DMD Fast Forward. Published on January 7, 2008 as doi:10.1124/dmd.107.018366

DMD # 18366

Title Page

Association of BCRP/ABCG2 Phenotypes and Novel promoter and intron 1 SNPs

Balasubramanian Poonkuzhali, Jatinder Lamba, Stephen Strom, Alex Sparreboom,

Kenneth Thummel, Paul Watkins, and Erin Schuetz

Department of Pharmaceutical Sciences (B.P, J.L, A.S., E.S.), St. Jude Children's Research Hospital, Memphis, TN, the Department of Pathology (S.S.), University of Pittsburgh, Pittsburgh, PA, Department of Pharmaceutics (K.T.), Seattle, WA, Department of Medicine, Division of Hepatology (P.W.), University of North Carolina

Running Title Page.

a) Sequence Diversity in the BCRP/ABCG2 Promoter

b) Corresponding Author: Dr. Erin Schuetz, St Jude Children's Research Hospital, 332 N Lauderdale S

Memphis, TN-38105, USA; PH(901)-495-2205;Email: Erin.schuetz@stjude.org

c)

Text Pages:

Tables: 2

Figures: 9

References: 40

Supplementary figure: 1

Words in Abstract: 241

Words in Introduction: 564

Words in results/discussion: 3,471

d) Abbreviations: TF, transcription factor; BCRP (ABCG2, MXR, ABCP), SNP, single nucleotide polymorphism

Abstract.

The hypothesis was tested that sequence diversity in 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 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 16702C>T was associated with high expression in livers; 1143G>A was associated with low expression in intestine; 12283T>C was associated with higher 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 oral imatinib. We next determined if BCRP expression was related to polymorphic alternative splicing or alternative promoter use. Liver polymorphically expressed an alternatively spliced mRNA (SV1) skipping exon 2. Although SV1+ livers did not uniformly carry the ex2 G34A allele, 90% of G34A livers expressed SV1 (vs. 4% of 34GG livers) and BCRP mRNA was significantly lower among Hispanic livers with the G34A variant genotype. 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 utilized multiple promoters, and livers differentially using alternative exon 1b had lower BCRP. Conclusion: BCRP expression in lymphoblasts, liver and intestine is associated with novel promoter and intron 1 SNPs.

Introduction.

The efflux transporter 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, topotecan, imatinib and methotrexate. BCRP also transports dietary carcinogens and endogenous substrates such as protoporphyrin IX and vitamin B2 (Jonker et al., 2007; van Herwaarden et al., 2006), 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 et al., 2005).

BCRP shows significant interindividual variation in expression (Zamber et al.,2003, Ross et al., 2000). BCRP overexpression has been described in drug resistant ovary, breast, colon and gastric cancer, fibrosarcoma cell lines, placental tissue, liver canalicular membranes, ducts and lobules of the breast, endothelium of veins and capillaries, epithelium of colon and small intestine and bile canaliculi (Maliepaard et al., 2001). Increased expression of BCRP has been associated with poor treatment outcome (increased risk of relapse, decreased disease free survival) in various leukemias, although this observation is controversial (Damiani et al.,2006; Ross et al.,2000).

Phenotypes associated with decreased, absent or drug inhibited BCRP can be predicted from studies in BCRP -/- mice. Mice lacking 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 due to variant alleles might be expected to show altered drug disposition, efficacy and toxicities including porphyrias and phototoxicity.

We previously identified several naturally occurring BCRP variants (Zamber et al.,2003). Among the coding SNPs, G34A (V12M) in exon2 and C421A (Q141K) in exon5 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 (deJong et al.,2004; Sparreboom et al.,2004), while 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, DHEAS, PAH (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 intron1 and could further explain variation in BCRP expression. Moreover, since 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.

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 USA (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 (NIH Contract #N01-DK-9-2310) and by the Cooperative Human Tissue Network. DNA from a set of 60 liver samples from three different racial groups (Whites [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 PDR44. 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) were processed as described previously (Mouly et al.,2005), and were genotyped for the common SNPs identified in the PDR44.

Patient DNA: Samples were obtained at steady-state from adult patients with c-kit positive gastrointestinal stromal tumors. Pharmacokinetic data has 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 ABCG2 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 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 (Guetens et al.,2003). Pharmacokinetic parameters, including area under curve (AUC), steady-state concentration (Css), and apparent oral clearance (C:/F) were obtained for each patient by non-compartmental 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 non-parametric Mann-Whitney U test (2-group comparison) or a Kruskal-Wallis test (multiple-group comparison). $P < 0.05$ was considered as statically significant.

In silico analysis of the BCRP/ABCG2 gene to identify potential regulatory regions to resequence: Several web based bioinformatic tools were used to screen 49 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 (LETFs), and for regions of high evolutionarily conservation between multiple species. Cister plot (http://zlab.bu.edu/~mfrith/cister.shtml), NUBISCAN (http://www.nubiscan.unibas.ch/) and Transfac (http://transfac.gbf.de/TRANSFAC/lists/matrix/matrixByName.html) were used to identify regions harboring DNA response elements for various transcription factors. [TF matrices included HNF1, HNF3, GATA, AP1, CDX, FOX, SP1, NF, GR, TATA, CAAT,YY1 DR3 And DR4, PXR, DR3, DR1, HNF4_DR1, HNF1, HNF3alpha, HNF4, HNF3beta, CEBPalpha, CEBP, CEBPdelta, HNF-6, CEBPbeta, GATA4, GR (glucocorticoid receptor), HNF3gamma, NF-1, NF-kappB, SP1 and TATA, CRE, ERE, NF-1, E2F, Mef-2, Myf, CCAAT, AP-1, Ets, Myc, GATA, LSF, SRF, and Tef.] The UCSC genome browser (http://genome.ucsc.edu), ECR (evolutionary conserved region (http://ecrbrowser.dcode.org/) and rVISTA (http://genome.lbl.gov/vista/rvista/submit.shtml) were used to identify regions of evolutionary conservation on the BCRP gene between humans and other species.

DNA sequencing of BCRP PCR 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 x 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). The PCR conditions include initial denaturation at 95^{o}C for 3 min followed by 32-34 cycles of

denaturation at 95^{o}C, annealing at appropriate temperatures, and synthesis at 72^{o}C, with final synthesis at 72^{0}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 (USB) and Exonuclease I (USB) 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 using the PCR primers or internal sequencing primers (Table 1). Sequences were assembled using the Phred-Phrap-Consed package (University of Washington, Seattle, http://droog.mbt.washington.edu/PolyPhred.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 (-3222- and -32252) with polymorphic repeats in some PDR44 DNAs and 12/60 liver samples. (2) A region approximately -13kb from the transcription start site (designated: ?insertion, 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.

RNA extraction and real time RT-PCR: RNA was isolated from sample sets I-III using Trizol® reagent (Life Technologies). The integrity of the isolated RNA was examined by quantitating the $A_{260/280}$ ratio and resolving the RNA on a 1% agarose gel. Total RNA (3-5 μg) was reverse-transcribed according to the manufacturer's instructions (Life Technologies). 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, Valencia, CA). Amplification

was done with the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Since we observed an additional dissociation curve (amplification of an additional alternative mRNA) in some of the liver samples with these real time primers (realtime F1/r), we redesigned the forward primer (realtime f2) (Table 1) to amplify only the wt mRNA for quantitation. (The rational for redesigning the primer was as follows: the additional dissociation curve we observed in some samples was due to insertion of part of intron7; to avoid amplifying the splice variant, 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 vs. GAPDH as a reference transcript (Zamber et al.,2003).

Detection of alternative first exons and splice variants 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 splice variants shown in the UCSC genome browser by designing primers in the flanking exons (Table 1). In order to check if these alternative first exons result in a full length cDNA, reverse primers in exon16 and a nested reverse primer in exon10 were used (Table 1).

Analysis of allelic imbalance: We used two different assays to determine 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 cDNA) were identified as having allelic imbalance. The results were confirmed using an exon 5 allele-specific quantitation kit from ABI to quantitate BCRP signal intensity from each allele in DNA vs. RNA. (Applied Biosystems, Foster City, CA- Assays on demand) and the ratio of signal intensity for the C vs. 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 (Whites n=15; Blacks 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 exon2 or exon5 of BCRP using the SYBR Green PCR kit (Qiagen, Chatsworth, California, USA) according to the manufacturers 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 delta delta Ct method and normalized relative to CYP3A4 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 intron1. Since the human BCRP promoter (105 kb) and intron 1 (19kb) 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 transcription factor binding sites. Fig 1A and 1B show snapshots from the UCSC and ECR (evolutionary conserved regions) genome browsers with global alignment of mammalian, amphibian, bird, dog, mouse, rat, chic, fugu and zebrafish. There is no synteny between the BCRP neighboring genes and the closest 5' neighbor is 160 kb and 105 kb in mice and humans, respectively. A conserved region located approximately -5kb from the transcription start site (-15782/-15923 in RefSeq AC084732) was identified as encoding ribosomal protein L31 (RPL31 mRNA). Because these two programs use global alignment to detect conserved regions, we confirmed this result using rVISTA, a program that uses gene-to-gene alignment. The ribosomal protein L31 was confirmed in the promoter of human and mouse BCRP by rVISTA analysis.

The same regions of BCRP were next screened for clusters of TF (transcription factor) 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 utilized 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 transcription factor) consistent with BCRP expression in liver and intestine; PPARgamma and C/EBPalpha and delta, consistent with the role of the ABCG proteins in lipid and sterol homeostasis; C/EBP and Cdx family TFs, important for regulating gene expression in the small

intestine and for the reported role of Cdx2 in regulating iron homeostasis since BCRP transports the heme precursor protoporphyrin IX; NKX25/CSX, a homeodomain factor important in development and found in many tissues that could be important for BCRP expression in embryonic stem cells; SRFQ2-5 , serum response transcription factors important for early growth response.

BCRP processed pseudogene in intron 1 of the NOX5 gene. Although BCRP is on chromosome 4q22, bioinformatic analysis revealed an intronless BCRP 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 since 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 that 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 Caucasians (CEPHS), indicated that the BCRP gene is framed by 7 LD blocks, with three LD blocks in the promoter region (Supplementary Figure 1). However, the degree of LD within each block was low with 6-8 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 RNA 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 to White (101.3+2534) and African American (291+6877) livers (p=.09) (note inset graph is Log2). The *mean* level of BCRP mRNA was higher in female (124.6+71) vs. male (95+78) intestines (not shown) similar to what we previously reported (Zamber C et al., 2003). However, BCRP mRNA was not different in female vs. male livers (105.43+9329 vs. 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 PDR44 (p=0.003), White (p=0.08) and Hispanic (p=0.05) livers as well as 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 towards significance in the PDR44 (p=0.16) (Fig 3B). The -15846A>C was associated with high BCRP in all livers (p=0.03) (Fig 3C); BCRP mRNA was higher in White livers with the -30477C>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 -13kb polymorphic genotype (heterozygotes could not be determined, see methods) nor the intron 1 polymorphic repeat genotype showed any association with BCRP expression in any cohort.

Intron1 SNPs associated with BCRP mRNA expression. Samples with the 16702C>T variant allele had significantly higher BCRP RNA levels compared to 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 12283T>C SNP showed significantly higher BCRP RNA levels compared to those with the wild type allele in the PDR44 (p=0.02) and a trend to significance in White livers (p=0.07) (Fig 4B). Samples with the BCRP 1143G>A SNP had substantially lower BCRP expression in intestines (p=0.06) and a trend towards 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 livers (p=0.05) (Fig 4D).

BCRP genotype vs. mRNA phenotype in the CEPH panel. BCRP genotypes for the HapMap CEPH trios was obtained from NCBI release 35 and was 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 -15622C>T (p=0.03) and -15994 C>T (p=0.09) were the only SNPs showing any association with BCRP expression in the CEPHs.

