Published in final edited form as:

Pharmacogenet Genomics. 2013 May ; 23(5): 269–278. doi:10.1097/FPC.0b013e32835ff536.

Expression of mRNA transcripts encoding membrane transporters detected with whole transcriptome sequencing of human brain and liver

Amy Webb^a, Audrey C. Papp^b, Jonathan C. Sanford^b, Kun Huang^a, Jeffrey D. Parvin^a, and Wolfgang Sadee^{b,c}

^aDepartment of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA

^bDepartment of Pharmacology, Program in Pharmacogenomics, College of Medicine, The Ohio State University, Columbus, Ohio, USA

^cDepartments of Pharmacy, Psychiatry, Internal Medicine, and Environmental Health Sciences, The Ohio State University, Columbus, Ohio, USA

Abstract

Background—Membrane transporters control the influx and efflux of endogenous and xenobiotic substrates, including nutrients and drugs, across cellular membranes.

Objective—Whole transcriptome sequencing enables simultaneous analysis of overall and allele-specific mRNA expression, and the detection of multiple RNA isoforms.

Methods—Here we characterize variation in RNA transcripts emanating from gene loci encoding transporters based on RNAseq data from 10 human brains (including cocaine overdose and normal brain tissues) and 12 normal livers.

Results—mRNA expression was detected in 65% of transporter genes in either tissue, with many genes generating multiple mRNA transcripts. Single-nucleotide polymorphisms within transporters with previous evidence for pharmacogenomics impact were detected. We also identified noncoding RNAs in the vicinity of transporter genes with potential regulatory functions.

Conclusion—The results obtained with RNAseq provide detailed information on transporter mRNA expression at the molecular level, affording new avenues for the study of membrane transport, with relevance to drug efficacy and toxicity.

Keywords

alternative splicing; gene expression; RNAseq; transporters

© 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins

Correspondence to Amy Webb, PhD, Department of Biomedical Informatics, The Ohio State University, 336A Biomedical Research Tower, 460W. 12th Ave. Columbus, OH 43210, USA, Tel: +1 502 641 3724; fax: +1 614 688 6600; amy.webb@osumc.edu.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.pharmacogeneticsandgenomics.com).

Conflicts of interest

There are no conflicts of interest.

Introduction

Membrane transporters controlling the influx and efflux of diverse substrates across the cellular membrane exist in two major classes. ATP-binding cassette (ABC) transporters, move substrates against a concentration gradient through the use of energy from ATP, with 50 ABC genes grouped in seven families (as annotated in hg19) based on sequence homology. ABC transporters have been implicated as drug resistance proteins, exporting drugs from the cell and removing them from the body through various modes of elimination [1–3].

The second class of transporter proteins includes solute carrier proteins (SLCs), which move substrates along a concentration gradient to establish a state of homeostasis; alternatively, SLCs can be coupled to ion transport (e.g. Na⁺ and H⁺) and serve as secondary active transporters, or serve as substrate exchangers (e.g. cystine/glutamate exchanger) [1,2]. Approximately 370 SLC genes have been identified in 47 gene families (as annotated in hg19), classified by function rather than homology, interacting with diverse substrates including ions, nutrients, neurotransmitters, and a majority of drugs in current clinical use. Expression levels and functionality of these transporter proteins can affect drug absorption, distribution, and excretion, thereby, modulating therapeutic and adverse responses, rendering ABC and SLC transporters germane to pharmacogenomics and personalized drug response [1,2].

A main goal of pharmacogenomics is to study the genetic variability of drug response. A nonsynonymous single-nucleotide polymorphism (SNP) in a coding region changes the amino acid sequence thus potentially affecting the functionality of a protein. However, most SNPs do not result in protein coding changes but can regulate transcription or RNA processing [4]. A *cis*-acting regulatory variant can change promoter, enhancer or silencer sites resulting in a change in transcription regulation [4–7]. Alternatively, a variant can affect RNA processing through changes in splicing, stability of function, or mRNA trafficking [4]. Potential *cis*-acting regulatory polymorphisms can be detected through differential expression of mRNA alleles at heterozygous SNPs used as markers in the expressed RNA transcripts.

The aim of this study was to characterize the expression levels and sequence variation of transporters through high-throughput RNAseq in two tissues relevant to pharmacology and drug response – the liver and brain, chosen because we expect distinct functions to dictate the expression of different sets of membrane transporters, for example: supporting elimination of xenobiotics in the liver or providing nutrients and signaling molecules in the brain. Specifically, we determined overall mRNA expression levels of ABC and SLC transporters, identified alternative splicing events, and detected SNPs. The results show unique expression patterns between the two tissues and multiple transcript isoforms.

Methods

Tissue sample description

Human prefrontal cortex (PFC) autopsy tissue was acquired from the Miami Brain Endowment Bank [8]. Of the 10 post-mortem brain samples, five were from cocaine overdose individuals and five from control individuals who had not been cocaine abusers. Liver tissue was acquired from 12 individuals and extracted from human post-mortem tissue samples obtained from The Cooperative Human Tissue Network Midwestern and Western Division.

