Supplemental Material

Allelic expression imbalance at high-density lipoprotein cholesterol locus *MMAB-MVK*

Marie P. Fogarty[1], Rui Xiao[2], Ludmila Prokunina-Olsson[3], Laura J. Scott[2] and Karen L. Mohlke[1]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA

[2]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

[3]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Gaithersburg, MD 20892, USA

Supplemental Methods

Protein isolation

For batch 1 and 2 samples, protein was isolated according to the manufacturer's instructions from the saved fraction left over after RNA extraction with Tri reagent (Applied Biosystems). Protein pellets were resuspended in a 1M Tris solution containing 0.1% SDS, 4M urea, and 1X protease inhibitor (Sigma, St. Louis, MO). Protein was isolated from batch 4 samples (cryopreserved cells) by lysing cells in buffer containing 50 mM HEPES, 2 mM EDTA, 0.5% IGEPAL CA-630, 175 mM NaCl, 10% glycerol (all from Sigma), and 1X protease inhibitor (Pierce, Rockford, IL). Protein content in lysates was quantitated by Nanodrop 1000 (Thermo Scientific).

Power simulation

To determine the power to detect AEI of a specified amount in samples heterozygous for each transcribed SNP, we simulated data based on (1) the number of cDNA samples with each HDL-C-associated SNP (rSNP) genotype, (2) the number of genotyped gDNA samples heterozygous for the transcribed SNP, (3) the standard deviation (SD) of the $\%B_n$ allele observed for gDNA and cDNA from specified sets of samples, and (4) the observed LD between the transcribed SNP and the rSNP. Because more than one statistical test can be considered for each transcribed SNP, for each SNP we simulated the %B allele for gDNA samples, for cDNA from samples homozygous for the rSNP (when consistent with the LD), and for cDNA from samples heterozygous for the rSNP.

For gDNA from samples heterozygous for the transcribed SNP, we simulated the $\%B_n$ allele from a normal distribution with mean 50% and SD equal to that observed for $\%B_n$ allele for gDNA from these samples.

For cDNA from samples homozygous for the rSNP, we simulated the %B allele from a normal distribution with mean of 50% and SD equal to that observed for $\%B_n$ allele for cDNA from these samples.

For cDNA from samples heterozygous for the rSNP, we simulated data based on the observed LD between the transcribed SNP and the rSNP. For SNPs with LD described in Scenarios 1 and 2 ($r^2 \leq 1$, D'=1) we simulated the %B allele from a normal distribution with mean 50% plus half of the specified amount of AEI (%B - %b), and SD equal to that observed for $\%B_n$ allele for cDNA from these samples. For SNPs with LD described in Scenarios 3 and 4 ($r^2 << 1$, D'<1) individuals doubly heterozygous for the rSNP and the transcribed SNP can have one of two possible haplotype pairs, ab/AB or Ab/aB. We used the expectation maximization (EM) algorithm and the observed genotypes from samples unselected for the rSNP or transcribed SNP genotype (including samples homozygous for the transcribed SNP that are not included in the testing of AEI) to estimate the frequency of each haplotype. We then estimated the frequency of the haplotype pair ab/AB in the doubly heterozygous samples. We simulated the number of doubly heterozygous samples with the ab/AB haplotype pair from a binomial distribution with number of trials equal to the number of doubly heterozygous samples and the frequency equal to the frequency of ab/AB haplotype pair. The remaining doubly heterozygous samples were assigned the other haplotype pair, Ab/aB. We simulated %B allele from a normal distribution with mean equal to 50% plus (ab/AB samples) or minus (Ab/aB samples) half a specified

amount of AEI (%B-%b), and SD equal to that observed for $\%B_n$ allele for cDNA from samples homozygous for the rSNP.

Using the observed data for each transcribed SNP, we performed 1000 simulations. For each simulation we calculated the test statistic and *P*-value for all plausible tests: gDNA t-test (Scenarios 1-4), cDNA test (Scenarios 2-4), and cDNA F-test (Scenarios 3 and 4). Power of a test was defined as the proportion of p-values < .007. The type 1 error rates in the simulations ranged from .041 - .061 before correction for multiple testing and from .005-.009 after multiple testing correction.

