UCSF

UC San Francisco Previously Published Works

ABCG2 regulatory single-nucleotide polymorphisms alter *in vivo* enhancer activity and expression

Rachel J. Eclov[a], Mee J. Kim[a,b], Aparna Chhibber[a], Robin P. Smith[a,b], Nadav Ahituv[a,b] and Deanna L. Kroetz[a,b]

Objectives The expression and activity of the breast cancer resistance protein (*ABCG2*) contributes toward the pharmacokinetics of endogenous and xenobiotic substrates. The effect of genetic variation on the activity of *cis*-regulatory elements and nuclear response elements in the *ABCG2* locus and their contribution toward ABCG2 expression have not been investigated systematically. In this study, the effect of genetic variation on the *in vitro* and *in vivo* enhancer activity of six previously identified liver enhancers in the *ABCG2* locus was examined.

Methods Reference and variant liver enhancers were tested for their ability to alter luciferase activity *in vitro* in HepG2 and HEK293T cell lines and *in vivo* using a hydrodynamic tail vein assay. Positive *in vivo* single-nucleotide polymorphisms (SNPs) were tested for association with gene expression and for altered protein binding in electrophoretic mobility shift assays.

Results Multiple SNPs were found to alter enhancer activity *in vitro*. Four of these variants (rs9999111, rs12508471, ABCG2RE1*2, and rs149713212) decreased and one (rs2725263) increased enhancer activity *in vivo*. In addition, rs9999111 and rs12508471 were associated with ABCG2 expression in lymphoblastoid cell lines, lymphocytes, and T cells, and showed increased HepG2 nuclear protein binding.

Conclusion This study identifies SNPs within regulatory regions of the *ABCG2* locus that alter enhancer activity *in vitro* and *in vivo*. Several of these SNPs correlate with tissue-specific ABCG2 expression and alter DNA/protein binding. These SNPs could contribute toward reported tissue-specific variability in ABCG2 expression and may influence the correlation between ABCG2 expression and disease risk or the pharmacokinetics and pharmacodynamics of breast cancer resistance protein substrates. *Pharmacogenetics and Genomics* 27:454–463 Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.

Introduction

The breast cancer resistance protein (BCRP), encoded by *ABCG2*, is a member of the ATP-binding cassette (ABC) membrane transporter family and is responsible for transport of its substrates across intestinal epithelial cells into the intestinal lumen, from the hepatocyte into the bile, into milk, away from the placenta and the brain, and into the lumen of the renal proximal tubule [1]. Reduced expression and function of MXR are associated with a variety of adverse events, such as pheophorbide-induced phototoxicity [2], urate-induced gout [3], gefitinib-induced diarrhea [4], as well as reduced chemotherapy response and increased susceptibility to cancer [5–8]. Although reduced function coding variants of BCRP exist [9], variability in ABCG2 expression and BCRP substrate pharmacokinetics cannot be accounted for solely by these nonsynonymous variants. Even in individuals without these variants, there is a wide range of ABCG2 expression [10]. Identification of regulatory regions of *ABCG2* and functional single-nucleotide polymorphisms (SNPs) within these regions may provide information on the mechanisms of genetic regulation of ABCG2 expression.

BCRP is one of many transporters important in drug absorption, distribution, metabolism, and excretion (ADME). Recently, the transcriptional regulation of ADME genes has been linked to *cis*-regulatory elements and alterations in ADME gene expression because of variants in these regulatory elements are becoming more evident [11,12]. In addition, expression quantitative trait loci (eQTL) studies of human genes have implicated proximal regulatory variation as a prevalent cause of population variation in gene expression [13,14]. *cis*-Regulatory elements include enhancers, suppressors, promoters, insulators, and locus control regions

that work to regulate the transcriptional activity of the basal transcription machinery. These genomic regions provide binding sites for transcription factors, either ubiquitous or tissue specific, that work through complex interactions with histones, RNA polymerase, and other transcription factors to determine gene transcription. In previous studies, we utilized comparative genomics along with *in vitro* and *in vivo* assays to identify six liver enhancers in the *ABCG2* gene locus [15].

In this study, the hypothesis that SNPs in regulatory regions of the *ABCG2* locus contribute to the variation in ABCG2 expression was tested. Variants in liver enhancer regions reported in publicly available databases were studied *in vitro* in kidney and liver cell lines, and variants with altered function were tested *in vivo* using a hydrodynamic tail vein assay. SNPs that significantly altered *in vivo* liver enhancer activity were then tested for association with ABCG2 expression in the human liver, kidney, placenta, breast, lymphocytes, T cells, and fibroblasts. On the basis of supporting ENCODE data indicating transcriptional activity for five of six of these enhancer regions, electrophoretic mobility shift assays (EMSAs) were used to confirm changes in protein/DNA binding for SNPs that correlate with gene expression. The findings from this study provide insights into how noncoding genetic variants may lead to altered exposure to BCRP substrates.

Methods
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]. 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 [18], for each population (CEU, YRI, and CHB + JBT) separately and a linkage analysis was carried out using the Haploview program, version 4.2 (Broad Institute, Cambridge, Massachusetts, USA)[19].

Variant enhancer plasmid construction
Reference enhancer plasmids in the pGL4.23 vector were described previously [15]. Site-directed mutagenesis on plasmids was performed using specific primers (Supplementary Table 1, Supplemental digital content 1, *http://links.lww.com/FPC/B278*) and Phusion High-Fidelity DNA Polymerase following the manufacturer's protocol. PCR reaction conditions are available in the Supplementary Materials and Methods (Supplemental digital content 1, *http://links.lww.com/FPC/B278*). Primers and PCR conditions for the deletion SNP rs36105707 were designed according to a large deletion protocol [20]. Endotoxin-free DNA for all vectors were isolated using the GenElute HP

Endotoxin-Free Maxiprep Kit (Sigma Aldrich, St. Louis, Missouri, USA) following the manufacturer's protocol.

Cell culture and transfections
HEK293T/17 and HepG2 cell lines were cultured and transfected for *in vitro* luciferase assays with lipofectamine 2000 following the manufacturer's protocol as described previously [15]. Cells were lysed 18–24 h 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, Wisconsin, USA) following the manufacturer's protocol. Each experiment also included the empty pGL4.23 vector and the *ApoE*-pGL4.23 [21] or pGL4.13 plasmids. Enhancer activity was expressed as the ratio of the plasmid firefly to *Renilla* luciferase activity; the activity of each variant plasmid was then normalized relative to the reference plasmid, setting the reference activity to 1 (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, was injected individually into the tail vein of 4–11 mice and hepatic luciferase activity was measured after 24 h as described previously [15]. 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 carried out 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, California, USA) [24] from Asterand (Detroit, Michigan, USA), Capital Biosciences (Rockville, Maryland, USA), and SRI International (Menlo Park, California, USA). DNA was extracted and purified from the tissues using a Qiagen AllPrep DNA/RNA Mini Kit (Valencia, California, USA) and a QIAquick PCR Purification Kit (Valencia, California, USA) following the manufacturer's protocols. RNA was extracted from the tissues following the protocol for Trizol reagent and cleaned up using the Qiagen RNeasy MinElute Cleanup Kit (Valencia, California, USA) following the manufacturer's protocol. High-quality RNA was isolated from 58 kidney 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 the SNP genotype with the total ABCG2 mRNA expression. RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, California, USA).