BCRP genotypes vs imatinib pharmacokinetics. No statistically significant associations were observed between the *ABCG2* -15846a>c, -15623c>t, 1143g>a, 16702c>t, 16823 delt, 12283t>c, and -30639delaaat variants and imatinib pharmacokinetic parameters (data not shown). In 90 patients with complete pharmacokinetic and genotypic information, patients with the *ABCG2* -

15994CT or TT genotype had a 38.7% increased CL/F compared with individuals carrying the CC genotype (P = 0.069) (Figure 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 transcription factor binding site (Figure 6, Table 2). Both of the promoter SNPs associated with higher BCRP expression in multiple tissues resulted in gain of transcription factor 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 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 transcription factor 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 (Zong et al.,2006 & Nakanishi et al.,2006). Review of the EST and Aceview (http://ncbi.nih.gov/IEB/Research/Acembly) databases and UCSC genome browser May2004 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 exon1 (Bailey-Dell et al.,2001). The alternative Exons 1b and 1c were located 369 bases and 72.2kb upstream, respectively, of the transcription start site.

To screen for the presence of BCRP transcripts utilizing the various exon1s we used forward primers in each exon 1 and a reverse primer in exon 2. We also confirmed that BCRP transcripts with alternative exon1s were full length BCRP transcripts. Exon 1a and exon1b transcripts were found in 85-100% of samples in each cohort; conversely the Exon 1c transcript was absent in intestines, and expressed in 45-48% of PDR and livers. There were no racial or gender differences in usage of any exon1s. Next we compared the usage of each exon 1 vs. BCRP RNA expression in the livers. Liver samples that generated a BCRP transcript utilizing exon1b had significantly lower BCRP RNA levels (Fig 7b) among all livers ($p=0.002$) and among Whites ($p=0.01$) and Hispanics ($p=0.02$). No SNPs within 1kb of Ex1b and Ex1c (Table 2, Fig 1) were related to differential use of the alternative first exons.

Polymorphic BCRP splice variants. Polymorphic splicing could also lead to human variation in BCRP expression. We amplified BCRP cDNA from all tissues utilizing 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 splice variant (SV1) transcript skipped exon 2 and was found in combination with all exon1s. The translation start site predictor program **http://research.i2r.a-star.edu.sg/DNAFSMiner/** revealed only two BCRP ATGs (in exon2 and exon 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 exon1a, 1b or 1c) could use the alternative ATG in exon3 (predicted initiation

scores of 0.589, 0.956 or 0.657, respectively) (Fig 8B) that would result in a BCRP variant protein that lacked the first 70 amino acids since 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, DA414147) supported exon2 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 +/- SV1 (Fig 8C), BCRP was low in 80% of SV1+ livers vs. 55% of SV1- livers, and the mean BCRP mRNA was strikingly different between the + SV1 livers (165 +/- 223) vs. –SV1 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 exon2 G34A was more prevalent in livers with SV1 (26%) vs. those without SV1 (3.8%) (p=0.01). Since Hispanics have a lower hepatic level of BCRP compared to Whites and Blacks (Fig 2) we determined whether splicing or the G34A genotype were associated with lower expression. Similar proportions of Hispanics (17/28) and non-Hispanics (15/32) had SV1 (Fishers 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 splice variant was identified (Fig 8D) that inserted 129 bases from intron7 (exon 7a) between exon7 and 8 (SV.2) 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 livers.

BCRP mRNA allelic expression imbalance (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 allelic expression imbalance (AEI), which can serve as an integrative quantitative measure of any and all *cis*-acting factors (or polymorphic splicing). AEI is measured by determining the number of genomic DNA molecules for each allele in comparison to 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 wt 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. It is possible that since exon 2 is subject to alternative splicing, that the exon 2 SNP was diminished due to exon 2 skipping in this transcript.

Importantly, BCRP mRNA expression was not significantly different between all PDR 421C homozygous wild-type vs. 421CA heterozygous samples. However, the 421CA imbalanced PDR samples tended to express lower levels of BCRP mRNA compared to heterozygous samples that were not imbalanced (p=0.06) (Fig 9B). However, when we compared the haplotypes of the BCRP heterozygous balanced vs. imbalanced samples, no cis-acting SNPs was significantly associated with BCRP allelic imbalance.

Analysis for BCRP DNA copy number variation. A recent screen of the CEPH, Yourba, Japanese and Chinese HapMap samples for CNV identified 2/60 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 CNV 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 intron1 showed significant association with BCRP RNA expression in the three tissues we analyzed (livers, intestines and EBV 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 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 transcription factors varied significantly across species, and after aligning the promoters of orthologous genes about two-third of the binding sites did not align (Odom et al 2007). This suggests that targeting genomic regions with clusters of transcription factors 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 demonstrated to date, this transcription factor is highly co-expressed with BCRP in several tissues (e.g., liver and intestine). Other SNPs in partial LD to -15994 (e.g., -15846) could also be

contributing to the BCRP phenotype. The intronic 12283T>C SNPs was also associated with higher hepatic BCRP in livers from White donors while 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, and intron one 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 response elements (PPARE); a functional estrogen response element and Hif1 binding site (Szatmari et al.,2006; Ee et al., 2004; Krishnamurthy et al.,2004). In addition, although the -790 bp BCRP promoter CTCA 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, over 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 of the SV1 livers, the SV1+ livers were significantly more likely to carry the exon 2 G34A and to have a lower (3.3x) mean level of BCRP mRNA compared to SV1 livers. Moreover, since 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 wt, 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 M12) cDNAs in cell lines and SF9 cell membranes (Kondo et al.,2004). However, the cDNA is already spliced it cannot assess how alternative splicing of this allele may affect the pool of wt BCRP mRNA.

SNPs that affect gene expression level in an allele specific manner are often located in the gene regulatory regions such as promoters, introns, and 5' and 3'-UTRs (Ponomarenko et al.,2002). AEI then serves as the phenotype that can be linked to the functional *cis*-acting polymorphisms by genotype scanning along the entire gene locus (Wang D and Sadee W, 2006). We detected BCRP AEI in the PDR44, but were unable to identify 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 there is an unidentified intron SNP linked to this event. It must also be considered that AEI could be due to epigenetic changes since it was recently reported that the BCRP promoter could be methylated (To et al.,2006). Likewise, AIE 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 positive selective pressure in the African American population, but this remains to be tested.

In total, the results from the present study indicate that there exists substantial genetic variability in potential regulatory regions of BCRP, that some SNPs are associated with altered BCRP expression, and that a number of these SNPs reside in or create or destroy putative transcription factor binding sites. The fact that the identical SNPs were associated with altered BCRP expression in multiple tissue types and imatinib clearance in vivo strongly suggests that they may be functionally important and need to be tested further for their relationship to BCRP mediated drug clearance in well controlled studies.

Acknowledgements.

We thank the Hartwell center at St Jude Children's Research Hospital for the DNA sequencing and oligo synthesis.

References.

Bailey-Dell KJ, Hassel B, Doyle LA, Ross DD (2001) Promoter characterization and genomic organization of the human breast cancer resistance protein (ATP-binding cassette transporter G2) gene. *Biochim Biophys Acta* 1520:234-241.

Burger H, van Tol H, Boersma AW, Brok M, Wiemer EA, Stoter G, Nooter K (2004) Imatinib mesylate (STI571) is a substrate for the breast cancer resistance protein (BCRP)/ABCG2 drug pump. *Blood* 104:2940-2942.

Damiani D, Tiribelli M, Calistri E, Geromin A, Chiarvesio A, Michelutti A, Cavallin M, Fanin R (2006) The prognostic value of P-glycoprotein (ABCB) and breast cancer resistance protein (ABCG2) in adults with de novo acute myeloid leukemia with normal karyotype. *Haematologica* 91:825-828.

de Jong FA, Marsh S, Mathijssen RH, King C, Verweij J, Sparreboom A, McLeod HL (2004) ABCG2 pharmacogenetics: ethnic differences in allele frequency and assessment of influence on irinotecan disposition. *Clin Cancer Res* 10:5889-5894.

Ee PL, Kamalakaran S, Tonetti D, He X, Ross DD, Beck WT (2004) Identification of a novel estrogen response element in the breast cancer resistance protein (ABCG2) gene. *Cancer Res* 64:1247-1251.

Gardner ER, Burger H, van Schaik RH, van OOsterom AT, de Bruijn EA, Guetens G, Prenen H, de Jong, FA, Baker SD, Bates SE, Figg WD, Verweij J, Sparreboom A, Nooter K. (2006) Association of enzyme and transporter genotypes with the pharmacokinetics of imatinib. *Clin Pharmacol Ther* 80:192-201.

Guetens G, De Boeck G, Highley M, Dumez H, Van Oosterom AT, de Bruijn EA. (2003) Quantification of the anticancer agent STI-571 in erythrocytes and plasma by measurement of sediment technology and liquid chromatography-tandem mass spectrometry *J Chromatogr* A 1020:27-34.

Ho AY, Deacon A, Osborne G, Mufti GJ (2003) Precipitation of porphyria cutanea tarda by imatinib mesylate? *Br J Haematol* 121:375-.

Imai Y, Nakane M, Kage K, Tsukahara S, Ishikawa E, Tsuruo T, Miki Y, Sugimoto Y (2002) C421A polymorphism in the human breast cancer resistance protein gene is associated with low expression of Q141K protein and low-level drug resistance. *Mol Cancer Ther* 1:611-616.

Jonker JW, Musters S, Vlaming ML, Plosch T, Gooijert KE, Hillebrand MJ, Rosing H, Beijnen JH, Verkade HJ, Schinkel AH (2007) Breast cancer resistance protein (Bcrp1/Abcg2) is expressed in the harderian gland and mediates transport of conjugated protoporphyrin IX. *Am J Physiol Cell Physiol* 292:C2204-C2212.

Kondo C, Suzuki H, Itoda M, Ozawa S, Sawada J, Kobayashi D, Ieiri I, Mine K, Ohtsubo K, Sugiyama Y (2004) Functional analysis of SNPs variants of BCRP/ABCG2. *Pharm Res*. 21:1895-1903.

Kobayashi D, Ieiri I, Hirota T, Takane H, Maegawa S, Kigawa J, Suzuki H, Nanba E, Oshimura M, Terakawa N, Otsubo K, Mine K, Sugiyama Y (2005) Functional assessment of ABCG2 (BCRP) gene polymorphisms to protein expression in human placenta. *Drug Metab Dispos* 33:94-101.

Krishnamurthy P, Ross DD, Nakanishi T, Bailey-Dell K, Zhou S, Mercer KE, Sarkadi B, Sorrentino BP, Schuetz JD (2004) The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme. *J Biol Chem* 279:24218-24225.

Lamba JK, Chen X, Lan LB, Kim JW, Wei Wang X, Relling MV, Kazuto Y, Watkins PB, Strom S, Sun D, Schuetz JD, Schuetz EG (2006) Increased CYP3A4 copy number in TONG/HCC cells but not in DNA from other humans. *Pharmacogenet Genomics* 16:415-427

Maliepaard M, Scheffer GL, Faneyte IF, van Gastelen MA, Pijnenborg AC, Schinkel AH, van De Vijver MJ, Scheper RJ, Schellens JH (2001) Subcellular localization and distribution of the breast cancer resistance protein transporter in normal human tissues, *Cancer Res* 61: 3458–3464.

Mathijssen RH, Marsh S, Karlsson MO, Xie R, Baker SD, Verweij J, Sparreboom A, McLeod HL (2003) Irinotecan pathway genotype analysis to predict pharmacokinetics. *Clin Cancer Res*. 9:3246-3253.

Meyer zu Schwabedissen HE, Grube M, Dreisbach A, Jedlitschky G, Meissner K, Linnemann K, Fusch C, Ritter CA, Völker U, Kroemer HK. (2006) Epidermal growth factor-mediated activation of the map kinase cascade results in altered expression and function of ABCG2 (BCRP). *Drug Metab Dispos.* 34:524-33.

