Supplement to

ABCG2 Regulatory SNPs Alter *In Vivo* Enhancer Activity and Expression

Rachel J. Eclov, Mee J. Kim, Aparna Chhibber, Robin P. Smith, Nadav Ahituv, Deanna L. Kroetz

Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA (R.J.E, M.J.K, R.S, A.C, N.A, D.L.K); Institute for Human Genetics, University of California San Francisco, San Francisco, CA (N.A., M.J.K, R.S., D.L.K)

Supplemental Materials and Methods

Below we provide details that extend the methods in the published paper. The references refer to the primary paper.

Chemicals and Materials. The vectors pGL4.23 [*luc2*/minP], pGL4.74 [*hRluc*/TK], pGL4.13 [*luc2*/SV40] and the Dual-Luciferase® Reporter Assay System were purchased from Promega (Madison, WI). The human embryonic kidney (HEK293T/17) and human hepatocellular carcinoma (HepG2) cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA). High-glucose Dulbecco's modified Eagle's medium (DMEM), Opti-Minimal Essential Medium (Opti-MEM), Lipofectamine 2000, 10% Tris-Borate-EDTA (TBE) gels and Trizol were all purchased from Invitrogen (Carlsbad, CA). Phosphate buffered saline (PBS), penicillin and streptomycin were purchased from the University of California San Francisco cell culture facility (San Francisco, CA). Phusion High-Fidelity DNA Polymerase and *Dpn*I were purchased from New England Biolabs (Ipswich, MA). AllPrep DNA/RNA Mini Kits, QIAquick PCR Purification Kits, and RNeasy MinElute Cleanup Kits were purchased from Qiagen (Valencia, CA). Fetal bovine serum (FBS) (Axenia BioLogix, Dixon, CA), GenElute HP Endotoxin-Free Maxiprep Kit (Sigma Aldrich, St. Louis, MO), TransIT EE *In Vivo* Gene Delivery System (Mirus Bio, Madison, WI), CD1 mice (Charles River Laboratories, Wilmington, MA), High Capacity cDNA Reverse Transcription Kits (Applied Biosystems, Foster City, CA), HepG2 nuclear extract (Abcam, Cambridge, MA), Odyssey EMSA Buffer Kit (Lincoln, NE), EMSA probe sets (Affymetrix, Santa Clara, CA) and Exonuclease I (GE Healthcare, Piscataway, NJ) were purchased from the indicated manufacturers.

Genetic Analysis of Enhancer Regions. SNPs in each of the *ABCG2 in vivo* enhancer regions were retrieved for all available ethnic populations from publicly available databases,

including 1000 Genomes 20120214 phase 1 release[16], dbSNP build 135 and HapMap release 28[17]. Haplotypes were determined by retrieving genotype and information files from the 1000 Genomes browser 20100804 phase 1 release for Caucasians (CEU), Nigerians (YRI), combined Chinese and Japanese (CHB+JBT) and all available ethnic groups combined (ALL), and linkage analysis was performed using the Haploview program version 4.2[18]. SNPs in linkage disequilibrium with rs12508471, rs72873421, rs149713212, rs9999111 and rs2725263 (r^2 threshold ≥ 0.8) were extracted from 1000 Genomes pilot 1 genotype data using the Broad Institute SNP annotation and proxy search (SNAP) version 2.2[19] for each population (CEU, YRI and CHB+JBT) separately.

Variant Enhancer Plasmid Construction. Reference enhancer plasmids in the pGL4.23 vector were previously described[15]. Site-directed mutagenesis (SDM) on plasmids was performed using specific primers (Supplemental Table 1) and Phusion High-Fidelity DNA Polymerase following the manufacturer's protocol. PCR reaction conditions for all primers except for rs36105707 were as follows: An initial cycle for 30 sec at 98°C, followed by 20 cycles of 10 sec at 98°C, melting temperature (varied per primer pair) for 30 sec and 3 min at 72°C, then a final extension for 10 min at 72°C. The SDM PCR reactions were then digested for at least 20 min at 37°C with 1 unit *Dpn*I, purified and transformed into competent cells, and the resulting plasmids were sequenced to confirm introduction of the SNP. In order to make the ABCG2RE1*2 haplotype, the SNPs for rs7287321 and rs12508471 were introduced into the rs12500008 plasmid. The ABCG2RE9*2 haplotype was made by mutating the rs41282399 SNP into the rs2622628 plasmid. The ABCG2RE8*2 haplotype was the genotype of the originally cloned ABCG2RE8 region and was mutated back to reference. Primers for the deletion SNP rs36105707 were designed according to a large deletion protocol[20]. PCR conditions for the

