Title
Transcriptional and Post-transcriptional Regulation of ATP-binding Cassette Transporter Expression

Author
Chhibber, Aparna

Publication Date
2014

Peer reviewed|Thesis/dissertation

Transcriptional and Post-transcriptional Regulation of ATP-binding Cassette Transporter Expression

by

Aparna Chhibber

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

[redacted]

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Deanna Kroetz. More than just a research advisor, Deanna has clearly made it a priority to guide her students to become better scientists, and I am grateful for the countless hours she has spent editing papers, developing presentations, discussing research, and so much more. I would not have made it this far without her support and guidance.

My thesis committee has provided valuable advice through the years. Dr. Nadav Ahituv in particular has been a source of support from my first year in the graduate program as my academic advisor, qualifying exam committee chair, and finally thesis committee member. Dr. Kathy Giacomini graciously stepped in as a member of my thesis committee in my 3rd year, and Dr. Steven Brenner provided valuable input as thesis committee member in my 2nd year.

My labmates over the past five years have been incredible colleagues and friends. Dr. Svetlana Markova first welcomed me into the lab and taught me numerous laboratory techniques, and has always been willing to act as a sounding board. Michael Martin has been my partner-in-crime in the lab from the beginning, and has made my days in lab fly by. Dr. Yingmei Lui has made the lab run smoothly, and has always been willing to jump in to help me at a moment's notice. All the graduate students, post-docs, and visiting scientists and students who have come through the lab in my time here have helped me along in innumerable ways.

I have also been lucky enough to collaborate with a number of incredible scientists in my time at UCSF, without whom the work presented here would not have been possible. Dr. Sook Wah has been a champion of my work from the beginning, and has been an incredible collaborator in numerous projects. Courtney French and Eric Gamazon, along with many others, made the RNA-Seq project possible, and Joel Mefford has provided valuable advice on many statistical questions through the years.

Finally, I would like to acknowledge my funding. The work presented in this thesis was funded by NIGMS Predoctoral Training Grant 2 T32 GM07175 and the Pharmacogenetics of Membrane Transporters project as a part of the Pharmacogenetics Research Network.

Abstract

ATP binding cassette (ABC) transporters are a family of proteins whose activity is vital to cell detoxification, protection against xenobiotics and oxidative stress, and maintenance of homeostasis of endogenous compounds. Subtle changes in endogenous expression level of these transporters can have significant clinical implications. In particular, the impact of such variation on the role of ABC transporters in transport of pharmaceutical agents is of interest; inter-individual differences in expression of ABC transporters can result in changes in exposure to pharmaceutical agents or their metabolites, leading to altered drug efficacy or drug-induced toxicities. Variation in gene expression can be caused by a number of factors that modulate the “normal” activity during transcription or translation. In this study, mechanisms that regulate mRNA or protein expression of ATP-binding cassette transporters were characterized in human tissues. First, the impact of DNA sequence variation on mRNA expression across individuals was evaluated in the human kidney. Several expression quantitative trait loci (eQTLs) were identified for ABC transporters in the kidney, and one eQTL for *ABCG2* (BCRP) was validated in an *in vitro* reporter-gene assay. Next, the role of alternative splicing in regulation of transporter expression was explored using transcriptome sequencing data from multiple individuals in several human tissues. Examples of splicing events that were found to modulate transporter expression across individuals include alternate 5 prime untranslated regions (5’ UTRs) in the genes *ABCC5* and *ABCA8.* Further, a splicing event in the gene *ABCC6* was identified that produces a premature termination codon, triggering the nonsense mediated

decay process; this event may be responsible for regulating inter-tissue expression of *ABCC6*. Finally, transcription factor regulators of ABC transporters were identified by searching for transcription factor binding motif enrichment in sets of genes co-expressed with ABC transporters in several human tissues. A number of potential transcriptional regulators of transporters were identified.

Table of Contents

List of Figures

List of Tables

Chapter 1: Background and Introduction

1.1 ATP-binding cassette transporters

ATP-Binding Cassette (ABC) transporters are a family of proteins responsible for the transport of both endogenous compounds and drugs across cell and organelle membranes. As a family, the ABC transporters play a central role in homeostasis and protection of cells in most organisms, including both prokaryotes and eukaryotes[1,2].

All ABC transporters contain at least one nucleotide binding domain (NBD) essential for its function, and most contain multiple transmembrane domains (TMDs) which allow the transporter to become embedded in a lipid bilayer. In humans, the ABC transporter family can be further divided into seven subfamilies. The general functions and structure of each subfamily are described in Table 1.1.

Table 1.1: Structure and Function of Human ABC Transporter Subfamilies

Subfamily	Function	Structure[1]	# of Genes
ABCA	Cholesterol and lipid transport	2 NBDs/2 TMDs	12
ABCB	Transport of peptides (antigens), iron, bile salts	2 NBDs/2 TMDs 1 NBD/1 TMD	4 7
ABCC	Transport of ions, anions, cyclic nucleotides	2 NBDs/2 TMDs	12
ABCD	Very long chain fatty acid transport in peroxisomes	1 NBD/1 TMD	4
ABCE/F	interferon action/translation initiation	2 NBDs	1/3
ABCG	Steroid and cholesterol transport	1 NBD/1 TMD	5

[1]NBD: nucleotide binding domain; TMD: transmembrane domain

Given the essential endogenous functions of ABC transporters described above, it is of no surprise that disruption of normal ABC transporter function can have

significant clinical implications. Mutations in ABC transporters cause a number of hereditary genetic disorders, including psuedoxanthoma elasticum, caused by mutations in *ABCC6*[3], cystic fibrosis caused by mutations in *ABCC7*[4]*?,* Tangier's disease caused by mutations in *ABCA1*[5], and Dubin-Johnson syndrome caused by mutations in *ABCC2*[6]. Polymorphisms in ABC transporters have also been associated with more common diseases and clinical measurements, including gout with polymorphisms in *ABCG2*[7], high density lipoprotein levels with polymorphisms in *ABCA1*[8]*,* and platelet levels with polymorphisms in *ABCC4*[9].

ABC transporters are also of particular interest in human health for their role in transporting drugs; in this capacity they play a key role in drug disposition. Transporters in the ABCA, ABCB, ABCC, and ABCG families have been shown to transport drugs. Changes in function or activity of ABC transporters have been associated with drug response, toxicity, and pharmacokinetic parameters[10,11]. While changes in transporter function or activity may be caused by disruption of protein or transcript structure, they may also be driven by changes in expression. For example, small differences in the expression of ABCA1*,* a cholesterol transporter, between individuals are associated with systematic changes in plasma high density lipoprotein (HDL) levels and thus increased risk of coronary artery disease[12].

1.2 Regulation of ABC transporter expression

Variation in ABC transporter expression can occur at the mRNA or protein level. Changes in transcript or protein levels may be driven by variation in transcription or translation initiation or changes in stability of transcripts or proteins.

1.2.1 Transcriptional regulation of ABC transporter expression

Transcription in eukaryotes is initiated by recruitment of RNA polymerase II (Pol II) to promoters (immediately 5' of the gene). The presence of a TATA box directly upstream (25-30 bases) of the transcription start site allows for binding of the TATA-binding protein (TBP), ultimately allowing for recruitment of Pol II. However, most ABC transporter promoters do not contain the common TATA-box[13]. Alternate transcription initiators, such as the presence of a specific initiator (Inr) sequence[14], have been identified in select transporter promoters[13,15]. Further, like most TATAless promoters, many transporter promoters contain CCAAT boxes[16,17], binding sites for the NF-Y transcription factor[18], and GC-rich elements, which bind Sp family transcription factors[19–24]. Other transcription factors are not constitutively active, but rather must be activated by phosphorylation in response to signaling cascades. For example, several transporters in the ABCC (MRP) family may be regulated by activator protein 1 (Ap-1)[25,26], a complex made up of c-Fos, c-Jun, ATF, and JDP families. Likewise, *ABCC1*[27] and *ABCG2*[28] are regulated by Notch signaling.

Transporters are also regulated by a number of inducible transcription factors, whose activation is dependent on an intra- or extra- cellular ligand. Binding sites

for these inducible transcription factors may be located in the promoter or many kilobases away as a part of distal transcriptional enhancers or silencers. In particular, nuclear receptor dimers regulate expression of a number of ABC transporters[13] in response to extracellular ligands. Retinoic acid responsive retinoic acid receptor (RAR) and retinoid X receptor (RXR) can form heterodimers with other nuclear receptors or with each other. RARα/RXRα heterodimers activate expression of ABCG1[29,30], ABCG2[31], and ABCC2[32]. Further, RXRα is involved in the regulation of expression of ABCC6[33], and RARα is involved in regulation of expression of ABCB1[34,35] and ABCA1[29,30]. Liver X receptor (LXR) is a key regulator of cholesterol, fatty acid, and glucose homeostasis and plays a role in regulating expression of other ABC transporters that are important in these processes. For example, LXRα, induced by oxysterols[36,37] (e.g. (22R)-Hydroxycholesterol), dimerizes with RXRα to regulate expression of ABC transporters such as *ABCA1* [38]. Pregnane X receptor (PXR) plays a role in regulating gene expression in response to the presence of foreign substances, and is important in regulation of the expression of ABC transporters involved in detoxification. For example, PXR, induced by a range of endogenous and exogenous compounds, including steroids, antibiotics, antimycotics, bile acids, hyperforin and others[39], regulates expression of ABCC2[40,41], ABCB1[42,43] and ABCC3[44]. Additional nuclear receptors, such as the peroxisome proliferator-activated receptor (PPAR) and hepatocyte nuclear factor (HNF) families, also play an important role in regulation of ABC transporter expression. As heterodimers with RXRα, PPARs regulate transcription of genes involved in fat

and glucose metabolism, including the ABC transporters ABCA1[45], ABCB1[46,47], ABCA12[48], ABCB4[49], and ABCG1-2[50–52]. Ligands of PPARs include fatty acids, eicosanoids, and a number of drugs[53]. HNFs regulate a broad range of genes, and while they are expressed at highest concentration in the liver, they can regulate transcription in other tissues as well. HNF4 in particular has been implicated in regulation of ABC transporters, including ABCA1[54], ABCG5/8[55], and ABCC2[56]. Other ligand inducible transcription factors involved in ABC transporter expression include the sterol regulatory element-binding proteins (SREBP), a family of sterol responsive transcription factors that regulate expression of transporters involved in cholesterol biosynthesis or efflux transporters, such as ABCA1[57].

Any change in the ability or availability of transcription factors to bind to DNA can modulate gene expression[58]. For example, variation in DNA sequence can modulate the binding of DNA and RNA binding factors that regulate transcription initiation or stability. Such changes in gene expression caused by polymorphisms can be evaluated via correlative genotype–expression studies, looking for statistically significant correlations between a variant and gene expression level in a population. For example, the *ABCG2* promoter variant –15994C>T is associated with higher mRNA expression in multiple human tissue types [59]. Likewise, promoter polymorphisms in the transporters *ABCA1* and *ABCD2* were associated with transporter expression in lymphoblastoid cell lines[60]. With the large number of genotype-expression studies that have been conducted across human tissue types, many such examples have been identified. A selection of

published polymorphisms associated with ABC transporter mRNA expression is included in Table 1.2.

Table 1.2: Polymorphisms Associated with ABC Transporter mRNA Expression

Tissue[1]	Transporter	Polymorphism	Reference
LCL	*ABCA1*	rs1800977	[60]
LCL	*ABCD2*	rs4072006	[60]
Whole Blood	*ABCA1*	rs2102121	[61]
Tibial Artery	*ABCA1*	rs2018930	[61]
Tibial Nerve	*ABCA2*	rs9410155	[61]
Whole Blood	*ABCA3*	rs4785914	[61]
Skeletal Muscle	*ABCA4*	rs7413585	[61]
Lung	*ABCA5*	rs7211418	[61]
Skeletal Muscle	*ABCA5*	rs7211418	[61]
Tibial Artery	*ABCA5*	rs7211418	[61]
Whole Blood	*ABCA5*	rs7211418	[61]
Lung	*ABCA6*	rs2716226	[61]
Lung	*ABCA7*	rs3087680	[61]
Thyroid	*ABCA8*	rs2716185	[61]
Sun Exposed Skin	*ABCA8*	chr17:66959267:D	[61]
Thyroid	*ABCA12*	rs6736127	[61]
Sun Exposed Skin	*ABCA13*	rs71549803	[61]
Skeletal Muscle	*ABCB1*	rs4728702	[61]
Tibial Artery	*ABCB1*	chr7:87735867:D	[61]
Sun Exposed Skin	*ABCB2*	rs147942062	[61]
Thyroid	*ABCB2*	rs114251877	[61]
Lung	*ABCB3*	rs2621323	[61]
Thyroid	*ABCB3*	rs2621323	[61]
Tibial Nerve	*ABCB3*	rs2621323	[61]
Heart (Ventricle)	*ABCB3*	rs241443	[61]
Sun Exposed Skin	*ABCB3*	rs2621323	[61]
Tibial Artery	*ABCB4*	rs4148812	[61]
Tibial Nerve	*ABCB4*	rs4148812	[61]
Whole Blood	*ABCB6*	rs11892048	[61]
Skeletal Muscle	*ABCB6*	rs34393607	[61]
Lung	*ABCB8*	rs6947821	[61]
Skeletal Muscle	*ABCB8*	rs6947821	[61]
Thyroid	*ABCB8*	rs2288652	[61]
Heart (Ventricle)	*ABCB9*	rs7296418	[61]
Sun Exposed Skin	*ABCB9*	rs7964876	[61]
Thyroid	*ABCB9*	rs28557411	[61]
Lung	*ABCB10*	rs476609	[61]
Sun Exposed Skin	*ABCB11*	rs853789	[61]
Thyroid	*ABCB11*	rs13387523	[61]
Whole Blood	*ABCC1*	rs35605618	[61]
Tibial Artery	*ABCC3*	rs9303560	[61]
Tibial Nerve	*ABCC3*	rs9303560	[61]
Skeletal Muscle	*ABCC3*	rs8070592	[61]
Whole Blood	*ABCC3*	chr17:48812872:I	[61]
Thyroid	*ABCC4*	chr13:95939546:I	[61]
Whole Blood	*ABCC4*	chr13:95874467:I	[61]
Skeletal Muscle	*ABCC4*	rs12865420	[61]
Tibial Nerve	*ABCC4*	rs7998388	[61]
Heart (Ventricle)	*ABCC4*	rs17189568	[61]
Whole Blood	*ABCC5*	rs7632670	[61]
Tibial Nerve	*ABCC5*	rs7632824	[61]
Adipose	*ABCC6*	rs12933479	[61]
Skeletal Muscle	*ABCC6*	rs118037182	[61]
Tibial Nerve	*ABCC6*	rs1377549	[61]
Sun Exposed Skin	*ABCC6*	rs11642570	[61]
Thyroid	*ABCC6*	rs212070	[61]
Whole Blood	*ABCC6*	rs3784870	[61]
Heart (Ventricle)	*ABCC7*	rs34654194	[61]
Tibial Artery	*ABCC8*	rs77889556	[61]
Heart (Ventricle)	*ABCC8*	rs985136	[61]
Lung	*ABCC8*	rs67445779	[61]
Skeletal Muscle	*ABCC8*	rs985136	[61]
Tibial Nerve	*ABCC8*	rs2237967	[61]
Sun Exposed Skin	*ABCC8*	rs77889556	[61]

Thyroid	*ABCC8*	rs77889556	[61]
Whole Blood	*ABCC9*	rs73070702	[61]
Skeletal Muscle	*ABCC10*	rs4530855	[61]
Lung	*ABCC13*	rs11910187	[61]
Thyroid	*ABCD3*	chr1:94878641:D	[61]
Tibial Nerve	*ABCD3*	rs587552	[61]
Whole Blood	*ABCD3*	rs11585564	[61]
Lung	*ABCD4*	rs17098989	[61]
Skeletal Muscle	*ABCE1*	chr4:146579204:D	[61]
Tibial Artery	*ABCF1*	rs141237570	[61]
Tibial Nerve	*ABCF1*	rs141237570	[61]
Adipose	*ABCF1*	rs112357947	[61]
Tibial Artery	*ABCF1*	rs141237570	[61]
Lung	*ABCF1*	rs116512495	[61]
Sun Exposed Skin	*ABCF1*	rs115769948	[61]
Thyroid	*ABCF2*	rs219233	[61]
Whole Blood	*ABCF2*	chr7:150944796:D	[61]
Skeletal Muscle	*ABCF3*	rs843360	[61]
Adipose	*ABCG1*	rs4920081	[61]
Lung	*ABCG1*	rs743542	[61]
Skeletal Muscle	*ABCG1*	rs12627125	[61]
Sun Exposed Skin	*ABCG1*	rs186531	[61]
Thyroid	*ABCG1*	rs28689993	[61]
Thyroid	*ABCG2*	chr4:89834373:D	[61]
Tibial Artery	*ABCG2*	rs34521406	[61]
Tibial Nerve	*ABCG4*	rs6589675	[61]
Thyroid	*ABCG4*	rs9667297	[61]
Adipose	*ABCG5*	rs34141057	[61]
Heart (Ventricle)	*ABCG8*	rs6749927	[61]
LCL	*ABCA2*	rs233465	[62]
Intestine	*ABCA13*	rs4130214	[63]
Intestine	*ABCA13*	rs10238679	[63]
Intestine	*ABCA13*	rs1526106	[63]
Intestine	*ABCA13*	rs7384280	[63]
Intestine	*ABCA2*	rs2784054	[63]
Intestine	*ABCA2*	rs2811787	[63]
Intestine	*ABCA2*	rs2811724	[63]
Intestine	*ABCC11*	rs8047091	[63]
Intestine	*ABCC11*	rs11076561	[63]
Intestine	*ABCF2*	rs219229	[63]
Intestine	*ABCA13*	rs6583476	[63]
Intestine	*ABCF2*	rs2608291	[63]
Intestine	*ABCF2*	rs310583	[63]
Skin	*ABCC3*	rs3785912	[64]
LCL	*ABCF2*	rs2374315	[65]
LCL	*ABCF3*	rs6766579	[65]
LCL	*ABCB9*	rs1316952	[65]
LCL	*ABCA7*	rs4807928	[65]
LCL	*ABCC1*	rs9930886	[65]
Liver	*ABCA10*	rs12938064	[66]
Liver	*ABCC11*	rs16946122	[66]
Liver	*ABCC5*	rs4973898	[66]
Liver	*ABCA5*	rs12938064	[66]
Liver	*ABCB9*	rs4275659	[66]
Liver	*ABCB2*	rs2071474	[66]
Liver	*ABCF2*	rs5174	[66]
Liver	*ABCE1*	rs11026697	[66]
Liver	*ABCG5*	rs3792009	[66]
Liver	*ABCB10*	rs1828591	[66]
Liver	*ABCA13*	rs2277849	[66]
Liver	*ABCA3*	rs877534	[66]
Liver	*ABCD3*	rs6681849	[66]
Liver	*ABCA1*	rs987870	[66]
Liver	*ABCB2*	rs241440	[66]
Brain (Cerebellum)	*ABCB2*	rs2071473	[67]
Brain (Frontal Cortex)	*ABCB2*	rs2071473	[67]
Brain (Temporal Cortex)	*ABCB2*	rs2071473	[67]
LCL	*ABCA1*	rs2472519	[67]

