). Of the six SNPs in the *RAB3GAP1-R3HDM1-LCT-MCM6* region, five SNPs were inside genes and one SNP was 4.2 kb upstream *MCM6*. The most significant SNP in this region was rs2322660 in intron 12 of *LCT* (Table 1). The *RAB3GAP1-R3HDM1-LCT-MCM6* region contained two other genes (*ZRANB3* and *UBXD2*) that did not have significant SNPs. Eleven other SNPs spanning a 1.23 Mb region (Figure 1A) that includes *RAB3GAP1-R3HDM1-LCT-MCM6* had p-values between 1.27×10^{-5} and 7.13×10^{-7} , including one SNP upstream *ACMSD*, one SNP in *ACMSD*, two SNPs in *YSK4*, one SNP in *R3HDM1*, one SNP in *UBXD2* (also named *UBXN4* according to NCBI

Table 1: Single-locus SNP effects for TC with genome control (GC) adjusted $P < 7.2 \times 10^{-8}$.

SNP	Chr	Position	Gene Region	Reported SNP effect	MAF	Effect Type & P value		
						Genotype	Additive	Effect Size
rs4970834	1	109814880	<i>CELSR2</i> ^a (intron 28)	Non-HDL-C [5-7] TC [8]	0.18	1.10E-08	1.75E-09	0.146 ± 0.023
rs599839	1	109822166	10 bases downstream <i>PSRC1</i> ^a	LDL-C [6,7,9,10] Non-HDL-C [5]	0.22	8.72E-14	2.46E-14	0.174 ± 0.021
rs4245791	2	44074431	<i>ABCG8</i> (intron 3)		0.32	1.82E-07	3.33E-08	-0.127 ± 0.021
rs6730157	2	135907088	<i>RAB3GAP1</i> (intron 17)	LDL-C: P = 0.018 [2] ^b	0.45	8.51E-08	2.16E-08	0.113 ± 0.019
rs12465802	2	136381348	<i>R3HDM1</i> (intron 7)	LDL-C: P = 0.022 [2] ^b	0.44	2.63E-08	7.98E-09	0.117 ± 0.019
rs4954280	2	136420690	<i>R3HDM1</i> (intron18)	LDL-C: P = 0.007 [2] ^b	0.33	1.49E-07	5.87E-08	0.114 ± 0.02
rs2322660	2	136557319	<i>LCT</i> (intron 12)	LDL-C: P = 0.055 [2] ^b TC: P = 0.003-0.005 [17] LDL-C: P = 0.002-0.0005 [17]	0.35	2.42E-08	7.08E-09	-0.120 ± 0.019
rs309180	2	136614255	<i>MCM6</i> (intron 11)	LDL-C: P = 0.057 [2] ^b	0.36	2.43E-08	8.39E-09	-0.119 ± 0.019
rs632632	2	136638216	4.2 kb upstream <i>MCM6</i>	LDL-C: P = 0.216 [2] ^b	0.36	2.50E-08	1.03E-08	-0.118 ± 0.019

^a This gene was reported to be associated with TC [1].

^b Available at <http://www.sph.umich.edu/csg/abecasis/public/lipids2008/>

Table 2: Single-locus SNP effects for HDL-C with genome control (GC) adjusted $P < 7.2 \times 10^{-8}$.

SNP	Chr	Position	Gene Region	Reported SNP effect	MAF	Effect Type & P value		
						Genotype	Additive	Effect Size
rs17482753	8	19832646	8 kb downstream <i>LPL</i> ^a	Triglyceride [7]	0.10	1.27E-08	3.50E-09	-0.191 ± 0.031
rs10503669	8	19847690	23 kb downstream <i>LPL</i> ^a	HDL-C [4,11]	0.09	3.75E-08	1.14E-08	-0.189 ± 0.031
rs17410962	8	19848080	23 kb downstream <i>LPL</i> ^a		0.12	2.75E-08	4.27E-09	0.173 ± 0.028
rs17489268	8	19852045	27 kb downstream <i>LPL</i> ^a		0.27	2.28E-10	5.97E-11	0.141 ± 0.02
rs17411031	8	19852310	28 kb downstream <i>LPL</i> ^a	HDL-C [7]	0.27	3.06E-10	7.21E-11	-0.14 ± 0.02
rs17489282	8	19852518	28 kb downstream <i>LPL</i> ^a		0.25	2.33E-09	6.54E-10	0.14 ± 0.021
rs4922117	8	19852586	28 kb downstream <i>LPL</i> ^a		0.25	2.28E-09	7.73E-10	-0.124 ± 0.019
rs17411126	8	19855272	31 kb downstream <i>LPL</i> ^a		0.27	6.14E-10	1.23E-10	0.138 ± 0.02
rs765547	8	19866274	42 kb downstream <i>LPL</i> ^a		0.27	1.60E-10	3.07E-11	-0.142 ± 0.02
rs11986942	8	19867445	43 kb downstream <i>LPL</i> ^a		0.33	2.22E-07	5.53E-08	0.112 ± 0.02
rs1837842	8	19868290	44 kb downstream <i>LPL</i> ^a		0.27	2.14E-10	4.09E-11	0.142 ± 0.02
rs1919484	8	19869676	45 kb downstream <i>LPL</i> ^a		0.27	4.12E-10	7.21E-11	-0.143 ± 0.021
rs17259942	12	77072077	<i>OSBPL8-ZDHC17</i>		0.12	8.61E-08	1.81E-08	0.168 ± 0.028
rs9989419	16	56985139	<i>HERPUD1-CETP</i> ^a	HDL-C [4,7]	0.40	4.57E-13	5.96E-14	-0.147 ± 0.019
rs1800775	16	56995236	0.5 kb upstream <i>CETP</i> ^a	HDL-C [3,12]	0.45	1.54E-29	1.64E-30	0.242 ± 0.02
rs7240405	18	47159090	<i>LIPG</i> ^a - <i>ACAA2</i> ^b		0.16	8.01E-08	1.58E-08	-0.145 ± 0.024
rs4939883	18	47167214	<i>LIPG</i> ^a - <i>ACAA2</i> ^b	HDL-C [1,2]	0.17	1.07E-07	1.85E-08	0.142 ± 0.024
rs1943981	18	47169815	<i>LIPG</i> ^a - <i>ACAA2</i> ^b		0.17	1.49E-07	2.50E-08	0.142 ± 0.024
rs2156552	18	47181668	<i>LIPG</i> ^a - <i>ACAA2</i> ^b	HDL-C [3,4]	0.16	5.87E-08	1.14E-08	-0.146 ± 0.024
rs6507945	18	47243912	<i>LIPG</i> ^a - <i>ACAA2</i> ^b		0.43	4.34E-08	6.91E-09	0.113 ± 0.018

^aThis gene was reported in [1-4].