ABCG2 mRNA expression and genotype in PMT liver and kidney tissues

ABCG2 gene expression was evaluated in 58 kidney and 60 liver samples using the Biotrove Open Array qPCR platform (Life Technologies, Carlsbad, California, USA) according to the manufacturer's protocol. ABCG2 mRNA expression was normalized to a geometric mean of the expression of glyceraldehyde 3-phosphate dehydrogenase, β2 microglobin, and β-actin and expressed as $2^{-\Delta\Delta C_t}$ per gene for each sample. All ΔC_t values for a given tissue type were quantile normalized across samples using the open-source R preprocess Core package [25,26]. Expression data were quality controlled using principal component analysis to identify outliers. Of these samples, 58 kidney and 34 liver samples were genotyped successfully on the Affymetrix Axiom genotyping platform using the Axiom Genome-Wide CEU 1 Array Plate (Affymetrix Inc., Santa Clara, California, USA). The samples were tested for quality control using sex check, identity by descent and call rate tests, of which six kidney samples failed and were excluded from further analysis. After the initial quality control, 52 kidney samples and 34 liver samples were included in subsequent analyses.

Association of single-nucleotide polymorphisms 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 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 correlations between the expression of ABCG2, PPM1K, and PKD2 and genotype using a linear regression and the Affymetrix Genotyping Console (Santa Clara, California, USA). The PMT liver and kidney linear regression were performed after adjusting for sex.

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 linkage disequilibrium (LD) ($r^2 > 0.8$, as determined above) with these SNPs. Using the GeneVar 3.2.0 eQTL analysis program, Spearman rank correlation coefficients (ρ) for 10 000 permutations per SNP between reference, heterozygous, and variant alleles were calculated.

Electrophoretic mobility shift assay

EMSA were performed using the Odyssey EMSA buffer kit (Lincoln, Nebraska, USA) following the manufacturer's protocol. EMSA probe sequences are available in the Supplementary Materials and Methods (Supplemental digital content 1, *http://links.lww.com/FPC/B278*). Competition assays were performed by adding 40-fold molar excess of unlabeled reference or SNP oligonucleotide. DNA/protein complexes were separated from free probe by gel electrophoresis and imaged using the Licor system (Odyssey, Lincoln, Nebraska, USA).

Statistical analysis

Normalized polymorphic enhancer activities were compared with the reference enhancer per transfection (3–8 wells/plasmid) using an analysis of variance analysis, followed by a Bonferroni's multiple-comparison *t*-test, with a *P* value less than 0.05 considered significant. Polymorphic enhancer constructs identified for *in vivo* testing showed either a two-fold increase or decrease in activity and a *P* value less than 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 the variant enhancer (for only one variant) or an analysis of variance analysis, followed by a Bonferroni's multiple-comparison *t*-test (for two or more variants); a *P* value less than 0.05 was considered significant for both tests. All statistics were run using the GraphPad Prism 5 software program (GraphPad Software Inc., San Diego, California, USA).

Results
Genetic variation in the ABCG2 locus enhancer regions

A total of 53 SNPs and three haplotypes in the six *in vivo* enhancer regions were obtained from publicly available databases (Table 1). There are three SNPs in ABCG2RE14, 10 in ABCG2RE26, and eight in ABCG2RE6. None of these SNPs are in LD with each other, and only rs9999111 and rs2725268 had a minor allele frequency (MAF) of at least 5% in at least one population. There are six SNPs in the ABCG2RE1 region. In addition, ABCG2RE1 has one haplotype (ABCG2RE1*2), which is a combination of the SNPs rs72873421, rs12500008, and rs12508471. These three SNPs have MAFs ranging from 8.4 to 38% and are in near-perfect LD ($r^2 = 0.96–0.98$). There are 12 SNPs in ABCG2RE8 and one haplotype, ABCG2RE8*2, which is a combination of the two most frequent SNPs rs2725263 and rs2725264. These two SNPs have MAFs ranging from 8 to 87% depending on the ethnic population and the ABCG2RE8*2 haplotype has a frequency similar to the individual SNPs (7–79%), with an r^2 equal to 0.46 between the SNPs. There are 14 SNPs and one haplotype in ABCG2RE9. The ABCG2RE9*2 (rs41282399 and rs2622628) haplotype has a frequency ranging from 1.7 to 3.7%, with an $r^2 = 0.13$ between SNPs.