Tissue preparation and sequencing

Total RNA was extracted from liver and PFC tissues with chloroform and isopropanol precipitation. Samples were collected less than 24 h *postmortem* and brain pH levels were acceptable. RNA quality was verified using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA), and all samples had RNA integrity number greater than 6.5. cDNA libraries were created with the NuGen-Ovation RNASEQ System using 10 ng of total RNA per sample (NuGen Technologies Inc., San Carlos, California, USA); this kit uses poly-dT and random hexamer primers for cDNA synthesis, the latter designed to reduce cDNAs from rRNA by at least 90%. Therefore, less 3'-untranslated region (UTR) bias is expected compared with the use of poly-dT alone, and intronic regions and all non-poly-A tagged RNAs are represented. However, as the poly-dT primer is more abundant than the random hexamer primers, there is still a bias in read counts towards the 3'-end in the 3'-UTR. Moreover, a linear amplification step enables RNAseq analysis on small amounts of RNA (we used 10–20 ng total RNA). Sequencing was performed using the SOLiD 4 Next-Generation Sequencing instrument (Life Technologies, San Francisco, California, USA) in a total of four experiments. Two sequencing runs were performed for the liver and brain; one with four barcoded samples and the other with eight barcoded samples, resulting in higher read coverage in the smaller set, which also contained the reverse sequences from paired-end sequencing.

Analysis of sequencing reads

RNAseq data were aligned to the International Human Genome Sequencing Consortium human reference sequence (hg19) using SOLiD Lifescope v2.5.1 (Life Technologies, San Francisco, California, USA). For each sample included in a four-barcode set, paired-end reads were aligned (with 50 bases sequenced in the forward direction and 35 bases for the reverse). For each sample included in an eight-barcode set, single-end alignment was performed for the 50 bases sequenced in the forward direction. Splice junctions were identified based on the alignment of reads across splice junctions between annotated exons – either by a single read that crosses the junction or by paired-end reads that flank the junction.

Single-nucleotide variants were identified with Samtools v0.1.16 [9]. SNPs were filtered based on a call quality greater than 10 (Phred scaled probability that the consensus matches the reference) and read coverage greater than 10. Identified SNP locations were annotated based on UCSC annotation databases and dbSNP using ANNOVAR annotation software

[10]. On the basis of annotation, each SNP was assigned to a location within a gene locus – whether exonic, intronic, intergenic, UTR, or upstream/downstream (within 1 kb of the coding region). In addition, SNPs were assigned to noncoding RNA (ncRNA), or RNA transcripts with no annotated coding regions.

Cufflinks v2.0.2 [11] was used to calculate gene specific expression levels based on the parsimonious set of reads that can be assembled into a gene. Expression level is reported as FPKM or fragments per kilobase per million reads, which normalizes the number of reads within a gene by the number of fragments per kilobase of exon per million mapped reads for a given sample.

Analysis of relative mRNA abundances and splice variants using alternative methods

Confirmation of RNAseq gene expression measurements was performed using qRT-PCR. Reactions were completed in duplicate using 25 ng of cDNA, 300 nmol/l gene specific primers and FAST-SYBR-Green Master Mix (Invitrogen and Life Technologies, Carlsbad, California, USA). Gene expression was measured using an ABI 7500 Fast PCR System (Applied Biosystems, Carlsbad, California, USA). Expression was normalized to the housekeeping gene GAPDH and converted to log₁₀ scale for analysis.

Predicted RNAseq splice variants were confirmed using fluorescent PCR. PCR was performed using 25 ng of cDNA, NEB 2 × Master Mix and a gene specific primer set (300 nm each) with fluorescence tagged forward primers (6FAM). Lengths of the products were determined using a 3730 DNA Analyzer (Life Technologies, San Francisco, California, USA) and analyzed in Gene-Mapper (Applied Biosystems and Life Technologies, Carlsbad, California, USA). Splice variant length and presence was considered accurate (> 65 bp and with a minimum peak height of 800). Observed product lengths were compared with predicted length of splice variants for conformation.

Sequencing data access

The RNAseq data will be available through the NCBI Sequence Read Archive.

Results

RNA transcript analysis at gene loci expressing membrane transporters

Whole transcriptome sequences were generated from 10 PFC samples and 12 liver samples. The number of mapped reads and mapping percentages per sample is listed in Table 1, with 70–80% of the generated reads for each sample mapping to the genome. Read alignments were concentrated at known exons; however, other regions also showed high read coverage. Figure 1 displays the distribution of reads across *SLC38A2* (a zinc transporter) for four liver samples. Although the largest peaks cover exonic regions, multiple peaks exist between exons with coverage ranging from 100–500 reads. The use of random hexamer primers leads to greater coverage in intronic regions; these could represent premature heterogeneous nuclear RNA (which is typically present at low levels), introns that are included in alternative splicing events, intronic RNA remnants, or ncRNAs. In the following analysis, we will focus on expression of mature mRNAs that encode the membrane transporters.