LCL	*ABCA1*	rs1582708	[67]
LCL	*ABCA1*	rs10991512	[67]
LCL	*ABCA1*	rs2487050	[68]
Adipose	*ABCA5*	rs12450167	[68]
LCL	*ABCA6*	rs11657804	[68]
LCL	*ABCA6*	rs6501723	[68]
LCL	*ABCA6*	rs740519	[68]
Skin	*ABCA6*	rs9911770	[68]
Adipose	*ABCA8*	rs4147950	[68]
Adipose	*ABCA8*	rs4968956	[68]
Skin	*ABCA8*	rs764425	[68]
Adipose	*ABCB3*	rs241454	[68]
LCL	*ABCB3*	rs241454	[68]
Skin	*ABCB3*	rs241454	[68]
LCL	*ABCB4*	rs17149647	[68]
LCL	*ABCB9*	rs949143	[68]
Skin	*ABCB9*	rs1727307	[68]
LCL	*ABCC1*	rs212093	[68]
Skin	*ABCC1*	rs212093	[68]
LCL	*ABCC4*	rs7335912	[68]
LCL	*ABCC5*	rs4148564	[68]
Skin	*ABCC5*	rs2139559	[68]
Adipose	*ABCC8*	rs10832794	[68]
Skin	*ABCC8*	rs4757527	[68]
LCL	*ABCC10*	rs2125739	[68]
LCL	*ABCF1*	rs2074504	[68]
Skin	*ABCF1*	rs9262145	[68]
Skin	*ABCF1*	rs3868078	[68]
Adipose	*ABCF2*	rs6464134	[68]
LCL	*ABCF2*	rs6464134	[68]
Skin	*ABCF2*	rs7786151	[68]
Skin	*ABCF3*	rs17804191	[68]
Skin	*ABCF3*	rs7653781	[68]
Adipose	*ABCG1*	rs7281720	[68]
Adipose	*ABCG1*	rs7281720	[68]
LCL	*ABCG1*	rs2839411	[68]
Fibroblast	*ABCA1*	rs2997729	[69]
Fibroblast	*ABCB1*	rs2718320	[69]
Fibroblast	*ABCB3*	rs241453	[69]
LCL	*ABCB3*	rs241453	[69]
T-Cell	*ABCB3*	rs241453	[69]
Fibroblast	*ABCB4*	rs6946119	[69]
T-Cell	*ABCD3*	rs3933012	[69]
LCL	*ABCB3*	rs4148876	[70]
Monocytes	*ABCB3*	rs2071474	[71]
Liver	*ABCB3*	rs2071474	[66]
Monocytes	*ABCA7*	rs2279796	[71]
Monocytes	*ABCA7*	rs12710103	[71]
Monocytes	*ABCC1*	rs212093	[71]
Monocytes	*ABCC3*	rs2189595	[71]
Monocytes	*ABCC3*	rs4794179	[71]
Monocytes	*ABCC4*	rs12485738	[71]
Monocytes	*ABCC4*	rs10508017	[71]
Monocytes	*ABCC4*	rs2274403	[71]
Monocytes	*ABCC4*	rs4148436	[71]
Monocytes	*ABCC4*	rs9524856	[71]
Monocytes	*ABCC5*	rs1000002	[71]
Monocytes	*ABCC5*	rs12054261	[71]
Monocytes	*ABCC5*	rs2278968	[71]
Monocytes	*ABCC5*	rs2606228	[71]
Monocytes	*ABCC5*	rs7620457	[71]
Monocytes	*ABCC5*	rs869335	[71]
Monocytes	*ABCC5*	rs9844777	[71]
Monocytes	*ABCC6*	rs2547851	[71]
Monocytes	*ABCC6*	rs2856553	[71]
Monocytes	*ABCC6*	rs4780600	[71]
Monocytes	*ABCF1*	rs6904236	[71]
Monocytes	*ABCD2*	rs11172848	[71]

Monocytes	*ABCA1*	rs2487050	[71]
LCL	*ABCD2*	rs1512943	[72]
LCL	*ABCF1*	rs13240743	[72]
Liver	*ABCF2*	rs5174	[66]
Liver	*ABCG5*	rs3792009	[66]
Liver	*ABCF3*	rs2272473	[66]
Liver	*ABCB10*	rs1828591	[66]
Liver	*ABCE1*	rs1828591	[66]
Liver	*ABCC10*	rs760370	[66]
Liver	*ABCA1*	rs987870	[66]
Liver	*ABCB3*	rs241440	[66]
Liver	*ABCB3*	rs2071474	[66]
Liver	*ABCF1*	rs9295843	[66]
Liver	*ABCE1*	rs11026697	[66]
Liver	*ABCC9*	rs4762901	[66]
Liver	*ABCA3*	rs877534	[66]
Brain	*ABCG4*	rs6823583	[73]
Liver	*ABCC11*	rs16946006	[74]
Liver	*ABCF2*	rs4726009	[74]
Liver	*ABCD4*	rs2301345	[74]
Liver	*ABCA5*	rs12941297	[74]
Liver	*ABCG2*	rs2728126	[74]
Liver	*ABCC9*	rs704215	[74]
Liver	*ABCA3*	rs17135889	[74]

[1]LCL: Lymphoblastoid Cell Line

1.2.2 Post-transcriptional and translational regulation of ABC transporter expression

Translation initiation in eukaryotes occurs when the small ribosomal subunit (40S), along with Methionine tRNA as part of the preinitiation complex (PIC), locates the initiation codon (AUG) on an mRNA transcript. The specific sequence context surrounding the start codon can regulate translation initiation[75–77]. Structural changes to the 5' untranslated region (UTR) that modify efficiency of ribosomal binding or scanning can affect translation initiation. Alternate start codons upstream (uAUG) of the true start codon can also reduce efficiency of translation. For example, multiple uAUGs are present in the *ABCC2* (MRP2) 5'UTR[78]. Inclusion of the uAUGs results in decreased expression of MRP2 [78]. Alternate transcription initiation sites in different tissues results in variable expression across tissues[78].

MicroRNAs, short (20-24 nucleotide) noncoding RNAs that bind to specific sequences, have also been shown to play a role in regulating expression of ABC transporters either by inhibiting translation by modulating recruitment of eukaryotic initiation factors (eIFs) [79] which prepare the mRNA transcript for PIC binding, or by promoting mRNA degradation[80]. For example, expression of the MRP4 transporter encoded by ABCC4 is downregulated by miR-124a and miR-506 expression in kidney tissue[81]. microRNAs can also mediate gene expression in response to particular cell conditions. For example, microRNA-379 mediates downregulation of ABCC2 in response to rifampicin exposure[82].

Changes in transcript and protein stability can modulate gene expression levels. Introduction of premature termination codons (PTCs) in a transcript can trigger the nonsense mediated decay (NMD) pathway, leading to transcript degradation[83,84]. For example, an insertion in *ABCC4* (MRP4) results in a PTC, triggering NMD; this PTC exon is conserved across species, suggesting that the insertion is a critical mechanism for regulating expression of MRP4[85]. Similarly, alternative splicing of *ABCC5* (MRP5) in the retina produces alternate transcripts with PTCs. These transcripts are targeted by NMD, and appear to be a mechanism for regulating expression of MRP5.[86] Variation in protein stability can occur when changes to protein folding[87] or localization[88] in the cell mark the protein for degradation. Sequence variation that occurs in the coding region can modify transporter stability, resulting in changes in protein levels and disruption of normal transporter function. For example, the *ABCG2* (BCRP) nonsynonomous polymorphism rs2231142 (421C>A, Gln141Lys) has been associated with decreased BCRP levels[89], possibly because of changes in stability[90] or mislocalization of the transporter[91], resulting in premature degradation. This change in expression results in reduced transport of BCRP substrates SN-38, mitoxantrone, topotecan, or diflomotecan[91,92]. Likewise, a nonsynonomous polymorphism in *ABCC4* (MRP4; rs3765534, 2269G>A) disrupts stability and membrane localization of MRP4; furthermore, patients with the 2269A variant have reduced MRP4 function[93]. Similarly, a nonsynonomous SNP in *ABCC11* disrupts stability of MRP8 protein, activating degradation and altering function of the transporter.[94] Perhaps the best recognized example of

regulation of ABC transporter expression by a coding region sequence change is the deletion of a phenylalanine (ΔF508) in the CFTR protein (*ABCC7*) which triggers premature degradation of the transporter and loss of functional CFTR; *ABCC7* ΔF508, is the cause of the vast majority of cystic fibrosis cases[95, 96].

1.3 Motivation for Thesis Research

Given the importance of ATP-binding cassette transporters in humans, and the functional impact that changes in transporter expression can have on both endogenous functions and response to drugs, understanding the mechanisms regulating transporter expression is important. Three mechanisms for regulation of transporter expression were examined:

(i) Variation in DNA sequence across individuals associated with ABC transporter expression in the human kidney.

(ii) Variation in alternative splicing across individuals and tissues in lymphoblastoid cell lines and the human kidney, liver, adipose tissue, and heart.

(iii) Transcription factor mediated variation in gene expression in lymphoblastoid cell lines and the human kidney, liver, adipose tissue, and heart.

The motivation for studying each mechanism is described in further detail below.

1.3.1 Variation in DNA sequence across individuals associated with ABC transporter expression in the human kidney

Variation in DNA sequence can modulate the binding of DNA and RNA binding factors that regulate transcription initiation or stability. While a number of large genotype-expression association, or expression quantitative trait loci (eQTL), studies have been conducted, most of these have been conducted in lymphoblastoid cell lines or a limited number of other human tissues or cell lines. In particular, for the study of ABC transporters, there is an acute lack of data available from the human kidney. The kidney is a key organ responsible for clearance of metabolic waste products and xenobiotics, and in maintaining homeostasis of endogenous compounds such as hormones and electrolytes. ABC transporters play a central role in these functions, and altered expression of ABC transporters in the kidney can result in changes in both systemic and localized (to the kidney) accumulation of xenobiotics and waste products, and imbalances in endogenous compounds. Finally, ABC transporters also appear to play a vital role in tissue regeneration following damage, especially in the kidney[97,98], and thus subtle changes in ABC transporter expression may change the propensity to long term kidney damage. Previous findings suggest that, for the same set of individuals, anywhere from 50 to 80% of eQTLs are tissue-specific[66,67,69,71,74,99–119]; thus to understand the genetic regulation of expression in the kidney, the analysis must be conducted directly in kidney tissue. Furthermore, while many eQTL studies have been conducted, identifying numerous putative expression regulatory loci, very few attempt to identify regulatory SNPs and evaluate these at a molecular level. In this study single nucleotide polymorphisms that regulate ABC transporter expression in the kidney

were identified and characterized, and proposed regulatory SNPs were further functionally validated. From these analyses, a set of high confidence variants were generated that will be valuable in understanding ABC transporter function and expression in the kidney.

1.3.2 Variation in alternative splicing across individuals and tissues in lymphoblastoid cell lines and the human kidney, liver, adipose tissue, and heart.

At least 74% of genes are alternatively spliced[120], with some estimates as high as 94%[121]; the Ensemble gene annotation documents anywhere from 1 to 28 putative isoforms for individual ABC transporters, and these are most certainly conservative estimates[122]. Most alternative splicing (AS) events occur in a tissue-specific pattern, and are a mechanism for regulating tissue specific expression[115,121,123]. However, differential expression of common isoforms across individuals can have significant biological implications as well**.** Estimates suggest that 10 to 30% of genes show differences in alternative splicing across individuals in the same tissue[121], and in the few studies that have attempted to study variation in splicing and expression at the population level, it is evident that there is significant variation in gene expression across individuals that is driven by AS[65,99,101,102,121,124]. Variation in alternative splicing can modify translation initiation through conformational changes in the 5' UTR or through changes in the sequence of 5' and 3' UTRs, or modulate stability through changes in RNA binding sequences in transcripts or the structure of the translated protein.

Early studies on alternative splicing relied on Expressed Sequence Tags (ESTs) (short cDNA sequences) and exon or splice junction arrays; however, these methods had limited resolution and in general could reliably detect only the most significant, 'switch-like' differences in expression or structure of alternate isoforms. In recent years, the introduction of whole genome transcriptome sequencing (RNA-seq) has greatly improved the ability to identify alternative splice variants[125]. Further, improvements to the technology, such as longer read lengths and paired end sequencing, have improved the accuracy and resolution of isoform identification. The available technology is now sensitive enough to enable the identification of relatively subtle differences in isoform expression between individuals and across tissues.

While a number of RNA-seq and Exon/Junction Array studies have been conducted to evaluate alternative splicing, relatively few have attempted a cross-population evaluation of alternative splicing in human cell lines and tissues. Several studies have examined alternative splicing across a panel of individuals using lymphoblastoid cell lines form Caucasian and African American populations using both RNA-seq[65,101,106] and exon/junction arrays[102,112,121,124], and one study evaluated AS across individuals in brain cortical and blood samples[115]. However, alternative splice variants are known to be highly tissue- and context–specific; for example about 50% of splicing eQTLs are specific to a single tissue type[115]. Applying the information from one tissue to another or from a cell line to a primary tissue will likely be misleading.

Further, while numerous putative alternative isoforms of ABC transporters have been identified through EST databases and microarray studies, few have been well characterized. Tissue specific expression of individual isoforms has been identified primarily by chance on a gene by gene basis in a limited number of tissues, but a detailed investigation of the extent and nature of tissue specific AS of ABC transporters is lacking.

In this study, ABC transporter splicing events were identified in four human tissues and lymphoblastoid cell lines using RNA-seq data from multiple individuals. Events that may be involved in regulation of transporter expression both across individuals and across tissues were functionally characterized.

1.3.3 Transcription factor mediated variation in gene expression in lymphoblastoid cell lines and the human kidney, liver, adipose tissue, and heart.

Variation in gene expression across tissues and across individuals is expected to be modulated primarily by changes in the initiation of transcription[126], and thus transcription factors play a central role in the regulation of gene expression. Characterizing the role that transcription factors play in modulating expression of ABC transporters across different tissues will be vital to understanding the expression patterns of the transporters across the body. Transcription factor binding sites are short, degenerate motifs that occur across the genome; given that the regulatory regions for a single gene may include many potential binding sites, it is difficult to predict functional transcription factor binding sites for a single gene. However, by searching within putative regulatory regions (such as

promoters) of a set of genes predicted to be under regulation by a common transcription factor, there is greater likelihood of successfully identifying functional binding sites.

Genes that show coordinated patterns of expression are predicted to share common transcriptional regulators of expression, and co-expression analyses use global gene expression data to gain insight into gene regulation in the context of other expressed genes. Regulatory "modules" are identified based on co-expression patterns in a dataset, and these "modules" – groups of co-expressed genes - can then be screened for enrichment of *cis*-regulatory motifs such as transcription factor binding sites (TFBS). Such methods have been applied extensively in yeast and organisms with "simple" genomes, as well as in large expression datasets in human cells and tissues[127–132]. By exploiting global expression data from a range of human tissues we can better understand the role and nature of regulation of ABC transporters by transcription factors. Further, co-expression has also been shown to be associated with common functional relationships between genes[133], and has been used to assign function to previously un-annotated genes; this analysis may provide clues to the function of ABC transporters that have not been well characterized.

While a number of large studies of gene expression in a range of human tissues have been conducted, most of these resources have not been exploited for the purpose of understanding gene expression and regulation from a global context. In addition, previous human co-expression studies have relied on microarray data, and comparisons across tissue type have often required comparisons using

datasets derived from different microarray platforms, giving highly variable and potentially misleading patterns of expression. Using a set of whole-transcriptome expression profiles across human tissues generated on a single platform with high accuracy and reproducibility is ideal for a study of this type.

In this study, sets of genes co-expressed with ABC transporters in the human kidney, liver, heart, adipose tissue, and lymphoblastoid cell lines were extracted using RNA-seq data generated under identical library preparation and sequencing conditions. Transcription factor motifs enriched in the promoters of co-expressed gene sets were identified; this analysis generated novel hypotheses about transcription factor mediated regulation of transporter expression.

1.4 References

1 Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* 2001; **42**: 1007–17.

2 Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 2001; **11**: 1156–66.

3 Bergen A, Plomp A, Schuurman E. Mutations in ABCC6 cause pseudoxanthoma elasticum. *Nat Genet* 2000; **25**: 228–31.

4 Rowntree RK, Harris A. The phenotypic consequences of CFTR mutations. *Ann Hum Genet* 2003; **67**: 471–85.

5 Rust S, Rosier M, Funke H, Real J. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat Genet* 1999; **22**: 352–5.

6 Paulusma CC, Kool M, Bosma PJ, Scheffer GL, ter Borg F, Scheper RJ, Tytgat GN, Borst P, Baas F, Oude Elferink RP. A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin-Johnson syndrome. *Hepatology* 1997; **25**: 1539–42.