^bThis gene was reported in [3].

[14]), two SNPs in *LCT*, and three SNPs downstream *DARS* (data not shown). These less significant results in the same neighborhood should add to the significance of the *RAB3GAP1-R3HDM1-LCT-MCM6* region to TC. Three SNPs identified two genes that were tightly linked with previously reported genes associated with TC [1]. These three SNPs were rs599839 that was 10 bases downstream *PSRC1* (chr01) and 3.498 kb downstream *CELSR2*, rs4970834 in intron 28 of *CELSR2*, and rs4245791 in intron 3 of *ABCG8* (chr02) that slightly overlapped with *ABCG5*, where *CELSR2* and *ABCG5* regions were reported to be associated with TC in a recent GWAS report [1]. *PSRC1* and *ABCG8* also were reported to affect low-density lipoprotein cholesterol (LDL-C) [2-

4,6,7,9,10]. The SNP (rs599839) that was 10 bases downstream *PSRC1* had the most significant single-locus effect on TC ($p = 3.7 \times 10^{-16}$), while the SNP inside *CELSR2* (rs4970834) had the second most significant single-locus effect on TC ($p = 1.29 \times 10^{-10}$).

For HDL-C, twenty SNPs with additive effects reached genome-wide significance (Table 2). SNP rs17259942 identified a new gene region associated with HDL-C, the *OSBPL8-ZDHC17* region (q21.2, Figure 1B), with rs17259942 being 117 kb downstream *OSBPL8* and 85 kb upstream *ZDHC17* [15]. According to NCBI [14], the *OSBPL8-ZDHC17* region contained three pseudo-genes (*RPL7AP59*, *RPL21P98* and *RPL7P43*) and rs17259942 was 18 kb downstream *RPL21P98* and 43 kb