Table 1 Variants in *ABCG2* locus enhancer elements

Region	SNP ID	Location[a]	ΔNT[c]	Conserved[c]	MAF%[b]		
					AA	EUR	AS
ABCG2RE1	rs72873421	chr4:88 923 906	G > A	Yes	21.1	8.4	38.1
	rs117741074	chr4:88`924 002	G > A	Yes	0.0	0.0	1.7
	rs12500008	chr4:88 924 176	C > A	Yes	21.1	8.4	38.3
	rs2728131	chr4:88 924 344	C > T	Yes	1.2	8.4	2.8
	rs12508471	chr4:88 924 371	A > G	Yes	21.1	8.4	38.3
	rs78901673	chr4:88 924 356	A > G	Yes	NR	NR	NR
	ABCG2RE1*2[d]	–	–	–	21.1	8.4	37.2
ABCG2RE6	rs45510401	chr4:89 011 422	A > G	No	4.5	0.0	0.0
	rs2725268	chr4:89 010 983	A > G	Yes	7.5	47.4	24.0
	rs57351915	chr4:89 011 141	A > –	Yes	NR	NR	NR
	rs58830217	chr4:89 011 309	– > A	No	NR	NR	NR
	rs186188962	chr4:89 011 051	G > A	No	0.0	0.5	0.0
	rs144180103	chr4:89 011 129	A > –	No	NR	NR	NR
	rs190754327	chr4:89 011 173	C > G	Yes	NR	NR	NR
	rs183322988	chr4:89 011 437	T > C	No	0.0	0.3	0.3
ABCG2RE8	rs139101431	chr4:89 026 007	A > G	No	1.6	0.0	0.0
	rs144062279	chr4:89 026 087	G > A	No	0.4	0.0	0.0
	rs2725264	chr4:89 026 109	C > T	Yes	83.7	8.2	21.5
	rs4148156	chr4:89 026 242	C > T	Yes	NR	NR	NR
	rs145932752	chr4:89 026 407	A > G	No	1.6	0.0	0.0
	rs6831395	chr4:89 026 420	G > A	Yes	2.2	0.0	0.0
	rs2725263	chr4:89 026 428	A > C	Yes	87.0	48.4	40.0
	rs192562676	chr4:89 026 490	C > T	No	0.0	0.0	1.0
	rs182159263	chr4:89 026 493	C > G	No	0.0	0.7	0.0
	rs187527722	chr4:89 026 537	A > G	No	0.0	0.0	0.2
	rs192781547	chr4:89 026 630	A > G	No	0.0	0.0	0.3
	rs184709106	chr4:89 026 642	C > T	No	0.0	0.0	0.3
	ABCG2RE8*2[d]	–	–	–	78.8	7.3	21.5
ABCG2RE9	rs2231148	chr4:89 028 478	T > A	Yes	4.3	40.2	18.5
	rs190738974	chr4:89 028 490	A > G	No	0.0	0.0	0.2
	rs117761897	chr4:89 028 542	C > T	Yes	0.0	0.0	0.2
	rs41282399	chr4:89 028 544	A > C	Yes	4.1	1.7	3.3
	rs113647079	chr4:89 028 578	C > G	Yes	NR	NR	NR
	rs2054576	chr4:89 028 775	A > G	Yes	1.2	7.9	25.0
	rs151266026	chr4:89 028 935	T > C	No	0.0	0.0	0.7
	rs183315559	chr4:89 028 979	G > A	No	0.0	0.0	0.3
	rs189214307	chr4:89 029 111	C > T	No	0.0	0.0	0.3
	rs2622628	chr4:89 029 252	A > C	Yes	45.5	4.4	20.8
	rs36105707	chr4:89 029 304	– > TTAAT	Yes	NR	NR	NR
	rs141635727	chr4:89 029 335	A > G	No	0.4	0.0	0.0
	rs190767980	chr4:89 029 361	C > T	No	0.0	0.0	0.0
	rs147070185	chr4:89 029 364	G > A	No	0.0	0.0	0.2
	ABCG2RE9*2[d]	–	–	–	3.7	1.7	3.1
ABCG2RE14	rs9999111	chr4:89 073 197	A > C	Yes	4.7	7.3	0.0
	rs138867860	chr4:89 073 397	C > A	No	0.0	0.0	0.0
	rs114916387	chr4:89 073 289	T > C	Yes	2.0	0.0	0.5
ABCG2RE26	rs137884075	chr4:89 189 556	C > T	No	0.0	0.0	0.2
	rs142621223	chr4:89 189 557	G > A	No	0.0	0.0	0.3
	rs139553964	chr4:89 189 571	G > A	No	0.0	0.0	1.2
	rs149713212	chr4:89 189 602	G > A	No	1.6	0.0	0.2
	rs144565932	chr4:89 189 634	G > A	No	0.0	0.0	1.2
	rs62309980	chr4:89 189 655	C > T	Yes	0.0	0.1	0.0
	rs76888829	chr4:89 190 325	T > C	Yes	4.3	1.3	1.2
	rs9998634	chr4:89 190 395	G > C	Yes	0.4	0.0	0.0
	rs77538297	chr4:89 189 971	C > T	Yes	0.2	0.0	0.0
	rs35696062	chr4:89 190 022	G > –	Yes	NR	NR	NR

NR, not reported; SNP, single-nucleotide polymorphism.
[a]Genomic location obtained from University of California Santa Cruz (UCSC) genome browser's hg19 build.
[b]Minor allele frequency (MAF) for African American (AA), European (EUR), and Asian (AS) populations as obtained from publically available databases or haplotype frequency as a percentage of all haplotypes.
[c]Nucleotide change and conservation of the reference allele to the variant allele as obtained from the UCSC genome browser.
[d]rs72873421/rs12500008/rs12508471 (ABCG2RE1*2); rs2725264/rs2725263 (ABCG2RE8*2); rs41282399/rs2622628 (ABCG2RE9*2).