mRNA abundance of SLC and ABC transporters in brain and liver tissue

mRNA abundance from RNAseq data is measured in FPKM, calculated from reads mapped to exonic regions. Thirty-five percent of the SLC and ABC transporters had sequence reads below 1FPKM and fall below the threshold for measurable mRNA abundance for the purpose of the present study. A higher percentage of transporters show measurable expression in the liver compared with brain. This finding may be a result of the greater diversity of cell types in PFC compared with liver, each with distinct expression profiles but representing only a small fraction of the tissues analyzed, or the tendency for the liver to express higher transporter levels in its primary function to cope with endogenous and exogenous substrates. High expression in specific cell types within the tissues, such as endothelial capillary cells forming the blood–brain barrier (e.g. *ABCB1* or *MDR1*), cannot be ascertained here as the RNAseq data are derived from whole tissue samples. Of transporter genes with expression greater than 5 FPKM, 32% were detected in liver and 30% in brain. Fifteen percent of genes were detected with expression at this level in both liver and brain. A detailed listing of gene specific expression can be found in Supplementary Table 1 (<http://links.lww.com/FPC/A578>)

PFC tissues were obtained from cocaine abusers and controls as part of a larger study to ascertain the effects of cocaine on expression in the brain. Differential gene expression analysis was applied to determine whether cocaine exposure affects the expression of membrane transporter proteins in the PFC. Four transporter genes differed between cases and controls with a nominal *P* value less than 0.05 (*t*-test), but only one of these had a fold change of 1.5 (*SLC24A2*, *P*=0.03). Adjusting for multiple hypotheses testing, none of these changes remain significant, and therefore, we combined the brain samples for the current study. Whole transcriptome analysis comparing cocaine and control tissues will be presented elsewhere.

Comparing gene expression from transporters in liver or brain between male and female yielded one gene that reached statistical significance after correcting for 420 comparisons. *ABCD4* had an average expression of 3.1 in female and 1.0 in male (*P* = 0.00004).

Comparing results obtained with NuGen kit cDNA preparation with previous RNAseq results using NuGen and only poly-dT selection, and with hybridization arrays

We performed RT-PCR on 85 genes (including eight ABC genes) for the liver samples included in this study showing a correlation of $r = 0.74$ ($P < 1e-8$) between the qRT-PCR derived expression and the RNAseq FPKM. These results validate the overall RNAseq results in the present study. In addition, we confirmed expression of 10 transporter genes showing a correlation of $r=0.94$ ($P = 0.00005$) in both the liver and brain separately. In each case, the $\log_{10}$ of the average expression measured from four brains of livers from RT-PCR was compared with the $\log_{10}$ of the average expression from all 10 brains or 12 livers. See Fig. 2 for comparison of the RNAseq and RT-PCR methods.

The NuGen kit was selected for library preparation over poly-A only selection as the NuGen kit captures all RNA transcripts at a given locus, including nonexonic regions, whereas using a set of hexamers designed to substantially reduce rRNA representation. Comparing

the NuGen preparation with poly-A selection and rRNA depletion, Tariq *et al.* [12] found it to be comparable in reproducibility and sensitivity but showed substantial difference in the genomic distribution of aligned reads. Although up to 69% of poly-A selected reads aligned to annotated exonic regions, the NuGen-Ovation prepared reads aligned with only 20% in exons, 34% in introns, and 36% in intergenic regions [12]. Using SNP calls as representation of read coverage, we see a similar distribution with 18% of SNPs in coding regions, 53% in introns and 21% in intergenic regions. Although using NuGen-Ovation for library preparation reduces the overall coverage in exonic regions compared with poly-A selection, it provides increased intronic read coverage and detection of potentially important regulatory regions within introns or within noncoding RNA. These data will be analyzed separately in the context of the whole transcriptome.

We compared these results with existing mRNA tissue expression results to test compatibility between methods. Tissue specific expression based on published hybridization microarray data was compiled using the GeneAtlas U133A microarray data from bioGPS.org. A total of 212 transporter genes overlapped between the two methods. A cutoff of 10 was chosen for microarray expression and a cutoff of 1 FPKM for RNAseq expression. RNAseq confirmed expression in 65 and 74% of genes with expression predicted by microarray (using our cutoff criteria) in the liver and PFC, respectively. A lack of detectable expression as shown by microarray data was confirmed for 51 and 56% of genes with RNAseq in the liver and PFC. On the basis of these comparisons, the sensitivity of liver and PFC expression in transporter genes was 0.60 and 0.58, whereas the specificity was 0.51 and 0.56.

Detection of SNPs

Sequencing of mRNA allows for the detection of polymorphisms in addition to expression levels. The number of SNPs with mapping quality greater than 10 (Phred scaled probability that the consensus matches the reference) and coverage greater than 10 reads per sample are listed in Table 1, both within transporter genes and genomewide. Tissues with a higher read coverage yielded a larger number of detected SNPs than lower covered samples, owing to the stringency criteria used to call a SNP. Similarly, as more transporter genes are highly expressed in the liver, SNP detection in transporter genes is greater in the liver.

Table 2 lists the locations of detected SNPs relative to annotated transporter genes. The majority of detected SNPs (60%) are located in intronic regions, with 5.8% in exonic regions, 13% in the 3'-UTR and 0.5% in the 5'-UTR, 0.8% in noncoding RNAs, and the remaining (< 5%) in downstream and upstream regions (within 1 kb of a coding region). qRT-PCR control experiments ruled out genomic DNA contamination as a source of reads mapping to nonexonic regions. SNPs in coding regions are divided into synonymous and nonsynonymous, and exonic SNPs located within two bases of a splice site are listed separately. Supplementary Table 2 (<http://links.lww.com/FPCIA579>) lists SNPs in coding regions within transporter genes. A small portion (0.34%) is located at the intronic side of a splice site, with potential impact on the splicing process. A comparison of SNP detection in 3' versus 5' UTRs reflects the increased length of 3'-UTR, but also a bias in detected reads mapping to these regions based on the library preparation, which includes oligo-dT priming

in addition to random hexamer priming. Genomewide, up to 5% of SNPs are located in annotated ncRNA regions highlighting the ability of the NuGen library preparation to detect ncRNAs.