deletion SDM were: an initial cycle of 5 min at 95°C, then 12 cycles of 95°C for 1 min, 46.5°C for 1 min and 72°C for 9 min, with a final cycle of 1 min at 36°C and 30 min at 72°C. Endotoxin-free DNA for the enhancer plasmids, empty pGL4.23, *ApoE*-pGL4.23[21], pGL4.13 and pGL4.74 vectors were isolated using the GenElute HP Endotoxin-Free Maxiprep Kit following the manufacturer's protocol.

Cell Culture and Transfections. HEK293T/17 and HepG2 cell lines were grown in high-glucose DMEM supplemented with 10% FBS, 100 U/mL of penicillin and 0.1 mg/mL of streptomycin, in a 5% CO_2 incubator at 37°C. For *in vitro* luciferase assays, the HEK293T/17 and HepG2 cells were seeded in a 96-well plate at 1.8 x 10^4 cells/well and transfected when they reached 80% confluency with 0.5 µL of Lipofectamine 2000 mixed with 0.08 µg plasmid plus 0.02 µg pGL4.74 following the manufacturer's protocol. Cells were lysed with passive lysis buffer 18-24 hr after transfection and measured for firefly and *Renilla* luciferase activity using the Dual-Luciferase® Reporter Assay System in a GloMax 96 microplate Dual Injector Luminometer (Promega, Madison, WI) following the manufacturer's protocol. Each experiment also included the empty pGL4.23 vector as the negative control and the *ApoE*-pGL4.23[21] or pGL4.13 plasmids as a positive control. Enhancer activity was expressed as the ratio of the plasmid firefly to *Renilla* luciferase activity; activity of each variant plasmid was then normalized relative to the reference plasmid, setting the reference activity to one (100%).

Hydrodynamic Tail Vein Assay. Selected positive *in vitro* variant enhancer elements were screened for their effect on *in vivo* liver enhancer activity using the hydrodynamic tail vein injection adapted for enhancer activity screening[22,23]. Each variant enhancer, along with their reference enhancer plasmid, the *ApoE*[21] positive control liver enhancer and an empty pGL4.23 vector, were injected individually into the tail vein of 4-11 mice using the TransIT EE *In Vivo*

Gene Delivery System following the manufacturer's protocol (Mirus Bio, Madison, WI). Briefly, 10 μg plasmid plus 2 μg of pGL4.74 (containing *Renilla* luciferase to correct for injection efficiency) were injected into the tail vein of CD1 mice. After 24 hr, mice were euthanized, their livers harvested, homogenized in passive lysis buffer and centrifuged at 4°C for 30 min at 21,000 g. The supernatant was diluted 1:20 with lysis buffer and firefly and *Renilla* luciferase activity measured using the Dual-luciferase® reporter assay system according to the manufacturer's protocol using a Synergy 2 (BioTek Instruments, Winooski, VT) microplate reader. Each plasmid's firefly activity was normalized to *Renilla* luciferase activity and expressed as fold activity relative to the negative control, empty pGL4.23. All mouse work was done following a protocol approved by the University of California San Francisco Institutional Animal Care and Use Committee.

Liver and Kidney Tissues. Kidney (n=60) and liver (n=60) samples were procured by the PMT research group at the University of California San Francisco (San Francisco, CA)[24]. These tissues were purchased from Asterand (Detroit, MI), Capital Biosciences (Rockville, MD) and SRI International (Menlo Park, CA). The Asterand samples included both postmortem tissues and surgical resections from donors; the Capital Biosciences specimens included surgical resections from normal tissue surrounding cancerous tissues and SRI International included postmortem tissues. All samples were stored frozen at -80°C until processing for DNA and RNA. Information on the age, sex, and ethnicity of the patient was available for all samples.