Effect of single-nucleotide polymorphisms on *in vitro* enhancer activity

All 53 SNPs in the six enhancer regions, along with the ABCG2RE1*2, ABCG2RE8*2 and ABCG2RE9*2 haplotypes, were tested for differential regulatory activity in HepG2 (Fig. 1) and HEK293T (Fig. 2) cell lines (Supplementary Table 2, Supplemental digital content 1, *http://links.lww.com/FPC/B278*). Activity of the variant and reference enhancers was determined by their ability to drive the expression of luciferase. Six ABCG2RE26
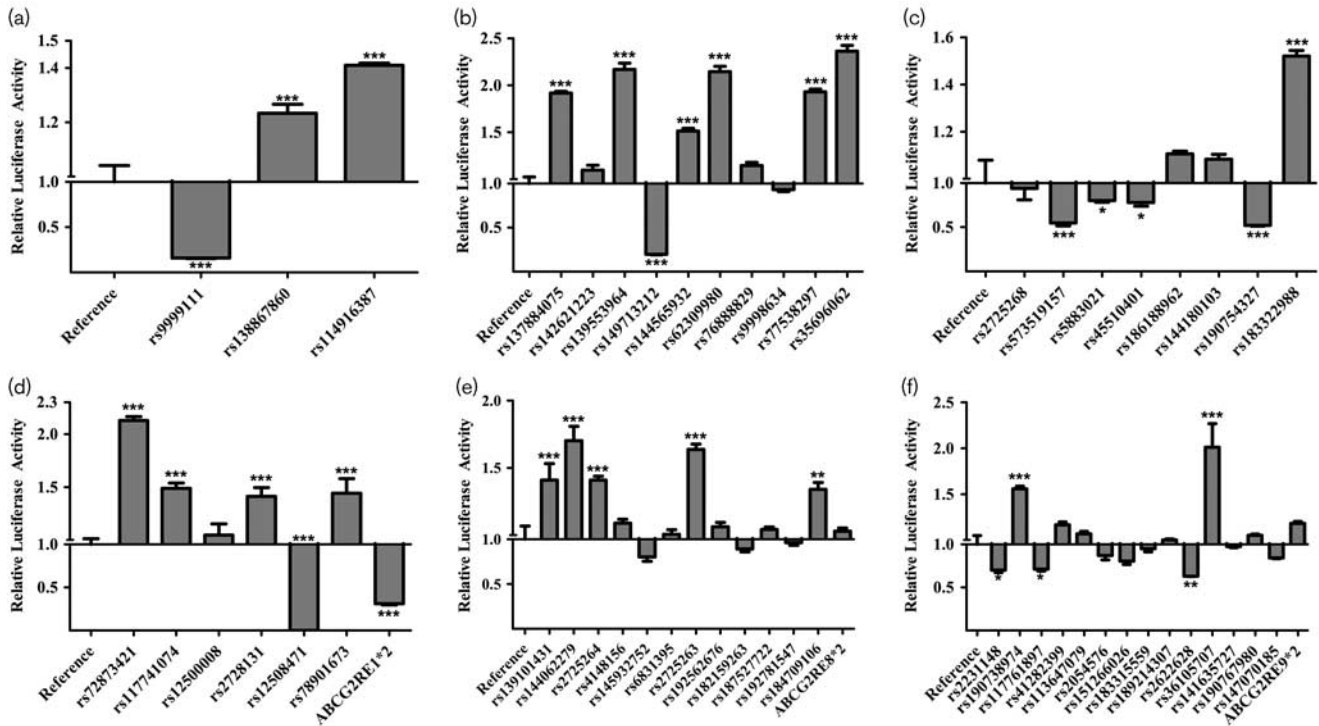
Fig. 1
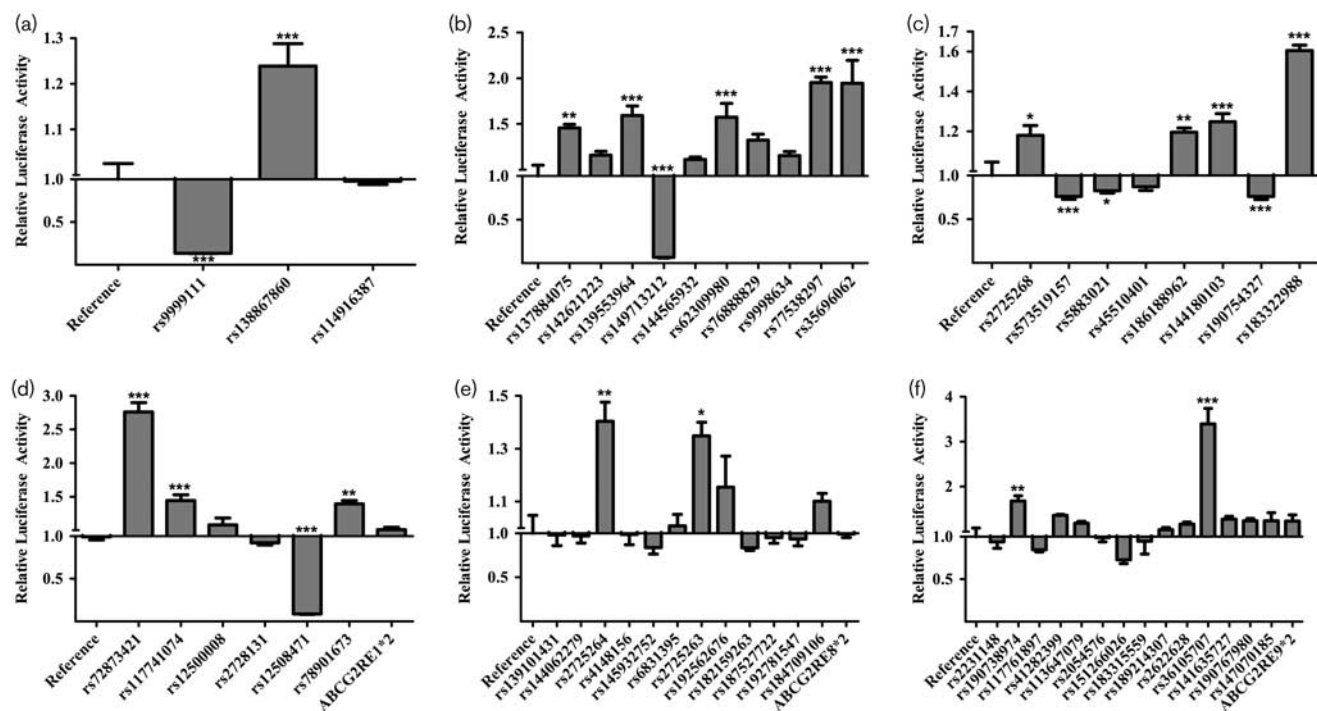
Effect of enhancer variants in HepG2 cells. The luciferase activities of reference and variant (a) ABCG2RE14, (b) ABCG2RE26, (c) ABCG2RE6, (d) ABCG2RE1, (e) ABCG2RE8, and (f) ABCG2RE9 enhancer regions were measured in a transiently transfected liver cell line. Enhancer activity is expressed as the ratio of firefly to *Renilla* luciferase activity normalized to the reference vector activity (reference is set to 1). Single-nucleotide polymorphisms are shown respective to their genomic orientation. Data are expressed as the mean ± SEM from a representative experiment with three to six wells per sequence. Differences between reference and variant enhancers were tested by an analysis of variance, followed by a post-hoc Bonferroni's multiple-comparison t-test; $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$.

SNPs and two ABCG2RE14 SNPs increased enhancer activity in both cell lines ($P < 0.001$). The ABCG2RE14 rs9999111 and ABCG2RE26 rs149713212 SNPs caused a reduction in activity in both cell lines to only 6–20% of the activity with the corresponding reference sequences ($P < 0.001$). Three ABCG2RE6 SNPs caused decreases in enhancer activity to levels that were 50–80% of the control in both cell lines ($P < 0.05$). The only ABCG2RE6 SNP to increase enhancer activity (~1.5-fold, $P < 0.001$) in both cell lines was rs183322988. ABCG2RE1 rs72873421 caused a greater than two-fold increase in enhancer activity and rs12508471 resulted in an almost complete loss of activity in both cell lines ($P < 0.001$). The ABCG2RE1*2 (rs72873421, rs12500008 and rs12508471) haplotype did not have a significant effect in HepG2 cells, but reduced enhancer activity ($P < 0.001$) in the HEK293T cell line. The ABCG2RE8 SNPs rs2725263 and rs2725264 and the ABCG2RE9 SNP rs190738974 increased enhancer activity in the hepatic and renal cell lines 1.35- to 1.68-fold; ABCG2RE9 rs36105707 was the most potent enhancer in HepG2 cells (3.39-fold above control; $P < 0.001$) and one of the most potent enhancers in HEK293 cells (2.01-fold above control; $P < 0.001$).

Although both ABCG2RE8 SNPs rs2725263 and rs2725264 caused a modest increase in enhancer activity (1.36- to 1.64-fold; $P < 0.05$), the combination of these SNPs as the ABCG2RE8*2 haplotype did not affect enhancer activity.