Detection of lincRNA flanking transporter genes

We used the Human Body Map lincRNA database to consider a specific type of ncRNA, which includes catalogued long intergenic noncoding RNA (lincRNA) from the UCSC hg19 annotation, and predicted lincRNA based on a survey of whole transcriptomes of various tissues [13]. One annotated lincRNA overlapped with the coding region of a transporter gene, however neither the gene (*ABCA17P*) nor the lincRNA (*linc-ABCA3-BP*) showed detectable expression within the liver or brain. Within 1 kb of a transporter gene, 8 of 16 annotated lincRNAs were detectable, mostly in samples with higher read coverage. One lincRNA, *linc-ALDH3A2-1*, a transcript of uncertain coding potential as described by Cabili *et al.* [13] showed robust expression in 11 livers (180–650 FPKM). Located downstream of *SLC47A1* and upstream of *ALDH3A2*, this lincRNA is considered an unprocessed pseudogene of *SLC47* as annotated by the Havana project for the Vega browser. RNAseq detected 3 SNPs from 2 liver samples within the region containing this lincRNA. Another lincRNA, *linc-RGPD8-2* detected in 9 brains is located directly upstream of *SLC20A1*. Also called *FLJ42351* or *AK124342*, this lincRNA is listed as an uncharacterized cDNA clone.

Alternative mRNA splicing

Splicing events are detected based on evidence from single reads that lie across an exon-exon junction and paired-end reads that flank a junction. Detection depends on the depth of coverage and thus is greater in samples with higher read coverage and in genes with higher expression. Across the two tissues, 351 alternative splicing events were detected in 133 transporter genes. Table 3 provides a listing of the number of splice events detected per gene, sorted by prevalence (for details see Supplementary Table 3, <http://links.lww.com/FPC/A580>). Splicing detection by Lifescope used here is restricted to junctions between annotated exons thus limiting the potential for the detection of novel isoforms with non-annotated exons, which will require additional analyses.

Many previously described alternative splicing events were detected in multiple samples. The most frequent potentially novel alternative splicing events occurred in *SLC22A1* (*OCT1*; organic cation transporter) between exons 9 and 11 detected in 10 samples and in *SLC22A10* (*OAT5*; organic anion transporter), between exons 2 and 5 detected in 10 samples. The highest number of distinct alternative splicing events within a gene was detected in *SLC4A10* showing 11 events and *ABCA6* showing 11 events. Multiple alternative splicing events were also detected in *SLC38A2* (*SNAT2*, zinc transporter), with several events occurring in multiple liver samples, involving 6 non-canonical splicing junctions. Many alternative splicing events detected here are not catalogued in the Alternative Splicing Database (listed on genecards.org). This discrepancy requires further computational and experimental analysis.

We explored the detection of alternative transcripts in 8 genes. Of these genes, *SLC22A1* showed the best evidence of replication. Cufflinks assembled 5 transcripts as seen in Fig. 3a. The main difference between transcripts was the inclusion or exclusion of exons 9 and 10. Within the livers, junctions were detected between exons 9 and 11 supporting exclusion of exon 10, between exons 8 and 10 supporting exclusion of exon 9, and between exons 8 and 11 supporting exclusion of both exons 9 and 10. Primers for fluorescent PCR were chosen within exon 8 and exon 11 to capture these differences. Fluorescent PCR detected peaks at 104, 207, 221, and 321 as seen in Fig. 3b. The peak at 207 was only detectable in one of the four livers tested, but the other peaks were present in all livers. We would expect peaks at 107, 206, 219, 318, and 1383 based on the RNAseq data. Fluorescent PCR confirms all but the longest transcript that includes expression across the intron. For *SLCO1B1*, *ABCB1*, and *SLC7A11* the results suggested the presence of the 2 transcripts predicted by RNAseq. For *SLC38A1*, *SLC38A2*, and *ABCC2* only one of the multiple transcripts predicted was detected. In *SLC7A2*, there were two transcripts suggested by RNAseq with a difference in length of 3 bases (with lengths of 240 and 243), this 3 base difference was confirmed by fluorescent PCR (with lengths of 246 and 249). *SLC8A1* showed no overlap of length between the methods.

Discussion

In this study, we have characterized mRNA transcript levels and sequence variations of membrane transporters by application of high-throughput RNAseq to human autopsy tissue samples taken from liver and brain (PFC). The results reveal expected differences in expression between the two tissues, formation of splice variants, and SNPs. The study represents a first pass analysis of the RNAseq data, providing examples of new splice variants and potential regulatory variants affecting mRNA expression and processing. The RNAseq data will be available through the Sequence Read Archive for others to extend the bioinformatics analysis and integrate with additional databases.