DNA was extracted and purified from the tissues using a Qiagen AllPrep DNA/RNA Mini Kit and QIAquick PCR Purification Kit following the manufacturer's protocols. RNA was extracted from the tissues following the protocol for Trizol reagent and cleaned-up with the Qiagen RNeasy MinElute Cleanup Kit following the manufacturer's protocol. High quality RNA

was isolated from 58 kidney samples and 60 liver samples and those with 260/280 >1.7, 260/230 >1.8, and RNA Integrity number from Bioanalyzer of 3-8 were used to correlate SNP genotype with total ABCG2 mRNA expression. RNA (2 µg) was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit and the following incubation conditions: 10 min at 22ºC, 2 hr at 37ºC, 5 min at 4ºC, 10 min at 75ºC and 5 min at 4ºC. Exonuclease I enzyme (10 U/mL) was added to each sample and the following incubation conditions were used to remove excess primers: 1 hr at 37ºC, 5 min at 4ºC and then 10 min at 85ºC to inactivate the exonuclease enzyme. Samples were then stored at -20ºC until assayed for gene expression or genotype.

ABCG2 mRNA Expression and Genotype in PMT Liver and Kidney Tissues. Gene expression for ABCG2 was evaluated in 58 kidney and 60 liver samples from surgical resection or postmortem collections in Caucasian males and females using the Biotrove Open Array™ qPCR platform (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. ABCG2 mRNA expression was normalized to a geometric mean of the expression of GAPDH, β-2 microglobin, and β-actin and expressed as $2^{-\Delta\Delta Ct}$ per gene for each sample. All ΔCt values for a given tissue type were quantile normalized across samples using the open source R preprocess Core package[25,26]. Expression data was quality controlled using principal component analysis to identify outliers. Of these samples, 58 kidney and 34 liver samples were successfully genotyped on the Affymetrix Axiom genotyping platform using the Axiom® Genome-Wide CEU 1 Array Plate (Santa Clara, CA). The samples were tested for quality control (QC) using sex-check, identity by descent and call rate tests, of which six kidney samples failed and were excluded from further analysis. After initial QC, 52 kidney samples and 34 liver samples were included in subsequent analyses.

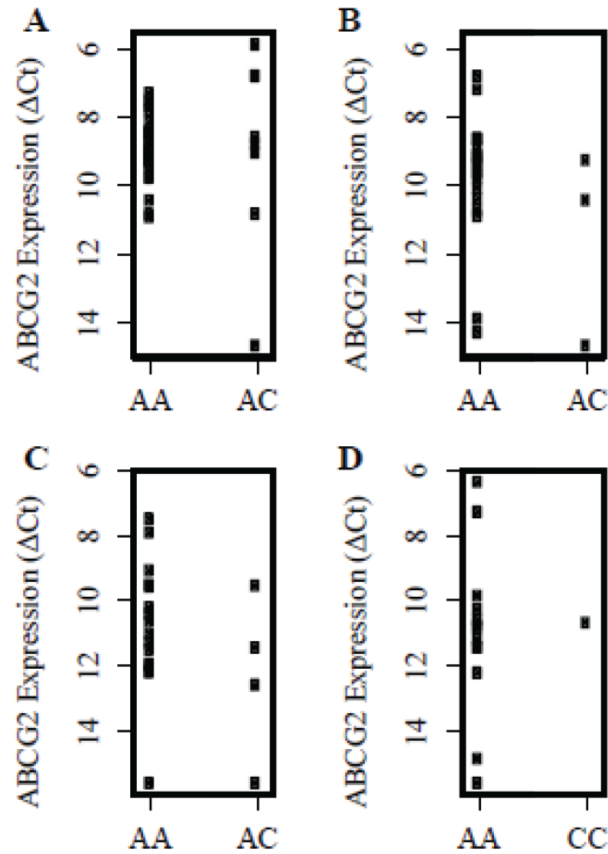
Association of SNPs with Gene Expression. Genotype and expression data from PMT liver and kidney tissue, 195 samples from Schadt *et al.*[27] liver tissue and 62 samples from The Cancer Genome Atlas (TCGA) for breast tissue[28] were analyzed for associations between enhancer variants and ABCG2 expression levels. Genotypes were imputed using 1000 Genomes data and then tested for correlation between the expression of ABCG2, PPM1K and PKD2 and genotype using a linear regression and the Affymetrix Genotyping Console (Santa Clara, CA). The PMT liver and kidney linear regression were performed after adjusting for gender.