For subsequent *in vivo* validation, SNPs were selected from among the most significant for each of the enhancer regions. The ABCG2RE14 rs9999111 and ABCG2RE26 rs149713212 SNPs were chosen because they reduced *in vitro* activity to less than 20% of the control ($P < 0.001$). Of the four SNPs that significantly altered the ABCG2RE1 enhancer activity in both cell lines, two of them, rs72873421 and rs12508471, and the ABCG2RE1*2 haplotype, were chosen. As ABCG2RE6 is a weak *in vivo* enhancer (~ two-fold) [15], only rs183322988 was chosen for *in vivo* validation because it increased enhancer activity *in vitro* ($P < 0.001$). The ABCG2RE8 rs2725263 SNP was chosen because it had a more consistent increase in enhancer function *in vitro* compared with rs2725264. Finally, the ABCG2RE9 rs190738974 SNP was chosen for *in vivo* follow-up over the rs36105707 SNP on the basis of predicted changes in TFBS (data not shown).

Fig. 2

Effect of enhancer variants in HEK293T cells. The luciferase activities of reference and variant (a) ABCG2RE14, (b) ABCG2RE26, (c) ABCG2RE6, (d) ABCG2RE1, (e) ABCG2RE8, and (f) ABCG2RE9 enhancer regions were measured in a transiently transfected kidney cell line. Enhancer activity is expressed as the ratio of firefly to *Renilla* luciferase activity normalized to the reference vector activity (reference is set to 1). Single-nucleotide polymorphisms are shown respective to their genomic orientation. Data are expressed as the mean ± SEM from a representative experiment with three to six wells per sequence. Differences between reference and variant enhancers were tested by an analysis of variance, followed by a post-hoc Bonferroni's multiple-comparison *t*-test; *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$.

Effect of single-nucleotide polymorphisms on enhancer activity *in vivo*

Seven SNPs and one haplotype were screened for their effect on *in vivo* liver enhancer activity using the hydrodynamic tail vein injection assay [22]. The *ApoE* liver-specific enhancer [21] showed a 50-fold activity and the six reference enhancers had a range of 2- to 50-fold activity relative to pGL4.23 (Fig. 3). Three of the variants resulted in decreased *in vivo* enhancer activity compared with their respective reference enhancers (Fig. 3, $P < 0.05$), Supplementary Table 2 (Supplemental digital content 1, *http://links.lww.com/FPC/B278*). Enhancer activity for ABCG2RE14 rs9999111 was only 30% of the reference sequence ($P < 0.05$). Enhancer activity for ABCG2RE1*2 was almost completely absent (6% of control) *in vivo* ($P < 0.001$). ABCG2RE1 rs12508471 and ABCG2RE26 rs149713212 SNPs also both resulted in an almost complete loss of enhancer activity ($P < 0.001$). The only SNP to increase enhancer activity *in vivo* was the ABCG2RE8 SNP rs2725263, which increased enhancer activity by 1.5-fold ($P < 0.05$). ABCG2RE6 rs183322988, ABCG2RE1 rs72873421, and ABCG2RE9 rs190738974 SNPs had no effect on enhancer activity

in vivo. All SNPs that significantly altered activity *in vivo* were consistent with their effect on activity *in vitro*.

Associations of single-nucleotide polymorphisms with mRNA expression levels

SNPs that showed an effect on *in vivo* liver enhancer activity were tested for their association with expression levels of ABCG2 in selected tissues as described in the Methods section. The ABCG2RE14 SNP rs9999111 was associated significantly with lower ABCG2 expression in human umbilical cord LCLs and T cells from GenCord [36] tissues ($P < 0.05$; $\rho = -0.24$, Fig. 4). However, rs9999111 had no association with ABCG2 expression in the PMT liver and kidney samples (Supplementary Fig. 1, Supplemental digital content 1, *http://links.lww. com/FPC/B278*). ABCG2 expression was also not correlated with rs9999111 in liver tissues in the study by Schadt *et al.* [27] (data not shown).

The ABCG2RE1 SNP rs12500008, which occurs only as the ABCG2RE1*2 haplotype and is in perfect LD with rs12508471, was associated with lower ABCG2 expression in LCLs of Chinese individuals ($\rho = -0.251$, $P = 0.025$; Fig. 5a) and trended toward significance with

Fig. 3
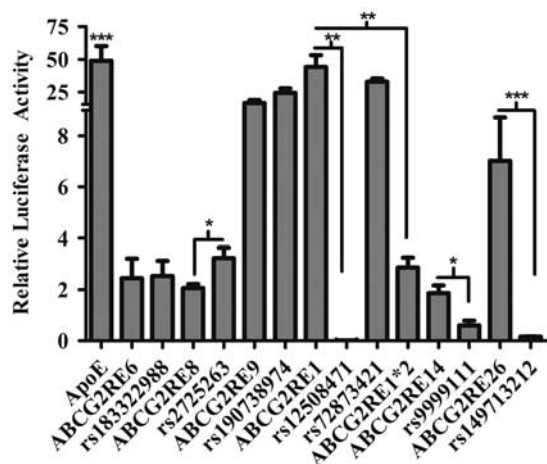
Effect of enhancer variants *in vivo*. The luciferase activity in mouse liver homogenates was measured 24 h after plasmid injection. Enhancer activity is expressed as the ratio of firefly to *Renilla* luciferase activity normalized to the empty vector (pGL4.23) activity. An enhancer for *ApoE* was used as the positive control [21]. Data are expressed as the mean ± SEM for 4–11 mice. Differences between reference and variant enhancer elements were tested using an unpaired Student's *t*-test (one variant) or an analysis of variance, followed by a Bonferroni's multiple-comparison *t*-test (for two or more variants); *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$.

Fig. 4
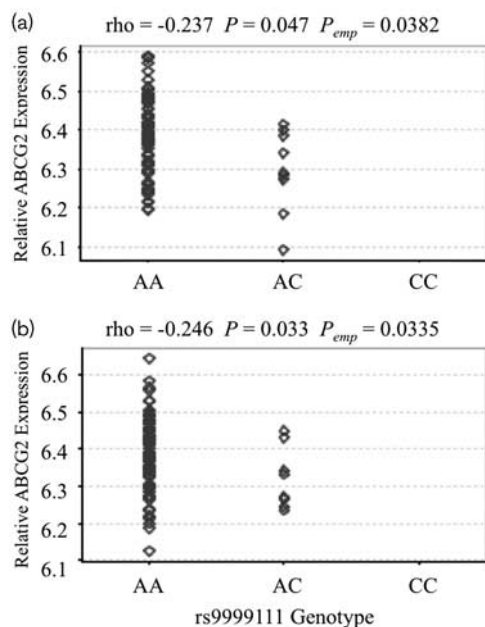
Association of rs9999111 with ABCG2 expression. Association of the ABCG2RE14 SNP rs9999111 with ABCG2 mRNA expression in 85 human umbilical cord (a) lymphocytes and (b) T cells from the GenCord study [36]. Analysis was carried out using the GeneVar [29] program with a linear regression and is shown as gene expression versus rs9999111 genotype with ρ (correlation coefficient), *P* value (*P*), and empirical *P* value (P_{emp}) indicated.