Cellular environment profoundly affects gene regulation, so that analysis of human liver and brain tissue sections includes the presence of diverse cell types. We expect significant differences between brain regions, but have selected here to focus on PFC tissues only. Nevertheless, the main transporters expressed in each tissue will be represented in this analysis, whereas extensive expression of select transporters in specific cells (for example those maintaining the blood-brain barrier) will be under-represented relative to ubiquitously expressed transporters. Sequencing to high read density is therefore important to detect relevant mRNA transcripts expressed in only a fraction of the cells.

Overall mRNA levels in brain and liver

Comparison with other available mRNA expression profiles indicates that the results reported here are compatible with earlier results, whereas substantial differences can occur with respect to specific genes. Comparison with a set of 80 mRNAs measured in our laboratory yielded a correlation of $r = 0.74$ and a set of 10 transporter genes yielded a correlation of $r = 0.94$ in both the liver and brain separately, showing that the RNAseq does yield reproducible results for the majority of genes. A rank order of transporter mRNA

expression in PFCs and liver documents the dramatic differences between the two tissues (Supplementary Table 1, (<http://links.lww.com/FPC/A578>). Of 8 genes tested for alternative splicing, *SLC22A1* showed clear evidence to support detection of transcripts that were not previously annotated. In the other cases, detection may not be completely confirmed because of low expression levels of selected alternative transcripts or low expression specifically in the samples chosen for exploration.

The NuGen preparation of cDNA provides similar relative quantitation compared with poly-dT primed cDNA as reported [12], but also yields numerous reads in introns and intergenic regions. We have performed a first analysis of lincRNAs, finding a few such transcripts near transporter gene loci, but we need to continue the analysis genomewide to address relevant questions such as any correlation between lincRNA and mRNA expression. Also, differences between RNA expression profiles between cocaine abusers and controls were not addressed here because the detectable transporter mRNA in PFCs did not differ substantially.

Pharmacogenomic relevant SNPs

Robustly detected SNPs were found in a number of genes relevant to pharmacogenomics, including *SLC1A2*, *ABCC2*, *SLCO1B1*, and *SLC22A1*. In some cases, specific SNPs were detected (or variants in high linkage disequilibrium with them), which have been previously implicated in diseases or treatment outcomes, including rs1042113 and rs1570216 in *SLC1A2*, rs3740066 in *ABCC2*, rs34059508 and rs683369 in *SLC22A1*, and rs4149056 in *SLCO1B1*.

SLC1A2 clears glutamate from the synapse. Abnormal activation of glutamate receptors is thought to play a role in numerous central nervous system disorders. GWAS results suggest that *SLC1A2* modulates response to antidepressants, but *P* values do not pass genomewide multiple testing correction [14]. Microarray analysis of individuals with major depressive disorder shows significantly lowered expression of *SLC1A2* mRNA [15]. SNP rs1570216 was included in a study of Deng *et al.* [16] to identify polymorphisms in *SLC1A2* associated with schizophrenia. A SNP located in intron 1 of *SLC1A2* was found to be in high linkage disequilibrium ($r^2 = 0.8$) with a novel SNP contained on a haplotype that was significantly associated with schizophrenia (Bonferroni's corrected, $P = 0.0052$). An additional SNP detected by RNAseq, rs1042113, also had been implicated in schizophrenia [16]. Molecular genetic studies are ongoing to test whether any of these variants has regulatory functions.

RNAseq also indicated the presence of two possible regulatory variants in *SLC22A1*, a cation transporter (also referred to as *OCT1*). SNP rs683369 is a missense mutation that was nominally associated with positive response to metformin in the treatment of diabetes [17] SNP rs34059508 was found to reduce the uptake of metformin and MPP⁺ in transfected cells [18,19] and change pharmacokinetic properties leading to higher plasma concentration, higher area under the plasma concentration-time curve, and lower volume of distribution resulting in reduced function of metformin [20].

The hepatic anion carrier *SLCO1B1* recognizes a number of drug substrates. The nonsynonymous SNP rs4149056 has been associated with reduced protein function,

lowering access of the drug to its target in hepatocytes while increasing exposure of peripheral muscles, a main target of statin toxicity [21–35]. However, the underlying mechanism had not yet been fully explored. Another nonsynonymous SNP, rs2306283, had been associated with decreased plasma concentration–time curve of pravastatin [36,37], implying a gain of function. Interactions between rs4149056 and rs2306283 may determine genetic effects on methotrexate pharmacokinetics. The rs2306283 minor allele was found to increase clearance of methotrexate only when adjusting for rs4149056 [37]. Although a number of *SLCO1B1* variants are thought to affect transporter function [38], the genetic regulation of *SLCO1B1* expression has been poorly defined. The presented results will guide further molecular genetic studies.

ABCC2 is a multidrug resistance protein that actively transports amphipathic anionic substrates across hepatocytes [39]. SNP calls were made in 17 locations in *ABCC2*. Two of the identified SNPs have been implicated in disease and drug response – rs3740066 and rs8187710. rs3740066 had been reported to increase risk of intrahepatic cholestasis of pregnancy four-fold [39]. Our results suggest that this synonymous SNP could play a regulatory role in the expression of *ABCC2*. Nguyen *et al.* [40] observed a modestly significant preference for the C allele mRNA of rs3740066 based on allele-specific expression analysis, supporting the evidence for regulatory mechanisms. The nonsynonymous SNP rs8187710 is associated with a higher lopinavir accumulation in HIV-treated patients [41], but regulatory processes should also be considered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by NIH Grant U01 GM092655 from the National Institute of General Medical Sciences.