Mouly SJ, Matheny C, Paine MF, Smith G, Lamba J, Lamba V, Pusek SN, Schuetz EG, Stewart PW, Watkins PB (2005) Variation in oral clearance of saquinavir is predicted by CYP3A5*1 genotype but not by enterocyte content of cytochrome P450 3A5. *Clin Pharmacol Ther* 78:605-618.

Nakanishi T, Bailey-Dell KJ, Hassel BA, Shiozawa K, Sullivan DM, Turner J, Ross DD (2006) Novel 5' untranslated region variants of BCRP mRNA are differentially expressed in drug-selected cancer cells and in normal human tissues: implications for drug resistance, tissue-specific expression, and alternative promoter usage. *Cancer Res* 66:5007-5011.

Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, Macisaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 39:730-732.

Ozvegy-Laczka C, Hegedus T, Várady G, Ujhelly O, Schuetz JD, Váradi A, Kéri G, Orfi L, Német K, Sarkadi B (2004) High-affinity interaction of tyrosine kinase inhibitors with the

ABCG2 multidrug transporter. *Mol Pharmacol*. ;65:1485-1495.

Ponomarenko JV, Orlova GV, Merkulova TI, Gorshkova EV, Fokin ON, Vasiliev GV, Frolov AS, Ponomarenko MP (2002) rSNP_Guide: an integrated database-tools system for studying SNPs and site-directed mutations in transcription factor binding sites. *Hum Mutat* 20:239-248. Review.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. *Nature* 444:444-454.

Ross DD, Karp JE, Chen TT, Doyle LA (2000) Expression of breast cancer resistance protein in blast cells from patients with acute leukemia. *Blood* 96:365-368.

Smalley MJ, Clarke RB (2005) The mammary gland "side population": a putative stem/progenitor cell marker? *J Mammary Gland Biol Neoplasia* 10:37-47. Review.

Sparreboom A, Gelderblom H, Marsh S, Ahluwalia R, Obach R, Principe P, Twelves C, Verweij J, McLeod HL (2004) Diflomotecan pharmacokinetics in relation to ABCG2 421C>A genotype. *Clin Pharmacol Ther* 76:38-44.

Sumi K, Tanaka T, Uchida A, Magoori K, Urashima Y, Ohashi R, Ohguchi H, Okamura M, Kudo H, Daigo K, Maejima T, Kojima N, Sakakibara I, Jiang S, Hasegawa G, Kim I, Osborne TF, Naito M, Gonzalez FJ, Hamakubo T, Kodama T, Sakai J (2007) Cooperative interaction between hepatocyte nuclear factor 4 alpha and GATA transcription factors regulates ATP-binding cassette sterol transporters ABCG5 and ABCG8. *Mol Cell Biol*. 27:4248-60.

Suvannasankha A, Minderman H, O'Loughlin KL, Nakanishi T, Ford LA, Greco WR, Wetzler M, Ross DD, Baer MR (2004) Breast cancer resistance protein (BCRP/MXR/ABCG2) in adult acute lymphoblastic leukaemia: frequent expression and possible correlation with shorter disease-free survival. *Br J Haematol* 127:392-398.

Szatmari I, Vamosi G, Brazda P, Balint BL, Benko S, Szeles L, Jeney V, Ozvegy-Laczka C, Szanto A, Barta E, Balla J, Sarkadi B, Nagy L (2006) PPARgamma regulated ABCG2 expression confers cytoprotection to human dendritic cells. *J Biol Chem* 281:23812-23823

Tamura A, Watanabe M, Saito H, Nakagawa H, Kamachi T, Okura I, Ishikawa T (2006) Functional validation of the genetic polymorphisms of human ATP-binding cassette (ABC)

transporter ABCG2: identification of alleles that are defective in porphyrin transport. *Mol Pharmacol* 70:287-296.

To KK, Zhan Z, Bates SE (2006) Aberrant promoter methylation of the ABCG2 gene in renal carcinoma. *Mol Cell Biol* 26:8572-8585.

van Herwaarden AE, Wagenaar E, Merino G, Jonker JW, Rosing H, Beijnen JH, Schinkel AH (2007) Multidrug transporter ABCG2/breast cancer resistance protein secretes riboflavin (vitamin B2) into milk. *Mol Cell Biol* 27:1247-1253.

Wang Z, Wang J, Tantoso E, Wang B, Tai AY, Ooi LL, Chong SS, Lee CG (2007) Signatures of recent positive selection at the ATP-binding cassette drug transporter superfamily gene loci. *Hum Mol Genet* 16:1367-1380.

Wang D, Sadee W (2006) Searching for polymorphisms that affect gene expression and mRNA processing: example ABCB1 (MDR1). *AAPS J* 18:E515-20

Watt AJ, Zhao R, Li J, Duncan SA (2007) Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev Biol* 7:37.

Zamber CP, Lamba JK, Yasuda K, Farnum J, Thummel K, Schuetz JD, Schuetz EG (2003) Natural allelic variants of breast cancer resistance protein (BCRP) and their relationship to BCRP expression in human intestine. *Pharmacogenetics* 13:19-28.

Zhou Q, Sparreboom A, Tan EH, Cheung YB, Lee A, Poon D, Lee EJ, Chowbay B (2005) Pharmacogenetic profiling across the irinotecan pathway in Asian patients with cancer.*Br J Clin Pharmacol* 59:415-424.

Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, Lagutina I, Grosveld GC, Osawa M, Nakauchi H, Sorrentino BP (2001). The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med* 7:1028-1034

Zong Y, Zhou S, Fatima S, Sorrentino BP (2006) Expression of mouse Abcg2 mRNA during hematopoiesis is regulated by alternative use of multiple leader exons and promoters. *J Biol Chem.*281:29625-29632

Footnotes.

This work is supported in part by the NIH/NIGMS Pharmacogenetics Research Network and Databas (U01GM61374, http://pharmgkb.org) under grant U01 GM61393, and the NIH P30 CA21765 Cance Center Support grant and by the American Lebanese Syrian Associated Charities (ALSAC).

Legends for Figures.

Figure 1: **In silico analysis of the BCRP promoter and intron 1.** (A) Snapshot from the UCSC genome browser displaying the BCRP promoter through exon2. The track indicated by "conservation" shows the degree of evolutionary conservation in 17 vertebrates. Heights of the peak are indicative of the degree of conservation. (B) Snapshot from the ECR browser with alignment of human BCRP with mouse and rat. (C) Results from the rVISTA browser with local alignment of mouse and human BCRP. (D) CISTER analysis of BCRP regions with clusters of binding sites for transcription factors chosen. Vertical colored lines indicate the probabilities that regulatory factors bind to cis-elements at these positions. The black curve indicates the overall probability of being within a cluster of *cis*-elements bound by their factors. Each color corresponds to a different TF binding site. Lines in the upper half of the plot indicate cis-elements on the sense strand, and lines in the lower half refer to the complementary strand.

Figure 2: **BCRP RNA expression in human livers.** Total BCRP RNA expression in 60 livers was measured by quantitative real time PCR and expressed as "relative BCRP RNA" levels (each sample normalized to its GAPDH RNA level and then expressed relative to a single human liver sample) were plotted from the lowest to highest relative BCRP in each racial group. The inset panel box plots show the median values of log2 transformed hepatic BCRP RNA expression in each racial group. Box plots indicate 1st and 3rd quartiles, with the bold line within the box representing the median value; the whiskers represent the range after excluding the outliers. The outliers are defined by the R package as data points which fall outside of the 1st and 3rd quartiles by more than 1.5 times the interquartile range, and circles falling outside the box

represent outliers. The p values from the Kruskal-Wallis nonparametric test comparing the significance in BCRP expression between the three races is shown.

Figure 3: **BCRP promoter SNPs showing significant association with BCRP RNA expression in different tissues**. Box plots for individual BCRP promoter SNPs (A-E) demonstrating statistically significant association with BCRP mRNA expression in livers, intestines and PDR samples. The Median $\pm$ 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 Wilcox test comparing the significance of three or two genotypes, respectively is shown for each of the SNPs.

Figure 4: **BCRP intronic SNPs showing significant association with BCRP RNA expression in different tissues**. Box plots for individual BCRP intron 1 SNPs (A-D) showing statistically significant association with BCRP mRNA expression are shown as described in Fig 3 legend.

Figure 5: **The BCRP -15994 genotype shows weak association with imatinib pharmacokinetics**. Plots show the relationship between apparent oral clearance of imatinib determined at steady state in patients. Each symbol represents an individual patient. CC genotype, N=61; CT genotype, N=25; TT genotype, N=4. Lines are the median with interquartile range (CC, 8.87 L/hr vs CT/TT, 12.3 L/hr).

Figure 6: **Summary of functionally significant BCRP SNPs.** Positions of BCRP promoter and intronic SNPs (numbered with reference to the transcription start site) showing significant

association with BCRP expression are mapped. Corresponding ***rs*** numbers of the SNPs identified in the dbSNP database are also indicated. Boxes below each SNP indicate the significant associations seen with BCRP mRNA expression in the PDR44 or livers or intestines with arrows representing increased or decreased BCRP mRNA expression. Potential functional consequences – resulting in either a gain (+), loss (-) or no TF change for each SNPs based on in silico analysis using TRANSFAC are also indicated in the boxes.

Figure 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 exon2 are indicated. (B) Cartoon of BCRP 5'UTR variants identified by Nakanishi et al., 2006 for comparison. (C) **Alternative ex1b usage is associated with lower BCRP expression in human livers.** Relative BCRP RNA expression in livers with presence or absence of exon1b was compared. Y- Axis indicates the log transformed relative BCRP RNA levels. Statistical significance was calculated using Wilcoxon test.

Figure 8. **BCRP splice variants**. (A) Left panel shows the agarose gel of RT-PCR products generated using forward primers in exon 1a, 1b or 1c and reverse primer in exon6. Arrows indicate exon2 skipped transcripts (SV.1) Right panel shows potential translation start sites used by the exon2 skipped transcripts. (B) Amino acid sequence alignment for BCRP wild type and SV.1 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 RT-PCR products generated using forward primer in exon 7 and reverse primer in exon8 showing higher MW alternative SV2 mRNA (left side) and schematic representation of inserted nucleotides from intron 7 (Exon 7a).

Figure 9: **Allelic imbalance in BCRP.** (A) Relative BCRP mRNA arranged from lowest to highest expression in PDR44 cells. The corresponding exon2 and exon5 BCRP genotypes sequenced in genomic DNA (g) vs. 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 vs. DNA. (B) Left panel, box plots of BCRP mRNA in samples with the exon5 wt or heterozygous genotypes. Right panel, box plots of BCRP mRNA in exon 5 wild-type vs. heterozygous genotypes +/- exon 5 allelic imbalance. The p value was generated by the Kruskal-Wallis test.

Supplementary Figure 1

Haplotype block structure of BCRP from the CEPH HAPMAP data generated with the Haploview program (**http://www.broad.mit.edu/mpg/haploview/)**. The colored blocks indicate the degree of linkage disequilibrium (D' values) between any two SNPs indicated (in the CEPHs). The darker the color of the box, the higher the LD. The triangular boxed regions indicate LD blocks identified. The rectangular boxed region indicates the BCRP gene boundary. Common haplotypes in each LD block, and their frequency, is shown in the bottom panel.