Fig. 5
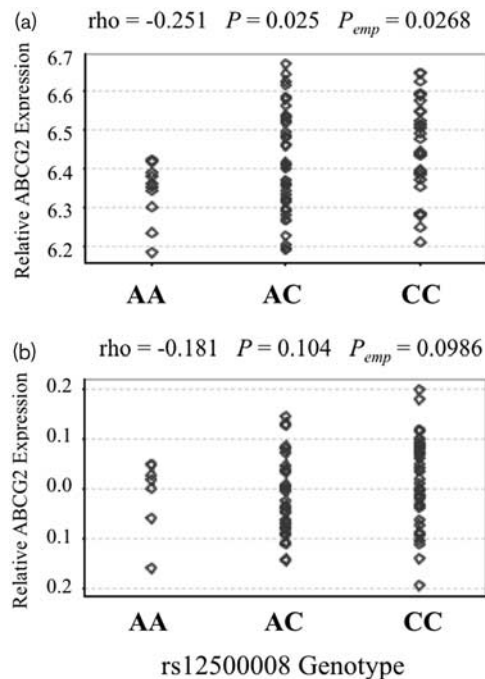
Association of rs12500008 with ABCG2 expression. Association of the ABCG2RE1 SNP rs12500008, which is in complete linkage disequilibrium with rs12508471 and part of the ABCG2RE1*2 haplotype, with ABCG2 mRNA expression in HapMap LCLs of (a) 80 Han Chinese and (b) 82 Kenyans from the Stranger study [35]. Analysis was carried out using the GeneVar [29] program with a linear regression and is shown as gene expression versus the rs12500008 genotype with ρ (correlation coefficient), *P* value (*P*), and empirical *P* value (P_{emp}) indicated.

ABCG2 expression in the Kenyan population ($\rho = -0.181$, $P = 0.1$; Fig. 5b). ABCG2RE1*2 was associated with lower PPM1K expression in adipose and skin ($\rho = -0.26$, $P = 0.015$ and $\rho = -0.32$, $P = 0.005$ respectively; Supplementary Fig. 2, Supplemental digital content 1, *http://links.lww.com/FPC/B278*). ABCG2RE1*2 was also associated with lower *PPM1K* expression in LCLs using two probes for PPM1K mRNA ($\rho = -0.22$, $P = 0.06$ and $\rho = -0.22$, $P = 0.047$; Supplementary Fig. 2, Supplemental digital content 1, *http://links.lww.com/FPC/B278*). All ABCG2RE1*2 associations with PPM1K and ABCG2 had similar effect sizes ($\rho \sim -0.25$).

Effect of genetic variants on binding of DNA to nuclear proteins

The ABCG2RE1 rs12508471 and ABCG2RE14 rs9999111 reference and variant DNA were tested for their ability to alter DNA binding to protein in an EMSA. ABCG2RE1 and ABCG2RE14 reference probes showed binding to HepG2 nuclear proteins, with specific DNA/protein interactions being susceptible to competition by unlabeled oligonucleotides (Fig. 6). Both rs12508471 and rs9999111 probes showed increased HepG2 nuclear

Fig. 6
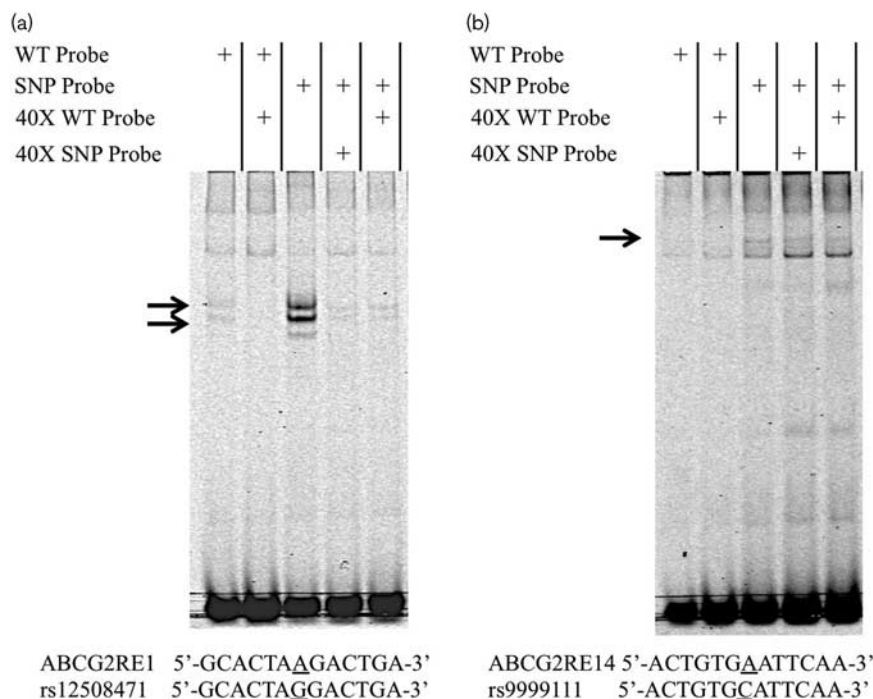
Effect of rs12508471 and rs9999111 on DNA-protein binding. Representative electrophoretic mobility shift assay using HepG2 nuclear extracts incubated with IRDye 700-labeled probes for (a) ABCG2RE1 reference or rs12508471 and (b) ABCG2RE14 reference or rs9999111 sequences. Competition assays were performed with 40-fold excess of unlabeled reference or variant oligonucleotides; specific DNA/protein bands are indicated by arrows. Reference and variant DNA sequences surrounding each single-nucleotide polymorphism (SNP) (± 6 bp) are shown below their respective gel.

protein binding compared with their reference sequence (Fig. 6). Proteins binding to the variant probes were less susceptible to competition by unlabeled reference oligonucleotides than they were to the unlabeled variant probe (Fig. 6).

Discussion

Although noncoding SNPs in the *ABCG2* gene locus have been correlated with drug response [37] and disease progression [38], relatively few have been correlated with gene expression [10,37] and none have confirmed mechanisms of action. Recent studies on other ADME genes, including transporters, have identified SNPs in *cis*-regulatory regions that are responsible for altering gene expression and contributing toward adverse drug effects [11,23,39,40]. To identify noncoding SNPs that correlate with ABCG2 expression, this study investigated the *in vitro* effect of 53 SNPs and three haplotypes on the activity of six previously identified *ABCG2* locus enhancer elements [15]. The SNP from each region with the largest *in vitro* effect was followed up *in vivo*; in the case of ABCG2RE1, two SNPs and the ABCG2RE1*2 haplotype were followed up *in vivo*.