References

1. DeGorter MK, Xia CQ, Yang JJ, Kim RB. Drug transporters in drug efficacy and toxicity. *Annu Rev Pharmacol Toxicol.* 2012; 52:249–273. [PubMed: 21942630]
2. Franke RM, Gardner ER, Sparreboom A. Pharmacogenetics of drug transporters. *Curr Pharm Des.* 2010; 16:220–230. [PubMed: 19835554]
3. Huang Y, Anderle P, Bussey KJ, Barbacioru C, Shankavaram U, Dai Z, et al. Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer Res.* 2004; 64:4294–4301. [PubMed: 15205344]
4. Jablonski KA, McAteer JB, de Bakker PI, Franks PW, Pollin TL, Hanson RL, et al. Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. *Diabetes.* 2010; 59:2672–2681. [PubMed: 20682687]
5. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet.* 2005; 76:8–32. [PubMed: 15549674]
6. Sadee W, Wang D, Papp AC, Pinsonneault JK, Smith RM, Moyer RA, et al. Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. *Clin Pharmacol Ther.* 2011; 89:355–365. [PubMed: 21289622]

7. Heap GA, Yang JH, Downes K, Healy BC, Hunt KA, Bockett N, et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet.* 2010; 19:122–134. [PubMed: 19825846]
8. Moyer RA, Wang D, Papp AC, Smith RM, Duque L, Mash DC, et al. Intronic polymorphisms affecting alternative splicing of human dopamine D2 receptor are associated with cocaine abuse. *Neuropsychopharmacology.* 2011; 36:753–762. [PubMed: 21150907]
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. [PubMed: 19505943]
10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38:e164. [PubMed: 20601685]
11. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
12. Tariq MA, Kim HJ, Jejelowo O, Pourmand N. Whole-transcriptome RNAseq analysis from minute amount of total RNA. *Nucleic Acids Res.* 2011; 39:e120. [PubMed: 21737426]
13. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011; 25:1915–1927. [PubMed: 21890647]
14. Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S, et al. A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry.* 2009; 66:966–975. [PubMed: 19736353]
15. Choudary PV, Molnar M, Evans SJ, Tomita H, Li JZ, Vawter MP, et al. Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression. *Proc Natl Acad Sci USA.* 2005; 102:15653–15658. [PubMed: 16230605]
16. Deng X, Shibata H, Ninomiya H, Tashiro N, Iwata N, Ozaki N, et al. Association study of polymorphisms in the excitatory amino acid transporter 2 gene (SLC1A2) with schizophrenia. *BMC Psychiatry.* 2004; 4:21. [PubMed: 15296513]
17. Johnson AD, Wang D, Sadee W. Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacol Ther.* 2005; 106:19–38. [PubMed: 15781120]
18. Shu Y, Brown C, Castro RA, Shi RJ, Lin ET, Owen RP, et al. Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. *Clin Pharmacol Ther.* 2008; 83:273–280. [PubMed: 17609683]
19. Shu Y, Leabman MK, Feng B, Mangravite LM, Huang CC, Stryke D, et al. Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. *Proc Natl Acad Sci USA.* 2003; 100:5902–5907. [PubMed: 12719534]
20. Shu Y, Sheardown SA, Brown C, Owen RP, Zhang S, Castro RA, et al. Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J Clin Invest.* 2007; 117:1422–1431. [PubMed: 17476361]
21. Voora D, Shah SH, Spasojevic I, Ali S, Reed CR, Salisbury BA, et al. The SLCO1B1 *5 genetic variant is associated with statin-induced side effects. *J Am Coll Cardiol.* 2009; 54:1609–1616. [PubMed: 19833260]
22. Nishizato Y, Ieiri I, Suzuki H, Kimura M, Kawabata K, Hirota T, et al. Polymorphisms of OATP-C (SLC21A6) and OAT3 (SLC22A8) genes: consequences for pravastatin pharmacokinetics. *Clin Pharmacol Ther.* 2003; 73:554–565. [PubMed: 12811365]
23. Pasanen MK, Miettinen TA, Gylling H, Neuvonen RJ, Niemi M. Polymorphism of the hepatic influx transporter organic anion transporting polypeptide 1B1 is associated with increased cholesterol synthesis rate. *Pharmacogenet Genomics.* 2008; 18:921–926. [PubMed: 18794729]
24. SEARCH Collaborative GroupLink E, Parish S, Armitage J, Bowman L, Heath S, et al. SLCO1B1 variants and statin-induced myopathy - a genomewide study. *N Engl J Med.* 2008; 359:789–799. [PubMed: 18650507]
25. Donnelly LA, Doney AS, Tavendale R, Lang CC, Pearson ER, Colhoun HM, et al. Common nonsynonymous substitutions in SLCO1B1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: a go-DARTS study. *Clin Pharmacol Ther.* 2011; 89:210–216. [PubMed: 21178985]