A database for integrated analysis and visualization of SNP-gene associations in eQTL studies, Genevar[29], includes data from several sequence and gene expression profiling studies: the MuTHER study[30,31] with data from female twins in adipose (166 samples), skin (160 samples) and lymphoblastoid cell lines (LCLs, 156 samples), the Stranger study[32-35] with data from 726 HapMap LCL samples, and the GenCord study[36] with data from 85 human umbilical fibroblasts, LCLs and T cell samples. The expression of ABCG2, and the neighboring 5' (PPM1K) and 3' (PKD2) genes was correlated to the *ABCG2* locus SNPs that altered enhancer activity *in vivo*, or with SNPs in LD ($r^2>0.8$, as determined above) with these SNPs. Using the GeneVar 3.2.0 eQTL analysis program, Spearman rank correlation coefficients (rho) for 10,000 permutations per SNP between reference, heterozygous and variant alleles were calculated.

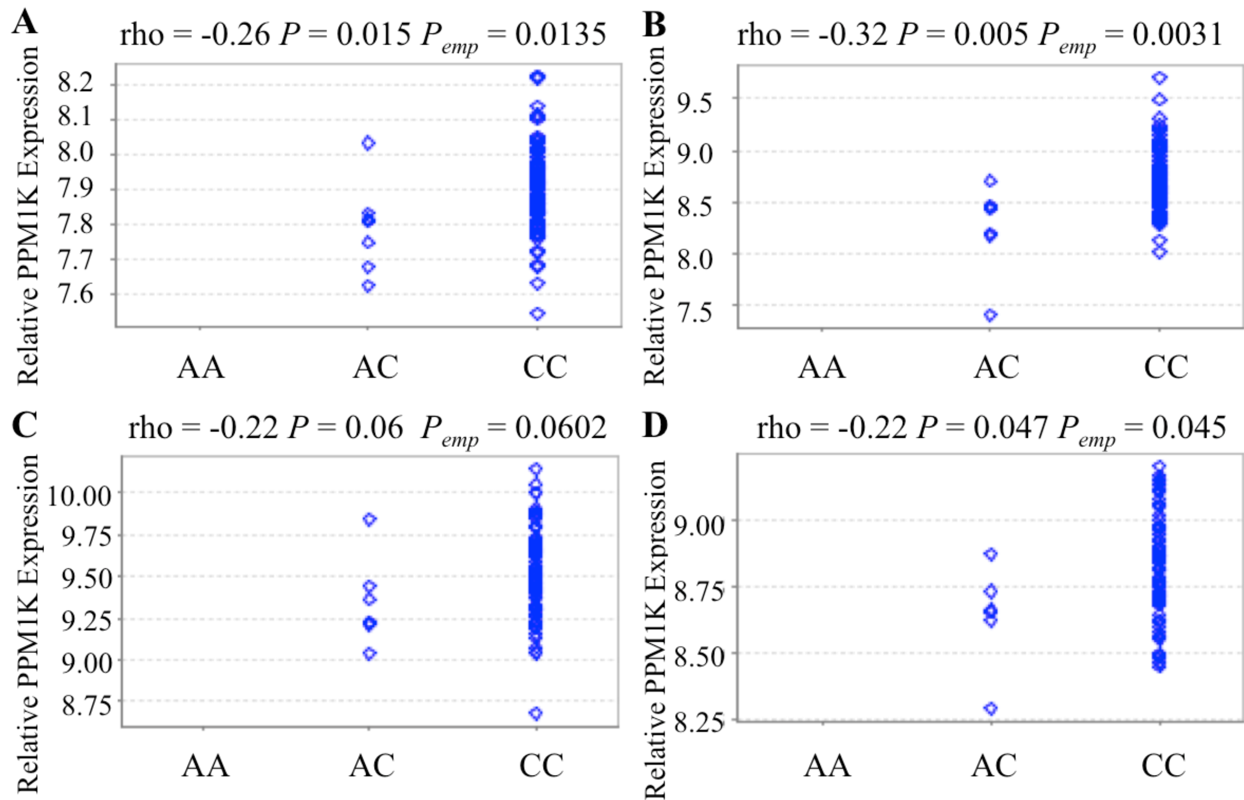
Electrophoretic Mobility Shift Assay. Electrophoretic mobility shift assays (EMSA) were performed at room temperature for 30 minutes by incubating 2.5-5 μg of HepG2 nuclear extract with 2.5 nM IRDye 700 labeled probe in 20 μL total volume using the Odyssey EMSA buffer kit, following the manufacturer's protocol. EMSA probes, with and without 5' IRDye 700, corresponding to enhancer sequence surrounding the variant (SNP locations are underlined) are as follows: ABCG2RE1, 5'-GTGGCACTAAGACTGAGGTGAGAT-3', rs12508471, 5'-