ABCG2RE14 SNP rs9999111, which is located in intron 1 of *ABCG2*, significantly decreased the activity of the ABCG2RE14 enhancer sequence *in vivo* to levels that

were only 30% of the reference. However, rs9999111 was not associated with liver or kidney expression of ABCG2 in the present study or with liver or intestinal expression of ABCG2 in a previous study [10]. The discordance between these findings could be the result of the combination of the low activity that ABCG2RE14 has as an *in vivo* liver enhancer and a change in the binding of a tissue-specific transcription factor or nuclear receptor that is not critical for the constitutive expression of ABCG2. Despite a small sample size ($n = 85$), the rs9999111 SNP associated with decreased ABCG2 expression in human umbilical cord LCLs and T cells. The β coefficient of rs9999111 in the umbilical cord (-0.24) and the relatively low frequency of rs9999111 (MAF = 0–7.3%) indicate that this SNP contributes modestly toward variation in ABCG2 expression. Using the transcription factor database (TRANSFAC) Match program, rs9999111 is predicted to reduce the binding of immune-related transcription factors (nuclear factor κB, Gfi, and FOXD3) and ADME-related transcription factors (vitamin D receptor and hepatic nuclear factor-3β), and increase binding of aryl hydrocarbon receptor (data not shown). EMSA analysis of rs9999111 with the HepG2 nuclear protein extract confirmed increased DNA/protein interaction compared with the reference sequence. Although we can speculate that the EMSA reflects increased

binding to a transcriptional repressor, consistent with the reduced *in vivo* enhancer activity, additional studies are needed to clearly define this interaction. BCRP is a part of the placental barrier important for protection of the fetus and reduction of BCRP function has the potential to increase fetal exposure to toxic compounds such as bile acids, topotecan, and PhIP [41–43]. The role of rs9999111 in placental regulation of ABCG2 should be investigated further as it could affect fetal exposure to BCRP substrates. Follow-up studies in a cohort with more tissues could help clarify the association of this SNP with ABCG2 expression in the placenta.

ABCG2RE1*2 is a common haplotype with a frequency of 8.4% in Caucasians, 21% in African Americans, and 38% in Asians; it is made up of rs12508471, rs72873421, and rs12500008. The rs12508471 and rs72873421 SNPs affected ABCG2RE1 enhancer activity *in vitro*, but only the rs12508471 eliminated ABCG2RE1 enhancer activity both *in vitro* and *in vivo*. The rs72873421 showed increased activity *in vitro*, which could limit the decrease in activity by rs12508471 when these SNPs occur together in the ABCG2RE1*2 haplotype. However, the ABCG2RE1*2 construct still showed a significant decrease in enhancer activity in HEK293T cells and an even greater loss of hepatic enhancer activity *in vivo*. Dexamethasone, a ligand for the glucocorticoid receptor (GR), is capable of decreasing ABCG2 expression through GR-dependent, progesterone receptor-dependent, and pregnane X-receptor-dependent mechanisms [44]. Using the transcription factor database (TRANSFAC) Match program, rs12508471 gains predicted binding sites for nuclear response element binding, including pregnane X-receptor, peroxisome proliferator-activated receptor, GR, p300, and vitamin D receptor (data not shown). EMSA analysis of the rs12508471 sequence confirmed increased DNA/protein binding compared with the reference sequence. However, oligonucleotide with a GR consensus sequence could not compete with the rs12508471 probe (data not shown), and additional studies will be needed to identify the specific transcription factor(s) whose binding is altered by this SNP. ABCG2RE1*2 was not associated with ABCG2 mRNA levels in liver and kidney tissue sets. However, these were composed primarily of Caucasians, a population that has a low frequency of ABCG2RE1*2. Utilizing publically available African and Asian LCL data-sets with higher ABCG2RE1*2 frequency, we found that ABCG2RE1*2 was associated with decreased expression of ABCG2 and PPM1K in several tissues, indicating that this haplotype may modestly influence the variability of ABCG2 and PPM1K expression. ABCG2RE1 is situated just upstream of the *PKD2* promoter, but it did not correlate with PKD2 expression in any of the tissue sets (data not shown). Enhancers can work as locus control regions to regulate the expression of neighboring genes and even modulate the tissue-specific expression of multiple genes [45]. Therefore, ABCG2RE1 regulates both ABCG2 and PPM1K expression, and the ABCG2RE1*2 haplotype contributes toward altered expression of these genes, with more effect in populations with a higher frequency of the ABCG2RE1*2 haplotype.

Of the eight variants tested *in vivo*, five significantly altered *in vivo* liver enhancer activity and two of these (rs9999111 and ABCG2RE1*2) were correlated with decreased ABCG2 gene expression in human tissues. Only one of four SNPs expected to increase enhancer activity was confirmed *in vivo*, but all four of the variants that decreased enhancer activity *in vitro* also did so *in vivo*, suggesting that SNPs decreasing enhancer activity *in vitro* might have a more consistent effect *in vivo*. Although there was some correlation between *ABCG2* locus enhancer SNPs and gene expression, further association studies in more diverse cohorts and in additional tissues are needed to validate these findings. In addition, follow-up is warranted on SNPs that correlate with ABCG2 expression in cohorts of patients receiving treatment with BCRP substrates.

Conclusion
Liver enhancers identified in the *ABCG2* gene locus have many genetic polymorphisms that alter their activity *in vitro*. Several of these SNPs, including rs9999111, rs12508471, rs72873421, rs2725263, and rs149713212, alter enhancer activity *in vivo*. The rs9999111 SNP and ABCG2RE1*2 haplotype were correlated with ABCG2 expression in a tissue-specific manner, with both the rs9999111 and rs12508471 SNPs showing increased binding to nuclear proteins. Taken together, these SNPs could account for some of the reported variability in ABCG2 expression in various tissues and may influence the correlation between ABCG2 and disease risk for cancers or gout. These novel regulatory SNPs may also influence the pharmacokinetics and pharmacodynamics of BCRP substrates. An estimate of the magnitude of this contribution toward variability in ABCG2 expression and function warrants further study.

Conflicts of interest
There are no conflicts of interest.

