Association of Breast Cancer Resistance Protein/ABCG2 Phenotypes and Novel Promoter and Intron 1 Single Nucleotide Polymorphisms

Balasubramanian Poonkuzhali, Jatinder Lamba, Stephen Strom, Alex Sparreboom, Kenneth Thummel, Paul Watkins, and Erin Schuetz

Department of Pharmaceutical Sciences, St. Jude Children’s Research Hospital, Memphis, Tennessee (B.P., J.L., A.S., E.S.); Department of Pathology, University of Pittsburgh, Pittsburgh, Pennsylvania (S.S.); Department of Pharmaceutics, University of Washington, Seattle, Washington (K.T.); and Department of Medicine, Division of Hepatology, University of North Carolina, Chapel Hill, North Carolina (P.W.)

Received August 16, 2007; accepted January 3, 2008

ABSTRACT:

The hypothesis was tested that sequence diversity in breast cancer resistance protein (BCRP)’s cis-regulatory region is a significant determinant of BCRP expression. The BCRP promoter and intron 1 were resequenced in lymphoblast DNA from the polymorphism discovery resource (PDR) 44 subset. BCRP single nucleotide polymorphisms (SNPs) were genotyped in donor human livers, intestines, and lymphoblasts quantitatively phenotyped for BCRP mRNA expression. Carriers of the −15622C>T SNP had lower BCRP expression in multiple tissues. The intron 1 SNP SNP 16702C>T was associated with high expression in livers; 1143G>A was associated with low expression in intestine; 12283T>C was associated with high expression in the PDR44 and White livers. The −15994C>T promoter SNP was significantly associated with higher BCRP expression in multiple tissues. Patients with the −15994C>T genotype had substantially higher clearance of p.o. imatinib. We next determined whether BCRP expression was related to polymorphic alternative splicing or alternative promoter use. Liver polymorphically expressed an alternatively spliced mRNA [splice variant (SV) 1] skipping exon 2. Although SV1+ livers did not uniformly carry the exon 2 G34A allele, 90% of G34A livers expressed SV1 (versus 4% of 34G livers). BCRP mRNA was significantly lower among Hispanic livers with the G34A variant genotype and may be due, in part, to polymorphic exon 2 splicing. Analysis of allele expression imbalance (AEI) showed that PDR44 samples with AEI had lower BCRP mRNA expression; however, no linked cis-polymorphisms were identified. BCRP used multiple promoters, and livers differentially using alternative exon 1b had lower BCRP. In conclusion, BCRP expression in lymphoblasts, liver, and intestine is associated with novel promoter and intron 1 SNPs.

The efflux transporter breast cancer resistance protein (BCRP) (ABCG2, MXR, ABCp) is expressed in many tissues, including intestine, placenta, mammary gland, and liver. BCRP plays an important role in the absorption, distribution, and elimination of a growing list of drugs that are its substrates, including the anticancer agents mitoxantrone, doxorubicin, vitamin B12 (Jonker et al., 2007; van Herwaarden et al., 2007) and can be inhibited by a growing number of drugs (e.g., gefitinib, nelfinavir) (Burger et al., 2004; Ozvegy-Laczka et al., 2004). Moreover, BCRP is highly expressed as a stem cell marker in a variety of cell types (Zhou et al., 2001; Smalley and Clarke, 2005).

This work is supported in part by the National Institutes of Health (NIH)/National Institute of General Medical Sciences Pharmacogenetics Research Network and Database (U01GM61374, http://pharmgkb.org) under Grant U01 GM61393, and the NIH P30 CA21765 Cancer Center Support grant, and by the American Lebanese Syrian Associated Charities (ALSAC).

Article, publication date, and citation information can be found at http://dmd.aspetjournals.org.

[1] The online version of this article (available at http://dmd.aspetjournals.org) contains supplemental material.

ABBREVIATIONS: BCRP, breast cancer resistance protein; SNP, single nucleotide polymorphism; PDR, polymorphism discovery resource; kb, kilobase; TF, transcription factor; HNF, hepatic nuclear factor; NF, nuclear factor; CEBP, CCAAT/enhancer binding protein; UCSC, University of California, Santa Cruz; ECR, evolutionary conserved regions; PCR, polymerase chain reaction; SV, splice variant; AEI, allele expression imbalance; LOH, loss of heterozygosity; PPAR, peroxisome proliferator-activated receptor; CEPH, Centre de’Etude du Polymorphism Humain; bp, base pair; ATG, translation start site; EST, expressed sequence tag; LD, linkage disequilibrium.
BCRP and exposed to a diet rich in chlorophyll experienced increased phototoxicity and protoporphyria. This is relevant to humans because phototoxicity, skin blisters, elevated porphyrins, and iron have now been reported in some humans treated with the BCRP substrate/ inhibitor imatinib mesylate (Ho et al., 2003). Thus, humans with altered BCRP function resulting from variant alleles may be expected to show altered drug disposition, efficacy, and toxicities including porphyrins and phototoxicity.

We previously identified several naturally occurring BCRP variants (Zamber et al., 2003). Among the coding single nucleotide polymorphisms (SNPs), G34A (V12M) in exon 2 and C421A (Q141K) in exon 5 occur in most racial groups but with a higher allele frequency in Asians and Hispanics. The C421A polymorphism is associated with similar levels of mRNA but decreased protein expression in PA317 cells (Imai et al., 2002). Several groups have reported that the C421A genotype is associated with altered pharmacokinetic parameters of some BCRP substrates (de Jong et al., 2004; Sparreboom et al., 2004), whereas others have not found an association (Mathijssen et al., 2003). Notably, the C421A allele has the signature of recent positive selection, and it is strongest in the Asian population (Wang et al., 2007), suggesting there is some advantageous property of this genotype. The BCRP G34A variant has been reported to have transport activity similar to wild-type BCRP in transport of methotrexate, dehydroepiandrosterone sulfate (Kondo et al., 2004), and porphyrin (Tamura et al., 2006). However, the frequency of the variant coding alleles cannot completely explain most human variation in BCRP expression or activity.