26. Johnson AD, Kavousi M, Smith AV, Chen MH, Dehghan A, Aspelund T, et al. Genome-wide association meta-analysis for total serum bilirubin levels. *Hum Mol Genet.* 2009; 18:2700–2710. [PubMed: 19414484]
27. Toms TE, Smith JP, Panoulas VF, Douglas KM, Saratzis AN, Kitas GD. Prevalence of risk factors for statin-induced myopathy in rheumatoid arthritis patients. *Musculoskeletal Care.* 2010; 8:2–9. [PubMed: 19642078]
28. Trevino LR, Shimasaki N, Yang W, Panetta JC, Cheng C, Pei D, et al. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol.* 2009; 27:5972–5978. [PubMed: 19901119]
29. Buch S, Schafmayer C, Volzke H, Seeger M, Miquel JF, Sookoian SC, et al. Loci from a genome-wide analysis of bilirubin levels are associated with gallstone risk and composition. *Gastroenterology.* 2010; 139:1942.e2–1951.e2. [PubMed: 20837016]
30. Brunham LR, Lansberg RJ, Zhang L, Miao F, Carter C, Hovingh GK, et al. Differential effect of the rs4149056 variant in SLCO1B1 on myopathy associated with simvastatin and atorvastatin. *Pharmacogenomics J.* 2012; 12:233–237. [PubMed: 21243006]
31. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol Rev.* 2011; 63:157–181. [PubMed: 21245207]
32. Aklillu E, Mugusi S, Ngaimisi E, Hoffmann MM, König S, Ziesenitz V, et al. Frequency of the SLCO1B1 388A >G and the 521T >C polymorphism in Tanzania genotyped by a new LightCycler(R)-based method. *Eur J Clin Pharmacol.* 2011; 67:1139–1145. [PubMed: 21630030]
33. Martin NG, Li KW, Murray H, Putt W, Packard CJ, Humphries SE. The effects of a single nucleotide polymorphism in SLCO1B1 on the pharmacodynamics of pravastatin. *Br J Clin Pharmacol.* 2012; 73:303–306. [PubMed: 21851379]
34. Santos PC, Soares RA, Nascimento RM, Machado-Coelho GL, Mill JG, Krieger JE, et al. SLCO1B1 rs4149056 polymorphism associated with statin-induced myopathy is differently distributed according to ethnicity in the Brazilian general population: Amerindians as a high risk ethnic group. *BMC Med Genet.* 2011; 12:136. [PubMed: 21992719]
35. Akao H, Polisecki E, Kajinami K, Trompet S, Robertson M, Ford I, et al. Genetic variation at the SLCO1B1 gene locus and low density lipoprotein cholesterol lowering response to pravastatin in the elderly. *Atherosclerosis.* 2012; 220:413–417. [PubMed: 22189199]
36. Mwinyi J, John A, Bauer S, Roots I, Gerloff T. Evidence for inverse effects of OATP-C (SLC21A6) 5 and 1b haplotypes on pravastatin kinetics. *Clin Pharmacol Ther.* 2004; 75:415–421. [PubMed: 15116054]
37. Ramsey LB, Bruun GH, Yang W, Trevino LR, Vattathil S, Scheet P, et al. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res.* 2012; 22:1–8. [PubMed: 22147369]
38. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol Rev.* 2011; 63:157–181. [PubMed: 21245207]
39. Sookoian S, Castano G, Burgueno A, Gianotti TF, Pirola CJ. Association of the multidrug-resistance-associated protein gene (ABCC2) variants with intrahepatic cholestasis of pregnancy. *J Hepatol.* 2008; 48:125–132. [PubMed: 17997497]
40. Nguyen TD, Markova S, Liu W, Gow JM, Baldwin RM, Habashian M, et al. Functional characterization of ABCC2 promoter polymorphisms and allele-specific expression. *Pharmacogenomics J.* 2012; 1–7. [Epub ahead of print].
41. Elens L, Yombi JC, Lison D, Wallemacq P, Vandercam B, Haufroid V. Association between ABCC2 polymorphism and lopinavir accumulation in peripheral blood mononuclear cells of HIV-infected patients. *Pharmacogenomics.* 2009; 10:1589–1597. [PubMed: 19842932]

Fig. 1. Read alignments across *SLC38A2*, using the Integrative Genomic Viewer. Read alignments for four liver samples are shown. Although peaks locate over the annotated exonic regions, peaks also occur consistently within the intronic regions. The range of the histogram is between 0 and 1000 reads for each sample. The height of the peak indicates the number of reads covering a base location. Exon locations are indicated as gray boxes along the line at the bottom of the figure.

Fig. 2. Correlation of expression estimated by RNAseq and RT-PCR. Plot of the $\log_{10}$ transform of the average measurement of expression by RT-PCR and RNAseq is shown. Brain and liver are plotted separately. Comparing each tissue separately yielded a correlation of $r=0.94$ for each.

Fig. 3.

(a) Transcripts assembled by Cufflinks for *SLC22A1*. Transcripts assembled by Cufflinks for *SLC22A1*, limited to unique transcripts highlighting the major differences such as extended 3'-UTR or missing exons are shown, (b) Sizes of PCR products detected from *SLC22A1*. Shown are the peaks detected from fluorescent PCR of sequences from *SLC22A1*. Peaks with heights greater than 800 were detected at 1 04, 207, 221, and 321, corresponding to transcripts including or excluding of exons 9 and 10, consistent with the transcript predicted from RNAseq (Fig. 3a). UTR, untranslated region.