7 Woodward OM, Köttgen A, Coresh J, Boerwinkle E, Guggino WB, Köttgen M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* 2009; **106**: 10338–42.

8 Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, Farlow DN, Folsom AR, Fornage M, Forrester T, Fox E, Haiman CA, Hartiala J, Harris TB, Hazen SL, Heckbert SR, Henderson BE, Hirschhorn JN, Keating BJ, Kritchevsky SB, Larkin E, Li M, Rudock ME, McKenzie CA, Meigs JB, Meng YA, Mosley TH, Newman AB, Newton-Cheh CH, Paltoo DN, Papanicolaou GJ, Patterson N, Post WS, Psaty BM, Qasim AN, Qu L, Rader DJ, Redline S, Reilly MP, Reiner AP, Rich SS, Rotter JI, Liu Y, Shrader P, Siscovick DS, Tang WHW, Taylor HA, Tracy RP, Vasan RS, Waters KM, Wilks R, Wilson JG, Fabsitz RR, Gabriel SB, Kathiresan S, Boerwinkle E. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet* 2011; **7**: e1001300.

9 Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, Serbanovic-Canic J, Elling U, Goodall AH, Labrune Y, Lopez LM, Mägi R, Meacham S, Okada Y, Pirastu N, Sorice R, Teumer A, Voss K, Zhang W,

Ramirez-Solis R, Bis JC, Ellinghaus D, Gögele M, Hottenga J-J, Langenberg C, Kovacs P, O'Reilly PF, Shin S-Y, Esko T, Hartiala J, Kanoni S, Murgia F, Parsa A, Stephens J, van der Harst P, Ellen van der Schoot C, Allayee H, Attwood A, Balkau B, Bastardot F, Basu S, Baumeister SE, Biino G, Bomba L, Bonnefond A, Cambien F, Chambers JC, Cucca F, D'Adamo P, Davies G, de Boer RA, de Geus EJC, Döring A, Elliott P, Erdmann J, Evans DM, Falchi M, Feng W, Folsom AR, Frazer IH, Gibson QD, Glazer NL, Hammond C, Hartikainen A-L, Heckbert SR, Hengstenberg C, Hersch M, Illig T, Loos RJF, Jolley J, Khaw KT, Kühnel B, Kyrtsonis M-C, Lagou V, Lloyd-Jones H, Lumley T, Mangino M, Maschio A, Mateo Leach I, McKnight B, Memari Y, Mitchell BD, Montgomery GW, Nakamura Y, Nauck M, Navis G, Nöthlings U, Nolte IM, Porteous DJ, Pouta A, Pramstaller PP, Pullat J, Ring SM, Rotter JI, Ruggiero D, Ruokonen A, Sala C, Samani NJ, Sambrook J, Schlessinger D, Schreiber S, Schunkert H, Scott J, Smith NL, Snieder H, Starr JM, Stumvoll M, Takahashi A, Tang WHW, Taylor K, Tenesa A, Lay Thein S, Tönjes A, Uda M, Ulivi S, van Veldhuisen DJ, Visscher PM, Völker U, Wichmann H-E, Wiggins KL, Willemsen G, Yang T-P, Hua Zhao J, Zitting P, Bradley JR, Dedoussis G V, Gasparini P, Hazen SL, Metspalu A, Pirastu M, Shuldiner AR, Joost van Pelt L, Zwaginga J-J, Boomsma DI, Deary IJ, Franke A, Froguel P, Ganesh SK, Jarvelin M-R, Martin NG, Meisinger C, Psaty BM, Spector TD, Wareham NJ, Akkerman J-WN, Ciullo M, Deloukas P, Greinacher A, Jupe S, Kamatani N, Khadake J, Kooner JS, Penninger J, Prokopenko I, Stemple D, Toniolo D, Wernisch L, Sanna S, Hicks AA, Rendon A, Ferreira MA, Ouwehand WH, Soranzo N. New gene functions in megakaryopoiesis and platelet formation. *Nature* 2011; **480**: 201–8.

10 Yee SW, Chen L, Giacomini KM. Pharmacogenomics of membrane transporters: past, present and future. *Pharmacogenomics* 2010; **11**: 475–9.

11 Franke RM, Gardner ER, Sparreboom A. Pharmacogenetics of drug transporters. *Curr Pharm Des* 2010; **16**: 220–30.

12 Frikke-schmidt R, Nordestgaard BG, Jensen GB, Tybjærg-hansen A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* 2004; **114**: 1343–53.

13 Scotto KW. Transcriptional regulation of ABC drug transporters. *Oncogene* 2003; **22**: 7496–511.

14 Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta* 1997; **1351**: 73–88.

15 Cornwell MM. The human multidrug resistance gene: sequences upstream and downstream of the initiation site influence transcription. *Cell Growth Differ* 1990; **1**: 607–15.

16 Goldsmith ME, Madden MJ, Morrow CS, Cowan KH. A Y-box consensus sequence is required for basal expression of the human multidrug resistance (mdr1) gene. *J Biol Chem* 1993; **268**: 5856–60.

17 Kauffmann HM, Vorderstemann B, Schrenk D. Basal expression of the rat, but not of the human, multidrug resistance protein 2 (MRP2) gene is mediated by CBF/NF-Y and Sp1 promoter-binding sites. *Toxicology* 2001; **167**: 25–35.

18 Mantovani R. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res* 1998; **26**: 1135–43.

19 Cornwell MM, Smith DE. SP1 activates the MDR1 promoter through one of two distinct G-rich regions that modulate promoter activity. *J Biol Chem* 1993; **268**: 19505–11.

20 Gromnicova R, Romero I, Male D. Transcriptional control of the multi-drug transporter ABCB1 by transcription factor Sp3 in different human tissues. *PLoS One* 2012; **7**: e48189.

21 Zhu Q, Center MS. Evidence that SP1 modulates transcriptional activity of the multidrug resistance-associated protein gene. *DNA Cell Biol* 1996; **15**: 105–11.

22 Takada T, Suzuki H, Sugiyama Y. Characterization of 5'-flanking region of human MRP3. *Biochem Biophys Res Commun* 2000; **270**: 728–32.

23 Bailey-Dell KJ, Hassel B, Doyle LA, Ross DD. Promoter characterization and genomic organization of the human breast cancer resistance protein (ATP-binding cassette transporter G2) gene. *Biochim Biophys Acta* 2001; **1520**: 234–41.

24 Davis W, Chen ZJ, Ile KE, Tew KD. Reciprocal regulation of expression of the human adenosine 5'-triphosphate binding cassette, sub-family A, transporter 2 (ABCA2) promoter by the early growth response-1 (EGR-1) and Sp-family transcription factors. *Nucleic Acids Res* 2003; **31**: 1097–107.

25 Kurz EU, Cole SP, Deeley RG. Identification of DNA-protein interactions in the 5' flanking and 5' untranslated regions of the human multidrug resistance protein (MRP1) gene: evaluation of a putative antioxidant response element/AP-1 binding site. *Biochem Biophys Res Commun* 2001; **285**: 981–90.

26 Tanaka T, Uchiumi T, Hinoshita E, Inokuchi A, Toh S, Wada M, Takano H, Kohno K, Kuwano M. The human multidrug resistance protein 2 gene: functional characterization of the 5'-flanking region and expression in hepatic cells. *Hepatology* 1999; **30**: 1507–12.

27 Cho S, Lu M, He X, Ee P-LR, Bhat U, Schneider E, Miele L, Beck WT. Notch1 regulates the expression of the multidrug resistance gene ABCC1/MRP1 in cultured cancer cells. *Proc Natl Acad Sci U S A* 2011; **108**: 20778–83.

28 Bhattacharya S, Das A, Mallya K, Ahmad I. Maintenance of retinal stem cells by Abcg2 is regulated by notch signaling. *J Cell Sci* 2007; **120**: 2652–62.

29 Ayaori M, Yakushiji E, Ogura M, Nakaya K, Hisada T, Uto-Kondo H, Takiguchi S, Terao Y, Sasaki M, Komatsu T, Iizuka M, Yogo M, Uehara Y, Kagechika H, Nakanishi T, Ikewaki K. Retinoic acid receptor agonists regulate expression of ATP-binding cassette transporter G1 in macrophages. *Biochim Biophys Acta* 2012; **1821**: 561–72.

30 Maitra U, Li L. Molecular mechanisms responsible for the reduced expression of cholesterol transporters from macrophages by low-dose endotoxin. *Arterioscler Thromb Vasc Biol* 2013; **33**: 24–33.

31 Hessel S, Lampen A. All-trans retinoic acid enhances the transport of phase II metabolites of benzo[a]pyrene by inducing the Breast Cancer Resistance Protein expression in Caco-2 cells. *Toxicol Lett* 2010; **197**: 151–5.

32 Denson LA, Auld KL, Schiek DS, McClure MH, Mangelsdorf DJ, Karpen SJ. Interleukin-1beta suppresses retinoid transactivation of two hepatic transporter genes involved in bile formation. *J Biol Chem* 2000; **275**: 8835–43.

33 Ratajewski M, Bartosz G, Pulaski L. Expression of the human ABCC6 gene is induced by retinoids through the retinoid X receptor. *Biochem Biophys Res Commun* 2006; **350**: 1082–7.

34 Stromskaya TP, Rybalkina EY, Zabotina TN, Shishkin AA, Stavrovskaya AA. Influence of RARalpha gene on MDR1 expression and P-glycoprotein function in human leukemic cells. *Cancer Cell Int* 2005; **5**: 15.

35 Stromskaya TP, Rybalkina EY, Shtil AA, Zabotina TN, Filippova NA, Stavrovskaya AA. Influence of exogenous RAR alpha gene on MDR1 expression and P-glycoprotein function in human and rodent cell lines. *Br J Cancer* 1998; **77**: 1718–25.

36 Janowski BA, Willy PJ, Devi TR, Falck JR, Mangelsdorf DJ. An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha. *Nature* 1996; **383**: 728–31.

37 Lehmann JM, Kliewer SA, Moore LB, Smith-Oliver TA, Oliver BB, Su JL, Sundseth SS, Winegar DA, Blanchard DE, Spencer TA, Willson TM. Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. *J Biol Chem* 1997; **272**: 3137–40.

38 Wågsäter D, Dimberg J, Sirsjö A. Induction of ATP-binding cassette A1 by all-trans retinoic acid: possible role of liver X receptor-alpha. *Int J Mol Med* 2003; **11**: 419–23.

39 Kliewer SA, Goodwin B, Willson TM. The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr Rev* 2002; **23**: 687–702.

40 Dussault I, Lin M, Hollister K, Wang EH, Synold TW, Forman BM. Peptide mimetic HIV protease inhibitors are ligands for the orphan receptor SXR. *J Biol Chem* 2001; **276**: 33309–12.

41 Fromm MF, Kauffmann HM, Fritz P, Burk O, Kroemer HK, Warzok RW, Eichelbaum M, Siegmund W, Schrenk D. The effect of rifampin treatment on intestinal expression of human MRP transporters. *Am J Pathol* 2000; **157**: 1575–80.

42 Synold TW, Dussault I, Forman BM. The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat Med* 2001; **7**: 584–90.

43 Geick A, Eichelbaum M, Burk O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* 2001; **276**: 14581–7.

44 Teng S, Jekerle V, Piquette-Miller M. Induction of ABCC3 (MRP3) by pregnane X receptor activators. *Drug Metab Dispos* 2003; **31**: 1296–9.

45 Chinetti G, Lestavel S, Bocher V, Remaley AT, Neve B, Torra IP, Teissier E, Minnich A, Jaye M, Duverger N, Brewer HB, Fruchart JC, Clavey V, Staels B. PPAR-alpha and PPAR-gamma activators induce cholesterol removal from human macrophage foam cells through stimulation of the ABCA1 pathway. *Nat Med* 2001; **7**: 53–8.

46 Weiss J, Sauer A, Herzog M, Böger RH, Haefeli WE, Benndorf RA. Interaction of thiazolidinediones (glitazones) with the ATP-binding cassette transporters P-glycoprotein and breast cancer resistance protein. *Pharmacology* 2009; **84**: 264–70.

47 Chen Q, Zhou J, Jiang C, Chen J. Reversal of P-glycoprotein-mediated multidrug resistance in SGC7901/VCR cells by PPARgamma activation by troglitazone. *J Huazhong Univ Sci Technol* 2010; **30**: 326–31.

48 Jiang YJ, Lu B, Kim P, Paragh G, Schmitz G, Elias PM, Feingold KR. PPAR and LXR activators regulate ABCA12 expression in human keratinocytes. *J Invest Dermatol* 2008; **128**: 104–9.

49 Shoda J, Okada K, Inada Y, Kusama H, Utsunomiya H, Oda K, Yokoi T, Yoshizato K, Suzuki H. Bezafibrate induces multidrug-resistance P-Glycoprotein 3 expression in cultured human hepatocytes and humanized livers of chimeric mice. *Hepatol Res* 2007; **37**: 548–56.

50 Hoque MT, Robillard KR, Bendayan R. Regulation of breast cancer resistant protein by peroxisome proliferator-activated receptor α in human brain microvessel endothelial cells. *Mol Pharmacol* 2012; **81**: 598–609.

51 Szatmari I, Vámosi G, Brazda P, Balint BL, Benko S, Széles L, Jeney V, Ozvegy-Laczka C, Szántó A, Barta E, Balla J, Sarkadi B, Nagy L. Peroxisome proliferator-activated receptor gamma-regulated ABCG2 expression confers cytoprotection to human dendritic cells. *J Biol Chem* 2006; **281**: 23812–23.

52 Chawla A, Boisvert WA, Lee C-H, Laffitte BA, Barak Y, Joseph SB, Liao D, Nagy L, Edwards PA, Curtiss LK, Evans RM, Tontonoz P. A PPARγ-LXR-ABCA1 Pathway in Macrophages Is Involved in Cholesterol Efflux and Atherogenesis. *Mol Cell* 2001; **7**: 161–71.

53 Apostoli A, Nicol C. PPAR Medicines and Human Disease: The ABCs of It All. *PPAR Res* 2012; **2012**: 504918.

54 Ohoka N, Okuhira K, Cui H, Wu W, Sato R, Naito M, Nishimaki-Mogami T. HNF4α increases liver-specific human ATP-binding cassette transporter A1 expression and cholesterol efflux to apolipoprotein A-I in response to cholesterol depletion. *Arterioscler Thromb Vasc Biol* 2012; **32**: 1005–14.

55 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. Cooperative interaction between hepatocyte nuclear factor 4 alpha and GATA transcription factors regulates ATP-binding cassette sterol transporters ABCG5 and ABCG8. *Mol Cell Biol* 2007; **27**: 4248–60.

56 Qadri I, Iwahashi M, Kullak-Ublick GA, Simon FR. Hepatocyte nuclear factor (HNF) 1 and HNF4 mediate hepatic multidrug resistance protein 2

up-regulation during hepatitis C virus gene expression. *Mol Pharmacol* 2006; **70**: 627–36.

57 Wong J, Quinn CM, Brown AJ. SREBP-2 positively regulates transcription of the cholesterol efflux gene, ABCA1, by generating oxysterol ligands for LXR. *Biochem J* 2006; **400**: 485–91.

58 Reményi A, Schöler HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol* 2004; **11**: 812–5.

59 Poonkuzhali B, Lamba J, Strom S, Sparreboom A, Thummel K, Watkins P, Schuetz E. Association of breast cancer resistance protein/ABCG2 phenotypes and novel promoter and intron 1 single nucleotide polymorphisms. *Drug Metab Dispos* 2008; **36**: 780–95.

60 Hesselson SE, Matsson P, Shima JE, Fukushima H, Yee SW, Kobayashi Y, Gow JM, Ha C, Ma B, Poon A, Johns SJ, Stryke D, Castro R a, Tahara H, Choi JH, Chen L, Picard N, Sjödin E, Roelofs MJE, Ferrin TE, Myers R, Kroetz DL, Kwok P-Y, Giacomini KM. Genetic variation in the proximal promoter of ABC and SLC superfamilies: liver and kidney specific expression and promoter activity predict variation. *PLoS One* 2009; **4**: e6942.

61 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–5.

62 Johnson A, Handsaker R, Pulit S, Nizzar M, O'Donnell C, de Bakker P. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938–9.

63 Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* 2013; **144**: 1488–96, 1496.e1–3.

64 Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, Weichenthal M, Ellinghaus E, Franke A, Cookson W, Nair RP, Elder JT, Abecasis GR. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet* 2010; **87**: 779–89.

65 Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 2010; **464**: 773–7.

66 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl C a, van Nas A, Mehrabian M, Drake T a, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107.

67 Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls M a, Lai S-L, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010; **6**: e1000952.

68 Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin S-Y, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, Montgomery SB, Parts L, Potter S, Surdulescu G, Tsaprouni L, Tsoka S, Bataille V, Durbin R, Nestle FO, O'Rahilly S, Soranzo N, Lindgren CM, Zondervan KT, Ahmadi KR, Schadt EE, Stefansson K, Smith GD, McCarthy MI, Deloukas P, Dermitzakis ET, Spector TD. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012; **44**: 1084–9.

69 Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M,

Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science (80-)* 2009; **325**: 1246–50.

70 Mangravite L, Engelhardt. Bayes Factor, LCLs, 480 ids from Cholesterol And Pharmacogenetics (CAP) study. *Rev*.

71 Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Mennerich D, Rust W, Perret C, Proust C, Nicaud V, Loscalzo J, Hübner N, Tregouet D, Münzel T, Ziegler A, Tiret L, Blankenberg S, Cambien F. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**: e10693.

72 Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007; **315**: 848–53.

73 Myers AJ, Gibbs JR, Webster J a, Rohrer K, Zhao A, Marlowe L, Kaleem M, Leung D, Bryden L, Nath P, Zismann VL, Joshipura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson J V, Holmans P, Heward CB, Reiman EM, Stephan D, Hardy J. A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**: 1494–9.

74 Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson D a, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD. Identification, replication, and functional fine-mapping of expression quantitative trait Loci in primary human liver tissue. *PLoS Genet* 2011; **7**: e1002078.