We tested the hypothesis that cis-polymorphisms affecting BCRP expression are present in DNA regulatory sequences in the promoter and intron 1 and could further explain variation in BCRP expression. Moreover, because variation in BCRP expression in tumors and stem cells involved differential use of alternative first exons (Nakanishi et al., 2006; Zong et al., 2006), we simultaneously determined whether BCRP expression (and associated SNPs) might be related to polymorphic splicing or differential promoter use.

Materials and Methods

Subjects. The Institutional Review Boards and Clinical Research Advisory Committees at St. Jude Children’s Research Hospital, the University of Pittsburgh, the University of North Carolina, and the University of Washington approved the use of these tissue samples for genotyping studies.

Sample Set I. A subset of 44 DNA samples from the polymorphism discovery resource (PDR-44) was purchased from the Coriell cell/DNA repository (http://ccr.coriell.org/nigms/products/pdr.html). These samples represent the major ethnic groups in the United States (European 26%, African 26%, Mexican 13%, Native American 6%, and Asian 26%) and were used for SNP discovery.

Sample Set II. Human liver tissue for sample set II was processed through the St. Jude Liver Resource at St. Jude Children’s Research Hospital and was provided by the Liver Tissue Procurement and Distribution System (National Institutes of Health contract N01-DK-92310) and by the Cooperative Human Tissue Network. DNA from a set of 60 liver samples from three different racial groups [White (n = 15; 10 males and 5 females), African Americans (n = 17; 9 males and 8 females), and Hispanics (n = 28; 19 males and 9 females)] were included for genotyping the common SNPs that were identified in the PDR-44. In addition to these SNPs, the full-length BCRP cDNA was amplified and sequenced to look for any other coding sequence variations in these samples.

Sample Set III. DNA from 28 White intestinal biopsy samples (10 from the University of Washington and 18 from the University of North Carolina) was processed as described previously (Mouly et al., 2005) and genotyped for the common SNPs identified in the PDR-44.

Patient DNA. Samples were obtained at steady state from adult patients positive for c-kit gastrointestinal stromal tumors. Pharmacokinetic data have been described previously on a subset of 82 patients (Gardner et al., 2006). None of the patients received any medication aside from imatinib that could possibly influence the activity of BCRP or the pharmacokinetic profile of imatinib. The study protocol was approved by the Institutional Review Boards (Leuven, Belgium and Rotterdam, The Netherlands), and written informed consent was obtained from each patient.

Genomic DNA was extracted from 1 ml of plasma using the UltraSens Virus Kit (Qiagen, Valencia, CA), and the REPLI-g mini/ midi kit (Qiagen) was used to amplify genomic DNA. Imatinib concentrations in plasma were determined by validated analytical methods using liquid chromatography with tandem mass-spectrometric detection (Guettens et al., 2003). Pharmacokinetic parameters, including area under curve, steady-state concentration, and apparent oral clearance, were obtained for each patient by noncompartmental analysis using WinNonlin version 5.0 (Pharsight, Mountain View, CA).

The association of variant genotypes with the pharmacokinetic parameters of imatinib was based on a nonparametric Mann-Whitney U test (two-group comparison) or a Kruskal-Wallis test (multiple-group comparison). A p < 0.05 was considered statistically significant.

In Silico Analysis of the BCRP/ABCG2 Gene to Identify Potential Regulatory Regions to Resequance. Several web-based bioinformatic tools were used to screen 49 kilobases (kb) of the BCRP gene (30 kb proximal promoter, introns, 18.9 kb intron 1) for the presence of DNA response elements for liver-enriched transcription factors (TFs) and for regions of high evolutionarily conservation between multiple species. Cister plot (http://zlab.bua.edu/~mfrieh/cister.html), NUBIScan (http://www.nubiscan.uniba.ch/), and Transfac (http://transfac.gbf.de/TRANSFAC/lists/matrixByName.html) were used to identify regions harboring DNA response elements for various TFs. [TF matrices included hepatic nuclear factor (HNF) 1, HNF3, GATA, activator protein-1, CDX, FOX, SP1, nuclear factor (NF), TATA, CAAT, YY1, DR3 and DR4, pregnane X receptor, DR3, DR1, HNF4, DR1, HNF1, HNF3α, HNF4, HNF3β, CCAAT/enhancer binding protein (CEBP)-α, CEBP, CEBPβ, HNF6, CEBPγ, GATA4, glucocorticoid receptor, HNF3γ, NF-1, NF-E2, SP1, and TATA, cAMP response element, estrogen response element, NF-1, E2F, Maf-F2, Myf, CCAAT, activator protein-1, Ets, Myc, GATA, LSF, SRF, and Tef.] The University of California, Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu), evolutionary conserved region (ECR) (http://ecrbrowser.dcode.org/), and rVISTA (http://genome.lbl.gov/vista/vista_submit.html) were used to identify regions of evolutionary conservation on the BCRP gene between humans and other species.