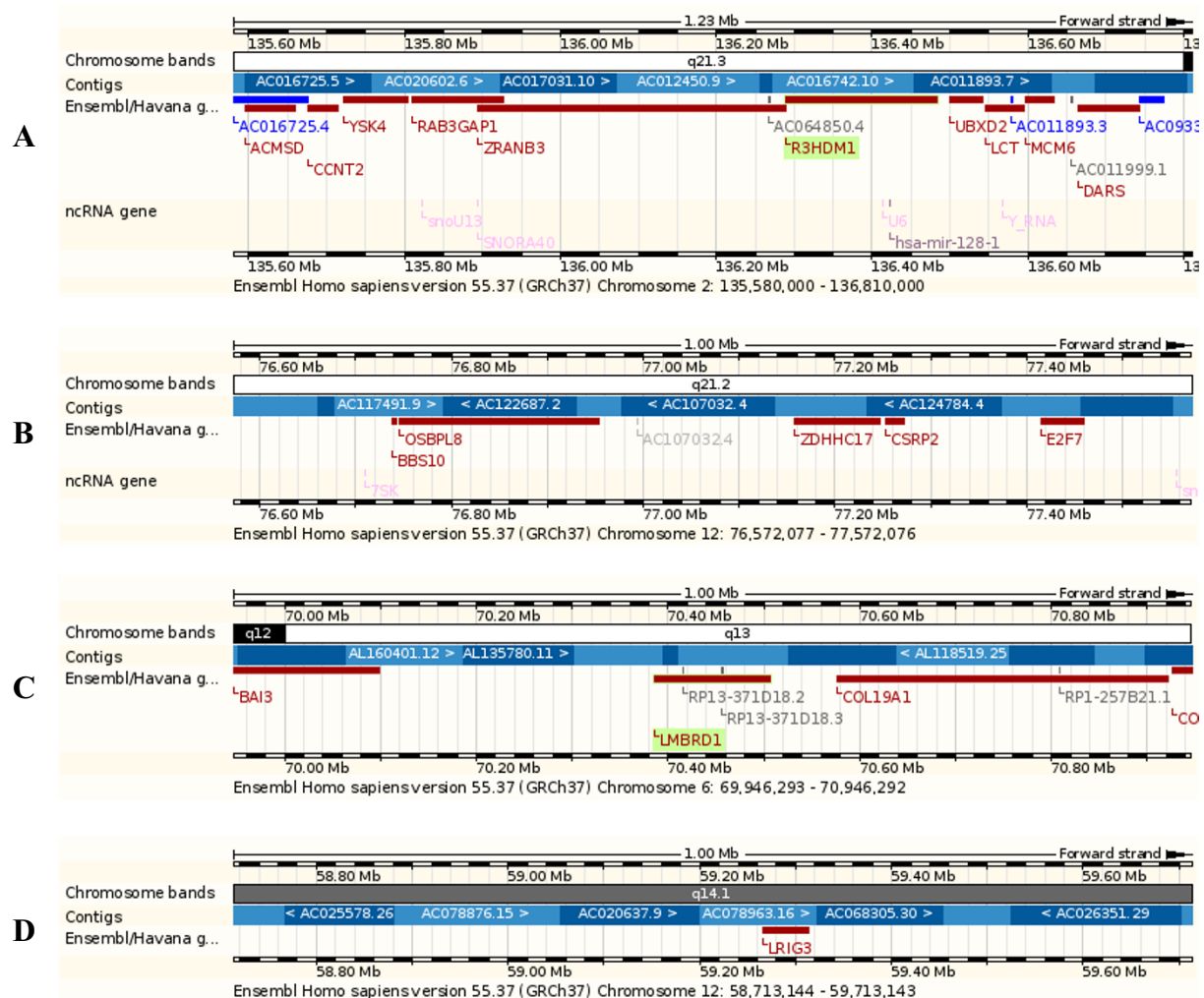


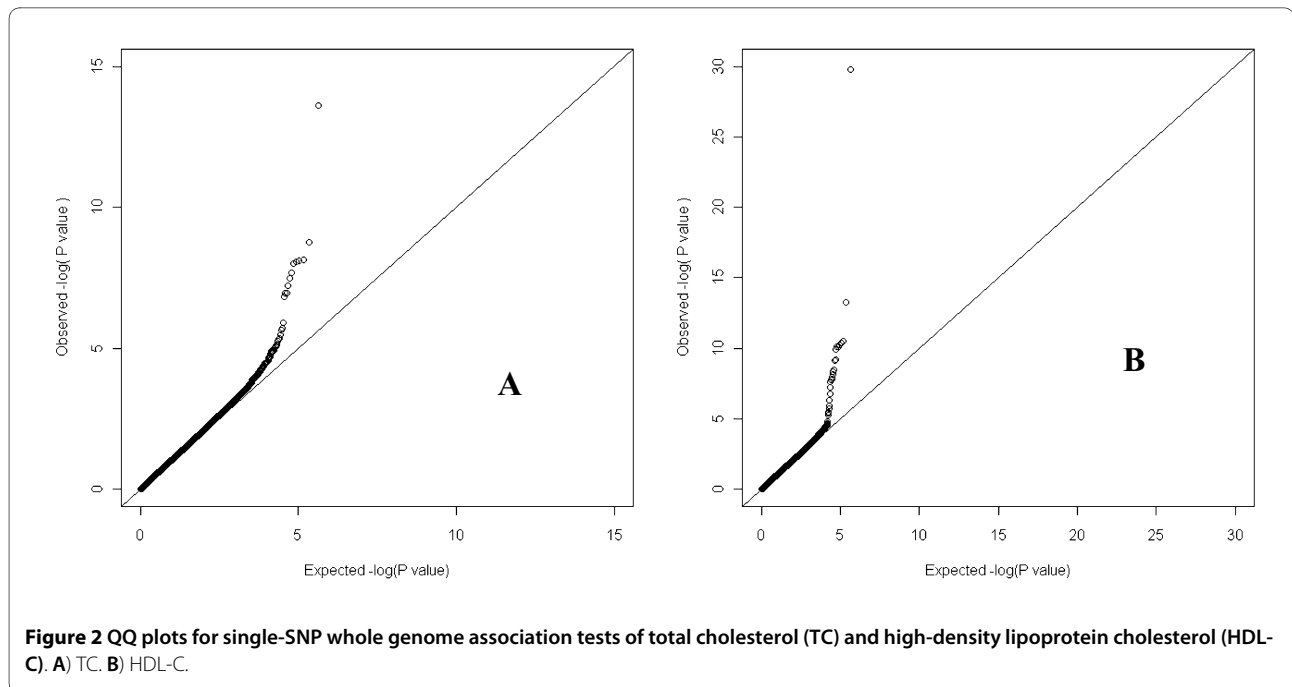
Figure 1 Gene regions associated with total cholesterol (TC). **A** A 1.23 Mb region containing *RAB3GAP1-R3HDM1-LCT-MCM6* with multiple SNP effects on TC. **B** One Mb region containing *OSBPL8-ZDHHC17* associated with HDL-C. **C** One Mb region containing *LMBRD1* that had multiple SNPs interacting with an SNP near *LRIG3* for TC. **D** One Mb region containing *LRIG3* which was near an SNP interacting with multiple SNPs in *LMBRD1* for TC.

upstream *RPL7P43*. The other nineteen SNPs confirmed previously reported gene regions, including twelve SNPs 8-45 kb downstream *LPL*, two SNPs 0.5-11 kb upstream *CETP*, and five SNPs in the *LIPG-ACAA2* region (39.812 kb downstream *LIG* and 65.963 kb upstream *ACAA2*) [1-4,7,11,12]. *LPL*, *CETP* and *LIPG* were reported to be associated with HDL-C in four recent GWAS reports [1-4] while *ACAA2* was reported in [3]. The SNP nearest to *CETP* (rs1800775) was the most significant effect ($p = 8.61 \times 10^{-34}$) in this study.

QQ plot for single SNP tests on TC and HDL-C showed that p-values of significant results all deviated from the expected p-values under the null hypothesis (Figure 2).

Epistasis effects

Two epistasis effects on TC and thirteen epistasis effects on HDL-C reached the significance of suggestive linkage defined in [16] (Table 3). The two epistasis effects on TC involved two different pairs of gene regions while the thirteen epistasis effects on HDL-C involved five different pairs of gene regions, so that the fifteen epistasis effects identified seven pairs of gene regions. Eight SNPs in introns 1, 5, 7, 9, and 14 of *LMBRD1* (chr06) (Figure 1C) interacted with a chr12 SNP about 53 kb from *LRIG3* (q14.1, Figure 1D) and all these eight pairs had $D \times D$ effects on HDL-C. One of the eight epistasis effects involving intron 14 of *LMBRD1* was the most significant epistasis effect that was close to reaching "significant



linkage" defined in [16] or genome-wide significance with 5% Bonferroni corrected type-I error.