75 Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* 1987; **196**: 947–50.

76 Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 1986; **44**: 283–92.

77 Kozak M. An analysis of 5’-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 1987; **15**: 8125–48.

78 Zhang Y, Zhao T, Li W, Vore M. The 5J-Untranslated Region of Multidrug Resistance Associated Protein 2 (MRP2 ; ABCC2) Regulates

Downstream Open Reading Frame Expression through Translational Regulation. *Science (80-)* 2010; **77**: 237–46.

79 Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 2009; **136**: 731–45.

80 Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* 2006; **20**: 515–24.

81 Markova SM, Kroetz DL. ABCC4 is regulated by microRNA-124a and microRNA-506. *Biochem Pharmacol* 2014; **87**: 515–22.

82 Haenisch S, Laechelt S, Bruckmueller H, Werk A, Noack A, Bruhn O, Remmler C, Cascorbi I. Down-regulation of ATP-binding cassette C2 protein expression in HepG2 cells after rifampicin treatment is mediated by microRNA-379. *Mol Pharmacol* 2011; **80**: 314–20.

83 Wu X, Brewer G. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* 2012; **500**: 10–21.

84 Guhaniyogi J, Brewer G. Regulation of mRNA stability in mammalian cells. *Gene* 2001; **265**: 11–23.

85 Lamba JK, Adachi M, Sun D, Tammur J, Schuetz EG, Allikmets R, Schuetz JD. Nonsense mediated decay downregulates conserved alternatively spliced ABCC4 transcripts bearing nonsense codons. *Hum Mol Genet* 2003; **12**: 99–109.

86 Stojic J, Stöhr H, Weber BHF. Three novel ABCC5 splice variants in human retina and their role as regulators of ABCC5 gene expression. *BMC Mol Biol* 2007; **8**: 42.

87 Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. *Nature* 2003; **426**: 895–9.

88 Hessa T, Sharma A, Mariappan M, Eshleman HD, Gutierrez E, Hegde RS. Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* 2011; **475**: 394–7.

89 Kondo C, Suzuki H, Itoda M, Ozawa S, Sawada J, Kobayashi D, Ieiri I, Mine K, Ohtsubo K, Sugiyama Y. Functional analysis of SNPs variants of BCRP/ABCG2. *Pharm Res* 2004; **21**: 1895–903.

90 Furukawa T, Wakabayashi K, Tamura A, Nakagawa H, Morishima Y, Osawa Y, Ishikawa T. Major SNP (Q141K) variant of human ABC

transporter ABCG2 undergoes lysosomal and proteasomal degradations. *Pharm Res* 2009; **26**: 469–79.

91 Morisaki K, Robey RW, Ozvegy-Laczka C, Honjo Y, Polgar O, Steadman K, Sarkadi B, Bates SE. Single nucleotide polymorphisms modify the transporter activity of ABCG2. *Cancer Chemother Pharmacol* 2005; **56**: 161–72.

92 Tamura A, Wakabayashi K, Onishi Y, Takeda M, Ikegami Y, Sawada S, Tsuji M, Matsuda Y, Ishikawa T. Re-evaluation and functional classification of non-synonymous single nucleotide polymorphisms of the human ATP-binding cassette transporter ABCG2. *Cancer Sci* 2007; **98**: 231–9.

93 Nakagawa H, Toyoda Y, Wakabayashi-Nakao K, Tamaki H, Osumi M, Ishikawa T. Ubiquitin-mediated proteasomal degradation of ABC transporters: a new aspect of genetic polymorphisms and clinical impacts. *J Pharm Sci* 2011; **100**: 3602–19.

94 Toyoda Y, Sakurai A, Mitani Y, Nakashima M, Yoshiura K, Nakagawa H, Sakai Y, Ota I, Lezhava A, Hayashizaki Y, Niikawa N, Ishikawa T. Earwax, osmidrosis, and breast cancer: why does one SNP (538G>A) in the human ABC transporter ABCC11 gene determine earwax type? *FASEB J* 2009; **23**: 2001–13.

95 Bobadilla JL, Macek M, Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. *Hum Mutat* 2002; **19**: 575–606.

96 Nishikawa S, Brodsky JL, Nakatsukasa K. Roles of molecular chaperones in endoplasmic reticulum (ER) quality control and ER-associated degradation (ERAD). *J Biochem* 2005; **137**: 551–5.

97 Huls M, van den Heuvel JJMW, Dijkman HBPM, Russel FGM, Masereeuw R. ABC transporter expression profiling after ischemic reperfusion injury in mouse kidney. *Kidney Int* 2006; **69**: 2186–93.

98 Huls M, Russel FGM, Masereeuw R. The Role of ATP Binding Cassette Transporters in Tissue Defense and Organ Regeneration. *Pharmacology* 2009; **328**: 3–9.

99 Coulombe-Huntington J, Lam KCL, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS Genet* 2009; **5**: e1000766.

100 Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second

generation sequencing in a Caucasian population. *Nature* 2010; **464**: 773–7.

101 Pickrell JK, Marioni JC, Pai A a, Degner JF, Engelhardt BE, Nkadori E, Veyrieras J-B, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010; **464**: 768–72.

102 Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 2008; **40**: 225–31.

103 Zhang W, Duan S, Bleibel WK, Wisel S a, Huang RS, Wu X, He L, Clark T a, Chen TX, Schweitzer AC, Blume JE, Dolan ME, Cox NJ. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* 2009; **125**: 81–93.

104 Kang HM, Ye C, Eskin E. Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots. *Genetics* 2008; **180**: 1909–25.

105 Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher JR, Reitman ML, Kong A, Schadt EE, Stefansson K. Genetics of gene expression and its effect on disease. *Nature* 2008; **452**: 423–8.

106 Lalonde E, Ha KCH, Wang Z, Bemmo A, Kleinman CL, Kwan T, Pastinen T, Majewski J. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res* 2011; **21**: 545–54.

107 Matsson P, Yee SW, Markova S, Morrissey K, Jenkins G, Xuan J, Jorgenson E, Kroetz DL, Giacomini KM. Discovery of regulatory elements in human ATP-binding cassette transporters through expression quantitative trait mapping. *Pharmacogenomics J* 2012; **12**: 214–26.

108 Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KCC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WOC. A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**: 1202–7.

109 Pickrell JK, Marioni JC, Pai A a, Degner JF, Engelhardt BE, Nkadori E, Veyrieras J-B, Stephens M, Gilad Y, Pritchard JK. Understanding

mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010; **464**: 768–72.

110 Veyrieras J-B, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 2008; **4**: e1000214.

111 Duarte CW, Zeng Z-B. High-confidence discovery of genetic network regulators in expression quantitative trait loci data. *Genetics* 2011; **187**: 955–64.

112 Zhang W, Duan S, Bleibel WK, Wisel S a, Huang RS, Wu X, He L, Clark T a, Chen TX, Schweitzer AC, Blume JE, Dolan ME, Cox NJ. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* 2009; **125**: 81–93.

113 Gamazon ER, Zhang W, Dolan ME, Cox NJ. Comprehensive survey of SNPs in the Affymetrix exon array using the 1000 Genomes dataset. *PLoS One* 2010; **5**: e9366.

114 Nembaware V, Lupindo B, Schouest K, Spillane C, Scheffler K, Seoighe C. Genome-wide survey of allele-specific splicing in humans. *BMC Genomics* 2008; **9**: 265.

115 Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna K V, Gabriel WN, Welsh-Bohmer K a, Hulette CM, Denny TN, Goldstein DB. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* 2008; **6**: e1.

116 Huang Y, Zheng J, Przytycka TM. Discovery of regulatory mechanisms from gene expression variation by eQTL analysis. *Biol Data Min* 2008; : 205.

117 Zou W, Zeng Z-B. Multiple interval mapping for gene expression QTL analysis. *Genetica* 2009; **137**: 125–34.

118 Kendziorski C, Wang P. A review of statistical methods for expression quantitative trait loci mapping. *Mamm Genome* 2006; **17**: 509–17.

119 Cheung VG, Nayak RR, Wang IX, Elwyn S, Cousins SM, Morley M, Spielman RS. Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol* 2010; **8**. doi:10.1371/journal.pbio.1000480.

120 Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey

of human alternative pre-mRNA splicing with exon junction microarrays. *Science (80-)* 2003; **302**: 2141–4.

121 Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008; **456**: 470–6.

122 Srinivasan S, Bingham JL, Johnson D. The ABCs of human alternative splicing: a review of ATP-binding cassette transporter splicing. *Curr Opin Drug Discov Devel* 2009; **12**: 149–58.

123 Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* 2002; **30**: 3754–66.

124 Kwan T, Benovoy D, Dias C, Gurd S, Serre D, Zuzan H, Clark T a, Schweitzer A, Staples MK, Wang H, Blume JE, Hudson TJ, Sladek R, Majewski J. Heritability of alternative splicing in the human genome. *Genome Res* 2007; **17**: 1210–8.

125 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; **10**: 57–63.

126 Maston G a, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006; **7**: 29–59.

127 Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science (80-)* 2003; **302**: 249–55.

128 Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998; **95**: 14863–8.

129 Carter SL, Brechbühler CM, Griffin M, Bond AT. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 2004; **20**: 2242–50.

130 Carlson MRJ, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* 2006; **7**: 40.

131 Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res* 2003; **13**: 773–80.

132 Prieto C, Risueño A, Fontanillo C, De las Rivas J. Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS One* 2008; **3**: e3911.

133 Oliver S. Guilt-by-association goes global. *Nature* 2000; **403**: 601–3.

Chapter 2: Expression Quantitative Trait Mapping Identifies Transcriptional Regulators of ATP-Binding Cassette Transporters in the Human Kidney

2.1 Abstract

ATP-Binding Cassette (ABC) transporters are a family of proteins responsible for the transport of endogenous compounds and drugs across cell membranes. In the kidney, depending on their localization to apical or basolateral membranes, ABC transporters are involved in both secretion of compounds into the proximal tubule and reabsorption back into the blood stream. Alterations in activity of ABC transporters, including changes in transporter expression, at the proximal tubule can have significant renal and systemic effects. Expression regulatory variation has been studied in a number of human tissues, but has not been explored in the human kidney. In this study, local or *cis*-eQTLs associated with ABC transporter expression in the kidney were identified, as well as distal transcription factors containing one or more SNPs associated with transporter expression.

2.2 Introduction

While they are expressed throughout the body, the function of ABC transporters is particularly essential in organs involved in excretory functions, such as the kidney. ABC transporters are expressed primarily in the renal proximal tubule epithelial cells, where they are involved in both secretion of compounds into the proximal tubule and reabsorption back into the bloodstream. ABC transporters

have also been implicated in protection against damage and regeneration in several tissues, including the kidney[1,2]. Thus, alterations in the activity of ABC transporters in the proximal tubule can have significant renal and systemic effects.

Changes in transporter function can be driven by changes in gene expression. Further, inter-individual variation in expression of ABC transporters in the kidney may explain differences between individuals in susceptibility to nephrotoxicity, systemic drug toxicities, or other conditions caused by disruptions in the excretion or homeostasis of metabolic wastes and other endogenous compounds. Genetic variation associated with changes in gene expression – known as expression quantitative trait loci (eQTL) – may be located proximal or distal to the gene of interest. Proximal variation, or *cis*-eQTLs, may be found within the gene itself or in a promoter or enhancer region outside of the coding region. Distal variation, or *trans*-eQTLs, are often located in genes that regulate expression of the gene of interest or are co-regulated with the gene of interest. For example, polymorphisms in transcription factors may modulate expression of the genes they regulate.[3]

While large eQTL studies have been conducted in lymphoblastoid cell lines[4–6] and in selected human tissues[7–11], variation in expression caused by genetic differences is often highly tissue specific[12,13]. To date, no eQTL studies have been published in the human kidney. In this chapter, genetic variants associated with ABC transporter expression in the renal cortex were identified.

2.3 Materials and Methods

2.3.1 Tissue Samples

Kidney samples were acquired from two commercial sources (Asterand; Detroit, MI and Capital Biosciences; Rockville, MD) and included both postmortem tissue and surgical resection from donors. Kidney samples are from the renal cortex. All samples originated from Caucasian males and females between ages 3-90 (average age 53.6 yr). Information regarding donors is included in Table 2.1.

Table 2.1: Tissue Donor Information

Sample	Vendor	Age	Sex	Ethnicity
1	Asterand	56	Female	Caucasian
2	Asterand	78	Male	Caucasian
3	Asterand	46	Female	Caucasian
4	Asterand	55	Female	Caucasian
5	Asterand	59	Male	Caucasian
6	Asterand	69	Female	Caucasian
7	Asterand	68	Male	Caucasian
8	Asterand	53	Male	Caucasian
9	Bioscience	74	Male	Caucasian
10	Asterand	41	Male	Caucasian
11	Asterand	81	Male	Caucasian
12	Asterand	55	Male	Caucasian
13	Asterand	55	Male	Caucasian
14	Asterand	75	Male	Caucasian
15	Asterand	52	Male	Caucasian
16	Asterand	90	Female	Caucasian
17	Asterand	71	Male	Caucasian
18	Asterand	38	Female	Caucasian
19	Asterand	50	Female	Caucasian
20	Asterand	54	Female	Caucasian
21	Asterand	4	Female	Caucasian
22	Asterand	57	Female	Caucasian
23	Asterand	40	Female	Caucasian
24	Asterand	3	Male	Caucasian
25	Asterand	55	Male	Caucasian
26	Asterand	72	Male	Caucasian
27	Asterand	48	Female	Caucasian
28	Asterand	53	Male	Caucasian
29	Asterand	8	Male	Caucasian
30	Asterand	45	Male	Caucasian
31	Asterand	63	Female	Caucasian
32	Asterand	44	Female	Caucasian

33	Asterand	37	Male	Caucasian
34	Asterand	4	Female	Caucasian
35	Asterand	54	Male	Caucasian
36	Asterand	47	Female	Caucasian
37	Asterand	60	Male	Caucasian
38	Asterand	43	Male	Caucasian
39	Asterand	51	Male	Caucasian
40	Asterand	58	Male	Caucasian
41	Asterand	49	Female	Caucasian
42	Asterand	6	Female	Caucasian
43	Asterand	48	Male	Caucasian
44	Asterand	51	Male	Caucasian
45	Asterand	36	Male	Caucasian
46	Asterand	58	Female	Caucasian
47	Asterand	80	Male	Caucasian
48	Asterand	52	Female	Caucasian
49	Asterand	52	Female	Caucasian
50	Asterand	49	Female	Caucasian
51	Asterand	46	Male	Caucasian
52	Asterand	62	Male	Caucasian
53	Bioscience	82	Female	Caucasian
54	Bioscience	61	Male	Caucasian
55	Bioscience	46	Male	Caucasian
56	Asterand	79	Female	Caucasian
57	Bioscience	68	Male	Caucasian
58	Asterand	72	Male	Caucasian
59	Asterand	70	Female	Caucasian
60	Asterand	81	Male	Caucasian

DNA was extracted from 60 kidney samples with the Qiagen AllPrep DNA/RNA Mini Kit and QIAquick PCR Purification Kit (Qiagen; Valencia, CA). RNA was extracted following the protocol for Trizol reagent (Invitrogen; Carlsbad, CA) and Qiagen (Valencia, CA) RNeasy MinElute Cleanup Kit; RNA with sufficient quality (260/280 >1.7 and 260/230 >1.8, RNA Integrity number 3-8) was isolated for 58 kidney samples.

2.3.2 Genotyping and qPCR

Transcript expression levels were quantified for 58 kidney samples using a custom array on the Biotrove OpenArray™ qPCR platform (Life Technologies;

Carlsbad, CA). Raw Ct values for each gene in each sample were normalized relative to the geometric mean of three housekeeping genes: Beta-Actin, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and Beta-2 microglobulin (B2M) to give a ΔCt value per gene for each sample. Because samples were run across separate arrays and in some cases separate days, all ΔCt values for a given tissue type were quantile normalized across samples using the R preprocessCore package[14]. To detect outliers in expression data, a principal components analysis was conducted on kidney samples; based on the first three principal components, two outliers were detected and removed from further analyses (Figure 2.1).

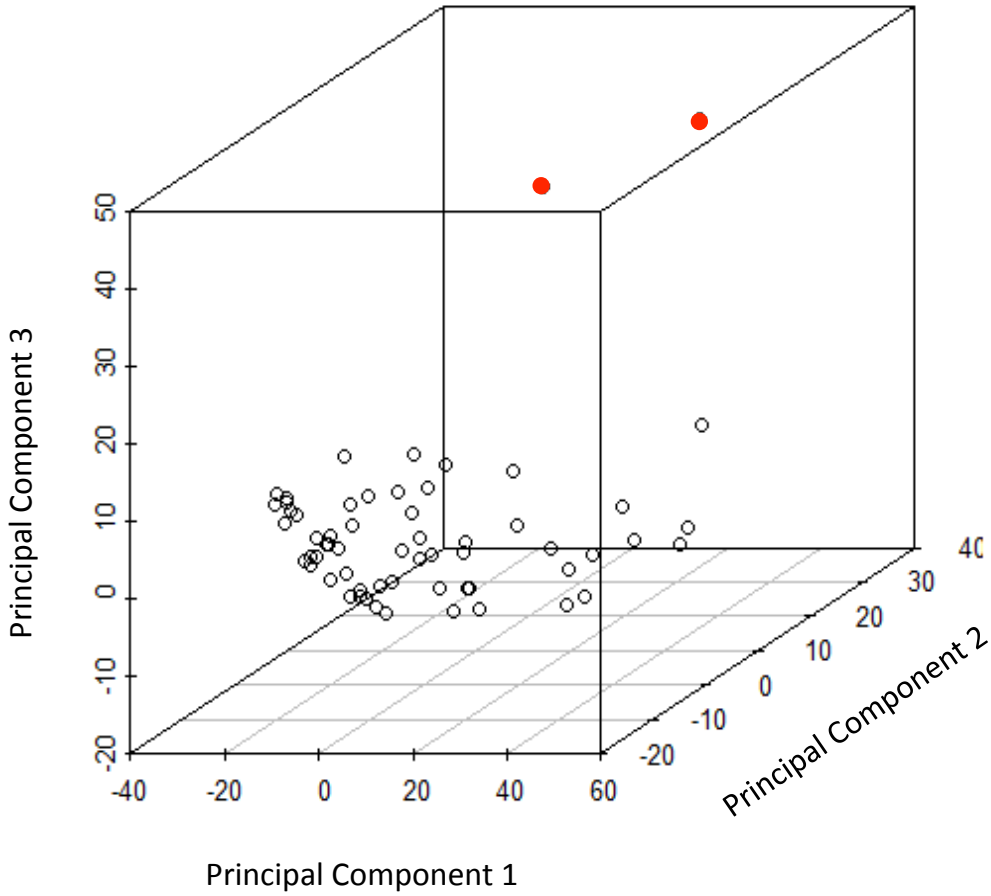


Figure 2.1: Principal components of renal cortex samples. Principal components 1 through 3 are shown for 58 renal cortex samples. Samples in red were considered outliers and excluded from further analyses.