DNA Sequencing of BCRP Polymere Chain Reaction Amplicons. The regions identified by in silico analysis were amplified from genomic DNA using specific primers (Table 1). Amplification was carried out in a 1 × polymerase chain reaction (PCR) buffer using 50 ng of DNA, 10 pmol each of forward and reverse primers, 0.2 mM dNTPs, and 1.5 units of Taq polymerase (Expand High Fidelity PCR System, Roche, Basel, Switzerland). The PCR conditions include initial denaturation at 95°C for 3 min, followed by 32 to 34 cycles of denaturation at 95°C, annealing at appropriate temperatures, and synthesis at 72°C, with final synthesis at 72°C for 10 min. PCR products were checked for the correct size by agarose gel electrophoresis. Before sequencing, unincorporated nucleotides and primers were removed by incubation with shrimp alkaline phosphatase (U.S. Biochemical Corp., Cleveland, OH) and exonuclease I (U.S. Biochemical Corp.) for 30 min at 37°C, followed by enzyme inactivation at 80°C for 15 min. Sequencing was carried out on an ABI Prism 3700 Automated Sequencer (Applied Biosystems, Foster City, CA) using the PCR primers or internal sequencing primers (Table 1). Sequences were assembled using the Phred-Phrap-Consed package (University of Washington, Seattle, WA; http://droog.mbt.washington.edu/PolymPhred.html), which automatically detects the presence of heterozygous single nucleotide substitutions by fluorescence-based sequencing of PCR products. Two regions were problematic to genotype: 1) a region in intron 1 (position 32222 to 32252 in AC084732) with polymorphic repeats in some PDR-44 DNAs and 12 of 60 liver samples; and 2) a region approximately 13 kb from the transcription start site (designated “insertion” in Table 2) did not produce any amplification product in 30% of the samples in each of the cohorts screened. None of the various strategies (e.g., repositioning the primers, long-range PCR) used to identify the nature of the genetic variation (large deletion or insertion) was successful. DNA samples that were homozygous, but not heterozygous, for this “variation” could be genotyped.
<table>
<thead>
<tr>
<th>Region Amplified</th>
<th>Relative to TSS</th>
<th>Position</th>
<th>Primer Sequence (S'→S')</th>
<th>Amplicon Size</th>
<th>Annealing/Extension Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC097484 Region I</td>
<td>71599–72533</td>
<td>137616F</td>
<td>AAGGAGAAAACCAGCAACCACCCTCT</td>
<td>934</td>
<td>59/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>96947R</td>
<td>TGGCGCCATCAAAACTGAAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region II</td>
<td>28916–30930</td>
<td>138550R</td>
<td>TTAGAGAAGGCTGCTACCTGGTGT</td>
<td>2014</td>
<td>59/68</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>96237F</td>
<td>TCTTGGCACTCTGGTTAAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region III</td>
<td>18132–14320</td>
<td>2227F</td>
<td>AAGGAGAAGCCCAACATCCT</td>
<td>3812</td>
<td>59/68</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>2792F</td>
<td>TTTAACCAAGGAGGCTGCAAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region IV</td>
<td>13890–14000</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>2435</td>
<td>59/68</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>15782R</td>
<td>TGGTGCCATCAAAACTGAAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region V</td>
<td>13601–11457</td>
<td>6758F</td>
<td>AATGGGCGAGTTCGTACCTGGTCTC</td>
<td>2144</td>
<td>59/68</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>17018F</td>
<td>GACATCCCATTTCACACCCAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region VI</td>
<td>6594–4577</td>
<td>2227F</td>
<td>AATGGGCGAGTTCGTACCTGGTCTC</td>
<td>2834</td>
<td>57/72, +MgCl₂</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>18990R</td>
<td>GCTCCATTGCCACCTGGTTCCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region VII</td>
<td>14203–1369</td>
<td>16156F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>2017</td>
<td>58/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>16900R</td>
<td>GAAACACATGACAAAGATGCTCTCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region VIII</td>
<td>814–1621</td>
<td>19545F</td>
<td>TGGATGTTGGAACGATTTAATA</td>
<td>2072</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21980R</td>
<td>GCACTTGGTGGAACGATTTAATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region IX</td>
<td>6711–7596</td>
<td>2227F</td>
<td>AAGGAGAAGCCCAACATCCT</td>
<td>885</td>
<td>59/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21980R</td>
<td>GCACTTGGTGGAACGATTTAATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region X</td>
<td>9296–10427</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>1131</td>
<td>59/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>15782R</td>
<td>TGGTGCCATCAAAACTGAAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region XI</td>
<td>11481–13553</td>
<td>16156F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>2027</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>17018F</td>
<td>GACATCCCATTTCACACCCAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region XII</td>
<td>13745–1704</td>
<td>16156F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>3659</td>
<td>59/68</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 2f</td>
<td>6711–7596</td>
<td>2227F</td>
<td>AAGGAGAAGCCCAACATCCT</td>
<td>2462</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 2r</td>
<td>9296–10427</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>1596</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 7F</td>
<td>6711–7596</td>
<td>2227F</td>
<td>AAGGAGAAGCCCAACATCCT</td>
<td>4221</td>
<td>63/72, +MgCl₂</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 7R</td>
<td>9296–10427</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>2072</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 10f</td>
<td>6711–7596</td>
<td>2227F</td>
<td>AAGGAGAAGCCCAACATCCT</td>
<td>4221</td>
<td>63/72, +MgCl₂</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 10r</td>
<td>9296–10427</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>2072</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 14f</td>
<td>61087F</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>3533</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 14r</td>
<td>86516R</td>
<td>2276F</td>
<td>GACAGGAGAGGATCCTCAATGCTC</td>
<td>3533</td>
<td>60/72</td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
<td>21332R</td>
<td>TTTAGTGCACTCCTGGTCTGATTC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RNA Extraction and Real-Time Reverse Transcription-PCR. RNA was isolated from sample sets I through III using TRIzol reagent (Invitrogen, Carlsbad, CA). The integrity of the isolated RNA was examined by quantitating the A260/A280 ratio and resolving the RNA on a 1% agarose gel. Total RNA was isolated from sample sets I through III using TRIzol reagent (Invitrogen). Relative quantitation of BCRP RNA using real-time PCR was performed as described previously (Zamber et al., 2003) using QuantiTect SYBR green PCR kit (Qiagen). Amplification was done with the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Because we observed an additional dissociation curve (amplification of an additional alternative mRNA) in some of the liver samples with these real-time primers (real-time F1/r), we redesigned the forward primer (real-time f2) (Table 1) to amplify only the wild-type mRNA for quantitation. (The rationale for redesigning the primer was as follows: the additional dissociation curve we observed in some samples was caused by insertion of part of intron 7; to avoid amplifying the splice variant (SV), we redesigned the forward primer flanking exons 7 and 8). The sequences of primers used for real-time PCR are given in Table 1. Relative BCRP RNA expression was calculated by the comparative Ct method and normalized relative to a reference transcript (Zamber et al., 2003).

Detection of Alternative First Exons and SVs of BCRP. The UCSC genome browser of the May 2004 Human Genome assembly identified at least three alternative first exons and various BCRP alternatively spliced mRNAs. We designed forward primers in these alternative first exons and a reverse primer in exon 6 of BCRP to check for the presence of these alternative first exons in our tissues of interest: liver, intestine, and lymphoblasts. We also tested for the presence of the SVs shown in the UCSC genome browser by designing primers in the flanking exons (Table 1). To check whether these alternative first exons result in a full-length cDNA, reverse primers in exon 16 and a nested reverse primer in exon 10 were used (Table 1).