Table 1

Mapping statistics and SNP counts for sequenced samples

Samples	Million reads mapped	% forward mapped	% reverse mapped	% reads exonic	Genome-wide SNPs	SNPs in transporters	Sex	Age
MB079	197	71	58	21	24 889	429	Male	28
MB093	225	74	57	22	32 690	557	Female	32
MB043	37	75	-	21	12 902	233	Male	39
MB044	38	78	-	21	13 824	244	Male	25
MB046	36	76	-	12	14 633	266	Female	42
MB065	40	74	-	26	14199	288	Male	28
MB067	44	74	-	19	18 019	296	Male	35
MB081	37	78	-	14	12 839	232	Female	35
MB094	56	59	-	19	21 736	361	Male	34
MB095	37	77	-	23	14 064	274	Male	39
LV065	166	57	52	20	34 967	1,064	Male	54
LV078	118	62	57	31	17271	461	Male	72
LV115	90	54	50	24	14418	456	Male	44
LV121	155	60	55	19	35 693	800	Male	68
LV034	29	79	-	24	12 024	353	Male	78
LV056	22	81	-	23	7940	244	Male	69
LV060	47	55	-	22	17 022	452	Male	79
LV063	18	80	-	22	6483	206	Female	75
LV082	24	77	-	28	7941	246	Male	71
LV090	19	80	-	25	7695	189	Female	78
LV091	21	79	-	17	7387	300	Male	28
LV101	19	73	-	16	6365	212	Male	81

Number of reads mapped, mapping percentage of total generated reads, number of detected SNPs in all genes, and number of detected SNPs in transporters genes are listed. Separate mapping percentages are listed for forward and reverse reads.

SNP, single-nucleotide polymorphism.

Table 2

Distribution of SNP locations

Location of SNP	%	Number of SNPs
Protein Coding total	5.8	243
Nonsynonymous	2.9	121
Synonymous	2.7	115
Stopgain	0.1	4
Splicing synonymous	0.1	3
Splicing nonsynonymous	0.05	2
3'-UTR	13	556
5'-UTR	0.5	23
Intronic	60	2532
Splicing	0.3	14
Noncoding RNA	0.8	33
Downstream	1.6	66
Upstream	0.2	10
Intergenic	18	737

Placement of SNPs was made based on UCSC annotation. Protein coding SNPs are divided into synonymous, nonsynonymous, those located within two bases of a splice site (whether synonymous or nonsynonymous), and those in UTRs. SNPs listed as 'splicing' alone are located on the intronic side of a splice site. An assignment of upstream or downstream is based on a location within 1 kb of the transcribed region.

SNP, single-nucleotide polymorphism; UTR, untranslated region.

Table 3

Number of alternative splice events per gene

Gene	Count of alternative junctions
<i>ABCA6</i>	11
<i>SLC4A10</i>	11
<i>ABCB4</i>	9
<i>SLC11A2</i>	8
<i>SLC38A2</i>	8
<i>SLC38A6</i>	8
<i>SLC2A2</i>	6
<i>SLC39A8</i>	5
<i>ABCA1</i>	5
<i>SLC33A1</i>	5
<i>SLC5A6</i>	5
<i>SLC3A2</i>	5
<i>SLC7A2</i>	5
<i>SLC16A4</i>	5
<i>ABCC2</i>	5
<i>ABCF1</i>	5
<i>SLC8A3</i>	5
<i>ABCC9</i>	5
<i>SLC01C1</i>	5
<i>ABCA9</i>	5
<i>ABCA10</i>	4
<i>ABCA5</i>	4
<i>SLC30A9</i>	4
<i>SLC9A8</i>	4
<i>SLC12A2</i>	4
<i>SLC22A10</i>	4
<i>SLC38A9</i>	4
<i>SLC39A10</i>	4
<i>ABCD3</i>	4
<i>SLC25A13</i>	4
<i>SLC35B3</i>	3
<i>SLC37A4</i>	3
<i>SLC30A7</i>	3
<i>SLC30A6</i>	3
<i>SLC27A2</i>	3
<i>SLC25A3</i>	3
<i>SLC25A26</i>	3
<i>SLC20A2</i>	3
<i>SLC29A1</i>	3

Gene	Count of alternative junctions
<i>SLC12A6</i>	3
<i>SLC10A7</i>	3
<i>ABCE1</i>	3
<i>SLC22A1</i>	3
<i>SLC39A6</i>	3
<i>SLC8A1</i>	3
<i>SLC6A6</i>	3
<i>SLC4A4</i>	3
<i>SLC9A6</i>	3
<i>SLC41A2</i>	3
<i>SLC6A15</i>	3
<i>SLCO1B1</i>	3
<i>SLC39A14</i>	3
<i>SLC4A8</i>	3
<i>SLC4A7</i>	3

Transporter gene containing three or more alternative splicing events are listed in the order of the number of observed alternative splicing events. Number of samples per event can be found in Supplementary Table 3 (<http://links.lww.com/FPC/A580>).