Fifty eight kidney samples were genotyped on an Affymetrix Axiom CEU genotyping array[15]. Initial quality control filters for genotype data included SNP and sample call rate, sex check, and relatedness (by IBD). Six kidney samples failed initial quality control for sex-check, relatedness, and call rate tests and were excluded from further analysis. Altogether, genotype and expression data were successfully generated for 52 kidney samples.

2.3.3 Association with Age and Sex

Associations between transporter mRNA expression and donor sex were detected using the Wilcoxon-Rank rum test, and associations between mRNA expression and donor age by linear regression.

2.3.4 cis-eQTL Analysis

ABC genes for which ≥ 50% of samples had expression values at the limit of detection of the qPCR instrument (Raw Ct = 30) were excluded from further analyses. For each of the remaining 42 ABC transporters, association between gene expression (normalized deltaCT) and genotype (number of minor alleles) was calculated by linear regression for all SNPs with minor allele frequency greater than 1% in two stages, first within 50 kb and second within 500 kb of the longest isoform (transcription start site - transcription stop site) for each gene. For transporters where association between transporter expression and age or sex has p-value ≤0.2, regressions were conducted including age and/or sex as a covariate. For each gene, p values from linear regression were corrected for multiple testing by dividing by the number of haplotype blocks + number of singletons within the region tested. Haplotype blocks were generated using the method of Gabriel at al.[16] as implemented in Haploview[17]. All *cis*-eQTL analyses were conducted in PLINK[18] unless otherwise noted.

2.3.5 trans-eQTL Analysis – transcription factors

Because of the small sample size, *trans*-eQTL analyses were conducted as gene-based tests, looking at the combined association of SNPs within a gene with transporter expression, and limiting this analysis to transcription factors expressed in the human kidney. Transcription factors were defined as genes associated with Gene Ontology set GO:0003700, "sequence-specific DNA binding transcription factor activity", and kidney expression was defined as genes with an FPKM>0 in the supplemental kidney expression dataset, for a total of 972 genes (Table 2.2). The gene-based analysis was conducted using the PLINK set test. This test calculates the mean of up to 10 significant ($p < 0.05$) per-SNP p-values after filtering for SNPs in linkage disequilibrium ($r^2 \geq 0.5$). An empirical p-value is applied to each set test by permuting phenotype labels across individuals (10,000 permutations). SNP p-values were calculated as described in the *cis*-eQTL analysis, and gene boundaries were defined as +/- 10 kb of the transcription start and stop site for the gene. A total of 796 transcription factor genes contained at least one SNP and were evaluated in the gene-based test.

Table 2.2: Human Transcription Factors Expressed[1] in the Kidney.

AATF	DBP	FOXF1	HOPX	LEF1	NFX1	POU6F1	SOX13	TGIF2	ZNF207
ADNP	DBX1	FOXF2	HOXA1	LEUTX	NFXL1	POU6F2	SOX15	THRA	ZNF213
ADNP2	DBX2	FOXG1	HOXA10	LHX1	NFYA	PPARA	SOX17	THRB	ZNF215
AEBP1	DDIT3	FOXH1	HOXA11	LHX2	NFYB	PPARD	SOX18	TLE4	ZNF217
AFF1	DLX1	FOXI1	HOXA13	LHX4	NFYC	PPARG	SOX2	TLX2	ZNF219
AFF4	DLX2	FOXI2	HOXA2	LHX5	NKX2-2	PRDM1	SOX30	TMEM229A	ZNF232
AHCTF1	DLX3	FOXJ1	HOXA3	LHX6	NKX2-3	PRDM2	SOX4	TP53	ZNF236
AHR	DLX4	FOXJ2	HOXA4	LHX8	NKX2-5	PROP1	SOX5	TP63	ZNF24
ALX1	DLX5	FOXJ3	HOXA5	LHX9	NKX2-6	PROX1	SOX6	TP73	ZNF256
ALX3	DLX6	FOXK1	HOXA6	LMO2	NKX2-8	PRRX1	SOX7	TPRX1	ZNF263
ALX4	DMBX1	FOXK2	HOXA7	LMO4	NKX3-1	PRRX2	SOX8	TRERF1	ZNF268
ANKRD30A	DMRT1	FOXL1	HOXA9	LMX1A	NKX3-2	PTGER3	SOX9	TRIM22	ZNF274
AOF2	DMRT2	FOXL2	HOXB1	LMX1B	NKX6-1	PTH	SP1	TRIM25	ZNF277
AR	DMRT3	FOXM1	HOXB13	LZTR1	NKX6-2	PTTG1	SP140	TRIM28	ZNF281
ARGFX	DMRTA1	FOXN1	HOXB2	LZTS1	NKX6-3	PURA	SP3	TRIM29	ZNF287
ARID3A	DMRTA2	FOXN2	HOXB3	MAF	NME2	PURB	SP4	TRPS1	ZNF292
ARID4A	DMRTB1	FOXN3	HOXB4	MAFA	NOBOX	RARA	SPDEF	TSC22D1	ZNF3
ARNT	DMRTC1	FOXN4	HOXB5	MAFB	NOTCH1	RARB	SPEN	TSC22D2	ZNF33A
ARNT2	DMRTC1	FOXO1	HOXB6	MAFF	NOTO	RARG	SPI1	TSC22D3	ZNF33B
ARNTL	DMRTC2	FOXO3	HOXB7	MAFG	NPAS1	RAX2	SPIB	TSC22D4	ZNF35
ARNTL2	DMTF1	FOXO4	HOXB8	MAFK	NPAS2	RB1	SPIC	TSHZ1	ZNF367
ARX	DPRX	FOXO6	HOXB9	MAX	NPAT	RBCK1	SREBF1	TSHZ2	ZNF37A
ASCL1	DRAP1	FOXP1	HOXC10	MAZ	NR0B1	RBP2	SREBF2	TSHZ3	ZNF394

ASCL2	DRGX	FOXP2	HOXC11	MBD1	NR0B2	RBPJ	SRF	TULP4	ZNF396
ATF1	DUX4C	FOXP3	HOXC12	MECP2	NR1D1	RBPJL	SRY	TWIST1	ZNF397
ATF2	DUXA	FOXP4	HOXC13	MED1	NR1D2	RCAN1	ST18	TWIST2	ZNF397OS
ATF3	E2F1	FOXQ1	HOXC4	MEF2A	NR1H2	RCOR1	STAT1	UBN1	ZNF41
ATF4	E2F2	FOXR1	HOXC5	MEF2B	NR1H3	RCOR2	STAT2	UBP1	ZNF438
ATF5	E2F3	FOXS1	HOXC6	MEF2C	NR1H4	REL	STAT3	UHRF1	ZNF444
ATF6	E2F4	FUBP1	HOXC8	MEF2D	NR1I2	RELA	STAT4	UNCX	ZNF445
ATF7	E2F5	GABPA	HOXC9	MEIS1	NR1I3	RELB	STAT5A	USF1	ZNF446
AXUD1	E2F6	GABPB1	HOXD1	MEIS2	NR2C1	RERE	STAT5B	USF2	ZNF449
BACH1	E2F7	GAS7	HOXD10	MEIS3	NR2C2	REST	STAT6	VAV1	ZNF45
BACH2	E2F8	GATA1	HOXD11	MEIS3P2	NR2E1	REXO4	STK16	VAX1	ZNF483
BARHL1	E4F1	GATA2	HOXD12	MEOX1	NR2E3	RFX1	STRN3	VAX2	ZNF488
BARHL2	EBF1	GATA3	HOXD13	MEOX2	NR2F1	RFX3	SUPT4H1	VDR	ZNF496
BARX1	ECSIT	GATA4	HOXD3	MESP1	NR2F2	RFX5	SUPT6H	VENTX	ZNF500
BARX2	EDF1	GATA5	HOXD4	MESP2	NR2F6	RFXANK	T	VSX1	ZNF518A
BATF	EGR1	GATA6	HOXD8	MGA	NR3C1	RFXAP	TADA2L	VSX2	ZNF518B
BATF2	EGR2	GATAD1	HOXD9	MITF	NR3C2	RHOXF1	TADA3L	WNT5A	ZNF576
BATF3	EGR3	GATAD2A	HR	MIXL1	NR4A1	RHOXF2	TAF10	WT1	ZNF628
BCL11B	EGR4	GATAD2B	HSF1	MIZF	NR4A2	RHOXF2	TAF12	XBP1	ZNF639
BCL3	EHF	GBX1	HSF2	MKL1	NR4A3	RHOXF2B	TAF13	YBX1	ZNF641
BCL6	ELF1	GBX2	HSF4	MKL2	NR5A1	RHOXF2B	TAF1B	YEATS4	ZNF69
BHLHB2	ELF2	GCM1	HSF5	MKX	NR5A2	RNF4	TAF4	YY1	ZNF70
BHLHB3	ELF3	GLI1	HSFX2	MLLT10	NR6A1	RORA	TAF4B	ZBTB17	ZNF71
BHLHB5	ELF4	GLI2	HSFY1	MLX	NRL	RORB	TAF5	ZBTB20	ZNF750
BLZF1	ELF5	GLI3	HSFY1	MLXIPL	OLIG2	RORC	TAF5L	ZBTB25	ZNF75D
BMPR1A	ELK1	GLIS2	ID1	MMP14	ONECUT1	RREB1	TAF6	ZBTB38	ZNF770
BNC1	ELK3	GLIS3	ID3	MNT	ONECUT2	RUNX1	TAF7	ZBTB48	ZNF80

BRD8	ELK4	GLP-1	IKZF1	MNX1	ONECUT3	RUNX1T1	TAL1	ZBTB7B	ZNF81
BSX	EMX1	GMEB1	IKZF3	MSC	OTP	RUNX2	TARDBP	ZC3H8	ZNF83
BTAF1	EMX2	GPBP1	IKZF4	MSL3L1	OTX1	RUNX3	TBPL2	ZEB1	ZNF85
BTBD14A	EN1	GPBP1L1	INSM1	MSRB2	OVOL1	RXRA	TBR1	ZEB2	ZNF90
BTG2	EN2	GRHL3	IRF1	MSX1	OVOL2	RXRB	TBX1	ZFAT	ZNF91
BUD31	ENO1	GSC	IRF2	MSX2	OVOL3	RXRG	TBX10	ZFHX2	ZNF92
C1orf85	EOMES	GSC2	IRF3	MTA1	PA2G4	SALL1	TBX15	ZFHX3	ZNF93
C2orf3	EPAS1	GSX2	IRF4	MTA2	PAX3	SALL2	TBX18	ZFHX4	ZRANB2
C5orf41	ERF	GTF2H2	IRF5	MTA3	PAX4	SATB1	TBX19	ZFP36L1	ZSCAN1
CBFA2T2	ESR1	GTF2H2	IRF6	MTF1	PAX5	SATB2	TBX2	ZFP36L2	ZSCAN10
CBFA2T3	ESR2	GTF2H3	IRF7	MXD1	PAX6	SCAND1	TBX20	ZFP37	ZSCAN12
CBFB	ESRRA	GTF2H4	IRF8	MYB	PAX7	SCAND3	TBX21	ZFP42	ZSCAN16
CBL	ESRRB	GTF2I	IRF9	MYBL2	PAX8	SCMH1	TBX22	ZFP90	ZSCAN18
CC2D1A	ESRRG	GTF2IRD1	IRX1	MYC	PBX1	SCML1	TBX3	ZFPM1	ZSCAN2
CC2D1B	ESX1	HAND2	IRX2	MYCN	PBX2	SCML2	TBX4	ZGPAT	ZSCAN20
CCRN4L	ETS1	HCFC1	IRX3	MYF5	PBX3	SCRT1	TBX5	ZHX1	ZSCAN21
CDX1	ETV1	HDAC1	IRX4	MYF6	PBX4	SEBOX	TBX6	ZHX2	ZSCAN22
CDX2	ETV2	HDAC2	IRX5	MYNN	PCGF2	SHOX	TCEAL1	ZHX3	ZSCAN23
CEBPA	ETV3	HDX	IRX6	MYOCD	PCGF6	SHOX	TCF12	ZIC1	ZSCAN29
CEBPB	ETV3L	HELT	ISL1	MYOD1	PDX1	SHOX2	TCF15	ZIC2	ZSCAN4
CEBPD	ETV4	HES1	ISL2	MYOG	PEG3	SIM1	TCF19	ZIC3	ZSCAN5A
CEBPE	ETV5	HES5	ISX	MYPOP	PFDN1	SIM2	TCF21	ZIC5	ZSCAN5B
CEBPG	ETV6	HES6	JARID1B	MYST2	PGBD1	SIN3A	TCF25	ZIM2	ZSCAN5C
CHCHD3	ETV7	HESX1	JDP2	MYT1	PGR	SIX1	TCF3	ZKSCAN1	ZXDA
CIR	EVI1	HEY1	JMJD1A	MYT1L	PHB	SIX2	TCF4	ZKSCAN2	ZXDC
CITED1	EVX1	HEY2	JUN	MZF1	PHF1	SIX3	TCF7	ZKSCAN3	
CITED2	EVX2	HEYL	JUNB	NANOG	PHF5A	SIX4	TCF7L1	ZKSCAN4	

CLOCK	FAM130A1	HHEX	JUND	NANOGNB	PHOX2A	SIX5	TCF7L2	ZKSCAN5
CNBP	FAM130A2	HIC1	KLF1	NANOGP1	PHOX2B	SLC26A3	TCFL5	ZNF117
CNOT7	FBXW7	HIF1A	KLF10	NCOR1	PHTF1	SLC2A4RG	TEAD1	ZNF131
CNOT8	FEV	HIF3A	KLF11	NEUROD1	PITX1	SLC30A9	TEAD2	ZNF132
CREB1	FLI1	HIRA	KLF12	NEUROD2	PITX2	SMAD1	TEAD3	ZNF133
CREB3	FOS	HLF	KLF15	NEUROG1	PITX3	SMAD2	TEAD4	ZNF134
CREB3L1	FOSB	HLX	KLF16	NEUROG3	PKNOX2	SMAD3	TEF	ZNF135
CREB3L2	FOSL1	HMBOX1	KLF17	NFAT5	PLAG1	SMAD4	TFAM	ZNF137P
CREB3L3	FOSL2	HMG20A	KLF2	NFATC1	PLAGL1	SMAD5	TFAP2A	ZNF138
CREB3L4	FOXA1	HMG20B	KLF3	NFATC2	PLAGL2	SMAD6	TFAP2B	ZNF140
CREB5	FOXA2	HMGA1	KLF4	NFATC3	PLSCR1	SMAD7	TFAP2C	ZNF148
CREBBP	FOXA3	HMGA2	KLF5	NFATC4	POU1F1	SMAD9	TFAP2D	ZNF154
CREBL1	FOXB1	HMGB1	KLF7	NFE2	POU2F1	SNAI1	TFAP2E	ZNF155
CREBL2	FOXC1	HMGB2	KLF9	NFE2L1	POU2F2	SNAI2	TFAP4	ZNF157
CREBZF	FOXC2	HMX2	L3MBTL	NFE2L2	POU2F3	SNAI3	TFCP2	ZNF165
CREM	FOXD2	HMX3	L3MBTL4	NFE2L3	POU3F1	SNAPC2	TFCP2L1	ZNF169
CRX	FOXD3	HNF1A	LASS2	NFIA	POU3F2	SNAPC4	TFDP1	ZNF174
CTBP1	FOXD4	HNF1B	LASS3	NFIB	POU3F3	SNAPC5	TFDP2	ZNF175
CTCF	FOXD4L1	HNF4A	LASS4	NFIC	POU3F4	SOHLH1	TFDP3	ZNF18
CTNNB1	FOXD4L3	HNF4G	LASS5	NFIL3	POU4F1	SOX1	TFE3	ZNF189
CUX1	FOXD4L5	HNRNPAB	LASS6	NFIX	POU4F3	SOX10	TFEB	ZNF19
CUX2	FOXE1	HNRNPK	LBX2	NFKB1	POU5F1	SOX11	TFEC	ZNF197
DACH1	FOXE3	HOMEZ	LCOR	NFKB2	POU5F1B	SOX12	TGIF1	ZNF202

[1] Transcription factors defined as genes associated with Gene Ontology set GO:0003700; kidney expression defined as genes with an FPKM >0 in the RNA-seq kidney expression dataset

2.3.6 trans-eQTL Analysis – distal genes

In addition to the transcription factor – QTL analysis, the analysis was expanded to all genes in the genome (17,787 genes). Because the gene-based PLINK set test is computationally intractable for large numbers of sets, the Versatile Gene-based Association Study tool was used[19]. Briefly, association between genome-wide SNPs and transporter expression (normalized ΔCT) was conducted by linear regression in PLINK. The Versatile Gene-based association test collapses genome-wide associations into a single p-value per gene (most significant p-value for each gene, defined as +/- 50 kb from the transcription start and stop site). A simulation for each gene is generated, matching number of SNPs and LD structure between SNPs to those observed in genotype data. Each gene is then assigned an empirical p-value, which represents the portion of simulated p-values for each gene that are larger than the observed most significant p-value. Linkage disequilibrium for SNPs in each gene was estimated using HapMap data from Caucasian populations (CEU).