Analysis of Allelic Imbalance. We used two different assays to determine allele expression imbalance (AEI). First, we sequenced the exon 2 G34A and exon 5 C421A SNPs in DNA and cDNA samples, and those showing disparity between the genomic and cDNA sequence [loss of heterozygosity (LOH) in the exon 5 C421A SNPs in DNA and cDNA samples, and those showing disparity between the genomic and cDNA sequence] were identified as having allelic imbalance. The results were confirmed in the promoter of human and mouse BCRP by using an exon 5 allele-specific quantitation kit from ABI to quantitate BCRP signal intensity from each allele in DNA versus RNA (Applied Biosystems Assays on Demand), and the ratio of signal intensity for the C versus A allele was calculated.

Quantification of BCRP Gene Copy Number by Real-Time PCR. BCRP gene copy number was determined by quantitative real-time PCR of genomic DNA as described previously (Lamba et al., 2006) from livers (White, n = 15; African Americans, n = 4; Hispanics, n = 3) and PDR samples representing the highest, medium, and lowest BCRP RNA expressors (n = 5 in each range/tissue) using primer pairs (Table 1) to amplify portions of exon 2 or exon 5 of BCRP using the SYBR Green PCR kit (Qiagen) according to the manufacturer’s instructions. Each sample was quantitated in duplicate, and specificity of amplification was determined by doing melt curve analysis. Standard curves were run with each plate. The quantitative values were determined for BCRP using the ΔΔCt method and normalized relative to CYF3A4 copy number values (Lamba et al., 2006). Statistical Analysis. Statistical analysis to evaluate possible genotype-phenotype relationships was carried out using the R statistical package (http://www.R-project.org) and the Wilcoxon or Kruskal-Wallis test.

Results

In Silico Analysis of the BCRP Promoter and Intron 1. Because the human BCRP promoter (105 kb) and intron 1 (19 kb) total 124 kb, we used bioinformatic tools to judiciously choose regions for resequencing. Selection criteria targeted those regions most likely to be functionally important: evolutionarily conserved sequences and regions containing clusters of TF binding sites. Figure 1, A and B, shows snapshots from the UCSC and ECR genome browsers with regions containing clusters of TF binding sites using matrices from Transfac and NUBIScan in the Rfam/Cluster Buster program (Fig. 1D), and additional regions were identified in the promoter of human and mouse BCRP by rVISTA analysis. The same regions of BCRP were next screened for clusters of TF binding sites using matrices from Transfac and NUBIScan in the Cister/Cluster Buster program (Fig. 1D), and additional regions were identified to resequence. In addition, we screened a 1-kb region that encompassed an alternative exon 1c located −72 kb upstream of BCRP and that is used in hematopoietic stem cells (Zong et al., 2006) (region 1, Table 1).

TFs found frequently in the ECR regions included HNF1 (liver- and gut-enriched TF), consistent with BCRP expression in liver and intestine; peroxisome proliferator-activated receptor (PPAR) γ and CEBPα and CEBPβ, consistent with the role of the ABCG proteins in lipid and sterol homeostasis; CEBP and Cdx family TFs, important for regulating gene expression in the small intestine and for the reported role of Cdx2 in regulating iron homeostasis because BCRP transports

<table>
<thead>
<tr>
<th>Region Amplified</th>
<th>Relative to TSS</th>
<th>Position</th>
<th>Primer Sequence (5′→3′)</th>
<th>Amplicon Size</th>
<th>Annealing/Extension Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse transcription-PCR primers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC084732</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 1b-f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM-004827</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 1a-f</td>
<td>437F</td>
<td></td>
<td>CCGAGGAGCCTTGGGATACCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 6R</td>
<td>1057R</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP real-time f1</td>
<td>1304F</td>
<td></td>
<td>GAGAAGGCTTTGGATACCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP real-time f2</td>
<td>1324F</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP real-time r</td>
<td>1411R</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 8F</td>
<td>1348F</td>
<td></td>
<td>GGGCTATTAAATTACCTCGACAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 10R</td>
<td>1739R</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 14F</td>
<td>2152F</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 16R</td>
<td>2431R</td>
<td></td>
<td>TCTGGATGCGTTGTCCTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC092408</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP exon 1c-f</td>
<td>70F</td>
<td></td>
<td>CTCGAGAAGCCTTCAATTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copy number primers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYF3A4 CNVF</td>
<td></td>
<td></td>
<td>CCCAGAAGTGGCCATGGAACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYF3A4 CNVR</td>
<td></td>
<td></td>
<td>CAGTAAAGGGAAGAGAGGAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 2

<table>
<thead>
<tr>
<th>Position from TSS</th>
<th>ATG Position in Refseq</th>
<th>dbSNP</th>
<th>Sequence</th>
<th>Region</th>
<th>AF in PDR44 Whites</th>
<th>Hispanics</th>
<th>African Americans</th>
<th>Effect on mRNA Expression and TRANSFAC Change</th>
<th>Region</th>
<th>Tag SNP</th>
</tr>
</thead>
</table>

**Numbering based on AC097484**

<table>
<thead>
<tr>
<th>#</th>
<th>Position</th>
<th>AF in Region</th>
<th>dbSNP</th>
<th>Sequence</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-79753</td>
<td>0.59</td>
<td>0.57</td>
<td>0.55</td>
<td>0.62</td>
</tr>
<tr>
<td>2</td>
<td>-71762</td>
<td>0.04</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>3</td>
<td>-71813</td>
<td>0.06</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>4</td>
<td>-71997</td>
<td>0.06</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>5</td>
<td>-71961</td>
<td>0.04</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>6</td>
<td>-72358</td>
<td>0.12</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>7</td>
<td>-69504</td>
<td>0.55</td>
<td>cagacc[c]tggagt</td>
<td>Upstream</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-30639</td>
<td>0.28</td>
<td>0.27</td>
<td>0.30</td>
<td>0.41</td>
</tr>
<tr>
<td>9</td>
<td>-30562</td>
<td>0.70</td>
<td>0.47</td>
<td>0.59</td>
<td>0.26</td>
</tr>
<tr>
<td>10</td>
<td>-30109</td>
<td>0.70</td>
<td>0.50</td>
<td>0.54</td>
<td>0.35</td>
</tr>
<tr>
<td>11</td>
<td>-15846</td>
<td>0.09</td>
<td>0.07</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td>12</td>
<td>-15756</td>
<td>0.22</td>
<td>0.07</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>13</td>
<td>-15672</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>-13680</td>
<td>0.07</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>15</td>
<td>-13000</td>
<td>0.09</td>
<td>0.07</td>
<td>0.13</td>
<td>0.06</td>
</tr>
</tbody>
</table>