CGGTGGCACTAGGACTGAGGTGAGAT-3’, ABCG2RE14, 5’-TTTTCTCACTGTGAATTCAATCAACA-3’ and rs9999111, 5’-TTTTCTCACTGTGCATTCAATCAACA-3’. Competition assays were preformed by adding 40-fold molar excess of reference or SNP oligonucleotide to the reaction mixture. Reactions were mixed with 10X orange loading dye and DNA/protein complexes separated from free probe by gel electrophoresis on 10% TBE gels. Gels were imaged using the Licor system (Odyssey, Lincoln, NE).

Statistical Analysis. Enhancer activity was expressed relative to the reference sequence as described above. Normalized polymorphic enhancer activities were compared to the reference enhancer per transfection (3-8 wells/plasmid) using an ANOVA analysis followed by a Bonferroni’s multiple comparison *t*-test with $P < 0.05$ considered significant. Polymorphic enhancer constructs identified for *in vivo* testing had either a 2-fold increase or decrease in activity and a $P < 0.0001$ in both cell lines. Results from the hydrodynamic tail vein injection were analyzed using an unpaired Student’s *t*-test between the reference and variant enhancer (for only one variant) or an ANOVA analysis followed by a Bonferroni’s multiple comparison *t*-test (for two or more variants); $P < 0.05$ was considered significant for both tests. All statistics were run using the GraphPad Prism 5 software program (San Diego, CA).



Supplemental Figure 1: Association of rs9999111 with ABCG2 expression. Association of the ABCG2RE14 SNP rs9999111 with ABCG2 in (A & B) 52 kidney and (C & D) 34 liver of (A & C) males and (B & D) females from PMT tissue samples. Linear regression analysis was done using the Affymetrix Genotyping Console and results are displayed as gene expression versus rs9999111 genotype.



Supplemental Figure 2: Association of rs12500008 with PPM1K expression. Association of the ABCG2RE1 SNP rs12500008 with PPM1K in (A) 166 adipose, (B) 160 skin and (C & D) 156 lymphoblastoid cell lines (two probes) from the MuTHER study. Linear regression analysis was done using the GeneVar program and results are displayed as gene expression versus rs12500008 genotype with rho (correlation coefficient), *P* value (P) and empirical *P* value (P_{emp}) indicated.