2.3.7 Functional Annotation

For each of the six *cis*-eQTL SNPs, all 1000 Genomes[20] Pilot 1 SNPs in LD (r^2 > 0.8) were extracted. Haploreg(v2)[21] and RegulomeDB[22] were used to prioritize SNPs based on available evidence for regulatory activity at and around the SNP locus. All selected SNPs had a RegulomeDB score > 4. Evidence of regulatory activity included protein binding predicted by CHIP-seq, open chromatin state by DNase-seq[23] or FAIRE-seq[24], histone modifications indicative of active

transcription by CHIP-seq, and/or putative change in transcription factor binding based on position weight matrix. Histone modifications were defined as markers of “active” rather than repressive state and with experimental evidence in human cell lines including histone 3 lysine 4 (H3K4) mono-, di-, and tri- methylation, H3K9 monomethylation or acetylation, H3K14 acetylation, H3K14 monomethylation or acetylation, H3K20 monomethylation, H3K79 mono- or di-methylation, and histone H2A variant H2AZ. [25,26] CHIP-seq, DNase-seq, and FAIRE-seq data come from the Encode[27] project. Transcription factor binding analysis was conducted using TRANSFAC® Match with the 2014.4 matrix library and high quality matrices from the vertebrate_nonredundant profile, using binding sites reported to minimize the sum of false positive and false negative error rates (minSUM setting).

2.3.8 Luciferase Assays

For each SNP, a DNA fragment including the 250 bases upstream and downstream (500 bases total) was purchased from Integrated DNA Technologies (Coralville, IA). DNA fragments and pGL4.23[*luc2/*minP] and pGL4.13[*luc2/*SV40] vectors (Promega; Madison, WI) were digested with EcoRV restriction enzyme (New England Biolabs; Ipswich, MA), ligated using T4 Ligase (New England Biolabs; Ipswich, MA), and transformed into DH5alpha competent cells. Plasmids containing DNA inserts in both forward and reverse orientations in pGL4.23 and pGL4.13 vectors were purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific; Waltham, MA).

Each plasmid or empty vector was transfected into HEK293 cells using 0.5 μl Lipofectamine 2000 (Life Technologies, Carlsbad CA) per well, along with pGL4.74 [*hRluc*/TK] (Promega; Madison, WI) control vector in a 96 well plate at a 20:1 ABC Luciferase Plasmid:Renilla ratio. Luciferase and Renilla bioluminescence was measured using the *Dual-Luciferase*® Reporter *Assay* System (Promega; Madison, WI), on a GloMax 96 microplate Dual Injector Luminometer (Promega, Madison, WI) 24 hours after transfection. For all luciferase assays, HEK293 cells were cultured in High-glucose Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum. Opti-Minimal Essential (OptiMEM®) reduced-serum medium (Life Technologies; Carlsbad, CA) was used during transfections.

Relative luciferase activity for each plasmid was measured by first normalizing luciferase bioluminescence (in relative light units, RLU) to *Renilla* bioluminescence for each well. Luciferase activity was then normalized to the mean of luciferase activity for six replicates of empty vector. To test statistical significance of the difference in relative luciferase activity between constructs and empty vector the nonparametric two-sided Wilcoxon rank sum (Mann-Whitney) test was used.

2.3.9 Supplemental Expression Dataset

To filter the list of transcription factors, an additional dataset consisting of gene expression (FPKM) estimates from a transcriptome sequencing study conducted on a subset of the kidney samples was used[28]. For this study, library preparation

was conducted on 100 bp, poly-A selected, paired end reads prepared using a strand specific (dUTP) protocol. Sequencing was conducted on an Illumina (Hayward, CA) HiSeq 2000. Reads were aligned using Tophat2(v2.1.1)[29], and gene expression quantitation was conducted using Cufflinks(v2.0.2). Further details regarding this dataset are included in Chapter 3.[30,31].

2.3.10 Liver Genotype and Expression Dataset

In addition to the kidney dataset, a liver dataset was generated consisting of 34 samples with genotype and ABC transporter expression data after quality control measures. All sample preparation and analysis details for this dataset are identical to those applied to kidney samples.

2.4 Results

2.4.1 Genotype and expression dataset for ABC transporters in human kidney

Gene expression was measured by qPCR for 58 healthy human renal cortex samples from post-mortem or post-surgical resection. Principal components analysis (PCA) identified two outlier samples that were excluded from further analysis (Figure 2.1). Further details regarding normalization are included in the Methods. Samples came from both males and females aged three to ninety years. Expression of no ABC transporters were significantly associated with gender after multiple testing correction. ABCC7 expression was significantly associated with age ($p = 7.51E-06$), with an increase in mRNA expression with age (Figure 2.2). Association between ABC transporter expression and age/gender with a nominal $p<0.05$ are included in Tables 2.3 and 2.4.

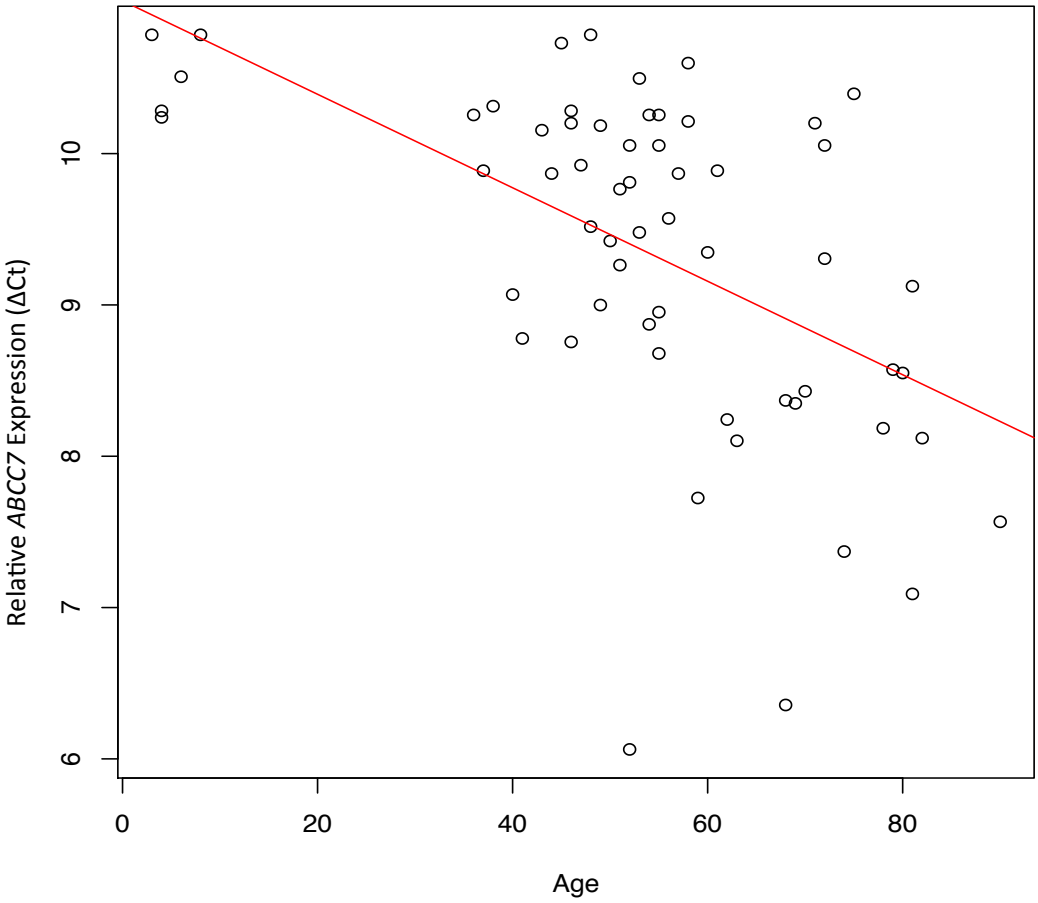


Figure 2.2: Association between *ABCC7* mRNA levels and age. mRNA levels are expressed relative to the geometric mean of three housekeeping genes (ΔCt) and are significantly associated with age of donor (p-value = 7.51E-06), with an increase in *ABCC7* expression with increasing age.

Table 2.3: Association of ABC transporter expression with age in renal cortex.

Gene	P-value	Beta[1]	Adjusted R^2
ABCA1	0.002	-5.72	0.133
ABCB10	0.032	3.83	0.061
ABCB6	0.003	-7.28	0.126
ABCC1	0.035	-6.12	0.058
ABCC3	0.044	-4.58	0.052
ABCC7	7.51e-06	-9.53	0.282
ABCF1	0.006	-8.98	0.108

[1]Negative Beta value represents increase in gene expression with increasing age

Table 2.4: Association of ABC transporter expression with gender in renal cortex.

Gene	P-value	Median Difference in Ct[1]
ABCA12	0.013	-1.95 [-3.34, -0.33]
ABCA5	0.009	0.60 [0.19, 1.04]
ABCA8	0.049	0.49 [6.00 e-05, 9.82e-01]
ABCB10	0.018	0.66 [0.12, 1.24]
ABCD3	0.040	-0.25 [-4.92e-01, -3.05e-05

[1]Median pairwise difference in Ct (Hodges-Lehmann estimator) between genders, expressed as Ct in males relative to females.

2.4.2 cis-eQTL analysis identifies genomic loci associated with expression of six ABC transporters

eQTL analysis was conducted by linear regression. Six *cis*-eQTLs (within 50 kb or 500 kb of the transcription start and stop site of each transporter) with adjusted p-value < 0.10 were identified in *ABCA9, ABCB8, ABCA10, ABCD1, ABCG1,* and *ABCG2* (Table 2.5, Figure 2.3). Minor allele frequencies for these eQTLs ranged from 7 to 50%, and percent of variance in transporter expression explained by the eQTL ranged from 9 to 24%.

Table 2.5: ABC transporter *cis*-eQTLs in renal cortex tissue

Analysis	Gene	Genotyped SNP	Location	MAF[1]	Adjusted R^2	Beta[5]	P (adj)[2]
50 kb	*ABCA9*	rs11077922	4.5 kb 5'	0.45	0.136	0.80	0.008
50 kb	*ABCB8*	rs7783954	50 kb	0.44	0.093	-0.48	0.100[3]
500 kb	*ABCA10*	rs16973316	340 kb	0.07	0.202	2.81	0.042
500 kb	*ABCD1*	rs12841019	400 kb	0.50	0.142	-0.60	0.013
500 kb	*ABCG1*	chr21:42428511	60 kb	0.18	0.237	-1.17	0.071[4]
500 kb	*ABCG2*	rs6532024	240 kb	0.36	0.138	-1.45	0.046[3]

[1] Minor allele frequency
[2] P value adjusted by the number of independent tests (#haplotype blocks+#singletons) in region tested
[3] Includes age as covariate
[4] Includes sex as covariate
[5] Beta for association with ΔCt; negative value corresponds to an association with *increase* in transporter expression

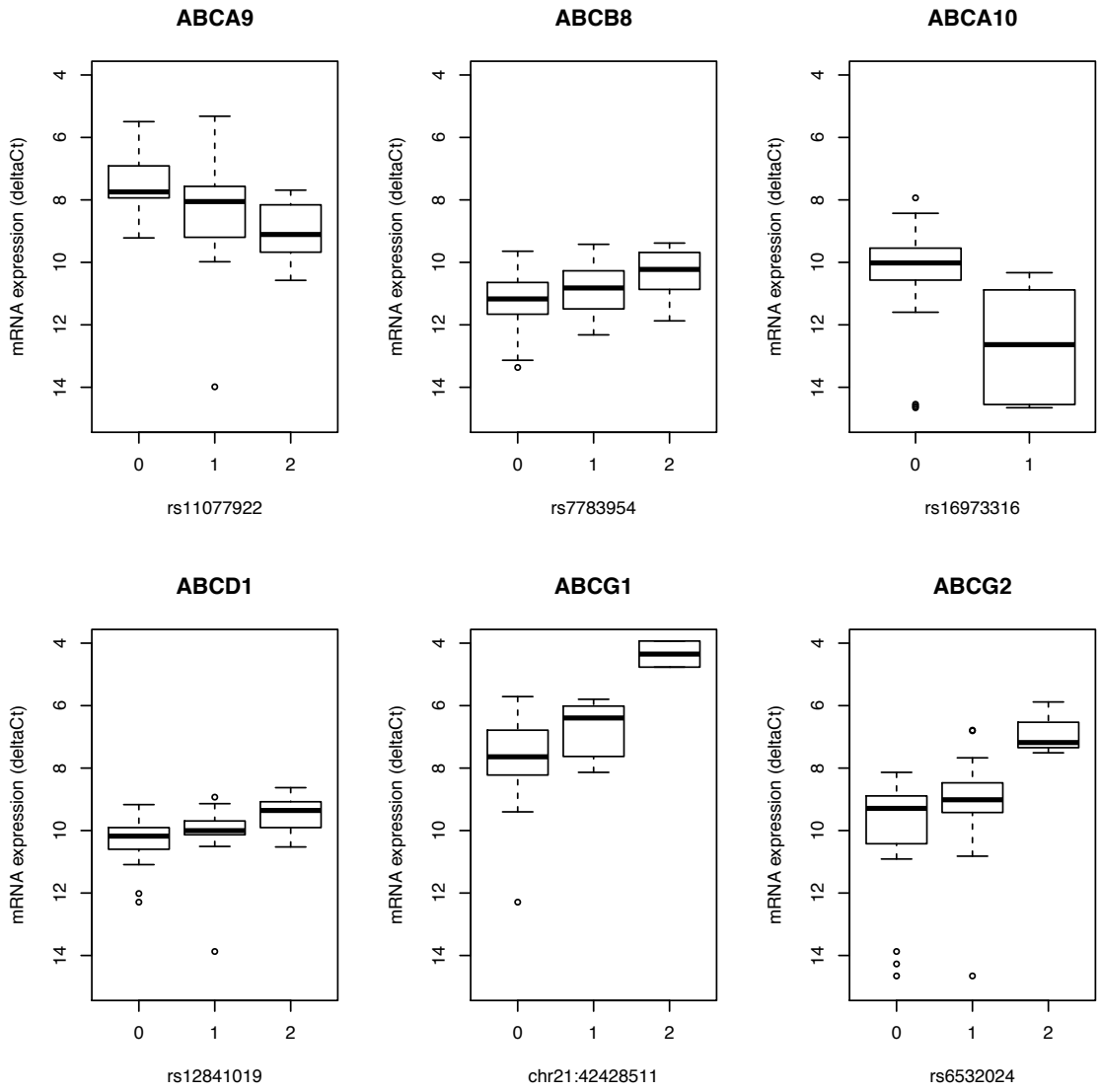


Figure 2.3. Human kidney mRNA expression by genotype for ABC transporters with significant *cis*-eQTLs. mRNA expression is given as the ΔCt and genotype is expressed as the number of minor alleles. Note that ΔCt is inversely correlated with mRNA expression.

2.4.3 rs1471400 modulates ABCG2 expression in vivo and transcriptional activity in vitro

For each of these eQTLs, the LD block around each SNP ($r^2 > 0.8$) was examined to identify regions with evidence suggestive of functional activity (protein binding predicted by ChIP-seq, open chromatin state by DNase-seq or FAIRE-seq, histone modifications indicative of active transcription by ChIP-seq, and/or putative change in transcription factor binding based on position weight matrix). Four SNPs in regions with the most evidence for functional activity were prioritized, three for the *ABCA10* eQTL (rs72631340, rs8078115, rs16967205) and one for the *ABCG2* eQTL (rs1471400). (Table 2.6 and Figure 2.4)

Table 2.6: Predicted ABC transporter enhancer regions

Gene	Genotyped SNP	SNP	R^2	Evidence for Enhancer Activity[+]
ABCG2	rs6532024	rs1471400	1.0	CTCF binding in 18 cell lines, SMARCA4 binding in 1 cell line, Sensitivity to DNase I in 19 cell lines, open chromatin in 5 cell lines (FAIRE-Seq), H2az mark in 11 cell lines, H3K4 monomethylation in 6 cell lines, H3K9 acetylation in 2 cell lines, H3K27 acetylation in 1 cell line, H3K4 dimethylation in 3 cell lines, H3K9 monomethylation in 1 cell line, H4K20 monomethylation in 3 cell lines, H3K79 dimethylation in 1 cell line, change in binding of 1 transcription factors (PWM)
ABCA10	rs16973316	rs72631340	1.0	MAFK binding in 1 cell line, sensitivity to DNase I in 1 cell lines, open chromatin in 1 cell lines (FAIRE-Seq), H2az mark in 9 cell lines, H3K4 monomethylation in 1 cell line, change in binding of 2 transcription factors (PWM)
ABCA10	rs16973316	rs8078115	1.0	CEBPB binding in 3 cell lines, STAT3 binding in 1 cell line, DNase I in 35 cell lines, H2az mark in 11 cell lines, H3K4 monomethylation in 8 cell lines, H3K4 trimethylation in 1 cell line, H3K27 acetylation in 5 cell lines, H3K4 dimethylation in 4 cell lines, H3K9 acetylation in 2 cell lines, open chromatin in 1 cell lines (FAIRE-Seq), change in binding of 4 transcription factors (PWM)
ABCA10	rs16973316	rs16967205	1.0	CTCF binding in 1 cell line, sensitivity to DNase I in 23 cell lines, H2az mark in 7 cell lines, H3K4 monomethylaion in 1 cell line, H3K27 acetylation in 1 cell line, H3K4 dimethylation in 1 cell line, H3K9 acetylation in 1 cell line, change in binding of 4 transcription factors (PWM))