Table S1. Evidence of AEI for SNPs in five genes at an HDL-C-associated locus at α=.05

Gene	Transcribed SNP	Number of heterozygotes tested	r^2/D' with rs7298565* †	Statistical test to detect AEI	Power to detect AEI (%B-%b) of 22% ‡	80% power to detect AEI (%B-%b)	AEI (%B-%b)	*P* value
MMAB	rs877710	20	.94 / 1.0	t-test gDNA	1.0	7.4	22	2.4×10^{-9}
MMAB	rs11067231	45	.94 / 1.0	t-test gDNA	1.0	4.7	22	1.4×10^{-13}
MMAB	rs11067233	27	.11 / .70	t-test gDNA	.87	16	15	8.6×10^{-4}
				t-test cDNA	.83	21		.046
				F-test cDNA	.26			.31
MMAB	rs2241201	22	.43 / 1.0	t-test gDNA	.96	16	20	7.1×10^{-4}
				t-test cDNA	.73	24		6.5×10^{-3}
MVK	rs7957619	16	.10 / 1.0	t-test gDNA	1.0	9.2	5.7	.10
				t-test cDNA	1.0	14		.21
KCTD10	rs1477117	21	.15 / 1.0	t-test gDNA	1.0	13	2.9	.68
				t-test cDNA	1.0	10		.47
UBE3B	rs7298565 †	19	1.0 / 1.0	t-test gDNA	1.0	3.8	1.4	.29
UBE3B	rs2058807	19	1.0 / 1.0	t-test gDNA	1.0	10	.07	.73
ACACB	rs3742023	34	.23 / .75	t-test gDNA	.99	6.7	3.7	.051
				t-test cDNA	.97	8.8		.44
				F-test cDNA	.85			.47
ACACB	rs7135947	14	.001 / .05	F-test cDNA	.86	20	1.7	.40
				t-test gDNA	.049	NA		.87

SSNP, single nucleotide polymorphism; AEI, allelic expression imbalance. *, r^2 and D' values calculated based on observed hepatocyte genotypes; †, rs7298565 is used as a representative HDL-C-associated SNP and also as a transcribed SNP in *UBE3B*; ‡, power to detect AEI size observed for *MMAB* SNP rs11067231. All possible statistical tests are shown.

Table S2. Evidence of AEI for SNPs in five genes at an HDL-C-associated locus for Caucasian ancestry only at $\alpha = .007$

Gene	Transcribed SNP	Number of heterozygotes tested	r^2/D' with rs7298565*†	Statistical test to detect AEI	Power to detect AEI (%B-%b) of 22% ‡	80% power to detect AEI (%B-%b)	AEI (%B-%b)	*P* value
MMAB	rs877710	16	.96 / 1.0	t-test gDNA	1.0	10.9	23	2.7×10^{-8}
MMAB	rs11067231	37	.96 / 1.0	t-test gDNA	1.0	6.6	21	7.5×10^{-11}
MMAB	rs11067233	19	.15 / .85	t-test gDNA	.78	22.5	16	2.0×10^{-3}
				t-test cDNA	.48	32.0		.14
				F-test	.045			.31
MMAB	rs2241201	18	.40 / 1.0	t-test gDNA	.61	25.1	20	2.5×10^{-3}
				t-test cDNA	.31	39.6		.020
MVK	rs7957619	11	.10 / 1.0	t-test gDNA	.98	11.2	3.3	.38
				t-test cDNA	.92	17.6		.51
KCTD10	rs1477117	15	.20 / 1.0	t-test gDNA	.61	25.0	5.2	.46
				t-test cDNA	.99	11.3		.42
UBE3B	rs7298565 †	16	1.0 / 1.0	t-test gDNA	1.0	4.9	.40	.72
UBE3B	rs2058807	16	1.0 / 1.0	t-test gDNA	.98	15.4	1.7	.47
ACACB	rs3742023	27	.15 / .62	t-test gDNA	.91	10.5	2.6	.20
				t-test cDNA	.88	11.1		.49
				F-test	.34			.51
ACACB	rs7135947	10	0 / .02	F-test	.22	39.2	.03	.34
				t-test gDNA	.005			.78