Table. 1. PCR and sequencing primers and cycle conditions used to for BCRP for resequencing and for real time quantification

Region amplified	Relative to TSS	Position in AC097484	Primer sequence (5'--> 3')	Amplicon size	Conditions (^{0}C)
Rg I	-71599 to -72533	137616F	AAGGAGAAGCCCAACATCCT	934 bp	Annealing 59, Extension 72
		138550R	CCTTCTTCTAGCCCCTTGCT		
Rg II	-28916 to -30930	94933F	AATGGGGCAGTTTCTTACCA	2014 bp	Annealing 59, Extension 68
		96947R	TGGTGCCATCAAACTGAAAG		
seq		95508R	TAGAGATGAGGTCTCACTGTGTTG		
		96237F	TCTTGCAGCTCCTTGGTTAAA		
		96257R	TTTAACCAAGGAGCTGCAAGA		
	Position in AC084732	**Position in AC084732**			
Rg III	-18132 to -14320	2227F	TTTGCTTCTCTTTTCCATAGCTC	3812 bp	Annealing 59, Extension 68
		6039R	AGTGGCCAATCTTTGTTTCA		
seq		3556R	GTAGTCTCCGCGTGCATTCT		
		4090F	GACGCAATCATGTAGAACATAACA		
		5273R	ATGGACAGCAGTGTGTCCTTGAGAAATT		
		2792F	AAAAAGCATTCTCTTTGGTTTCA		
Rg IV	-13890 to -13400	6469F	AAAAACCCCTCCTGCAAAAT	490 bp	Annealing 59, Extension 72
		6959R	GCCTAGTGGCTCATGCCTAT		
Rg V	-13601 to -11457	6758F	AACCTCTGCCTCCTTGGTTC	2144 bp	Annealing 59, Extension 72
		8902R	GGGCCTAATCCTGACACAAA		
seq		7475R	TTGAGGGTGGAGGTTGAGAG		
		8055F	CGTGTCATGAGGGTTTGTTG		
Rg VI	-6594 to -4577	13765F	GGGAGGCAGAAGAAAATCAA	2017 bp	Annealing 58, Extension 72
		15782R	TGAAATGGGATCCAGAGGAA		
seq		14273F	AAGACCAGCTGAGGCAACAT		
Rg VII	-4203 to -1369	16156F	CGGACTGAATACATCTGTTATAT	2834 bp	Annealing 57, Extension 72, + MgCl2
		18990R	GGCCGTTAACGACTGTTTGCAA		
seq		16900R	AAAGACACATGACAATAGAGTCCTCA		
		17018F	GACATTCCATGTTACTACCCAGAA		
Rg VIII	-814 to +1621	19545F	TGCATTGGGTGAACCATTAATA	2435 bp	Annealing 59, Extension 68,

					+7.5% DMSO
		21980R	CAATGAAAGGCTGAGGAACTG		
seq		21332R	TTTAGCAAACCATCCAAAGC		
		19937F	AAATGGGTGGTTTCTGGTGA		
		21308F	AAATTATTGGTCAATCCCTTTAAAG		
Rg IX	6711 to 7596	27070F	TCAAATGAAACCCTGCCAGT	885 bp	Annealing 59, Extension 72
		27955R	TTACCCAAAACCACACAGCA		
Rg X	9296 to 10427	29655F	CAAAGTGCTGGCATTACAGG	1131 bp	Annealing 59, Extension 72
		30786R	TGATCAGGAGCTAAGACTTGACC		
seq		30583R	GGAAGGGATCCTATGCCAAC		
Rg XI	11481 to 13553	31840F	GGCAAAGAGCATGAAGAAG	2072 bp	Annealing 60, Extension 72
		33912R	GTTGATCCATGGGCTTTCAT		
seq		32409F	TCTGTTGCTTAGGCTGGATTC		
Rg XII	13745 to 17404	34104F	TTTATGCAGCATGCCCTATG	3659 bp	Annealing 59, Extension 68
		37763R	GAGGAGCCCTGGAAAGAGAT		
seq		34500R	TGAGAAGTTTAACCATTTGTCTTC		
		35645F	TCCTATCACAGGAAAGGCAAA		
		36776F	TTGAATTTGCTGGGTGTCAG		
Intron2f		44341F	GGAAAGCTTTTCTGACAGTGG	2462 bp	Annealing 60, Extension 72
Intron2r		46803R	GCGTTGCAAATGCTCAATAA		
seq		45246R	AGACTTCAGGGTGGGGTCTG		
		45750R	CGTGGTACATACATAAGCACGTT		
Intron7F		61087F	CACCTTATTGGCCTCAGGAA	1596 bp	Annealing 60, Extension 72
Intron7R		62683R	AATCAACTGCTGTGCTGCTG		
seq		61756R	GCAGTCCCAGCTACTCAGGA		
Intron10f		69358F	GTGCCAAGGGTTTTCTGAAC	4221 bp	Annealing 63, Extension 72, +MgCl2
Intron10r		73579R	GGAGCACAGTATCTGCCACA		
seq		70728F	CCAGCCCATTTCTTGTTTTT		
		71488R	GGGGAGAGTGGTAGGACACA		
		71476F	TACCACTCTCCCCAAAGCAC		
		72283R	TGGATGGTGGTGATAGTTGC		
Intron14f		82983F	CAATGCCAGGTGTATTGGTG	3533 bp	Annealing 60, Extension 72
Intron14r		86516R	AAAGGAGCCTAAAATTGAACCA		

seq		83688R	AGGAGGTATCTCCCCTAGCC		
		83665F	AAAGGGCTAGGGGAGATACC		
		84434R	CAAAACCCATTTTGACACTGAA		
		84612F	CATGCAGCAATGTTTCTTACG		
		85419R	AGCCTCATCTTCCGTTACCA		
RT-PCR primers					
		Position in AC084732			
bcrp exon 1b-f		20049F	AACCCAGCTAGGTCAGACGA		
		Position in NM-004827			
bcrp exon 1a-f		437F	CCTGAGCCTTTGGTTAAGACC		
bcrp exon 6R		1057R			
bcrp realtime f1		1304F	CAGGAGGCCTTGGGATACTT		
bcrp realtime f2		1324F	TGAATCAGCTGGTTATCACTG		
bcrp realtime r		1411R	TGCCACAGCAGTGGAATCT		
bcrp exon 8F		1348F	GGCCTATAATAACCCTGCAGAC		
bcrp exon 10R		1739R	GCCCAAAGTAAATGGCACCT		
bcrp exon 14F		2152F	TCTGTTGGTCAATCTCACAACC		
bcrp exon 16R		2431R	CAGGTAGGCAATTGTGAGGAA		
		Position in BC092408			
bcrp exon 1c-f		70F	CTCGGAAAGCCTCAATGTTC		
Copy number primers					
CYP3A4 CNVF			CCCAGACTTGGCCATGGAAACC		
CYP3A4 CNVR			CAGATAAGGGAAAGAGAGGC		

TSS=transcription start site

Table 2. Allele frequencies of SNPs(in/dels) identified in the promoter and intronic regions of BCRP.

Position from TSS	Position From ATG	Position in Refseq*	dbSNP	Sequence	Region	AF in PDR44	AF in Whites	AF in Hispanics	AF in AAs	Effect on mRNA expression** and TRANSFAC change	Region	Tag SNP ***
Numbering based on AC097484												
-79753	-98670	145770	rs1114568	tcact[a/g]atgct	Upstream							
-71762	-90679	137779		tcatg[t/a]gttca	Upstream	0.04	na	na	na		I	
-71813	-90730	137830		caaaa[a/g]tgcaa	Upstream	0.06	na	na	na		I	
-71997	-90914	138014	rs10032109	tgaca[c/t]atttg	Upstream	0.06	na	na	na		I	
-71961	-90878	137978		atatg[c/t]aagtt	Upstream	0.04	na	na	na		I	
-72258	-91175	138275		ttggg[g/a]cagga	Upstream	0.12	na	na	na		I	
-69504	-88421	135521	rs9307048	cagacc[c/t]gggat	Upstream							
-30639	-49556	96656		aataa[-ataa]gaaat	Upstream	0.28	0.27	0.30	0.41	Decreased mRNA (L); loss of 1Hnf1site	II	
-30562	-49479	96579	rs2127862	gccac[t/c]gtact	Upstream	0.70	0.47	0.59	0.26		II	
-30477	-49394	96494	rs2127861	tgctt[c/g]tattc	Upstream	0.94	0.80	0.82	0.85	Decreased mRNA (L)	II	
-30109	-49026	96126		acaac[ins-a]aaaaa	Upstream	0.70	0.50	0.54	0.35		II	
Numbering based on AC084732												
-17824	-36741	2535	rs13135956	caagg[c/t]gagag	Upstream	0.19	0.23	0.09	0.47		III	
-17750	-36667	2609	rs13111149	aaaaa[a/g]aaaaa	Upstream	0.19	0.23	0.09	0.44		III	
-17243	-36160	3116	rs1481016	aaaaa[a/c]tgaga	Upstream	0.16	0.27	0.14	0.59		III	
-17242	-36159	3117		aaaaa[t/c]gagaa	Upstream	0.56	0.63	0.55	0.76		III	
-17084	-36001	3275	rs1481014	tggag[g/a]attgg	Upstream	0.00	0.23	0.14	0.38		III	YRI
-15994	-34911	4365	rs7699188	tacac[c/t]ttata	Upstream	0.22	0.07	0.09	0.06	Increased mRNA (L,I,P);gain of 1HNF4 site	III	
-15846	-34763	4513		gcaac[a/c]aaagc	Upstream	0.09	0.07	0.13	0.06	Increased mRNA (L,I,P)	III	
-15756	-34673	4603		ctcac[g/a]cctgt	Upstream	0.22	0.07	0.14	0.03		III	
-15672	-34589	4687		tggag[a/g]aaccc	Upstream	0.21					III	
-15622	-34539	4736		ggtgg[c/t]gcatg	Upstream	0.20	0.03	0.14	0.06	Decreased mRNA (L,I,P)	III	
-13680	-32597	6679		aatta[-a]ttttt	Upstream	0.07	0.07	0.05	0.06	loss 2LHX3, gain 1FoxD3	IV	

-13000	-31917	7924 to 8055		?insertion	Upstream	hets not known	hets not known	hets not known	hets not known		V	
-13359	-32276	7345		tagca[c/t]actgt	Upstream	0.01					V	
-12968	-31885	7391		gcttc[g/a]acctt	Upstream	0.14					V	
-12945	-31862	7414		gatta[c/g]aggca	Upstream	0.08					V	
-12870	-31787	7489		aggtg[t/a]cttaa	Upstream	0.02				gain of 1 AP1 site	V	
-11868	-30785	8491		gagcc[g/a]gtcgc	Upstream	0.01					V	
-6055	-24972	14304	rs11383890	acccc[+c]atctc	Upstream	0.59	0.57	0.55	0.62		VI	
-6219	-25136	14485		acaaa[a/c]aacaa	Upstream	0.02					VI	
-5267	-24184	15092	rs2725226	tgagg[t/c]agtag	Upstream	0.57	0.57	0.55	0.62		VI	
-5113	-24030	15246		ctttt[-t]aaaaa	Upstream	0.07					VI	
-3602	-22519	16757	rs3114020	tccca[g/a]tgtaa	Upstream	0.59	0.57	0.55	0.62		VII	YRI
-3335	-22252	17024	rs6846742	acatt[c/g]catgt	Upstream	0.01					VII	
-3243	-22160	17116		ggctc[g/a]ctgca	Upstream	0.07					VII	
-2959	-21876	17400		tcaca[c/g]cagtt	Upstream	0.06					VII	
-2728	-21645	17631		acaca[g/t]gcaca	Upstream	0.59	0.57	0.55	0.62		VII	
-790	-19707	19705	rs4148162	cactca[-ctca]caaag	Upstream	0.59	0.57	0.55	0.62		VIII	
-212	-19129	20147		gcggg[+g]agtgt	Upstream	0.06	0.03	0.05	0.00		VIII	
681	-18236	21040	rs2622605	tatgc[g/t]gttta	Intron 1	0.57	0.50	0.55	0.50		VIII	CHB
1143	-17774	21502	rs2622604	aatac[g/a]ccaga	Intron 1	0.15	0.33	0.29	0.15	Decreased mRNA (L,I,P)	VIII	CEU
1450	-17467	21809		ataac[a/c]tgggc	Intron 1	0.04	0.17	0.00	0.00		VIII	
6861	-12056	27220	rs9999111	tgaat[t/c]cacag	Intron 1	0.09					IX	
7250	-11667	27609		ttcac[g/t]ccatt	Intron 1	0.26					IX	
7438	-11479	27797		agcca[c/a]tgcac	Intron 1	0.08					IX	
9496	-9421	29855		tctcc[g/a]ttccc	Intron 1	0.01				loss 1CDX, 1Hnf1, gain 1Hnf3	X	
9666	-9251	30025		tcttt[t/g]attta	Intron 1	0.06	0.00	0.02	0.00		X	
9908	-9009	30267		atgaa[a/g]aaaca	Intron 1	0.01					X	
9932	-8985	30291		gtctt[a/c]tgttc	Intron 1	0.01					X	
11560	-7357	31919	rs2725246	taaag[t/c]tgtta	Intron 1	0.54					XI	
11603	-7314	31962	rs17731799	aggta[a/c]gtttt	Intron 1	0.44					XI	CEU
12283	-6634	32642		tgagg[t/c]tggga	Intron 1	0.06	0.00	0.04	0.07	Increased mRNA (L,P)	XI	
12638	-6279	32997	rs11287117	aggaa[-c]ccagc	Intron 1	0.33					XI	
12674	-6243	33033		aagga[-t]ttttt	Intron 1	0.24					XI	
13070	-5847	33429	rs10632140	ctgcc[-tcagc]gtccc	Intron 1	0.37					XI	