**Numbering based on AC084732**

<table>
<thead>
<tr>
<th>#</th>
<th>Position</th>
<th>AF in Region</th>
<th>dbSNP</th>
<th>Sequence</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>-17824</td>
<td>0.19</td>
<td>0.23</td>
<td>0.09</td>
<td>0.47</td>
</tr>
<tr>
<td>17</td>
<td>-17750</td>
<td>0.19</td>
<td>0.23</td>
<td>0.09</td>
<td>0.44</td>
</tr>
<tr>
<td>18</td>
<td>-17243</td>
<td>0.16</td>
<td>0.27</td>
<td>0.14</td>
<td>0.59</td>
</tr>
<tr>
<td>19</td>
<td>-17242</td>
<td>0.56</td>
<td>0.63</td>
<td>0.55</td>
<td>0.76</td>
</tr>
<tr>
<td>20</td>
<td>-17084</td>
<td>0.00</td>
<td>0.23</td>
<td>0.14</td>
<td>0.38</td>
</tr>
<tr>
<td>21</td>
<td>-15994</td>
<td>0.22</td>
<td>0.07</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>22</td>
<td>-15846</td>
<td>0.50</td>
<td>0.59</td>
<td>0.57</td>
<td>0.62</td>
</tr>
<tr>
<td>23</td>
<td>-15756</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-15672</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-13680</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Bold text indicates significant association with BCRP expression in one or more tissues.**
the heme precursor protoporphrin IX; NKX25/Csx, a homeodomain factor important in development in many tissues that could be important for BCRP expression in embryonic stem cells; and SRFQ2-5, serum response TFs important for early growth response.

**BCRP Processed Pseudogene Intron 1 of the NOX5 Gene.** Although BCRP is on chromosome 4q22, bioinformatic analysis revealed an intron-less pseudogene pseudogene residing within intron 1 of the Nox5 gene on chromosome 15. The pseudogene showed 89% similarity to BCRP in the coding region but lacked exons 1, 15, and 16. It appeared to be a processed pseudogene because there is a transcribed mRNA (CR610432) in GenBank with this sequence.

**BCRP Sequence Variations.** Ninety BCRP SNPs (41 in the promoter and 49 in the introns) were identified (Table 2). Forty-three SNPs were novel, and 47 were in the dbSNP database and/or genotyped in the Centre de’Etude du Polymorphism Humain (CEPH) samples in the HapMap project.

A comparison of LD between pairs of BCRP loci using the “Haploview” analysis of HapMap data in White (CEPHS) indicated that the BCRP gene is framed by seven LD blocks, with three LD blocks in the promoter region (Supplementary Fig. 1). However, the degree of LD within each block was low, with six to eight haplotypes in many of the blocks. Given the size of our cohorts, it was not possible to determine the effect of BCRP haplotypes on BCRP phenotype.

**Phenotyping BCRP mRNA Expression.** BCRP mRNA expression was quantified by real-time PCR in the various cohorts. Figure 2 shows the range of BCRP RNA expression in livers between three different racial groups: the median (relative BCRP) was lower in Hispanics (81.2 ± 953) compared with White (101.3 ± 2534) and African American (291 ± 6877) livers ($p = 0.09$) (note inset graph is Log2). The mean level of BCRP mRNA was higher in female (124.6 ± 71) versus male (95 ± 78) intestines (not shown), similar to what we previously reported (Zamber et al., 2003). However, BCRP mRNA was not different in female versus male livers (105.43 ± 9329 versus 95.6 ± 604, t test, $p = 0.33$).

**Promoter SNPs Associated with BCRP mRNA Expression.** Samples carrying the −15994T variant allele had significantly higher BCRP mRNA in the PDR4 ($p = 0.003$). White ($p = 0.08$), and Hispanic ($p = 0.05$) livers and intestines ($p = 0.02$) (Fig. 3A). Samples carrying the −15622C>T variant allele showed significant association with low BCRP expression in intestines ($p = 0.036$).
Hispanic livers ($p = 0.04$), and a trend toward significance in the PDR44 ($p = 0.16$) (Fig. 3B). The $15846\text{A}/11002\text{C}$ was associated with high BCRP in all the livers ($p = 0.03$) (Fig. 3C); BCRP mRNA was higher in White livers with the $30477\text{C}/11022\text{G}$ ($p = 0.01$) (Fig. 3D); and White livers with a deletion of nucleotides AAAT at $30639$ had substantially lower BCRP mRNA (Fig. 3E). Neither the $15\text{-kb}$ polymorphic genotype (heterozygotes could not be determined, see under Materials and Methods) nor the intron 1 polymorphic repeat genotype showed any association with BCRP expression in any cohort.

**Intron 1 SNPs Associated with BCRP mRNA Expression.** Samples with the $16702\text{C}/11022\text{T}$ variant allele had significantly higher BCRP RNA levels compared with those with the homozygous wild-type allele ($p = 0.008$ among all the livers and $p = 0.01$ in African American livers) (Fig. 4A). Samples having the variant allele for the $12283\text{T}/11022\text{C}$ SNP showed significantly higher BCRP RNA levels compared with those with the wild-type allele in the PDR44 ($p = 0.02$) and a trend toward significance in White livers ($p = 0.07$) (Fig. 4B). Samples with the BCRP $1143\text{G}/11022\text{A}$ SNP had substantially lower BCRP expression in intestines ($p = 0.06$) and a trend toward lower BCRP mRNA in the PDR44 ($p = 0.1$) and African American livers (Fig. 4C). A T-deletion at position 16823 from the translation start site showed association with significantly lower BCRP RNA levels in all the livers ($p = 0.05$) (Fig. 4D).