SNP, single nucleotide polymorphism; AEI, allelic expression imbalance. *, r^2 and D' values calculated based on observed hepatocyte genotypes; †, rs7298565 is used as a representative HDL-C-associated SNP and also as a transcribed SNP in *UBE3B*; ‡, power to detect AEI size observed for *MMAB* SNP rs11067231. All possible statistical tests are shown.

Table S3. Evidence of AEI for SNPs in five genes at an HDL-C-associated locus for Caucasian ancestry only at α = .05

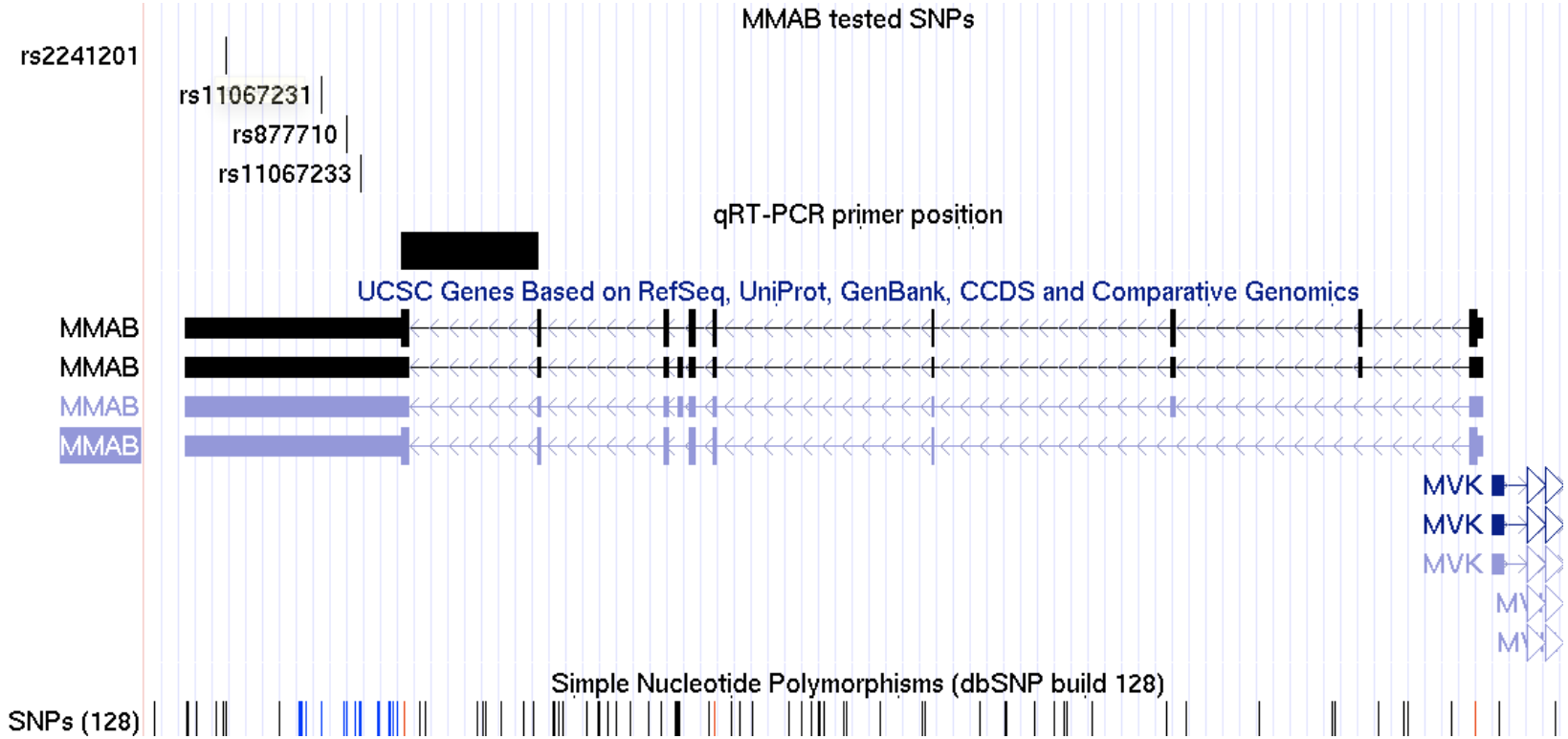
Gene	Transcribed SNP	Number of heterozygotes tested	r^2/D' with rs7298565*†	Statistical test to detect AEI	Power to detect AEI (%B-%b) of 22% ‡	80% power to detect AEI (%B-%b)	AEI (%B-%b)	*P* value
MMAB	rs877710	16	.96 / 1.0	t-test gDNA	1.0	8.5	23	2.7×10^{-8}
MMAB	rs11067231	37	.96 / 1.0	t-test gDNA	1.0	5.2	21	7.5×10^{-11}
MMAB	rs11067233	19	.15 / .85	t-test gDNA	.92	12.7	16	2.0×10^{-3}
				t-test cDNA	.88	17.8		.14
				F-test	.15			.31
MMAB	rs2241201	18	.40 / 1.0	t-test gDNA	.87	18.8	20	2.5×10^{-3}
				t-test cDNA	.61	23.6		.020
MVK	rs7957619	11	.10 / 1.0	t-test gDNA	1.0	10.4	3.3	.38
				t-test cDNA	.99	15.3		.51
KCTD10	rs1477117	15	.20 / 1.0	t-test gDNA	.93	17.4	5.2	.46
				t-test cDNA	.99	11.2		.42
UBE3B	rs7298565 †	16	1.0 / 1.0	t-test gDNA	1.0	3.9	.40	.72
UBE3B	rs2058807	16	1.0 / 1.0	t-test gDNA	.98	12.0	1.7	.47
ACACB	rs3742023	27	.15 / .62	t-test gDNA	.98	7.2	2.6	.20
				t-test cDNA	1.0	9.0		.49
				F-test	.60			.51
ACACB	rs7135947	10	0 / .02	F-test	.38	32	.03	.34
				t-test gDNA	.072			.89