Supplemental Table 1: Site-Directed Mutagenesis Primers

Region	SNP ID	Δnt[1]	Primer Sequence[2]	Tm[3]
ABCG2RE1	rs72873421	G<A	CCAAATCTATCATGAA[A]AAGGCCACAAATCCTAGC	60
			GCTAGGATTTGTGGCCTT[T]TTCATGATAGATTTGG	
	rs117741074	C<T	GTTCTTCTCATAAA[A]CCCAAAACACCAGA	60
			TCTGGTGTTTTGGG[T]TTTATGAGAAGAAC	
	rs2728131	C<T	CCTTTTAAAATGG[T]TCCTTCCAGCGTC	60
			GACGCTGGAAGGA[A]CCATTTTAAAAGG	
	rs12508471	A<G	ACGGTGGCACTA[G]GACTGAGGTGAG	60
			CTCACCTCAGT[C]CTAGTGCCACCGT	
	rs12500008	G<T	GTCTCCACCTC[T]AGAGGAAATCC	62
			GGATTTCCTCT[A]GAGGTGGAGAC	
	rs78901673	A<G	ACCGTGG[G]CGCTGGAAGGA	62
			TCCTTCCAGCG[C]CCACGGT	
ABCG2RE6	rs45510401	A<G	AAAGAATACATAAAATAGGAT[G]TAATTAAATTCTCATTTAT	60
			ATAAATGAGAATTTAATTA[C]ATCCTATTTTATGTATTCTTT	
	rs573519157[4]	-<A	CAAAAAAAAAAAA[A]TTAGCCGGGCGTG	60
			CACGCCCGGCTAA[T]TTTTTTTTTTTTG	
	rs5883021	A<-	CAAAAAAAAAAAAAAAAAAAA[-]GTCCTTAATTTTAAAATGG	60
			CCATTTTAAAATTAAGGAC[-]TTTTTTTTTTTTTTTTTTTTG	
	rs186188962	C<T	CCACCCGCTT[T]GGCCTCCCA	65.4
			TGGGAGGCC[A]AAGCGGGTGG	
	rs144180103	T<-	GCTAATTTTTTTTTTTT[-]GTATTTTTAGTAGAGAC	58
			GTCTCTACTAAAAATAC[-]AAAAAAAAAAAATTAGCC	
	rs190754327	G<C	GTAATGGCTGG[C]ACTACAGGCTCC	64.9
			GGAGCCTGTAGT[G]CCAGCCATTAC	
	rs2725268	A<G	AAAGAATACATAAAATAGGAT[G]TAATTAAATTCTCATTTAT	56
			ATAAATGAGAATTTAATTA[C]ATCCTATTTTATGTATTCTTT	
	rs183322988	A<G	ATTAACTTTTTAAATAA[G]TGAGAATTTAATTATATCCT	58
			AGGATATAATTAAATTCTCA[C]TTATTTAAAAAGTTAAT	
ABCG2RE8	rs139101431	T<C	CTGAAGAAACACCTAAGGTTCTTC[T]TTATTTCTC	59.2
			GAGAAATAA[A]GAAGAACCTTAGGTGTTTCTTCAG	
	rs144062279	G<A	CAAAGTTCCAGTTG[A]AGACCAGGTTACTCC	65
			GGAGTAACCTGGTC[T]TCAACTGGAACTTTG	
	rs2725264[4]	C<T	GTTACTCCATGTCCT[C]TCCAAATGCTTCCTG	60
			CAGGAAGCATTTGGA[G]AGGACATGGAGTAAC	
	rs4148156	G<A	GTCTGGAAATAATCT[A]GATACCTCAGCCC	59.6
			GGGCTGAGGTATC[T]AGATTATTTCCAGAC	

Region	SNP ID	Δnt[1]	Primer Sequence[2]	Tm[3]
	rs145932752	A<G	CTTATTCTTTAAAAAATA[G]TCAGCCTTTCCAGACATC	59.2
			GATGTCTGGAAAGGCTGA[C]TATTTTTAAAGAATAAG	
	rs6831395	G<A	ATAATCAGCCTTTCCA[A]ACATCAAAATAGGCTGC	62.1
			GCAGCCTATTTTGATGT[T]TGGAAAGGCTGATTAT	
	rs2725263[4]	A<C	CTTTCCAGACATCAA[A]ATAGGCTGCACATAAG	60
			CTTATGTGCAGCCTAT[T]TTGATGTCTGGAAAG	
	rs192562676	C<T	GATACTACCATCTA[T]CCCCTCTAAATCAC	55
			GTGATTTAGAGGGG[A]TAGATGGTAGTATC	
	rs182159263	C<G	CTACCATCTACCC[G]CTCTAAATCACTGG	62.1
			CCAGTGATTTAGAG[C]GGGTAGATGGTAG	
	rs187527722	A<G	GACAGCAAGC[G]CTACGGAGCAC	65
			GTGCTCCGTAG[C]GCTTGCTGTC	
	rs192781547	A<G	GCTTATGTTCAGA[G]CACAATACAGTCG	65
			CGACTGTATTGTG[C]TCTGAACATAAGC	
	rs184709106	C<T	CAGAACACAATACAGT[T]GATAAAAAGTCCCCTC	59.6
			GAGGGGACTTTTTATC[A]ACTGTATTGTGTTCTG	
ABCG2RE9	rs2231148	T<A	CAATTAGAGATAAAAACTTA[A]ACACACCATTTATTAGTATA	57
			TATACTAATAAATGGTGTGT[T]TAAGTTTTTATCTCTAATTG	
	rs190738974	A<G	AACTTATACACACCATTT[G]TTAGTATAATATATGGT	57
			ACCATATATTATACTAA[C]AAATGGTGTGTATAAGTT	
	rs117761897	C<T	CATTAAGATAAACACT[T]AATGGCTTGGCCAACG	62.7
			CGTTGGCCAAGCCATT[A]AGTGTTTATCTTAATG	
	rs41282399	A<C	GATAAACACTCA[C]TGGCTTGGCCAAC	61
			GTTGGCCAAGCCA[G]TGAGTGTTTATC	
	rs113647079	C<G	CCCTGGGAGAAAATAAAA[G]AGCATACATTATTTAGAC	61
			GTCTAAATAATGTATGCT[C]TTTTATTTTCTCCCAGGG	
	rs2054576	A<G	CAATGCAAGTATGT[G]GCAAAGCAAAGTC	61
			GACTTTGCTTTGC[C]ACATACTTGCATTG	
	rs151266026	T<C	AAAAATTTTAAAGCACACA[C]TAAAAAAATTCTAACAATGG	61
			CCATTGTTAGAATTTTTTTA[G]TGTGTGCTTTAAAATTTTT	
	rs183315559	G<A	GTGAGGAAATAG[A]GGTGAGATGGAGC	61
			GCTCCATCTCACC[T]CTATTTCCTCAC	
	rs189214307	C<T	CACTACCCATCTC[T]TGTCACTGCTTC	61
			GAAGCAGTGACA[A]GAGATGGGTAGTG	
	rs36105707[4,5]	-< TTAAA	ATAACGTAACCC[TTAAA]TTAACCCTTTGCTTATTGAA	62
			TTTAAGGGTTACGTTATGATAT[AATTT]ATCTGAGAAAATCC	
	rs141635727	A<G	ACGTTATGATATAATTT[G]TCTGAGAAAATCCTATTT	57
			AAATAGGATTTTCTCAGA[C]AAATTATATCATAACGT	
	rs190767980	C<T	AGAAAATCCTATTTATATTTA[T]TCGTGAGTTAAATATTAAAA	57