**BCRP Genotype versus mRNA Phenotype in the CEPH Panel.** BCRP genotypes for the HapMap CEPH trios were obtained from National Center for Biotechnology Information release 35 and were compared with BCRP mRNA expression data quantitated on Affymetrix Focus Arrays (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1485). No BCRP HapMap genotypes were significantly associated with BCRP mRNA expression in the CEPH samples. We then genotyped the CEPHs for the SNPs that were significantly associated with BCRP expression in our study. The $15622\text{C}/11022\text{T}$ ($p = 0.03$) and $15994\text{C}/11022\text{T}$ ($p = 0.09$) were the only SNPs showing any association with BCRP expression in the CEPHs.

**BCRP Genotypes versus Imatinib Pharmacokinetics.** No statistically significant associations were observed between the BCRP $15846\text{A}/11022\text{C}$, $15623\text{C}/11022\text{T}$, $1143\text{G}/11022\text{A}$, $16702\text{C}/11022\text{T}$, $16823\text{del}$,
12283t>c, and −30639delaaat variants and imatinib pharmacokinetic parameters (data not shown). In 90 patients with complete pharmacokinetic and genotypic information, patients with the BCRP 15994CT or TT genotype had a 38.7% increased CL/F compared with individuals carrying the CC genotype (p = 0.069) (Fig. 5).

**Functional Significance of Novel BCRP SNPs.** The potential functional effect of each SNP was determined by in silico analysis using Transfac to identify whether the SNPs were creating or disrupting any TF binding site (Fig. 6; Table 2). Both of the promoter SNPs associated with higher BCRP expression in multiple tissues resulted in gain of TF binding sites. The 15994C>T variant resulted in the gain of one HNF4 binding site, and the 16702C>T variant resulted in a gain of one GATA4 binding site, a TF important for both intestine and liver development (Watt et al., 2007) and for expression of ABCG5 and ABCG8 (Sumi et al., 2007). The 4-base pair (bp) (AAAT) −30639 promoter deletion associated with low BCRP RNA levels resulted in loss of one HNF1 site. None of the other significant SNPs showed any Transfac change. Although a variety of other SNPs were predicted to alter TF binding sites (Table 2), none of these SNPs was associated with BCRP expression in the tissues examined.

**Differential Exon 1b Usage Is Associated with Decreased BCRP Hepatic mRNA.** Human variation in BCRP expression could also result from the differential use of alternative BCRP promoters and SNPs influencing promoter activity. Indeed, two reports described tissue-specific use of multiple BCRP alternative first exons in mice and in drug-selected cancer cells (Nakanishi et al., 2006; Zong et al., 2006). Review of the expressed sequence tag (EST) and Aceview (http://ncbi.nih.gov/IEB/Research/Acembly) databases and UCSC genome browser May 2004 assembly revealed at least three different alternative first exons in BCRP (Fig. 7A). There were multiple EST clones showing BCRP mRNAs with an alternative first exon 1. Exon 1a represented the known first exon 1 (Bailey-Dell et al., 2001). The alternative exons 1b and 1c were located 369 bases and 72.2 kb upstream, respectively, of the transcription start site.

To screen for the presence of BCRP transcripts utilizing the various exon 1s, we used forward primers in each exon 1 and a reverse primer in exon 2. We also confirmed that BCRP transcripts with alternative exon 1s were full-length BCRP transcripts. Exon 1a and 1b transcripts were found in 85 to 100% of samples in each cohort; conversely, the exon 1c transcript was absent in intestines and expressed in 45 to 48% of PDR and livers. There were no racial or gender differences in usage of any exon 1s. Next we compared the usage of each exon 1 versus BCRP RNA expression in the livers. Liver samples that generated a BCRP transcript using exon 1b had significantly lower BCRP RNA levels (Fig. 7C) among all the livers (p = 0.002) and among Whites (p = 0.01) and Hispanics (p = 0.02). No SNPs within 1 kb of exons

![Fig. 2. BCRP RNA expression in human livers.](http://dmd.aspetjournals.org/)
1b and 1c (Table 2; Fig. 1) were related to differential use of the alternative first exons.

**Polymorphic BCRP SVs.** Polymorphic splicing could also lead to human variation in BCRP expression. We amplified BCRP cDNA from all the tissues using forward primers in exons 1a, 1b, or 1c and a reverse primer in exon 6. An additional low molecular weight band was seen in some of the livers (Fig. 8A) but not in intestines. This SV1 transcript skipped exon 2 and was found in combination with all the exon 1s. The translation start site predictor program http://research.i2r.a-star.edu.sg/DNAFSMiner revealed only two BCRP ATGs (in exons 2 and 3) that produced in-frame transcripts that did not prematurely terminate. The exon 2 ATG had a predicted initiation score of 0.878. The alternative transcripts skipping exon 2 (formed from exon 1a, 1b, or 1c) could use the alternative ATG in exon 3 (predicted initiation scores of 0.589, 0.956, or

---

**Fig. 3.** BCRP promoter SNPs showing significant association with BCRP RNA expression in different tissues. Box plots for individual BCRP promoter SNPs (A–E) showing statistically significant association with BCRP mRNA expression in livers, intestines, and PDR samples. The median ± S.D. values of log2-normalized BCRP relative RNA are plotted on the y-axis as described in Fig. 2 legend. The $p$ values from the Kruskal-Wallis nonparametric test or the Wilcoxon test comparing the significance of three or two genotypes, respectively, is shown for each of the SNPs.
0.657, respectively) (Fig. 8B) that would result in a BCRP variant protein that lacked the first 70 amino acids because the initiating methionine is in exon 2. This variant retains the Walker A and B motifs, and the functional consequence of this transcript is unknown. Three EST clones (DA367705, DB168974, and DA414147) supported exon 2 skipping, and the alternative ATG in exon 3 was conserved in different species.