References

1 Robey RW, To KK, Polgar O, Dohse M, Fetsch P, Dean M, Bates SE. ABCG2: a perspective. *Adv Drug Deliv Rev* 2009; **61**:3–13.
2 Jonker JW, Buitelaar M, Wagenaar E, van der Valk MA, Scheffer GL, Scheper RJ, *et al*. The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria. *Proc Natl Acad Sci USA* 2002; **99**:15649–15654.
3 Reginato AM, Mount DB, Yang I, Choi HK. The genetics of hyperuricaemia and gout. *Nat Rev Rheumatol* 2012; **8**:610–621.
4 Cusatis G, Gregorc V, Li J, Spreafico A, Ingersoll RG, Verweij J, *et al*. Pharmacogenetics of ABCG2 and adverse reactions to gefitinib. *J Natl Cancer Inst* 2006; **98**:1739–1742.
5 Yoh K, Ishii G, Yokose T, Minegishi Y, Tsuta K, Goto K, *et al*. Breast cancer resistance protein impacts clinical outcome in platinum-based chemotherapy for advanced non-small cell lung cancer. *Clin Cancer Res* 2004; **10**:1691–1697.
6 Noguchi K, Katayama K, Mitsuhashi J, Sugimoto Y. Functions of the breast cancer resistance protein (BCRP/ABCG2) in chemotherapy. *Adv Drug Deliv Rev* 2009; **61**:26–33.
7 Tsunoda S, Okumura T, Ito T, Kondo K, Ortiz C, Tanaka E, *et al*. ABCG2 expression is an independent unfavorable prognostic factor in esophageal squamous cell carcinoma. *Oncology* 2006; **71**:251–258.
8 Usuda J, Tsunoda Y, Ichinose S, Ishizumi T, Ohtani K, Maehara S, *et al*. Breast cancer resistant protein (BCRP) is a molecular determinant of the outcome of photodynamic therapy (PDT) for centrally located early lung cancer. *Lung Cancer* 2010; **67**:198–204.
9 Mao Q, Unadkat JD. Role of the breast cancer resistance protein (BCRP/ABCG2) in drug transport–an update. *AAPS J* 2015; **17**:65–82.
10 Poonkuzhali B, Lamba J, Strom S, Sparreboom A, Thummel K, Watkins K, Schuetz E. Association of breast cancer resistance protein/ABCG2 phenotypes and novel promoter and intron 1 single nucleotide polymorphisms. *Drug Metab Dispos* 2008; **36**:780–795.
11 Georgitsi M, Zukic B, Pavlovic S, Patrinos GP. Transcriptional regulation and pharmacogenomics. *Pharmacogenomics* 2011; **12**:655–673.
12 Smith RP, Lam ET, Markova S, Yee SW, Ahituv N. Pharmacogene regulatory elements: from discovery to applications. *Genome Med* 2012; **4**:45.
13 Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, *et al*. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* 2009; **41**:1216–1222.
14 Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat Rev Genet* 2010; **11**:533–538.
15 Eclov RJ, Kim MJ, Smith RP, Liang X, Ahituv N, Kroetz DL. In vivo hepatic enhancer elements in the human ABCG2 locus. *Drug Metab Dispos* 2017; **45**:208–215.
16 Abecasis GR, Auton A, Brooks LD, dePristo MA, Durbin RM, Handsaker RE, *et al*. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**:56–65.
17 Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, *et al*. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**:851–861.
18 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**:2938–2939.
19 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**:263–265.
20 Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* 2008; **8**:91.
21 Simonet WS, Bucay N, Lauer SJ, Taylor JM. A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein E and C-I genes in transgenic mice. *J Biol Chem* 1993; **268**:8221–8229.
22 Kim MJ, Ahituv N. The hydrodynamic tail vein assay as a tool for the study of liver promoters and enhancers. *Methods Mol Biol* 2013; **1015**:279–289.
23 Kim MJ, Skewes-Cox P, Fukushima H, Hesselson S, Yee SW, Ramsey LB, *et al*. Functional characterization of liver enhancers that regulate drug-associated transporters. *Clin Pharmacol Ther* 2011; **89**:571–578.
24 Kroetz DL, Yee SW, Giacomini KM. The pharmacogenomics of membrane transporters project: research at the interface of genomics and transporter pharmacology. *Clin Pharmacol Ther* 2010; **87**:109–116.
25 Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19**:185–193.
26 Boes T, Neuhäuser M. Normalization for Affymetrix GeneChips. *Methods Inf Med* 2005; **44**:414–417.
27 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, *et al*. Mapping the genetic architecture of gene expression in human liver. *Plos Biol* 2008; **6**:e107.
28 Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**:61–70.
29 Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, *et al*. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010; **26**:2474–2476.
30 Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, *et al*. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 2011; **7**:e1002003.
31 Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, *et al*. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012; **44**:1084–1089.
32 Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, *et al*. Population genomics of human gene expression. *Nat Genet* 2007; **39**:1217–1224.
33 Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, *et al*. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1**:e78.
34 Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, *et al*. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007; **315**:848–853.
35 Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, *et al*. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* 2012; **8**:e1002639.
36 Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, *et al*. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009; **325**:1246–1250.
37 Rudin CM, Liu W, Desai A, Karrison T, Jiang X, Janisch L, *et al*. Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *J Clin Oncol* 2008; **26**:1119–1127.
38 Campa D, Butterbach K, Slager SL, Skibola CF, de Sanjosé S, Benavente Y, *et al*. A comprehensive study of polymorphisms in the ABCB1, ABCC2, ABCG2, NR1I2 genes and lymphoma risk. *Int J Cancer* 2012; **131**:803–812.
39 McLeod HL, Sargent DJ, Marsh S, Green EM, King CR, Fuchs CS, *et al*. Pharmacogenetic predictors of adverse events and response to chemotherapy in metastatic colorectal cancer: results from North American Gastrointestinal Intergroup Trial N9741. *J Clin Oncol* 2010; **28**:3227–3233.
40 Matsson P, Yee SW, Markova S, Morrissey K, Jenkins G, Xuan J, *et al*. Discovery of regulatory elements in human ATP-binding cassette transporters through expression quantitative trait mapping. *Pharmacogenomics J* 2012; **12**:214–226.
41 Blazquez AG, Briz O, Romero MR, Rosales R, Monte MJ, Vaquero J, *et al*. Characterization of the role of ABCG2 as a bile acid transporter in liver and placenta. *Mol Pharmacol* 2012; **81**:273–283.
42 Jonker JW, Merino G, Musters S, van Herwaarden AE, Bolscher E, Wagenaar E, *et al*. The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk. *Nat Med* 2005; **11**:127–129.
43 Jonker JW, Smit JW, Brinkhuis RF, Maliepaard M, Beijnen JH, Schellens JH, Schinkel AH. Role of breast cancer resistance protein in the bioavailability and fetal penetration of topotecan. *J Natl Cancer Inst* 2000; **92**:1651–1656.
44 Honorat M, Mesnier A, di Pietro A, Lin V, Cohen P, Dumontet C, Payen L. Dexamethasone down-regulates ABCG2 expression levels in breast cancer cells. *Biochem Biophys Res Commun* 2008; **375**:308–314.
45 Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. *Blood* 2002; **100**:3077–3086.