SNP, single nucleotide polymorphism; AEI, allelic expression imbalance. *, r^2 and D' values calculated based on observed hepatocyte genotypes; †, rs7298565 is used as a representative HDL-C-associated SNP and also as a transcribed SNP in *UBE3B*; ‡, power to detect AEI size observed for *MMAB* SNP rs11067231. All possible statistical tests are shown.

Table S4. Primers used to measure total mRNA levels

Gene	Primer Sequence 5'-3'
MMAB	Fwd_TTGTCCAGATGGGAGAGACC
	Rvs_CTCCAAGCTCCCACTTTCTG
MVK	Fwd_GCTCAAGTTCCCAGAGATCG
	Rvs_ATGGTGCTGGTTCATGTCAA
KCTD10	Fwd_GTGGCATAGACGATGGAGGT
	Rvs_ATAAGCCAGCCGTGAAGTTG
UBE3B	Fwd_CATCTGATGGCACATTTTCG
	Rvs_CCCAGAGCCAGATGATGACT
ACACB	Fwd_GGAACTTAACCGGATGCGTA
	Rvs_TCAGGTCAGAGTGCCTGATG
B2M	Fwd_TGTCTGGGTTTCATCCATCCGACA
	Rvs_TCACACGGCAGGCATACTCATCTT
GUSB	Fwd_GGTAGTGGCTGGTACGGAAA
	Rvs_AGCCAGTTCCTCATCAATGG

Figure S1. Location of *MMAB* SNPs tested and *MMAB* qRT-PCR primers - Screen shot from UCSC Genome browser http: //genome.ucsc.edu created using March 2006 genome assembly (1).



REFERENCE

1 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996-1006.