Among the seven different pairs of gene regions with epistasis effects, four pairs had $A \times A$ effects, one pair had $A \times D$ effect, and two pairs had $D \times D$ effects (Table 3). For the $A \times A$ effect on TC involving chr04 and chr10, the A-T gamete had the highest TC value while the A-G gamete had the lowest TC value (Table 4). This showed that the G and T alleles of rs705169 on chr10 had significantly different effects when combined with the A allele of rs4437278 on chr04, noting that rs705169 did not have significant single-locus effect. The same phenomenon was also observed for the other three $A \times A$ effects in Table 4. For the $A \times D$ effect, the A-GG allele-genotype combination had the highest TC value while the G-GG allele-genotype combination had the lowest TC value. The two $D \times D$ effects were on HDL-C. For the $D \times D$ effect of rs4706271 \times rs6581219 representing the eight pairs of $D \times D$ effects involving the same gene regions, GT-GG had the highest HDL-C value while GG-GG had the lowest HDL-C value. For the remaining $D \times D$ effect of rs12596869 \times rs6506699 representing the two $D \times D$ effects of the same gene regions, CC-AG had the highest HDL-C value while CC-AA had the lowest HDL-C value (Table 4).

Discussion

The single-locus results in this study had strong confirmations with existing studies. For TC, we confirmed *CELSR2* and *ABCG5* reported in [1,8]. These confirmed TC results should be considered as strong confirmation

because our study had no overlapping samples with studies of [1,8]. We detected seven effects on TC in the *RAB3GAP1-R3HDM1-LCT-MCM6* region with the SNP in *LCT* being the most significant. Six of these seven effects had p-values for LDL-C in the range of 0.007-0.056 from a meta-analysis ([2], Table 1). This could be an indication about the significance on TC from a meta-analysis because LDL-C is calculated from TC [17]. A study in FINRISK cohorts with 14,140 individuals reported that *LCT* was associated with both TC and LDL-C with p-values in the range of 0.0005-0.005 [18]. *In Silico* replication using 1231 Italian subjects from the InCHIANTI cohort [19] generally lacked confirmation for the TC results in Table 1. The first three markers had p-values in the range of 0.005-0.07 while the other effects had p-values greater than 0.14 from the InCHIANTI cohort. The biological function of *LCT* for digesting lactose could be a reason for agreements and disagreements in replicating *LCT* effects on cholesterol. *LCT* affects lactose digestion and long-term consumption of lactose in rats was found to affect aortic cholesterol levels [20]. Therefore, dietary lactose levels that have not been considered by human GWAS could have affected the *LCT* results of different studies. *MCM6* contains two of the regulatory regions for *LCT* [21] so that the significant effects in or near *MCM6* (Table 1) could be due to *MCM6*'s regulatory role to *LCT*. HDL-C had twenty significant SNP effects, but only one SNP identified a new gene region (*OSBPL8-ZDHC17*) while all the other SNPs confirmed previously reported gene regions, although only seven of the twenty significant SNPs for HDL-C were reported previ-

Table 3: Epistasis effects for TC and HDL-C with the significance of suggestive linkage.

SNP1	Chr1	Pos1	Gene1	MAF1	SNP2	Chr2	Pos2	Gene2	MAF2	P value		Effect	
										Genotype	Epistasis		
TC													
rs4437278	4	12488199	<i>U6</i> ^a (174 kb)	0.15	rs705169	10	125285443	<i>GRP26</i> (140 kb)	0.49	4.49E-10	AA	5.93E-12	0.249 ± 0.036
rs4738150	8	72607907	<i>U8</i> ^a (120 kb)	0.40	rs16918936	9	33009027	<i>APTX, LOC646808</i>	0.04	4.68E-13	AD	1.29E-12	-1.348 ± 0.188
HDL-C													
rs10476539	5	91991628	<i>AC026781.5</i> ^a (62 kb)	0.18	rs2392885	8	129003117	<i>PVT1</i> ^a	0.28	2.63E-10	AA	2.99E-12	-0.269 ± 0.037
rs4706271	6	70390132	<i>LMBRD1</i> (intron14)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	7.67E-11	DD	5.72E-13	-0.412 ± 0.056
rs7741758	6	70412380	<i>LMBRD1</i> (intron9)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	5.73E-10	DD	4.60E-12	-0.388 ± 0.054
rs9346333	6	70426479	<i>LMBRD1</i> (intron8)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	1.39E-10	DD	1.16E-12	-0.398 ± 0.054
rs9351772	6	70428200	<i>LMBRD1</i> (intron8)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	5.38E-10	DD	4.25E-12	-0.390 ± 0.055
rs7762400	6	70445634	<i>LMBRD1</i> (intron7)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	4.41E-10	DD	3.52E-12	-0.388 ± 0.054
rs9294851	6	70457629	<i>LMBRD1</i> (intron5)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	2.83E-10	DD	2.22E-12	-0.392 ± 0.054
rs9354890	6	70504296	<i>LMBRD1</i> (intron1)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	5.54E-10	DD	4.18E-12	-0.388 ± 0.054
rs9364063	6	70514750	<i>LMBRD1</i> (8 kb)	0.41	rs6581219	12	59213144	<i>LRIG3</i> (53 kb)	0.42	4.65E-10	DD	4.46E-12	-0.386 ± 0.054
rs2787520	6	106821428	<i>ATG5</i> (48 kb)	0.35	rs7236739	18	20800715	<i>CABLES1</i> (intron4)	0.31	2.98E-10	AA	2.48E-12	-0.211 ± 0.029
rs2842169	10	128330713	<i>AL583860.7</i> ^a (85 kb)	0.10	rs4756344	11	36765284	<i>C11orf74</i> (84 kb)	0.26	3.56E-10	AA	3.73E-12	0.348 ± 0.049
rs17623128	16	77630108	<i>AC092724.2</i> ^a (114 kb)	0.33	rs6506699	18	9775566	<i>RAB31</i>	0.49	4.65E-10	DD	2.78E-12	-0.449 ± 0.062
rs12596869	16	77630380	<i>AC092724.2</i> ^a (114 kb)	0.33	rs6506699	18	9775566	<i>RAB31</i>	0.49	3.81E-10	DD	2.30E-12	-0.451 ± 0.062

^aThis was an RNA gene.

Table 4: Frequency and effect of gamete, allele-genotype or genotype-genotype combination in each epistasis effect with statistical significance of suggestive linkage.