13989	-4928	34348	rs6857600	gttgc[g/a]caatt	Intron 1	0.29	0.32	0.20	0.53	loss of 1 cebp delta site	XII	
15463	-3454	35822	rs3109823	tttcc[a/g]tcttg	Intron 1	0.28	0.27	0.29	0.71		XII	
15484	-3433	35843	rs3114018	ctatt[g/t]aaatt	Intron 1	0.49	0.63	0.50	0.74		XII	CEU YRI
16080	-2837	36439	rs2725250	gtagc[t/c]gaaac	Intron 1	0.28	0.32	0.43	0.21		XII	
16214	-2703	36573	rs2622620	aaagt[t/g]ctggg	Intron 1	0.52	0.57	0.57	0.85		XII	
16702	-2215	37061	rs2046134	ttttg[c/t]tgcat	Intron 1	0.10	0.59	0.05	0.00	Increased mRNA (L); gain of 1 GATA4 site	XII	YRI
16711	-2206	37070		atttt[c/t]ccttc	Intron 1	0.04					XII	
16823	-2094	37182	rs5860120	ttttt[-t]ccccc	Intron 1	0.24	0.03	0.04	0.00	Decreased mRNA (L)	XII	
18951	34	39310	rs2231137	tccca[g/a]tgtca	Exon2	0.13	0.06	0.20	0.06		Exon2	
19156	239	39515	rs4148152	tttta[a/g]tttac	Intron 2	0.13	0.06	0.20	0.06			
24353	5436	44712		ttagg[g/c]agctg	Intron 2	0.01					Intron2	
24487	5570	44846		tagat[c/t]ttgct	Intron 2	0.01					Intron2	
24580	5663	44939	rs2725255	ttcac[c/t]cttgt	Intron 2	0.05					Intron2	
24686	5769	45045	rs17731538	tcact[c/t]ataat	Intron 2	0.12					Intron2	
25049	6132	45408		cccac[c/t]accgt	Intron 2	0.02					Intron2	
25398	6481	45757	rs4148155	cttca[t/c]attct	Intron 2	0.09	0.10	0.21	0.00		Intron2	CHB
25579	6662	45938	rs17013859	ctgaa[g/a]tgctt	Intron 2	0.11	0.00	0.14	0.12		Intron2	
26347	7430	46706	rs2231138	gtata[a/g]gagag	Intron 2	0.06					Intron3	
27742	8825	48101	rs2231142	actta[c/a]agttc	Exon 5	0.12					Exon5	
40981	22064	61340	rs1481012	ccagc[t/c]tgtta	Intron 7	0.13	0.00	0.13	0.07		Intron 7	CHB
50510	31593	70869	rs2622628	caaga[g/t]tggtg	Intron 9	0.19	0.10	0.18	0.21		Intron 9	
50987	32070	71346	rs2054576	tttgc[t/c]acata	Intron 9	0.14					Intron 9	
51218	32301	71577		agcca[t/g]tgagt	Intron 9	0.04					Intron 9	
51284	32367	71643	rs2231148	tgtgt[a/t]taagt	Intron 9	0.24					Intron 9	CEU
51493	32576	71852	rs2231150	tccat[t/g]aagaa	Intron 10	0.01					Intron 10	
51526	32609	71885		tgtaa[a/c]tgtca	Intron 10	0.02					Intron 10	
51538	32621	71897	rs2231151	tcttt[t/c]attga	Intron 10	0.01					Intron 10	

51922	33005	72281	rs7681519	accat[c/g]caact	Intron 10	0.20					Intron 10	
62989	44072	83348	rs2231162	tgact[c/t]ttagt	Intron 13	0.22					Intron 13	
63217	44300	83576		aactc[c/t]ctttt	Intron 14	0.01					Intron 14	
63596	44679	83955	rs2725267	aagaa[t/c]gaaag	Intron 14	0.28	0.10	0.29	0.74		Intron 14	
63914	44997	84273	rs2231164	ttctt[a/g]aaatt	Intron 14	0.43	0.06	0.30	0.68		Intron 14	CHB
64152	45235	84511	rs2231165	ttttc[c/t]gagcc	Intron 15	0.07					Intron 15	
64681	45673	84949		aaggc[c/t]gcata	Intron 15	0.39					Intron 15	
64984	45691	84967	rs4148159	gttgt[t/a]gtttt	Intron 15	0.04					Intron 15	
64590	45764	85040	rs4148160	aataa[g/a]ttgag	Intron 15	0.01					Intron 15	
64608	46067	85343		tggtc[a/g]ggatg	Intron 15	0.12					Intron 15	

*RefSeq=AC097484 or AC084732; AF=allele frequency; hets=heterozygotes; shaded boxes=significant association with BCRP expression in one or more tissues.

**Effect of variant allele on mRNA expression in Liver (L), PDR44 (P) and Intestine (I).

***Tag SNPs:Abbreviations used:CHB, Han Chinese in Beijing, China; CEU, Utah residents with ancestry from northern and western Europe; YRI, Yoruba in Ibadan, Nigeria.