[+]H3K* - histone 3 lysine #; H2az - histone 2 variant

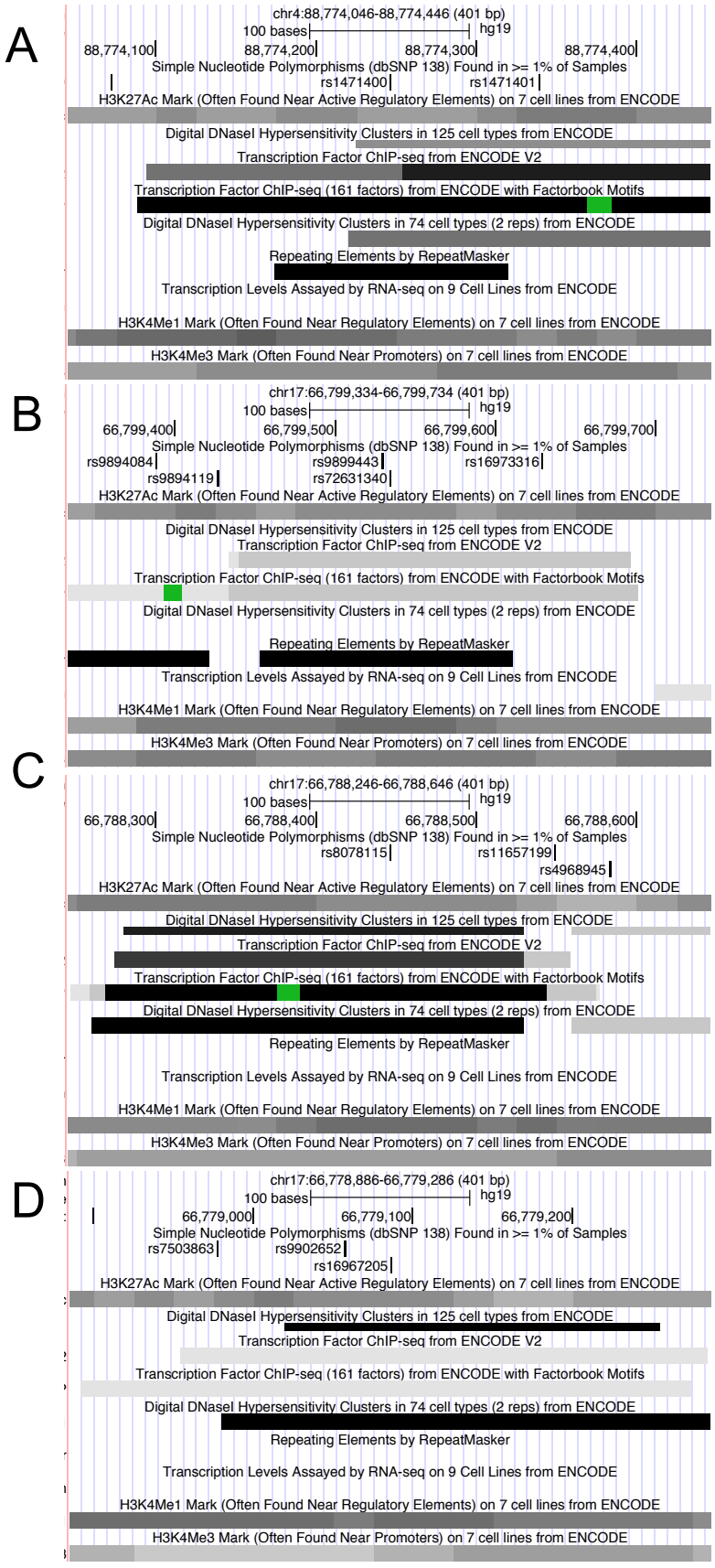
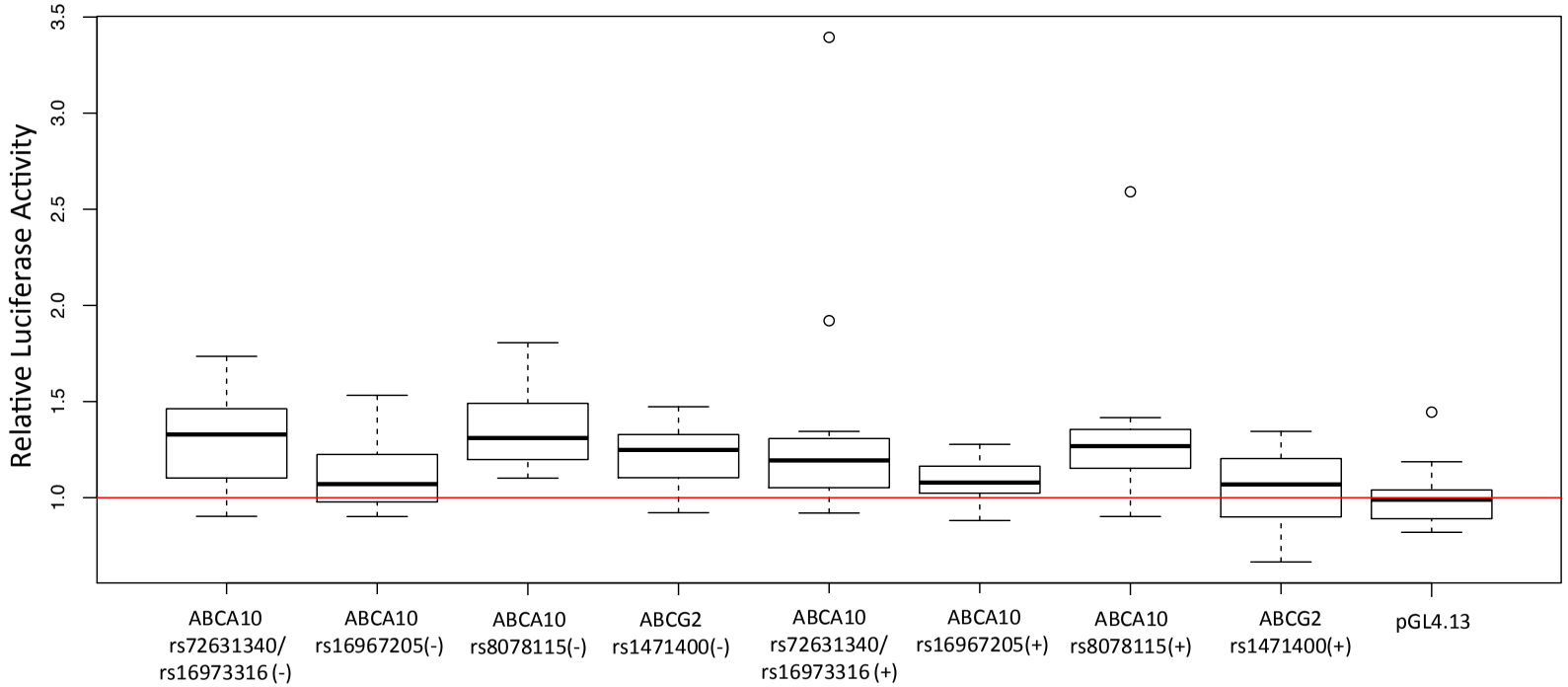


Figure 2.4. Evidence of enhancer activity for A: SNP rs1471400, B: SNP rs72631340, C: rs8078115, and D: rs16967205. Images come from UCSC Genome Browser GRCH37/hg19 assembly with tracks showing dbSNP 138 polymorphisms with minor allele frequency > 1% and Integrated Regulation from ENCODE.

For each of these regions, the 500 bp region around the SNP was cloned into a luciferase vector with a minimal promoter (pGL4.23) and an SV40 promoter (pGL4.13) and transfected into a human embryonic kidney cell line (HEK293) to look for increases or decreases in luciferase activity. No putative regulatory regions showed silencer activity in pGL4.13 vectors. (Figure 2.5) The *ABCG2* eQTL region in pGL4.23 significantly increased luciferase activity, indicating enhancer function (Figure 2.6).



* Expressed Relative to pGL4.13 empty vector and pGL4.74 Renilla control vector

Figure 2.5. Silencer activity of eQTL regions *in vitro*. Luciferase activity was measured for plasmids containing a 500 bp region around putative eQTL SNPs in both forward (+) and reverse (-) orientations relative to the direction of transcription of the luciferase gene. Luciferase activity is normalized to *Renilla* and expressed relative to empty vector (pGL4.13). SNPs rs72631340 and rs16973316 are 97 bases apart and fall within the same tested enhancer region; the enhancer block is 250 bp 5' and 3' of rs16973316. No tested regions showed silencer activity. Results shown for six replicates on one day; results from subsequent experiments support these results.

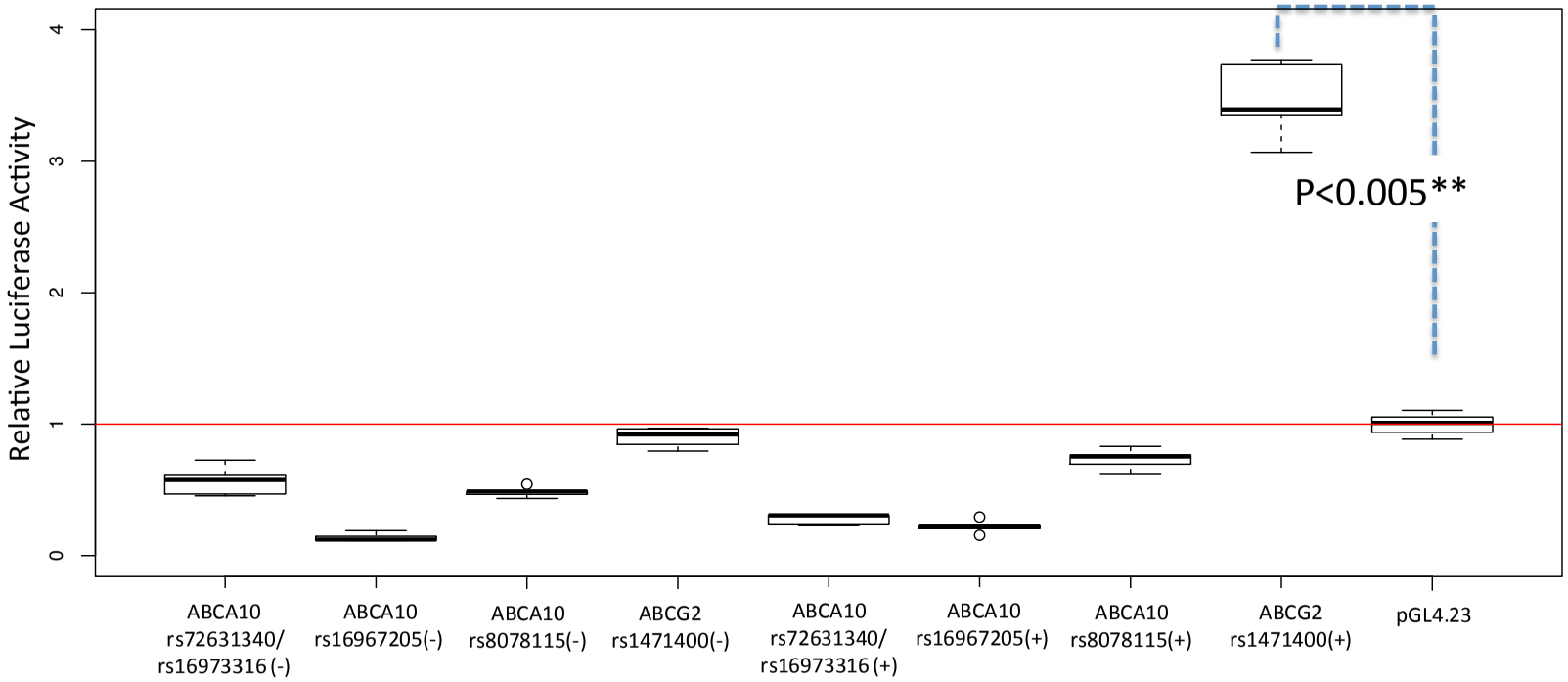


Figure 2.6. Enhancer activity of eQTL regions *in vitro*. Luciferase activity is shown for plasmids containing a 500 bp region around putative eQTL SNPs in both forward and reverse orientations relative to the direction of transcription of the luciferase gene, expressed relative to empty vector (pGL4.23) and *Renilla* activity. SNPs rs72631340 and rs16973316 are 97 bases apart and fall within the same tested enhancer region; the enhancer block is 250 bp 5' and 3' of rs16973316. The region around SNP rs1471400 in forward orientation shows significant increase in luciferase activity relative to empty vector. Results shown for six replicates on one day; results from subsequent experiments support these findings.

Enhancer activity of *ABCG2* eQTL region with rs1471400 minor allele (G>A) was also examined. Enhancer activity was slightly but significantly lower for the rs1471400 allele G (pGL4.23.ABCG2.rs1471400G) relative to allele A (pGL4.23.ABCG2.rs1471400A) (Figure 2.7), in accordance with eQTL results.

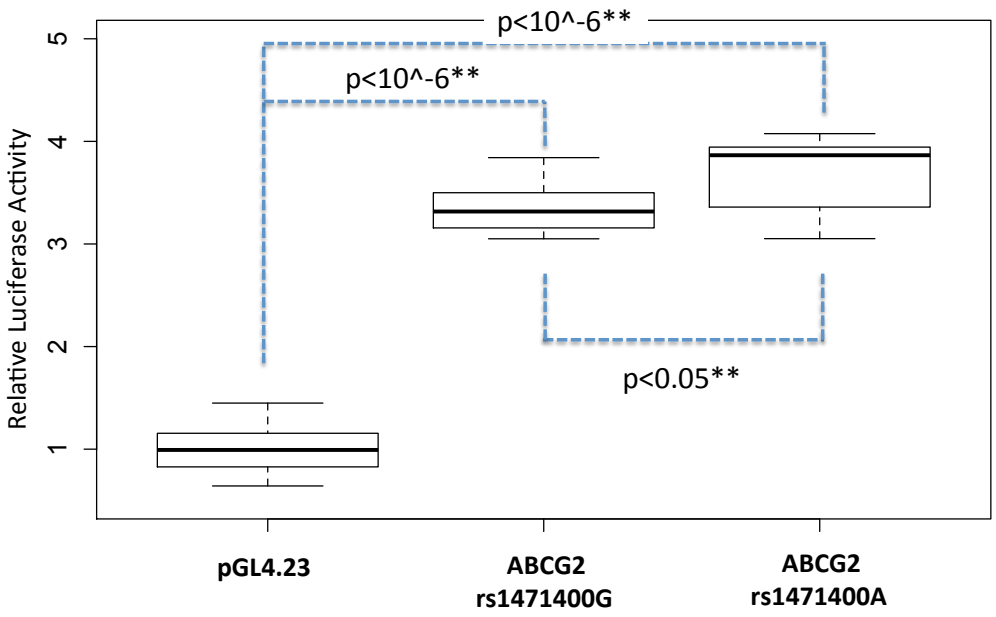


Figure 2.7. Effect of rs1471400 on enhancer activity *in vitro*. Luciferase activity is shown for plasmids containing a 500 bp region around putative ABCG2 eQTL SNP rs1471400, expressed relative to empty vector (pGL4.23) and *Renilla* activity. Plasmids containing both reference (G) and minor (A) alleles for rs1471400 were tested. Plasmids containing reference allele (rs1471400G) result in a decrease in luciferase activity relative to alternate allele (rs1471400A). Results shown for six replicates on one day; results from subsequent experiments support these results.

2.4.4 Polymorphisms in distal genes modulate expression of ABC transporters

Two gene-based *trans*-eQTL analyses were also conducted, looking for combined effects of SNPs in distal genes on ABC transporter expression. First, a set of 796 transcription factors expressed in the human kidney was evaluated, looking at the combined effect of linkage disequilibrium filtered SNPs in each transcription factor on ABC transporter expression. Significance of each transcription factor-transporter association was evaluated by permutation. Transcription factor-QTLs (tfQTLs) are listed in Table 2.7.

Table 2.7: ABC-transporter transcription factor eQTLs in renal cortex

Transporter	Transcription Factor	# SNPs Used	Empirical P[1]
ABCA1	ZSCAN16	1	<0.001
ABCA10	SLC26A3	1	<0.001
ABCA10	TSHZ3	1	<0.001
ABCA10	USF1	1	<0.001
ABCA12	MYOCD	1	<0.001
ABCA2	IRF2	1	<0.001
ABCA2	ZHX1	1	<0.001
ABCA3	T	1	<0.001
ABCA4	ASCL2	1	<0.001
ABCA4	ATF1	1	<0.001
ABCA6	ASCL2	1	<0.001
ABCA6	GRHL3	1	<0.001
ABCA6	NKX6-1	1	<0.001
ABCA6	ZHX1	1	<0.001
ABCA7	SMAD3	1	<0.001
ABCA7	ZBTB7B	1	<0.001
ABCA7	ZFP36L2	1	<0.001
ABCA8	AFF1	1	<0.001
ABCB1	HSF5	1	<0.001
ABCB10	MYOCD	1	<0.001
ABCB4	POU4F1	1	<0.001
ABCB7	ATF5	1	<0.001
ABCB7	PBX1	6	<0.001
ABCC2	IRF2	1	<0.001
ABCC3	RORB	1	<0.001
ABCC4	SIN3A	1	<0.001
ABCC6	BTBD14A	1	<0.001
ABCC6	T	1	<0.001
ABCC7	RELB	1	<0.001
ABCC8	ZNF140	1	<0.001
ABCD1	BTG2	1	<0.001
ABCF1	FOXR1	1	<0.001
ABCF1	VDR	1	<0.001
ABCF1	ZHX1	1	<0.001
ABCG4	EGR2	1	<0.001
ABCG4	ZFHX2	1	<0.001

[1] p<0.001 is the lowest detectable p-value with 1000 permutations

Next, all genes in the genome were evaluated for an association with ABC transporter expression, looking for the effect of the most significant SNP in each gene. Two gene-eQTLs were identified that met the Bonferroni-corrected threshold of 2.8E-06 (0.05/17,787) (Table 2.8).

Table 2.8: ABC-transporter gene-eQTLs in renal cortex

Gene	Gene-eQTL	Empirical P	Top SNP in Gene	Top SNP P
ABCC10	*ARHGAP27*	$< 10^{-6}$	rs16939964	1.8E-06
ABCB7	*ERN1*	1e-06	rs8076809	3.3E-06

Because the Bonferroni-corrected thresholds are quite conservative, all transcription factor and gene based eQTLs down to an empirical p-value of 1E-04 are included in Table 2.9.