Region	SNP ID	Δnt[1]	Primer Sequence[2]	Tm[3]
			TTTTAATATTTAACTCACGA[A]TAAATATAAATAGGATTTTCT	
	rs147070185	G<A	AAATCCTATTTATATTTACTCA[T]GAGTTAAATATTAAAAAAC	57
			GTTTTTTAATATTTAACTC[A]TGAGTAAATATAAATAGGATTT	
	rs2622628[4]	C<A	CTGGACAAACAC[C]AATCTTGTTTCTAGG	56
			CCTAGAAACAAGATT[G]GTGTTTGTCCAG	
ABCG2RE14	rs9999111	T<G	CCTTTTCTCACTGTG[C]ATTCAATCAACAGA	58
			TCTGTTGATTGAAT[G]CACAGTGAGAAAAGG	
	rs138867860	G<T	CTACCAATTTTACTT[A]TTTCCCATAAGAGACT	58
			AGTCTCTTATGGGAAA[T]AAGTAAAATTGGTAG	
	rs114916387	A<G	CTTACCAGAGCCTAA[C]AGATAGAAGCTCAC	60
			GTGAGCTTCTATCT[G]TTAGGCTCTGGTAAG	
ABCG2RE26	rs137884075	C<T	GGAGGGCAGATCA[T]GAGGTCAGGAGATC	60
			GATCTCCTGACCTC[A]TGATCTGCCCTCC	
	rs142621223	G<A	GAGGGCAGATCAC[A]AGGTCAGGAGATC	60
			GATCTCCTGACCT[T]GTGATCTGCCCTC	
	rs139553964	G<A	GGTCAGGAGATC[A]AGACCATCCTGG	58
			CCAGGATGGTCT[T]GATCTCCTGACC	
	rs149713212	G<A	CACGGTGAAACCCC[A]TCTCTACTAAAAAAC	58
			GTTTTTTAGTAGAGA[T]GGGGTTTCACCGTG	
	rs144565932	G<A	CAAAAAATTAGCCGAGC[A]TGTTGGCAGGC	60
			GCCTGCCAACA[T]GCTCGGCTAATTTTTTG	
	rs62309980	C<T	GGCGCCTGTAGT[T]CCAGCTACTCAG	60
			CTGAGTAGCTGG[A]ACTACAGGCGCC	
	rs76888829	T<C	GCATCACTATCTACAAA[C]GGCCTCTATTCATATC	60
			GATATGAATAGAGGCC[G]TTTGTAGATAGTGATGC	
	rs9998634	C<A	GGATATCTGGTGTCCA[T]ACTGAAAGTATTAAA	60
			TTTAATACTTTCAGT[A]TGGACACCAGATATCC	
	rs77538297	C<T	CACAAAATAGCC[T]GGCTGTCCCCAAC	60
			GTTGGGGACAGCC[A]GGCTATTTTGTG	
	rs35696062	G<-	CATCTCGCAATA[-]GGCTACTGTTGCAGTAG	60
			CTACTGCAACAGTAGCC[-]TATTGCGAGATG	