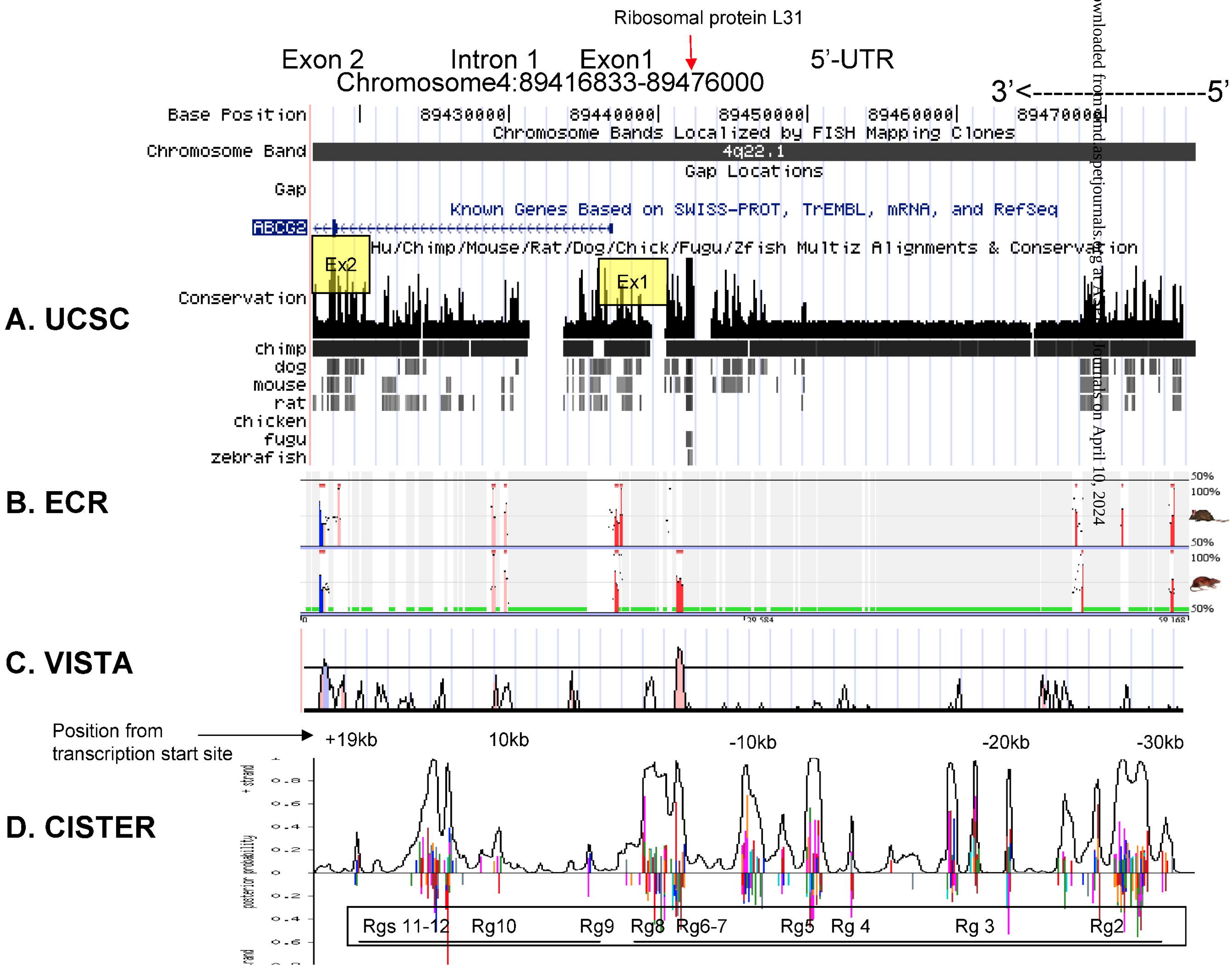


Figure 1

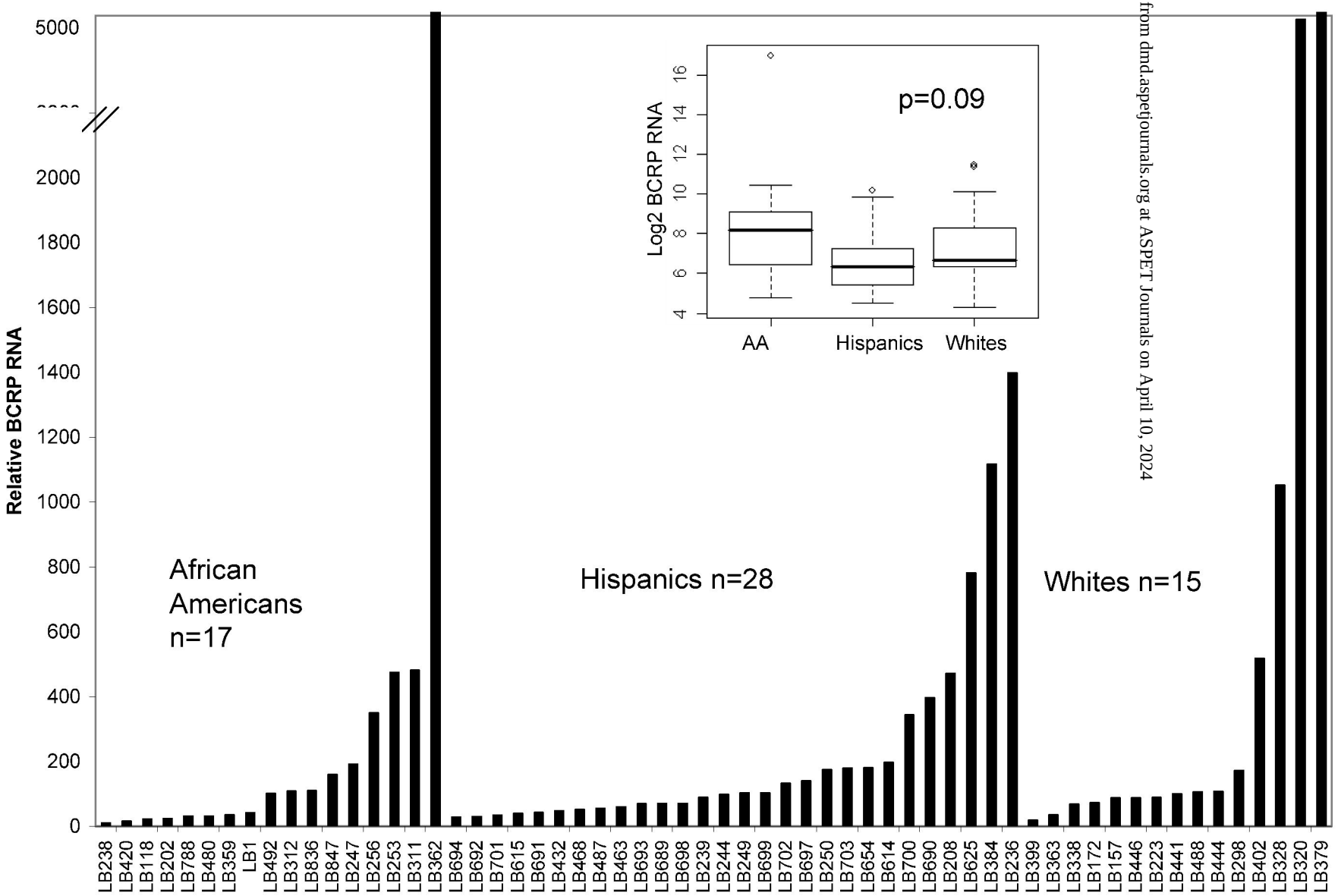


Figure 2

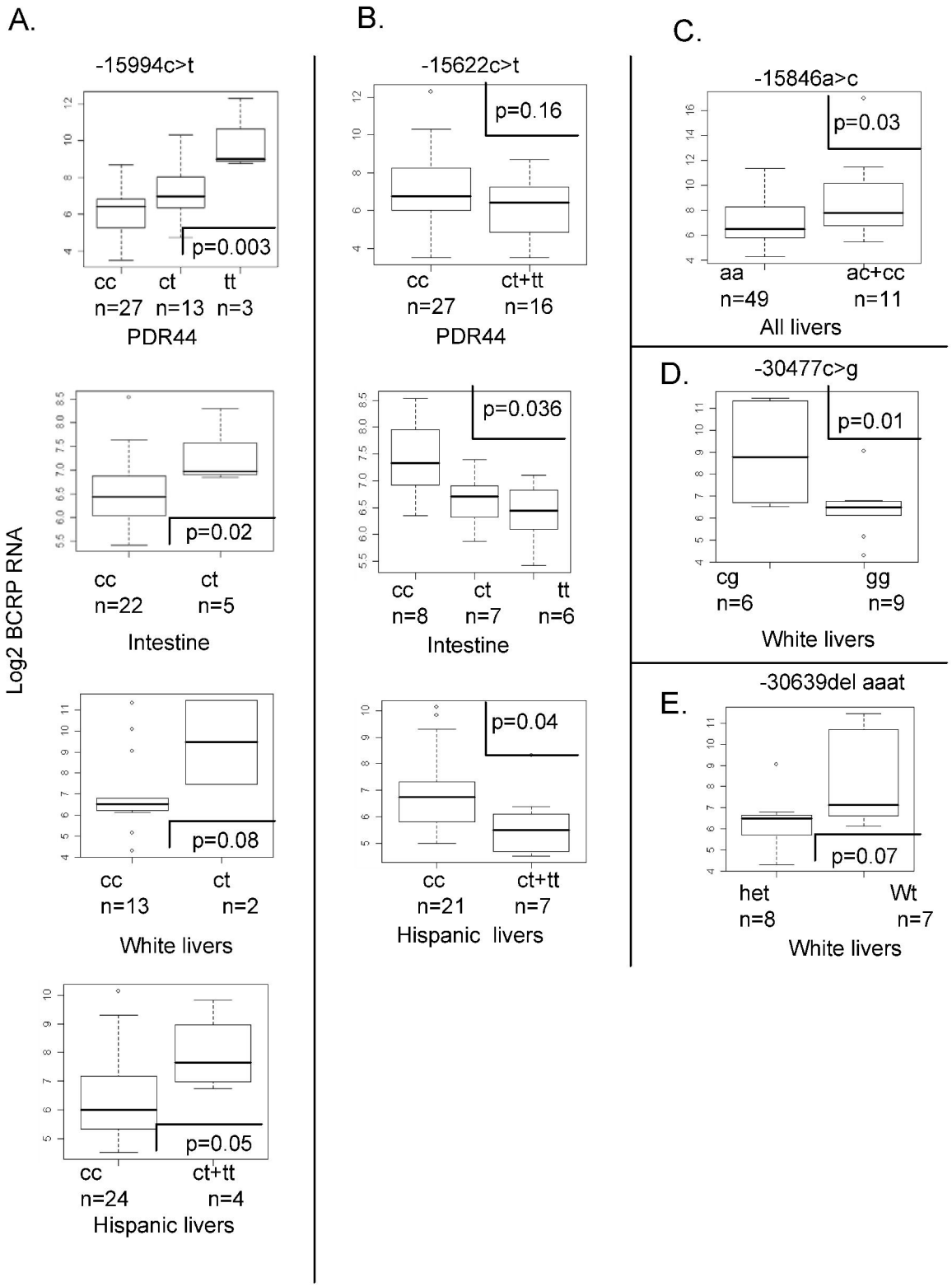

A.
B.
C.
D.
E.
Log2 BCRP RNA
-15994c>t
p=0.003
cc
ct
tt
n=27
n=13
n=3
PDR44
p=0.02
cc
ct
n=22
n=5
Intestine
p=0.08
cc
ct
n=13
n=2
White livers
p=0.05
cc
ct+tt
n=24
n=4
Hispanic livers
-15622c>t
p=0.16
cc
ct+tt
n=27
n=16
PDR44
p=0.036
cc
ct
tt
n=8
n=7
n=6
Intestine
p=0.04
cc
ct+tt
n=21
n=7
Hispanic livers
-15846a>c
p=0.03
aa
ac+cc
n=49
n=11
All livers
-30477c>g
p=0.01
cg
gg
n=6
n=9
White livers
-30639del aaat
p=0.07
het
Wt
n=8
n=7
White livers