Trait	SNP1	Chr1	SNP2	Chr2											
TC	rs4437278	4	rs705169	10	Gamete	A-T	G-G	G-T	A-G						
					Frequency	0.0785	0.413	0.434	0.0742						
					Effect	0.103	0.0195	-0.0186	-0.108						
	rs4738150	8	rs16918936	9	Allele-genotype	A-GG	G-GT	A-TT	G-TT	A-GT	G-GG				
					Frequency	0.0015	0.0284	0.557	0.374	0.0387	0.001				
					Effect	0.961	0.0917	0.00219	-0.00313	-0.0636	-1.42				
HDL-C	rs10476539	5	rs2392885	8	Gamete	A-C	G-T	G-C	A-T						
					Frequency	0.0538	0.593	0.226	0.127						
					Effect	0.15	0.0137	-0.0359	-0.0635						
	rs4706271	6	rs6581219	12	Genotype-genotype	GT-GG	GG-AG	GT-AA	TT-AG	TT-AA	TT-GG	GT-AG	GG-AA	GG-GG	
					Frequency	0.0817	0.0887	0.161	0.169	0.112	0.0619	0.24	0.055	0.0298	
					Effect	0.137	0.127	0.0694	0.066	-0.0495	-0.0908	-0.0932	-0.103	-0.189	
	rs2787520	6	rs7236739	18	Gamete	G-G	T-A	G-A	T-G						
					Frequency	0.106	0.448	0.241	0.206						
					Effect	0.0946	0.0224	-0.0416	-0.0488						
	rs2842169	10	rs4756344	11	Gamete	C-A	T-G	T-A	C-G						
					Frequency	0.0751	0.237	0.661	0.0272						
					Effect	0.0801	0.0253	-0.00908	-0.221						
rs12596869	16	rs6506699	18	Genotype	CC-AG	CT-AA	CT-GG	TT-AG	TT-GG	TT-AA	CT-AG	CC-GG	CC-AA		
				Frequency	0.0501	0.103	0.115	0.208	0.131	0.115	0.215	0.0344	0.0279		
				Effect	0.223	0.109	0.0979	0.0535	-0.0433	-0.0494	-0.103	-0.167	-0.197		

ously (Table 2). *OSBPL8* encodes a group of intracellular lipid receptors and suppresses *ABCA1* [22], and *ABCA1* was found to affect HDL-C level [23,24]. For HDL-C, the InCHIANTI cohort did not confirm the effects in the *LIPG-ACAA* region ($p > 0.55$) but confirmed the other effects. In light of different samples and different methods of data analysis between our study and those in previous reports, the confirmations of gene results we observed for TC and HDL-C should be considered strong confirmations. This study used log-transformed TC and HDL-C values while recent GWAS on TC [1] and HDL-C [1-4] used the original observations of TC and HDL-C that deviated from normal distribution. However, single-locus effects from our study and previous studies [1-4] had remarkable mutual confirmation, indicating that single-locus analysis was somewhat robust to data distribution and possibly to methods of analysis.

Epistasis effects on TC and HDL-C were not reported in other GWAS so that a comparison between our epistasis results and those from others was unavailable. We detected eight SNP pairs indicating the interaction between gene *LMBRD1* and gene *LRIG3* with the significance of suggestive linkage. Both *LMBRD1* and *LRIG3* encode membrane proteins. *LMBRD1* gene is involved in the transportation and metabolism of vitamin B12 which is important for metabolism of branched chain amino acids and odd chain fatty acids [25]. Replication using the InCHIANTI cohort did not confirm the epistasis results ($p > 0.15$).

The statistical power of epistasis testing is less than that for testing a single-locus effect, particularly for epistasis effects involving dominance such as $A \times D$ and $D \times D$ effects, with $D \times D$ effect being the most difficult to detect. The reason for this difficulty was due to the fact that higher-order effects explain less phenotypic variation even if the effect sizes were the same as lower-order effects [13]. The reduced power for epistasis testing could have contributed to the fact that the epistasis effects we detected only reached 'suggestive linkage' although the sample size was over 6000. The data analysis of this study showed that pairwise analysis was sensitive to outliers. This was due to the fact that artificially significant epistasis effects could occur when rare combinations of loci had extreme genotypic values by chance. This may happen when outliers exist due to the large number of pairwise effects arising from the large number of pairwise combinations. For example, over 466 billion pairwise effects ($93,353,260,560$ pairs \times 5 effects per pair = $466,766,302,800$ pairwise effects) were tested per trait in this study. A small fraction of random association between rare frequencies and outliers in opposite directions among a large number of pairs could yield a long list of artificially significant epistasis results. Therefore, dealing with outliers such as removing outliers and using data

transformation is important in pairwise analysis. Pairwise analysis is computationally intensive but timely analysis is possible using parallel computing. Using 784 processor cores on the SGI Altix XE 1300 Linux cluster system with 2.66 GHz Intel Clovertown processor at the Minnesota Supercomputer Institute, the completion of pairwise epistasis analysis required about 15 hours per trait.

Conclusions

Genome-wide association analysis of the FHS data detected new single-locus and epistasis effects on TC and HDL-C and confirmed some previously reported effects. Additive effects were the primary single-locus effects of TC and HDL-C while epistasis effects involved allele \times allele, allele \times genotype (or genotype \times allele), and genotype \times genotype interactions.