[1]Change in reference to variant nucleotide of the anti-strand

[2]Forward and reverse primers per SNP with mutagenized nucleotide in brackets

[3]Melting temperature used for annealing step of SDM PCR

[4]Primer mutates enhancer region back to reference

[5]This primer is used with special PCR conditions

Supplemental Table 2: Variants in *ABCG2* Locus Enhancer Elements

Region	SNP ID	Relative Luciferase Activity[1]		
		HepG2	HEK293T	In Vivo
ABCG2RE1	rs72873421	2.76*	2.13*	0.74
	rs117741074	1.44*	1.49*	
	rs12500008	1.08	1.05	
	rs2728131	0.92	1.41*	
	rs12508471	0.09*	0.002*	0.0004*
	rs78901673	1.39*	1.44*	
	ABCG2RE1*2[2]	1.01	0.31*	0.065*
ABCG2RE6	rs45510401	0.87	0.78*	
	rs2725268	1.18*	0.94	
	rs57351915	0.76*	0.55*	
	rs58830217	0.83*	0.80*	
	rs186188962	1.20*	1.11	
	rs144180103	1.25*	1.08	
	rs190754327	0.76*	0.52*	
	rs183322988	1.60*	1.52*	1.03
ABCG2RE8	rs139101431	0.98	1.41*	
	rs144062279	0.97	1.70*	
	rs2725264	1.40*	1.41*	
	rs4148156	0.99	1.09	
	rs145932752	0.84	0.80	
	rs6831395	1.01	1.01	
	rs2725263	1.35*	1.64*	1.56*
	rs192562676	1.15	1.06	
	rs182159263	0.84	0.89	
	rs187527722	0.95	1.04	
	rs192781547	0.94	0.96	
	rs184709106	1.10	1.34*	
	ABCG2RE8*2[2]	0.99	1.03	
ABCG2RE9	rs2231148	0.94	0.70*	
	rs190738974	1.68*	1.56*	1.46
	rs117761897	0.85	0.71*	
	rs41282399	1.36	1.17	
	rs113647079	1.19	1.07	
	rs2054576	0.99	0.87	
	rs151266026	0.73	0.81	

	rs183315559	0.95	0.95	
	rs189214307	1.04	1.00	
	rs2622628	1.16	0.63*	
	rs36105707	3.39*	2.01*	
	rs141635727	1.28	0.98	
	rs190767980	1.24	1.05	
	rs147070185	1.24	0.84	
	ABCG2RE9*2[2]	1.23	1.18	
ABCG2RE14	rs9999111	0.14*	0.16*	0.32*
	rs138867860	1.24*	1.23*	
	rs114916387	0.98	1.41*	
ABCG2RE26	rs137884075	1.46*	1.92*	
	rs142621223	1.16	1.10	
	rs139553964	1.59*	2.17*	
	rs149713212	0.066*	0.20*	0.021*
	rs144565932	1.12	1.52*	
	rs62309980	1.58*	2.15*	
	rs76888829	1.32	1.15	
	rs9998634	1.16	0.93	
	rs77538297	1.95*	1.93*	
	rs35696062	1.94*	2.36*	

[1]Enhancer activity in pGL4.23 expressed as the ratio of firefly to Renilla luciferase activity and normalized to the reference pGL4.23 vector. Values are the mean activities from a representative experiment with 3-6 wells or 4-11 mice per plasmid.

[2]rs72873421/rs12500008/rs12508471 (ABCG2RE1*2); rs2725264/rs2725263 (ABCG2RE8*2); rs41282399/rs2622628 (ABCG2RE9*2)

* Indicates $P < 0.05$