There was no gender-specific difference in the presence of SV1. Although the median BCRP levels were not significantly different between livers (Fig. 8C), BCRP was low in 80% of SV1/H11001 livers versus 55% of SV1/H11002 livers, and the mean BCRP mRNA was strikingly different between the SV1/H11001 livers (165/223) versus SV1/H11002 livers (544/767) (unequal variance t test, \(p = 0.025\)). The region in and around exon 2 was sequenced to identify whether any SNPs were associated with SV1. The exon 2 G34A was more prevalent in livers with SV1 (26%) versus those without SV1 (3.8%) (\(p = 0.01\)). Because Hispanics have a lower hepatic level of BCRP compared with Whites and African Americans (Fig. 2), we determined whether splicing or the G34A genotype was associated with lower expression. Similar proportions of Hispanics (17/28) and non-Hispanics (15/32) had SV1 (Fisher’s exact test, \(p = 0.22\)). However, BCRP hepatic RNA was significantly lower among Hispanics who
had the G34A variant genotype (unpaired t test, \( p = 0.02 \)), and this may be related, in part, to polymorphic splicing of exon 2.

A second higher molecular weight BCRP SV was identified (Fig. 8D) that inserted 129 bases from intron 7 (exon 7a) between exons 7 and 8 (SV2) but resulted in a premature termination codon. There was no difference in BCRP expression between samples with and without the insertion, and no SNPs in and around this insertion associated with its appearance. Thus, although polymorphic BCRP splicing was seen, it was not correlated with BCRP expression or genotype among all the livers.

BCRP mRNA AEI. Individuals heterozygous for cis-acting BCRP polymorphisms that affect gene expression or mRNA processing would be predicted to show a different level of mRNA expression originating from one allele compared with the other. This is called AEI, which can serve as an integrative quantitative measure of any and all the cis-acting factors (or polymorphic splicing). AEI is measured by determining the number of genomic DNA molecules for each allele compared with the number of allelic mRNA molecules in the target tissue.

We used two frequent coding SNPs residing in the transcribed region of the BCRP gene (exon 2 G34A and exon 5 C421A) to quantitatively analyze the BCRP allelic DNA and mRNA abundance in samples heterozygous for each marker SNP. These SNPs were used to find allelically imbalanced subgroups among each genotype, which could then be linked to cis-acting polymorphisms by assessing genotypes shared between the alleles in the imbalanced subgroups.

Among the 12 PDR G34A heterozygotes, two showed consistent loss of the variant allele in the cDNA (Fig. 9A). In contrast, among the 10 PDR exon 5 C421A heterozygotes and 10 CEPH heterozygotes showing LOH (six PDR and two CEPH), half lost the wild-type and half lost the variant allele. Samples with LOH had statistically lower BCRP mRNA \(( p = 0.017 \)) compared with heterozygotes with equal expression of both alleles. Interestingly, several of the samples were compound heterozygotes for C421A and G34A. Although AEI was expected for both exon 2 and 5 genotypes, there was one sample where the AEI was apparent with the exon 2 but not exon 5 genotype.

Importantly, BCRP mRNA expression was not significantly different between all the PDR 421C homozygous wild-type versus 421CA heterozygous samples. However, the 421CA-imbalanced PDR samples tended to express lower levels of BCRP mRNA compared with heterozygous samples that were not imbalanced \(( p = 0.06 \)) (Fig. 9B). However, when we compared the haplotypes of the BCRP heterozy-
FIG. 7. A, usage of alternative first exons in BCRP. Organization of the human ABCG2 gene and the positions of alternative first exons in BCRP are indicated relative to the exon 1a transcription start site. At least three different first exons with the common translation start site in exon 2 are indicated. B, cartoon of BCRP 5'-untranslated region variants identified by Nakanishi et al. (2006) for comparison. C, alternative exon 1b usage is associated with lower BCRP expression in human livers. Relative BCRP RNA expression in livers with presence or absence of exon 1b was compared. The y-axis indicates the log-transformed relative BCRP RNA levels. Statistical significance was calculated using the Wilcoxon test.
FIG. 8. BCRP SVs. A, left shows the agarose gel of reverse transcription-PCR products generated using forward primers in exon 1a, 1b, or 1c and reverse primer in exon 6. Arrows indicate exon 2–skipped transcripts (SV1). Right shows potential translation start sites used by the exon 2–skipped transcripts. B, amino acid sequence alignment for BCRP wild-type and SV1 lacking the 70 amino terminal amino acids. C, relative BCRP RNA levels in livers arranged by liver with lowest to highest BCRP mRNA in samples with SV1 (left) and without SV1 (right) and the corresponding genotype of exon 2 G34A and intron 2 (19146a/g) in each liver (white, homozygous wild-type; light gray, heterozygous; black, homozygous variant allele). D, agarose gel of reverse transcription-PCR products generated using forward primer in exon 7 and reverse primer in exon 8 showing higher molecular weight alternative SV2 mRNA (left) and schematic representation of inserted nucleotides from intron 7 (exon 7a).

POONKUZHALI ET AL.

Downloaded from dmd.aspetjournals.org at ASPET Journals on May 2, 2016
Analysis for BCRP DNA Copy Number Variation. A recent screen of the CEPH, Yourba, Japanese, and Chinese HapMap samples for copy number variation identified 2 of 60 CEPH samples (NA10863 and NA12234) with one copy number gain for BCRP (http://projects.tcag.ca/variation/) (Redon et al., 2006). We screened DNA from liver and PDR samples representing the highest, lowest, and median BCRP expressors (n = 5 in each range/tissue) but detected no copy number variation that might have explained variable BCRP expression.

Discussion

SNPs in regulatory regions represent an important but relatively unexplored class of genetic variation. We resequenced potential regulatory regions in the 5′ region and intron 1 of BCRP and identified 90 SNPs. Several SNPs in the 5′ region and three in intron 1 showed significant association with BCRP RNA expression in the three tissues we analyzed (livers, intestines, and Epstein-Barr virus-immortalized lymphoblasts), as well as imatinib clearance in vivo. Although evolutionary conservation analysis is one strategy used to identify intragenic regions likely to have functional importance, most of the SNPs were not in ECRs, but all the SNPs resided in regions identified by screening for clusters of TF binding sites. This result is of interest because it was recently shown that binding sites for highly conserved TFs varied significantly across species, and after aligning the promoters of orthologous genes, about two thirds of the binding sites did not align (Odom et al., 2007). This suggests that targeting genomic regions with clusters of TFs is a useful alternative approach in identifying promoter regions to resequence.