Figure 3

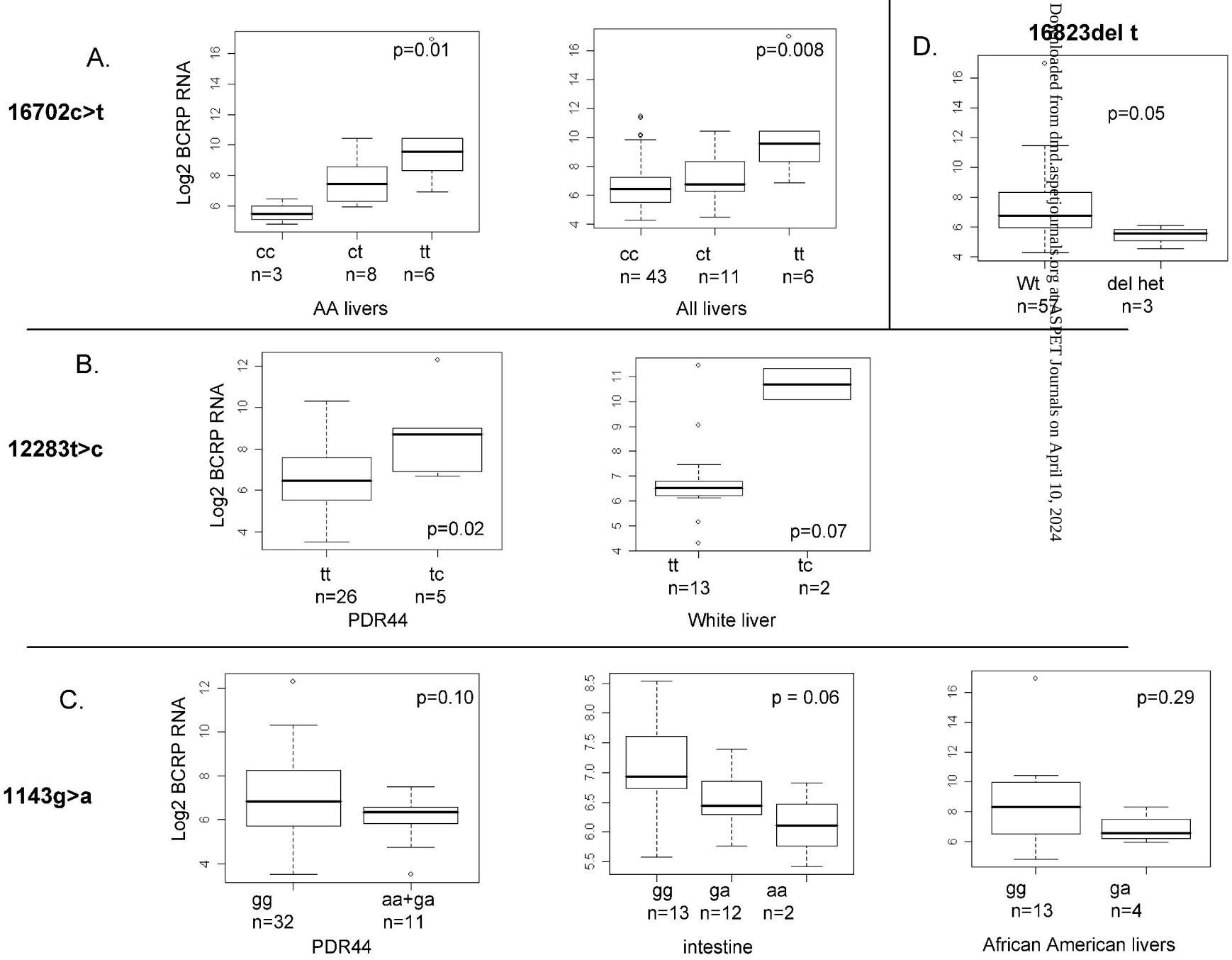


Figure 4

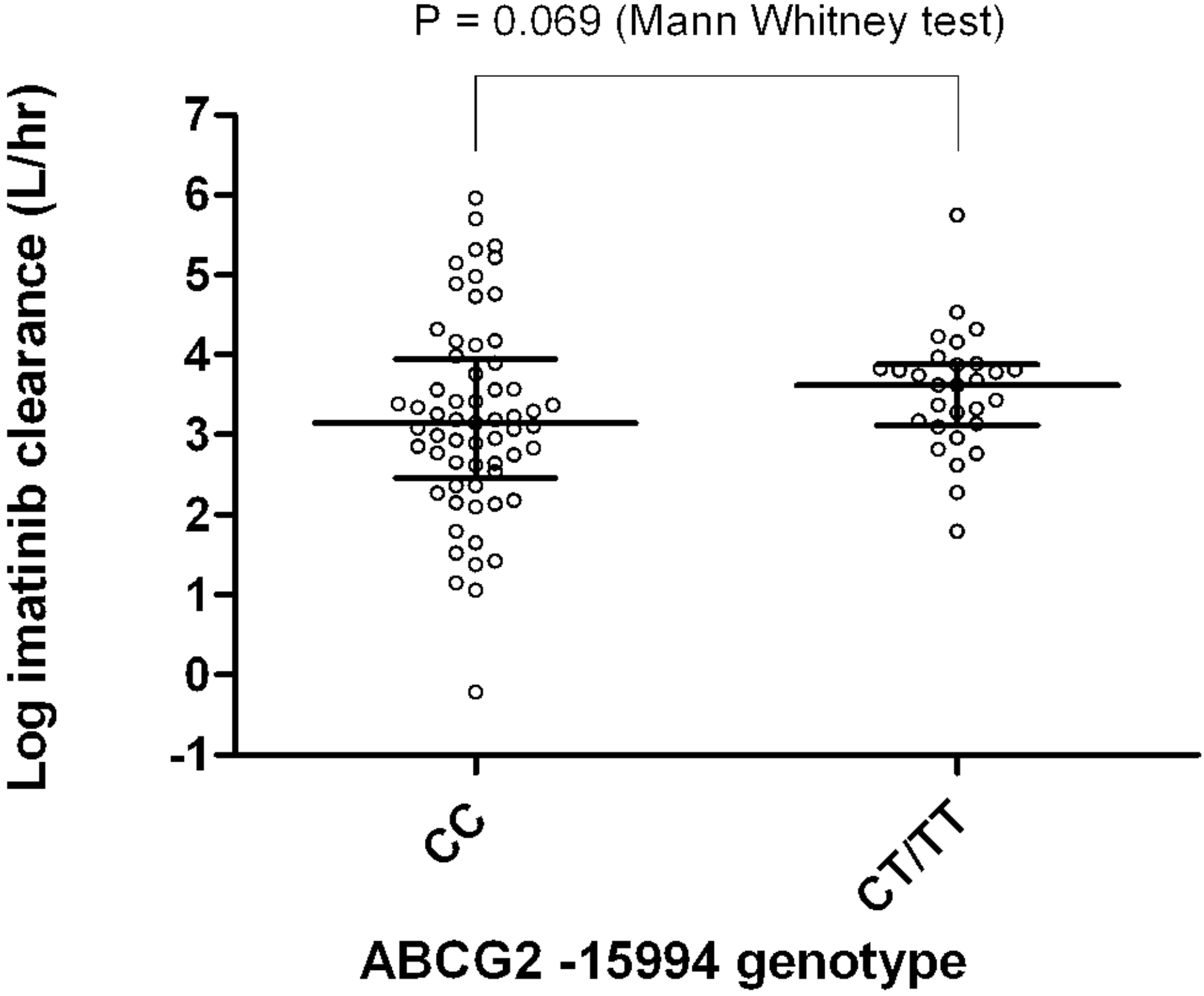

P = 0.069 (Mann Whitney test)
Log imatinib clearance (L/hr)
7
6
5
4
3
2
1
0
-1
CC
CT/TT
ABCG2 -15994 genotype

Figure 5

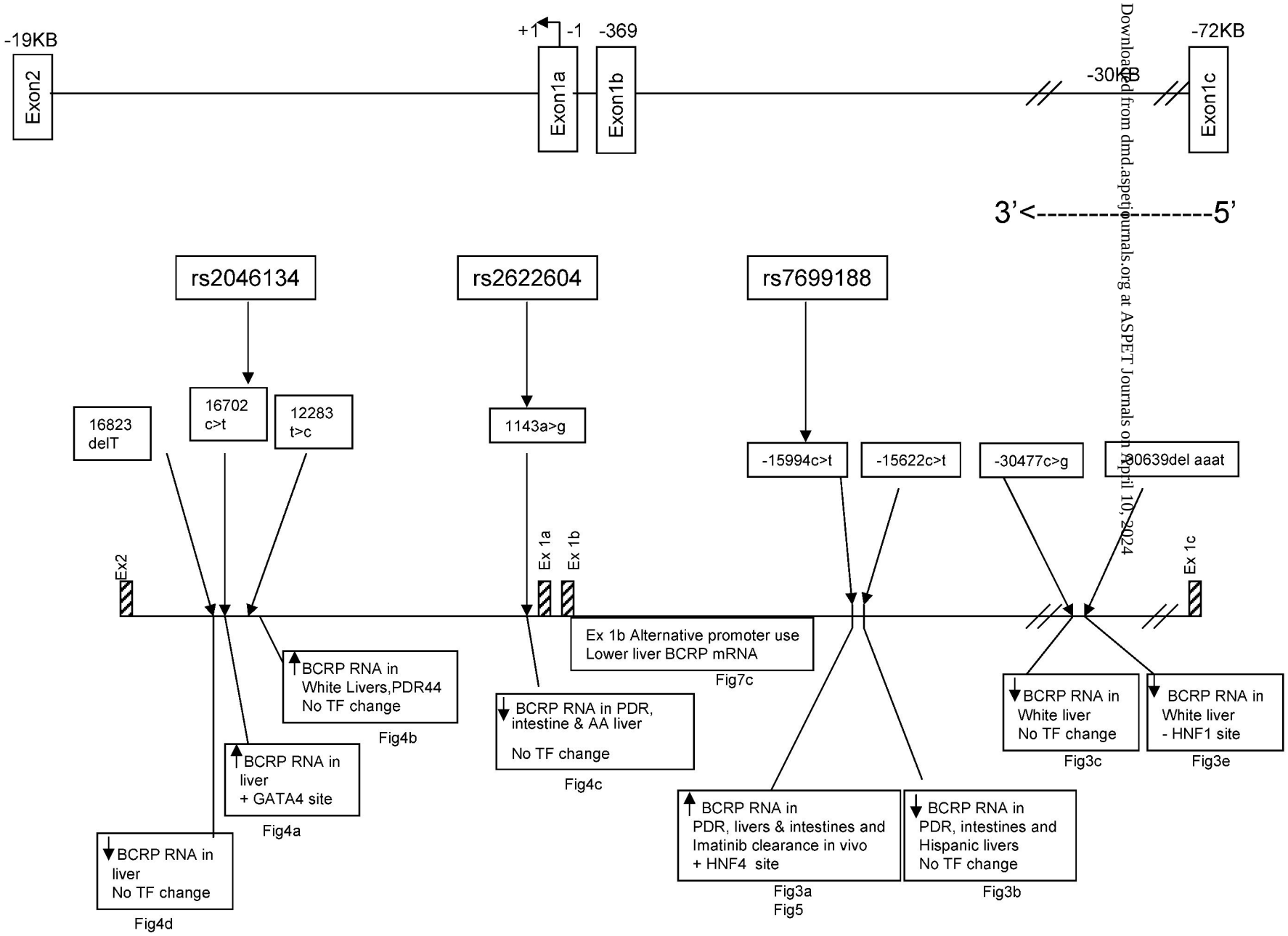
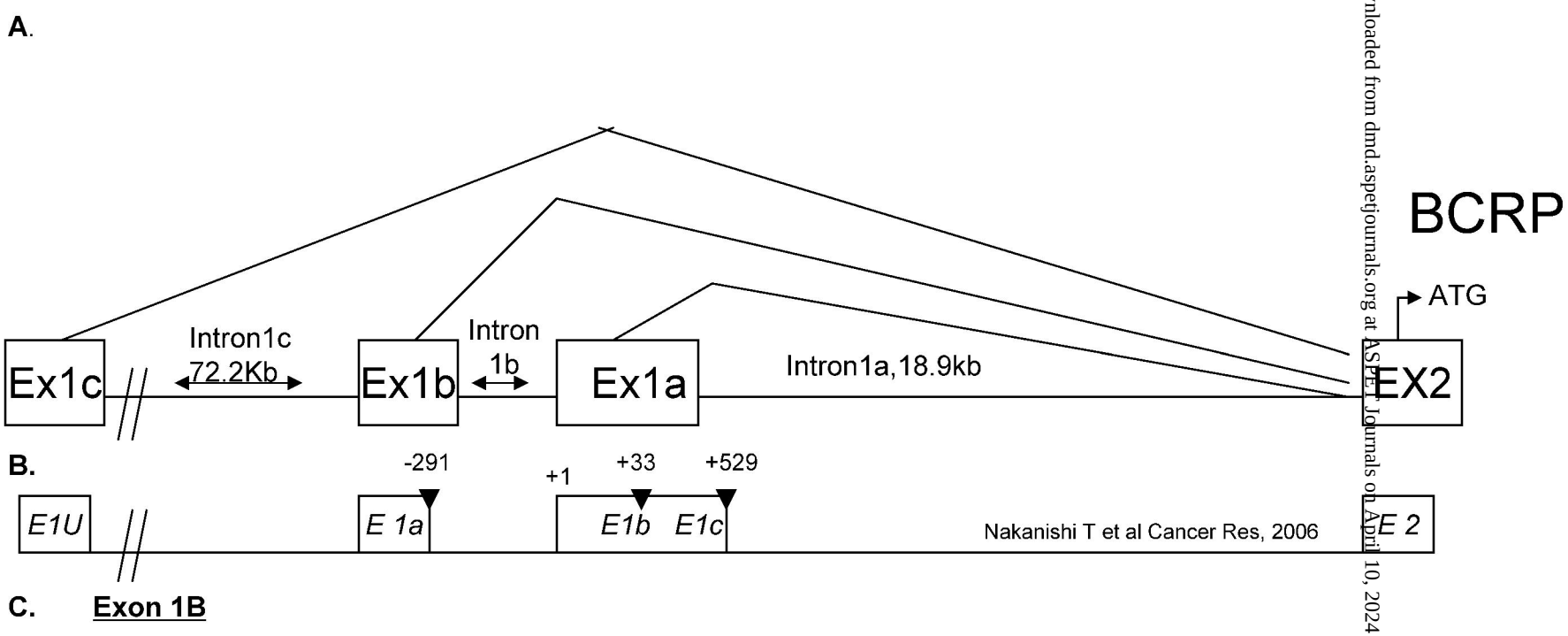