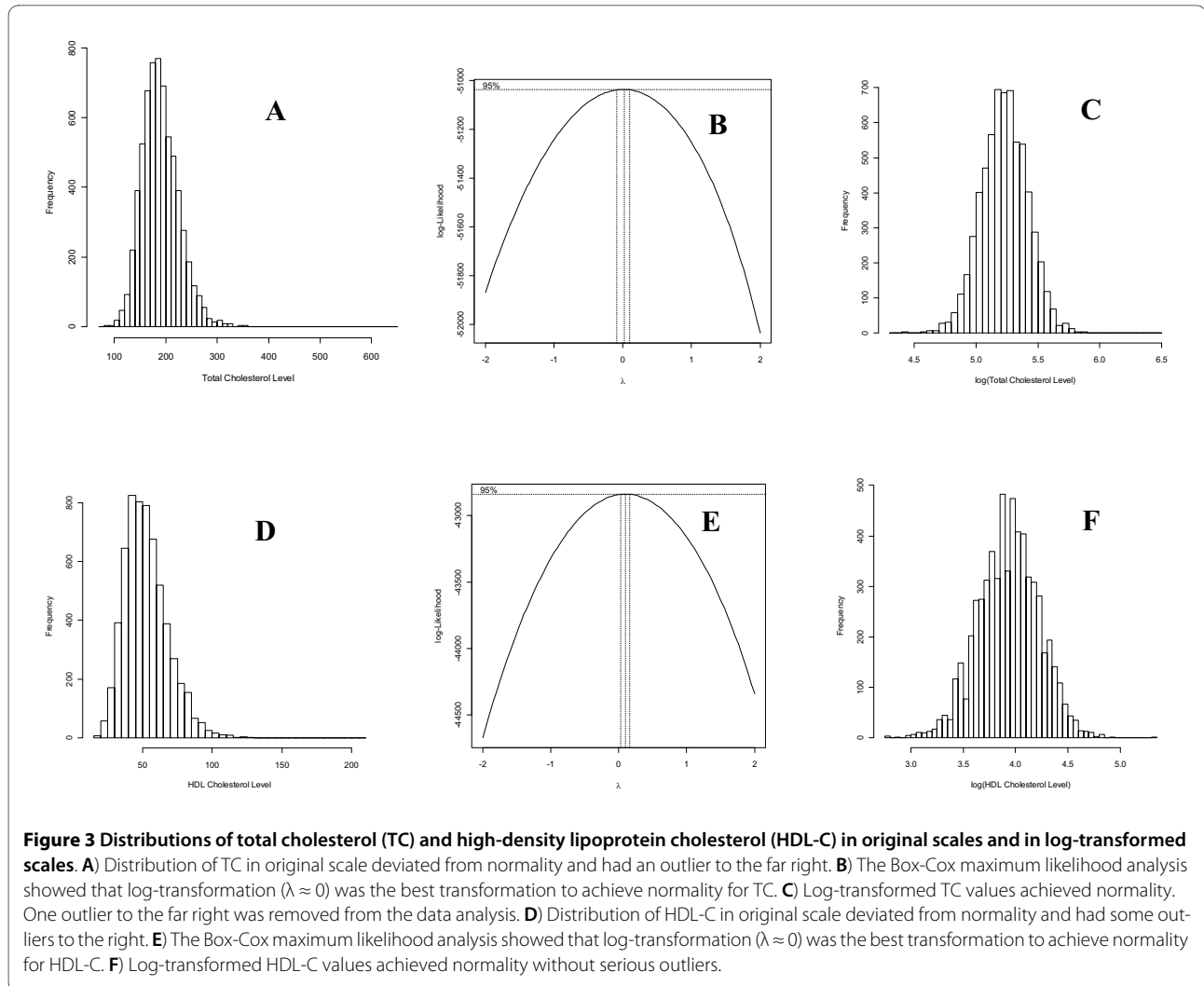
Methods

Phenotype and SNP data

The FHS GWAS data (version 2) had 6575 individuals with SNP genotypes of the 500 k SNP panel from dbGAP. Of the 6575 individuals, 6431 had observations on TC and 6078 individuals had observations on HDL-C. A total of 496,858 SNP markers had known chromosome locations and 432,096 of these SNP markers with minor allele frequencies greater than or equal to 0.01 were analyzed.

Statistical Analysis

Original TC and HDL-C observations deviated from normality and had outliers (Figure 3A, D). The Box-Cox transformation analysis [26] implemented by the R statistical package [27] showed that the log-transformation was approximately the best transformation to achieve normality for those two traits (Figure 3B-C, E-F). One TC outlier, the highest TC value, was removed from the data analysis while no HDL-C outlier was removed. Log-transformed TC values were adjusted for blood sugar, body mass index, smoking status, and sex that had significant effects on log(TC). Age, age-squared, cholesterol treatment, and alcohol consumption were also tested for significant effects on log(TC) but were not included in the phenotypic model because they were insignificant. Log(HDL-C) was adjusted for age, age-squared, cholesterol treatment, blood sugar, body mass index, smoking status, number of cigars smoked, alcohol consumption and sex. Age was insignificant for HDL-C but was included because age-squared was nearly significant ($p < 0.0543$). Single-locus and epistasis effects for both traits were tested using the extended Kempthorne model that allows Hardy-Weinberg disequilibrium and linkage disequilibrium [13]. For each SNP, three effects were tested, genotypic, additive (A) and dominance (D) effects. For each SNP pair, five effects were tested, two-locus genotypic effect, $A \times A$, $A \times D$, $D \times A$, and $D \times D$ epistasis



effects. The EPISNPmpi parallel computing program [28] with a modification to implement a generalized least squares (GLS) analysis to account for sib correlations [29] was used to implement the statistical tests of single-locus and pairwise epistasis effects. For single-locus tests, $p = 7.2 \times 10^{-8}$ was used as the threshold p-value to declare genome-wide significance [30]. To assess genome-wide significance of pairwise epistasis results, we used 5% type-I error with the Bonferroni correction as the genome-wide significance. The 500 k SNP data was estimated to have 276,666 independent SNPs [31]. Each pairwise test was considered to have four independent tests although five effects were tested, because the two-locus marker genotypic effect was confounded with one of the four epistasis effects in reporting significant results. Therefore, the genome-wide 5% type-I errors with the Bonferroni correction was calculated as $p = 0.05 [4(276,666)(276,665)/2]^{-1} = 3.266 \times 10^{-13}$. This 5% significance level is equivalent to "significant linkage" defined in [16]. Since the Bonferroni correction is generally consid-