In the present study, roughly one third of the livers showed high BCRP RNA expression. This was similar to previous reports in which BCRP RNA was more highly expressed in 30% of acute leukemia patients (Suvannasankha et al., 2004). Several SNPs were associated with higher BCRP hepatic expression. Among Whites and Hispanics, the 15994C>T promoter SNP was associated with higher BCRP RNA expression. This SNP was predicted by in silico analysis to result in the gain of one HNF4 site. Although a role for HNF4 in BCRP regulation has not been shown to date, this TF is highly coexpressed with BCRP in several tissues (e.g., liver and intestine). Other SNPs in partial LD to 15994C>T promoter SNP could also contribute to the BCRP phenotype. The intronic 12283T>C SNPs were also associated with higher hepatic BCRP expression. Among Whites and Hispanics, the −15994C>T promoter SNP was associated with higher BCRP RNA expression. This SNP was predicted by in silico analysis to result in the gain of one HNF4 site. Although a role for HNF4 in BCRP regulation has not been shown to date, this TF is highly coexpressed with BCRP in several tissues (e.g., liver and intestine). Other SNPs in partial LD to −15994 (e.g., −15846) could also contribute to the BCRP phenotype. The intronic 12283T>C SNPs were also associated with higher hepatic BCRP in livers from White donors, whereas BCRP RNA was higher in African American livers with the intronic 16702C>T SNP, which potentially leads to a gain of a GATA4 site. Conversely, SNPs associated with lower BCRP mRNA included the −15622C>T, −30477C>G, and the AAAT deletion at −30639, as well as intron 1 SNPs 122893 T>C and a T deletion at −16823.

Notably, although several cis-regulatory sites have been identified in the BCRP promoter, we failed to find SNPs in any of them (e.g., a 150-bp conserved enhancer region, containing three functional PPAR

![Fig. 9. Allelic imbalance in BCRP. A, relative BCRP mRNA arranged from lowest to highest expression in PDR44 cells. The corresponding exon 2 and exon 5 BCRP genotypes sequenced in genomic DNA (g) versus cDNA (c) are shown below the graph (white, homozygous wild-type; light gray, heterozygous; black, homozygous variant allele). Boxed genotypes indicate samples with discordant genotypes in the cDNA versus DNA. B, left, box plots of BCRP mRNA in samples with the exon 5 wild-type or heterozygous genotypes. Right, box plots of BCRP mRNA in exon 5 wild-type versus heterozygous genotypes ± exon 5 allelic imbalance. The p value was generated by the Kruskal-Wallis test.](http://projects.tcag.ca/variation/)
response elements, a functional estrogen response element, and Hif1 binding site) (Ee et al., 2004; Krishnamurthy et al., 2004; Szatmari et al., 2006). In addition, although the ~790-bp BCRP promoter CTCGA deletion was shown to be associated with the relative extent of irinotecan conversion to SN38 (Zhou et al., 2005), in our study this SNP did not show any association with BCRP RNA expression. Nevertheless, it will be important to validate the potential clinical importance of BCRP SNPs in vivo.

In a previous study of human intestines, we failed to find any alternatively spliced BCRP mRNAs. In contrast, more than half of the livers examined polymorphically expressed an alternative BCRP mRNA (SV1) that skipped exon 2. Although the exon 2 G34A was not present in all the SV1 livers, the SV1 + livers were significantly more likely to carry the exon 2 G34A and to have a lower (3.3 times) mean level of BCRP mRNA compared with SV1 livers. Moreover, because 95% of persons with G34A had SV1, the G34A genotype could be used to identify a subset of SV1 livers and whether SV1 had a functional consequence. Indeed, the G34A is more frequent in Hispanics, who have a lower level of BCRP, suggesting the G34A could contribute to this phenotype. Mechanistically, polymorphic BCRP alternative splicing may be diminishing the pool of wild type, and hence functional BCRP mRNA. Moreover, association of the G34A genotype with lower BCRP expression may be liver-specific because we did not detect SV1 in intestines, and alternative splicing can be tissue-specific. Whether the G34A change is associated with lower BCRP mRNA in other tissues that express BCRP, such as brain or placenta, remains to be determined. Finally, the fact that 95% of G34A livers were SV1 + has implications for designing in vitro studies to determine the functional consequence of BCRP variant alleles. Several groups have expressed the 34G and 34A (V12 and binding site) (Ee et al., 2004; Krishnamurthy et al., 2004; Szatmari et al., 2006). AEI then serves as the phenotype that can be linked to the cis-SNPs (including intronic SNPs) that were uniquely linked with this event. Likewise, others (Kobayashi et al., 2005) have reported BCRP AEI in human placenta but failed to find linked SNP. However, neither our study nor others have totally resequenced the introns, so it is possible that there is an unidentified intron SNP linked to this event. It must also be considered that AEI could result from epigenetic changes because it was recently reported that the BCRP promoter could be methylated (To et al., 2006). Likewise, AEI could be influenced by other processes known to regulate BCRP expression such as kinases (Meyer zu Schwabedissen et al., 2006).

It was recently shown that BCRP 421A coding variant has the signature of recent positive selection in the Asian population (Wang et al., 2007). It is equally reasonable to consider that there could be evolutionary constraints on the promoter. Intriguingly, one SNP associated with BCRP expression showed a large frequency difference between populations (16702), suggesting it might be under recent evolutionary constraints on the promoter. Intriguingly, one SNP associated with BCRP expression showed a large frequency difference between populations (16702), suggesting it might be under recent evolutionary constraints on the promoter.


Address correspondence to: Erin Schuetz, St. Jude Children’s Research Hospital, 332 North Lauderdale Street, Memphis, TN 38105. E-mail: erin.schuetz@stjude.org zdd004084361XX