Figure 6



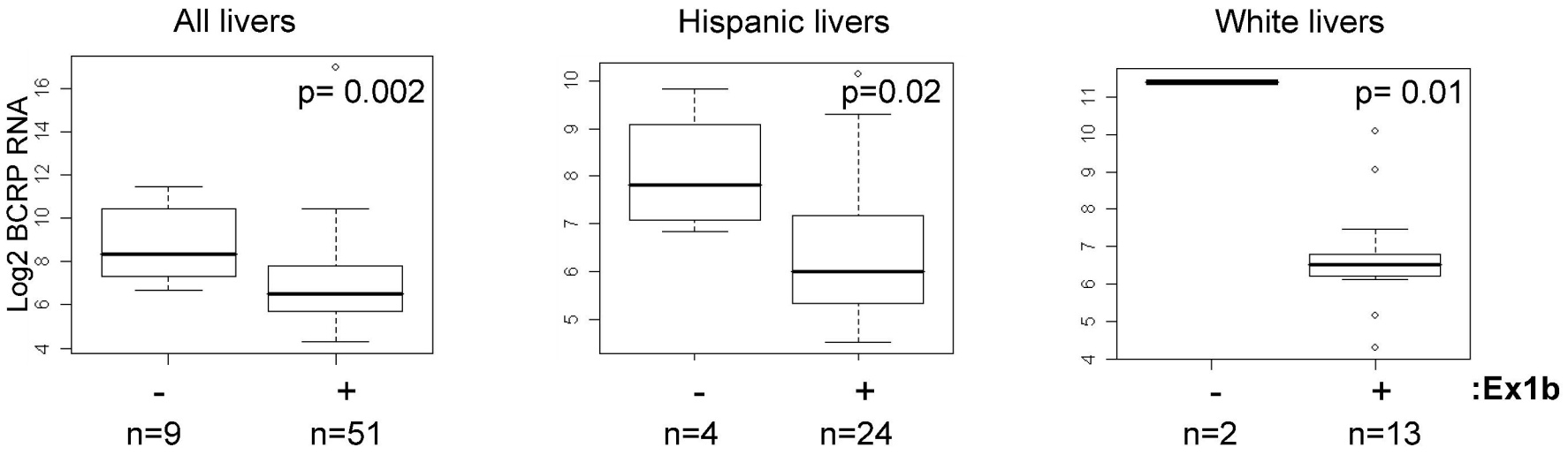
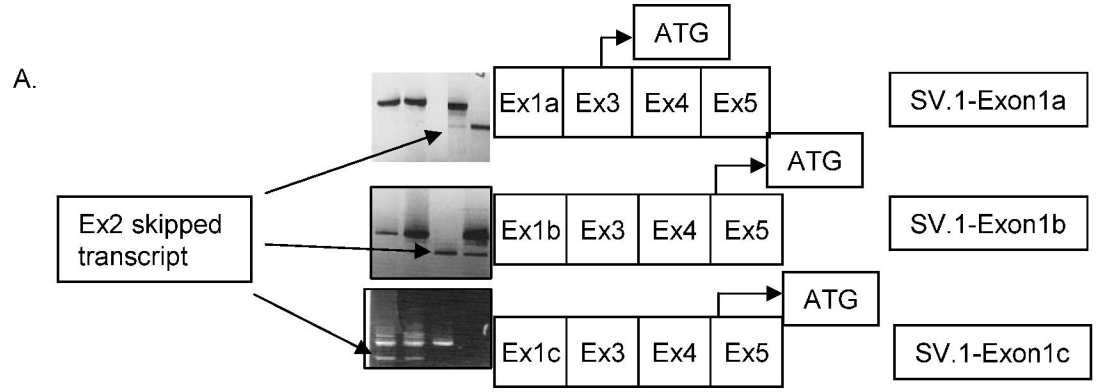
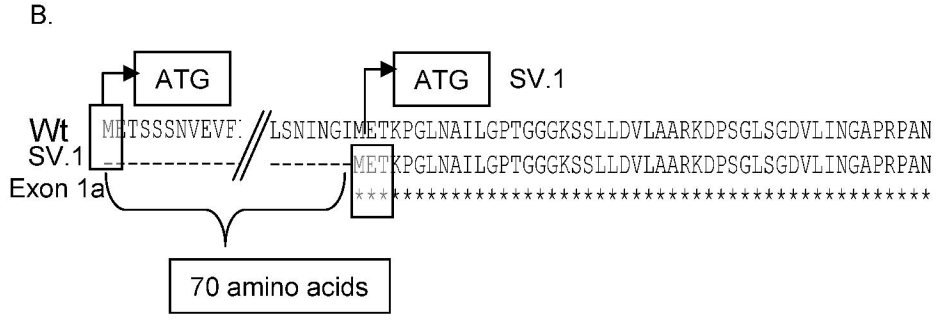
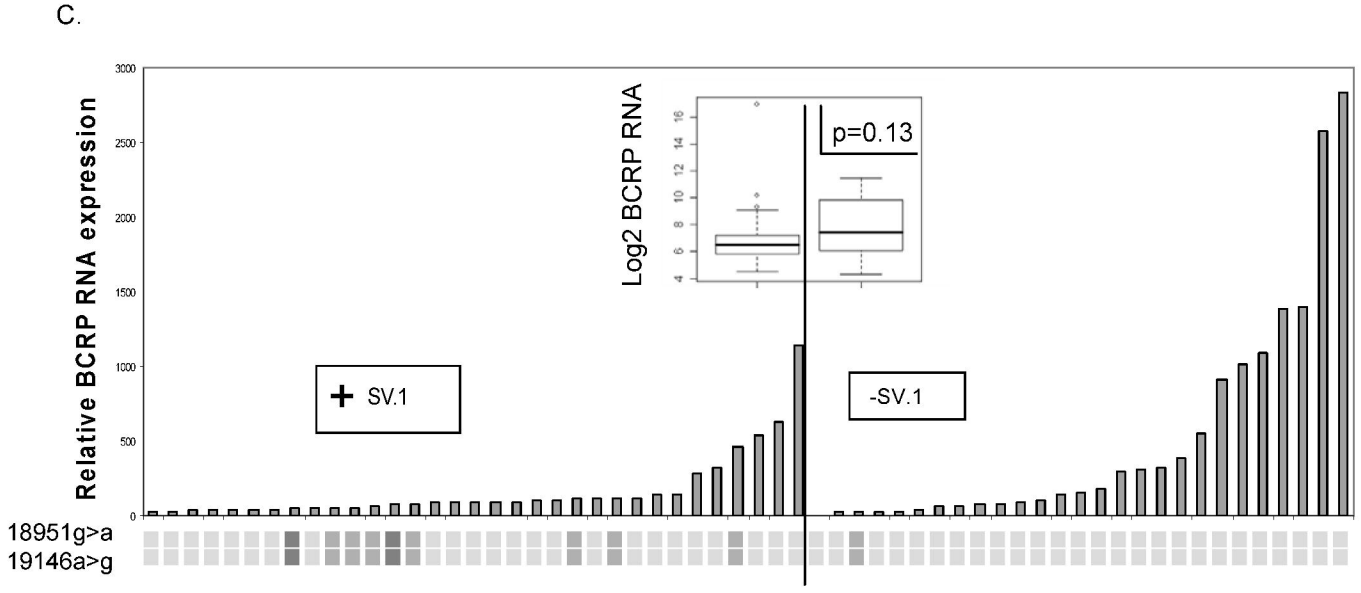


Figure 7


A.
ATG
Ex1a
Ex3
Ex4
Ex5
SV.1-Exon1a
ATG
Ex2 skipped transcript
Ex1b
Ex3
Ex4
Ex5
SV.1-Exon1b
ATG
Ex1c
Ex3
Ex4
Ex5
SV.1-Exon1c


B.
ATG
ATG
SV.1
Wt
SV.1
Exon 1a
70 amino acids


C.
Relative BCRP RNA expression
Log2 BCRP RNA
p=0.13
+ SV.1
-SV.1
18951g>a
19146a>g

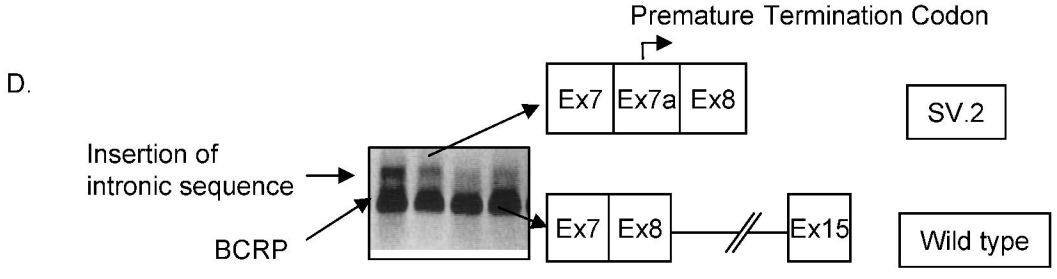
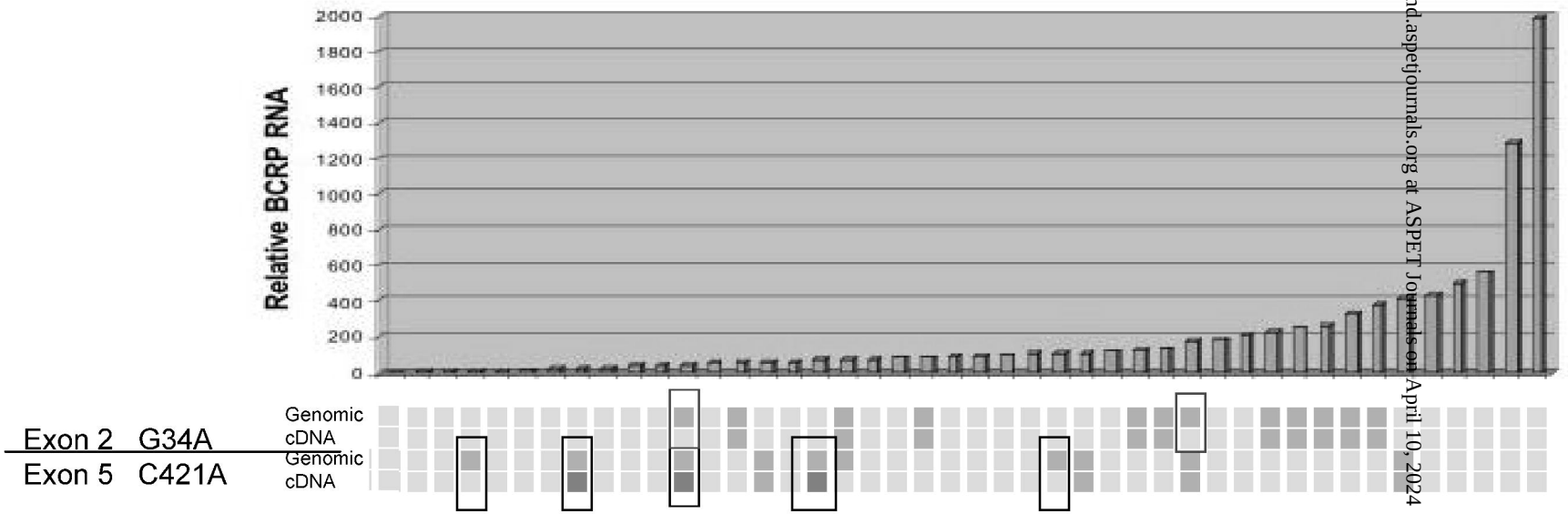

D.
Premature Termination Codon
Ex7
Ex7a
Ex8
SV.2
Insertion of intronic sequence
BCRP
Ex7
Ex8
Ex15
Wild type

Figure 8

A. PDR 44



B. Exon 5 C421A

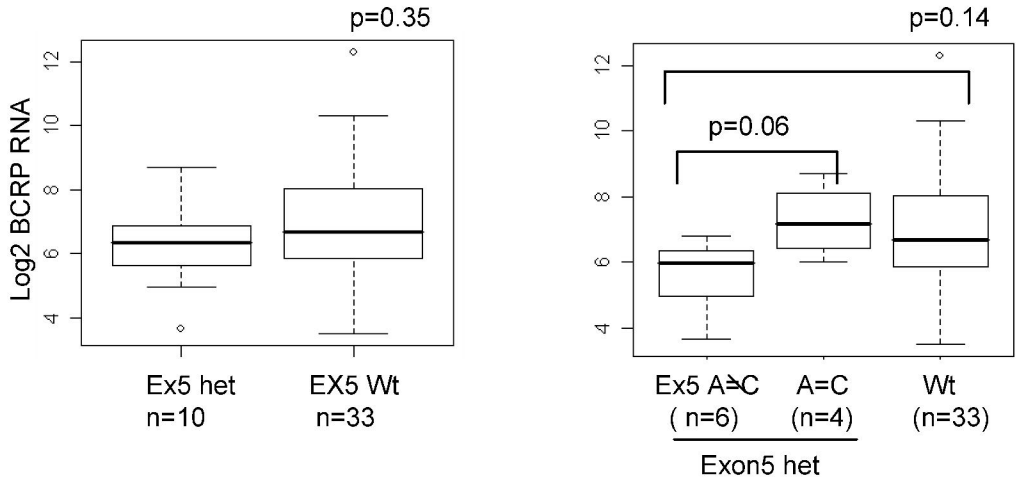


Figure 9