ered too severe, we also reported epistasis effects reaching "suggestive linkage" with statistical evidence that would be expected to occur one time at random in a genome-wide analysis [16]. In addition to the GLS method to account for sib correlation, the genomic control (GC) method [32] was used to account for potential sub-population structures in the three generation cohort of the FHS data set. For single-locus tests, all p-values were used to estimate inflation parameters for TC and HDL-C, yielding inflation parameter estimates of 1.14 and 1.11 respectively, and test statistics from the GLS tests were then adjusted by the estimates of inflation parameters and p-values were recalculated using the GC adjusted test statistics, which resulted in fewer significant effects. For the pairwise epistasis testing, we randomly selected 50,000 p-values and test statistics from over 466 billion pairwise tests for computational efficiency. Then we estimated the inflation parameters using two samples of 50,000 data points each for TC and HDL-C, yielding inflation parameter estimates of 1.01 and 1.05 respec-

tively. All p-values were then adjusted using the inflation parameters and such adjustments also resulted in fewer significant epistasis results. Frequency of each subclass in an epistasis effect was calculated and each subclass was required to have a minimal number of five observations. After GC adjustment, QQ plots were made to show deviations of the observed p-values from the expected p-values under the null hypothesis for significant test results for single-locus tests only. QQ plot for epistasis effects were not made because the number of p-values for epistasis tests was too large. Gene locations of significant SNPs were identified according to ENSEMBL [15] and NCBI [14] based on Build 37.0 of the human genome.

List of abbreviations

GWAS: genome-wide association study. **SNP:** single nucleotide polymorphism. **TC:** total cholesterol. **HDL-C:** high-density lipoprotein cholesterol. **LDL-C:** low-density lipoprotein cholesterol. **LD:** linkage disequilibrium. **HWD:** Hardy-Weinberg disequilibrium. **A × A:** additive × additive epistasis effect. **A × D:** additive × dominance epistasis effect. **D × A:** dominance × additive epistasis effect. **D × D:** dominance × dominance epistasis effect. **GLS:** generalized least squares.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LM processed the FHS data, implemented the data analysis and prepared the manuscript with YD. JY contributed to the data analysis and manuscript preparation. HBR directed the parallel computing implementation of the GWAS analysis. TT, LF and SB conducted the *In Silico* replication using the InCHIANTI cohort. YD coordinated this research and prepared the manuscript with LM. All authors read and approved this manuscript.

Acknowledgements

This research was supported in part by project MN-16-043 of the Agricultural Experiment Station at the University of Minnesota and by a grant from the Minnesota Supercomputer Institute. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the NHLBI. The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).

Author Details

¹Department of Animal Science, University of Minnesota, USA, ²Supercomputer Institute, University of Minnesota, USA, ³Medstar Health Research Institute, Baltimore MD, USA, ⁴Longitudinal Study Section, National Institute on Aging, Baltimore MD, USA and ⁵Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy

Received: 23 June 2009 Accepted: 6 April 2010

Published: 6 April 2010

References

1. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BWJH, Janssens ACJW, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllenstein U,

- Campbell H, Rudan I, Johansson A, Marroni F, Hayward C, Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, De Geus EJC, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJJ, Uitterlinden AG, Witteman JCM, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Döring A, Wichmann HE, Smit JH, McCarthy MI, Duijn CM, Leena: **Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.** *Nat Genet* 2008, **41**:47-55.
2. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, De Bakker PIW, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altschuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA: **Common variants at 30 loci contribute to polygenic dyslipidemia.** *Nat Genet* 2008, **41**:56-65.
3. Kathiresan S, Melander O, Guiducci C, Surti A, Burt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen M-R, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altschuler DM, Orho-Melander M: **Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans.** *Nat Genet* 2008, **40**:189-197.
4. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duran WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR: **Newly identified loci that influence lipid concentrations and risk of coronary artery disease.** *Nat Genet* 2008, **40**:161-169.
5. Karvanen J, Silander K, Kee F, Tiret L, Salomaa V, Kuulasmaa K, Wiklund PG, Virtamo J, Saarela O, Perret C, Perola M, Peltonen L, Cambien F, Erdmann J, Samani NJ, Schunkert H, Evans A: **The impact of newly identified loci on coronary heart disease, stroke and total mortality in the MORGAM prospective cohorts.** *Genet Epidemiol* 2009, **33**:237-246.
6. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, Inouye M, Luben R, Sims M, Hadley D, McArdle W, Barter P, Kesäniemi YA, Mahley RW, McPherson R, Grundy SM, Wellcome Trust Case Control Consortium, Bingham SA, Khaw KT, Loos RJ, Waeber G, Barroso I, Strachan DP, Deloukas P, Vollenweider P, Wareham NJ, Mooser V: **LDL-cholesterol concentrations: a genome-wide association study.** *Lancet* 2008, **371**:483-491.
7. Wallace C, Newhouse SJ, Braund P, Zhang F, Tobin M, Falchi M, Ahmadi K, Dobson RJ, Marcano AC, Hajat C, Burton P, Deloukas P, Brown M, Connell JM, Dominiczak A, Lathrop GM, Webster J, Farrall M, Spector T, Samani NJ, Caulfield MJ, Munroe PB: **Genome-wide association study identifies genes for biomarkers of cardiovascular disease: Serum urate and dyslipidemia.** *Am J Hum Genet* 2008, **82**:139-149.
8. Samani NJ, Braund PS, Erdmann J, Gotz A, Tomaszewski M, Linsel-Nitschke P, Hajat C, Mangino M, Hengstenberg C, Stark K, Ziegler A, Caulfield M, Burton PR, Schunkert H, Tobin MD: **The novel genetic variant predisposing to coronary artery disease in the region of the PSRC1 and CELSR2 genes on chromosome 1 associates with serum cholesterol.** *J Mol Med* 2008, **86**:1233-1241.
9. Nakayama K, Bayasgalan T, Yamanaka K, Kumada M, Gotoh T, Utsumi N, Yanagisawa Y, Okayama M, Kajii E, Ishibashi S, Iwamoto S, The Jichi Community Genetics Team (JCOG): **Large scale replication analysis of loci associated with lipid concentrations in a Japanese population.** *J Med Genet* 2009, **46**:370-374.
10. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, Guha Thakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom

- SC, Schuetz E, Rushmore TH, Ulrich R: **Mapping the genetic architecture of gene expression in human liver.** *PLoS Biol* 2008, **6**:e107.
11. Kathiresan S, Manning AK, Demissie S, D'Agostino RB, Surti A, Guiducci C, Gianniny L, Burt NP, Melander O, Orho-Melander M, Arnett DK, Peloso GM, Ordovas JM, Cupples LA: **A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study.** *BMC Med Genet* 2007, **19**:8(Suppl 1):S17.
 12. Murray A, Cluett C, Bandinelli S, Corsi AM, Ferrucci L, Guralnik J, Singleton A, Frayling T, Melzer D: **Common lipid-altering gene variants are associated with therapeutic intervention thresholds of lipid levels in older people.** *Eur Heart J* 2009, **30**:1711-1719.
 13. Mao Y, London NR, Ma L, Dvorkin D, Da Y: **Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model.** *Physiol Genomics* 2007, **28**:46-52.
 14. National Center for Biotechnology Information: [<http://www.ncbi.nlm.nih.gov>]. Last accessed 2/13/2010
 15. ENSEMBL Genome Browser: [<http://www.ensembl.org/index.html>]. Last accessed 2/13/2010
 16. Lander ES, Kruglyak L: **Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.** *Nat Genet* 1995, **11**:241-247.
 17. Friedewald WT, Levy RI, Fredrickson DS: **Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge.** *Clin Chem* 1972, **18**:499-502.
 18. Silander K, Alanne M, Kristiansson K, Saarela O, Ripatti S, Auro K, Karvanen J, Kulathinal S, Niemelä M, Ellonen P, Vartiainen E, Jousilahti P, Saarela J, Kuulasmaa K, Evans A, Perola M, Salomaa V, Peltonen L: **Gender Differences in Genetic Risk Profiles for Cardiovascular Disease.** *PLoS ONE* 2008, **3**:e3615.
 19. Tanaka T, Shen J, Abecasis GR, Kisiailiou A, Ordovas JM, Guralnik JM, Andrew Singleton A, Bandinelli S, Cherubini A, Arnett D, Tsai MY, Ferrucci L: **Genome-Wide Association Study of Plasma Polyunsaturated Fatty Acids in the InCHIANTI Study.** *PLoS Genet* 2009, **5**:e1000338.
 20. Wostmann BS, Bruckner-Kardoss E: **The effect of long-term feeding of 10% dietary lactose on serum, liver and aortic cholesterol of the rat and the gerbil.** *J Nutr* 1980, **110**:82-89.
 21. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Leena Peltonen L, Järvelä I: **Identification of a variant associated with adult-type hypolactasia.** *Nat Genet* 2002, **30**:233-237.
 22. Yan D, Mäyränpää MI, Wong J, Perttilä J, Lehto M, Jauhiainen M, Kovanen PT, Ehnholm C, Brown AJ, Olkkonen VM: **OSBP-related protein 8 (ORP8) suppresses ABCA1 expression and cholesterol efflux from macrophages.** *J Biol Chem* 2007, **283**:332-340.
 23. Wang J, Burnett JR, Near S, Young K, Zinman B, Hanley AJ, Connelly PW, Harris SB, Hegele RA: **Common and rare ABCA1 variants affecting plasma HDL cholesterol.** *Arterioscler Thromb Vasc Biol* 2000, **20**:1983-1989.
 24. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: **Multiple rare alleles contribute to low plasma levels of HDL cholesterol.** *Science* 2004, **305**:869-872.
 25. Rutsch F, Gailus S, Miousse IR, Suormala T, Sagné C, Toliat MR, Nurnberg G, Wittkamp T, Buers I, Sharifi A, Stucki M, Becker C, Baumgartner M, Robenek H, Marquardt T, Hohne W, Gasnier B, Rosenblatt DS, Fowler B, Nurnberg P: **Identification of a putative lysosomal cobalamin exporter altered in the cblF defect of vitamin B(12) metabolism.** *Nat Genet* 2009, **41**:234-239.
 26. Box GEP, Cox DR: **An analysis of transformations (with discussion).** *J Roy Stat Soc B* 1964, **26**:211-252.
 27. R Development Core Team: **R: A language and environment for statistical computing.** [<http://www.R-project.org>]. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, 2008. Last accessed 6/22/09
 28. Ma L, Runesha HB, Dvorkin D, Garbe JR, Da Y: **Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies.** *BMC Bioinformatics* 9:315.
 29. Ma L, Amos CI, Yang Da Y: **Accounting for correlations among individuals for testing SNP single-locus and epistasis effects in Genome-wide association analysis (Abstract).** *Plant & Animal Genomes XVI Conference, January 12-16, San Diego, CA 2008* [http://www.intl-pag.org/16/abstracts/PAG16_P11_903.html]. Last accessed 6/22/09
 30. Dudbridge F, Gusnanto A: **Estimation of significance thresholds for genomewide association scans.** *Genet Epidemiol* 2008, **32**:227-234.
 31. Moskvina V, Schmidt K: **On multiple-testing correction in genome-wide association studies.** *Genet Epidemiol* 2008, **32**:567-573.
 32. Devlin B, Roeder K: **Genomic control for association studies.** *Biometrics* 1999, **55**:997-1004.
- Pre-publication history**
The pre-publication history for this paper can be accessed here:
<http://www.biomedcentral.com/1471-2350/11/55/prepub>
- doi: 10.1186/1471-2350-11-55
Cite this article as: Ma et al., Genome-wide association analysis of total cholesterol and high-density lipoprotein cholesterol levels using the Framingham Heart Study data *BMC Medical Genetics* 2010, **11**:55

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