Table 2.9: Expanded list of ABC-transporter gene eQTLs in renal cortex

Gene[1]	Gene-eQTL	Number of SNPs	Empirical P	Top SNP in Gene	Top SNP P
ABCC10	*ARHGAP27*	5	<1E-06	rs16939964	1.80E-06
ABCB7	*ERN1*	10	1.00E-06	rs8076809	3.30E-06
ABCC6	*CBX8*	1	3.00E-06	rs8080723	5.20E-06
ABCA8	*C10orf141*	16	4.00E-06	rs2489391	7.60E-07
ABCC6	*CBX2*	1	6.00E-06	rs8080723	5.20E-06
ABCB7	*C9orf89*	1	9.00E-06	rs10761192	1.10E-05
ABCB7	*SUSD3*	1	9.00E-06	rs10761192	0.000011
ABCB10	*ACOT4*	3	1.00E-05	rs6574129	0.0000018
ABCB7	*NINJ1*	1	1.10E-05	rs10761192	0.000011
ABCA6	*ACOT4*	3	2.00E-05	rs6574129	0.000011
ABCA6	*VGLL3*	13	2.20E-05	rs9863011	0.0000015
ABCA13	*GPR45*	19	2.20E-05	rs10496390	0.000021
ABCC2	*ARFGEF1*	6	2.30E-05	rs891589	0.000012
ABCA10	*FAM46A*	18	2.70E-05	rs3736841	6.40E-09
ABCA10	*DLX1*	6	2.90E-05	rs12692981	0.000024
ABCB9	*PLEKHO1*	2	3.00E-05	rs12124389	0.0000032
ABCB7	*LINGO1*	7	3.00E-05	rs13313467	0.0000047
ABCA13	*TKT*	3	3.00E-05	rs6769824	0.0001171
ABCA5	*ACOT4*	3	3.10E-05	rs6574129	0.0000036
ABCA10	*TACC1*	13	3.40E-05	rs7816768	0.000015
ABCB7	*ZNF425*	3	3.70E-05	rs1202453	0.0001407
ABCC10	*PADI1*	10	4.00E-05	rs2977268	0.0000014
ABCB11	*GSTA1*	7	4.20E-05	rs6917325	0.000026
ABCB9	*CRMP1*	45	4.30E-05	rs3885409	0.000022
ABCD3	*C7orf54*	4	4.70E-05	rs17151639	0.000021
ABCB7	*FAM59A*	41	4.70E-05	rs16963326	0.0000039
ABCD2	*SF3B3*	4	5.20E-05	rs7190039	0.000075
ABCC2	*DDT*	5	5.20E-05	rs4822458	0.000045
ABCA10	*DLX2*	10	5.50E-05	rs12692981	0.000024
ABCB5	*ZNF558*	8	6.10E-05	rs2967747	0.000046
ABCB11	*GSTA5*	12	6.20E-05	rs6917325	0.000026
ABCA10	*SLC7A1*	23	6.60E-05	rs3011617	0.0000053
ABCC7	*6-Mar*	5	6.70E-05	rs1875017	0.0002833
ABCB1	*PFN1*	5	6.70E-05	rs17706762	0.000047
ABCC7	*RELB*	2	6.80E-05	rs35891370	0.000336
ABCA10	*FLJ33790*	4	6.80E-05	rs4348932	0.0000002
ABCE1	*FBXL5*	6	7.20E-05	rs12649803	0.0005051
ABCB1	*RNF167*	5	7.20E-05	rs17706762	0.000047
ABCB1	*GP1BA*	5	7.30E-05	rs17706762	0.000047

ABCD4	*CD5L*	9	7.50E-05	rs17690088	0.0006457
ABCB1	*SLC25A11*	5	7.50E-05	rs17706762	0.000047
ABCC2	*GSTT2-1*	6	7.60E-05	rs4822458	0.000045
ABCB1	*ENO3*	5	8.00E-05	rs17706762	0.000047
ABCC2	*GSTT2B*	6	8.10E-05	rs4822458	0.000045
ABCD4	*FCRL1*	6	8.30E-05	rs17690088	0.0006457
ABCB7	*HPS3*	22	8.40E-05	rs2681092	0.0001323
ABCC2	*DDTL*	6	8.70E-05	rs4822458	0.000045
ABCG8	*FGF10*	18	8.80E-05	rs6883600	0.000245
ABCB1	*INCA1*	6	9.10E-05	rs17706762	0.000047
ABCA6	*ZHX1*	6	9.40E-05	rs4871353	0.000054
ABCB7	*CNPY1*	12	9.60E-05	rs11760611	0.000013
ABCA4	*TRIM39*	142	1.00E-04	rs3130139	0.0002185

[1]All eQTLs with a p < 1E-04 are shown.

2.4.5 cis-eQTL analysis in liver replicates published associations

A liver genotype and expression dataset for 34 samples was also generated. Several large liver eQTL studies have been published previously. *cis*-eQTL (50 kb upstream and downstream of each ABC transporter) analyses were conducted to replicate published findings and validate the methodology used in the kidney analysis. In the liver samples, associations between rs12941297 and ABCA5 expression and rs4437575 and ABCB4 expression were detected, validating the earlier findings from Innocenti et al.[8] (Table 2.10).

Table 2.10: Previously published cis-eQTLs identified in liver samples

Analysis	Gene	Published SNP	Genotyped SNP	LD[1] (R^2)	MAF[2]	Beta[3]	P (adj)[4]
50 kb	*ABCA5*	rs12941297	rs12938097	1.00	0.49	0.87	0.0004
50 kb	*ABCB4*	rs4437575	rs2235038	0.93	0.44	-0.86	0.0366[5]

[1] Linkage disequilibrium with published SNP
[2] Minor allele frequency
[3] Beta for association with ΔCt; negative value corresponds to an association with *increase* in transporter expression
[4] P value adjusted by the number of independent tests (#haplotype blocks + #singletons) in region tested
[5] Includes age as a covariate

2.5 Discussion

Genetic markers associated with ABC transporter expression in the human kidney were identified in the current study. In the first stage of analysis, SNPs within 50 kb of each transporter were considered. For example, a *cis*-eQTL for *ABCA9* was identified in the promoter region of the gene. Similar to the other genes in the ABCA gene family, *ABCA9* is a cholesterol responsive gene that is predicted to be involved in lipid homeostasis[32]. rs11077922 is associated with decreased expression of ABCA9, and is predicted to create a binding site for kidney-specific transcription factor Evi-1. In the second stage of this analysis, SNPs within 500 kb of each transporter were examined. An eQTL for *ABCB8* was identified ~50 kb downstream in the *AGAP3* gene. *ABCB8* is a mitochondrial membrane protein involved in iron transport[33], while *AGAP3* is an enzyme involved in receptor trafficking[34]. While there are no known interactions between the two genes, a unique eQTL (rs2288652) for *ABCB8* was identified in *AGAP3* in thyroid tissue as a part of the Genotype-Tissue Expression (GTEx) project[35], suggesting that there may be some relationship between these genes.

Functional validation of *cis*-eQTLs revealed one region containing a SNP (rs1471400) associated with expression of ABCG2 with evidence of enhancer activity in *in vitro* luciferase assays. The magnitude of effect in the *in vitro* assay may be muted because the effect of age in the system was not incorporated into the assay (the association in tissues is corrected for age of patients). The variant allele at the polymorphic position was associated with an increase in ABCG2 expression in kidney samples and an increase in enhancer activity *in vitro*,

suggesting that this variant contributes to inter-individual differences in expression of ABCG2. Computational analysis of predicted transcription factor binding showed a predicted decrease in binding of transcription factor ING4, a transcriptional repressor[36]. Further, there is some evidence that ING4 expression is inversely correlated with age through the actions of one of its regulators, miR-650[37], possibly explaining the role of age in association of ABCG2 expression with genotype.

ABCG2 encodes the breast cancer resistance protein (BCRP); BCRP is involved in transport of a broad range of substrates, including natural compounds such as urate[38] and folic acid[39], as well as a number of drugs and drug metabolites, including several antivirals[40,41], statins[42], and chemotherapeutics[43,44]. In the kidney, BCRP is expressed on the apical membrane of proximal tubule epithelial cells, and plays a key role in renal elimination of a number of compounds. The transporter has also been implicated in protection of cells against hypoxic conditions[45] and in regeneration after tissue damage[46]. Subtle differences in expression of this transporter between individuals may be responsible for differences in systemic or local exposure to drugs excreted by BCRP or susceptibility to tissue damage. While most studies of clinical impact of *ABCG2* polymorphisms have focused on variants within or immediately adjacent to the gene, the current findings suggest that distal variants, such as the one identified, may also play a role in modulating clinical outcomes associated with BCRP activity. Such distal polymorphisms may account for some of the unexplained variation between individuals in disposition and toxicity of BCRP substrates.

tfQTLs and gene-QTLs were also identified for renal ABC transporters. An association between polymorphisms in the *RELB* transcription factor and expression of the transporter *ABCC7,* which encodes the cystic fibrosis transmembrane conductance regulator (CFTR), was identified in both transcription factor- and genome wide *trans*-eQTL analyses. CFTR is a chloride transporter, and in the kidney plays a key role in chloride secretion, regulation of salt and water balance, and potentially transport of drugs[47]. RELB is a member of the NF-κB family, and acts as a transcriptional regulator when bound to p50 or p52[48]. While NF-κB is a known transcriptional regulator of *ABCC7,* the experimental evidence for this association has been generated with the more common RELA-p50 heterodimer[49]. These results suggest that RELB heterodimers also play a role in transcriptional regulation of *ABCC7*, and polymorphisms in RELB, like the one identified, may modulate inter-individual expression of NF-κB targets.

It is important to note that the sample size was small, and thus it is likely there are additional kidney eQTLs for the ABC transporter family that were not detected in this study. Further, because of this small sample size, an alpha (significance threshold) of 10% was applied, rather than the more standard 5%, as these results are intended to generate hypotheses which can be further explored in additional datasets or experimental studies. Despite this limitation, several *cis*-eQTLs for ABC transporters were identified in the human kidney, and one of these eQTLs was validated *in vitro* for the transporter BCRP (*ABCG2)*. Further, the ability to detect previously identified SNP-gene associations in a

smaller liver tissue set suggests that such associations can be detected even with small sample sizes. In addition, while the current study was not powered to conduct a complete *trans*-eQTL analysis, transcription factor-QTLs for ABC transporters were identified. The transcription factor-QTLs provide novel hypothesis about the transcriptional regulation of ABC transporters in the kidney.

2.6 References

1 Huls M, van den Heuvel JJMW, Dijkman HBPM, Russel FGM, Masereeuw R. ABC transporter expression profiling after ischemic reperfusion injury in mouse kidney. *Kidney Int* 2006; **69**: 2186–93.

2 Huls M, Russel FGM, Masereeuw R. The Role of ATP Binding Cassette Transporters in Tissue Defense and Organ Regeneration. *Pharmacology* 2009; **328**: 3–9.

3 Kedenko L, Lamina C, Kiesslich T, Kapur K, Bergmann S, Waterworth D, Heid IM, Wichmann H-E, Kedenko I, Kronenberg F, Paulweber B. Genetic polymorphisms of the main transcription factors for adiponectin gene promoter in regulation of adiponectin levels: association analysis in three European cohorts. *PLoS One* 2012; **7**: e52497.

4 Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* 2012; **8**: e1002639.

5 Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1**: e78.

6 Göring HHH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole S a, Jowett JBM, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007; **39**: 1208–16.

7 Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls M a, Lai S-L, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010; **6**: e1000952.

8 Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson D a, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD. Identification, replication, and functional fine-mapping of expression quantitative trait Loci in primary human liver tissue. *PLoS Genet* 2011; **7**: e1002078.

9 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107.

10 Padi S, Shi X, Zhao Y, Ruff M, Baichoo N, Pert C, Zhang J. Attenuation of rodent neuropathic pain by an orally active peptide, RAP-103, which potently blocks CCR2-and CCR5-mediated monocyte chemotaxis and inflammation. *Pain* 2012; **153**: 95–106.

11 Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* 2013; **144**: 1488–96, 1496.e1–3.

12 Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, Ingle C, Nestle FO, di Meglio P, Min JL, Wilk A, Hammond CJ, Hassanali N, Yang T-P, Montgomery SB, O'Rahilly S, Lindgren CM, Zondervan KT, Soranzo N, Barroso I, Durbin R, Ahmadi K, Deloukas P, McCarthy MI, Dermitzakis ET, Spector TD. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 2011; **7**: e1002003.

13 Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science (80-)* 2009; **325**: 1246–50.

14 Bolstad BM, Irizarry R., Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19**: 185–93.

15 Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, Cawley S, Chung E, Connell S, Eshragh J, Ewing M, Gollub J, Henderson M, Hubbell E, Iribarren C, Kaufman J, Lao RZ, Lu Y, Ludwig D, Mathauda GK, McGuire W, Mei G, Miles S, Purdy MM, Quesenberry C, Ranatunga D, Rowell S, Sadler M, Shapero MH, Shen L, Shenoy TR, Smethurst D, Van den Eeden SK, Walter L, Wan E, Wearley R, Webster T, Wen CC, Weng L, Whitmer RA, Williams A, Wong SC, Zau C, Finn A, Schaefer C, Kwok P-Y, Risch N. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* 2011; **98**: 79–89.

16 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science (80-)* 2002; **296**: 2225–9.

17 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–5.

18 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

19 Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–45.

20 Consortium 1000 Genomes Project. A map of human genome variation from population scale sequencing. *Nature* 2010; **473**: 1061–73.

21 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**: D930–4.

22 Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–7.

23 Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, Fields S, Stamatoyannopoulos JA. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 2009; **6**: 283–9.

24 Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* 2007; **17**: 877–85.

25 Koch CM, Andrews RM, Flicek P, Dillon SC, Karaöz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetrie D, Dunham I. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* 2007; **17**: 691–707.

26 Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007; **129**: 823–37.

27 Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.

28 Gamazon ER, French C, Yee SW, Chhibber A. PGRN Network-wide Project: Transcriptome Analysis of Pharmacogenes in Human Tissues. *Present Am Soc Hum Genet Annu Meet* 2013.

29 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–11.

30 Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; : 1–5.

31 Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; **28**: 511–5.

32 Piehler A, Kaminski WE, Wenzel JJ, Langmann T, Schmitz G. Molecular structure of a novel cholesterol-responsive A subclass ABC transporter, ABCA9. *Biochem Biophys Res Commun* 2002; **295**: 408–16.

33 Ichikawa Y, Bayeva M, Ghanefar M, Potini V, Sun L, Mutharasan RK, Wu R, Khechaduri A, Jairaj Naik T, Ardehali H. Disruption of ATP-binding cassette B8 in mice leads to cardiomyopathy through a decrease in mitochondrial iron export. *Proc Natl Acad Sci U S A* 2012; **109**: 4152–7.

34 Oku Y, Huganir RL. AGAP3 and Arf6 regulate trafficking of AMPA receptors and synaptic plasticity. *J Neurosci* 2013; **33**: 12586–98.

35 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–5.

36 Garkavtsev I, Kozin S V, Chernova O, Xu L, Winkler F, Brown E, Barnett GH, Jain RK. The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. *Nature* 2004; **428**: 328–32.

37 Zeng Z-L, Li F-J, Gao F, Sun D-S, Yao L. Upregulation of miR-650 is correlated with down-regulation of ING4 and progression of hepatocellular carcinoma. *J Surg Oncol* 2013; **107**: 105–10.

38 Woodward OM, Köttgen A, Coresh J, Boerwinkle E, Guggino WB, Köttgen M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* 2009; **106**: 10338–42.

39 Chen Z-S, Robey RW, Belinsky MG, Shchaveleva I, Ren X-Q, Sugimoto Y, Ross DD, Bates SE, Kruh GD. Transport of Methotrexate, Methotrexate Polyglutamates, and 17{beta}-Estradiol 17-({beta}-D-glucuronide) by ABCG2: Effects of Acquired Mutations at R482 on Methotrexate Transport. *Cancer Res* 2003; **63**: 4048–54.

40 Wang X, Furukawa T, Nitanda T, Okamoto M, Sugimoto Y, Akiyama S-I, Baba M. Breast cancer resistance protein (BCRP/ABCG2) induces cellular resistance to HIV-1 nucleoside reverse transcriptase inhibitors. *Mol Pharmacol* 2003; **63**: 65–72.

41 Kim H-S, Sunwoo YE, Ryu JY, Kang H-J, Jung H-E, Song I-S, Kim E-Y, Shim J-C, Shon J-H, Shin J-G. The effect of ABCG2 V12M, Q141K and Q126X, known functional variants in vitro, on the disposition of lamivudine. *Br J Clin Pharmacol* 2007; **64**: 645–54.

42 Huang L, Wang Y, Grimm S. ATP-dependent transport of rosuvastatin in membrane vesicles expressing breast cancer resistance protein. *Drug Metab Dispos* 2006; **34**: 738–42.

43 Ozvegy C, Litman T, Szakács G, Nagy Z, Bates S, Váradi A, Sarkadi B. Functional characterization of the human multidrug transporter, ABCG2, expressed in insect cells. *Biochem Biophys Res Commun* 2001; **285**: 111–7.

44 Elkind NB, Szentpétery Z, Apáti A, Ozvegy-Laczka C, Várady G, Ujhelly O, Szabó K, Homolya L, Váradi A, Buday L, Kéri G, Német K, Sarkadi B. Multidrug transporter ABCG2

prevents tumor cell death induced by the epidermal growth factor receptor inhibitor Iressa (ZD1839, Gefitinib). *Cancer Res* 2005; **65**: 1770–7.

45 Krishnamurthy P, Schuetz JD. Role of ABCG2/BCRP in biology and medicine. *Annu Rev Pharmacol Toxicol* 2006; **46**: 381–410.

46 Israeli D, Ziaei S, Gonin P, Garcia L. A proposal for the physiological significance of mdr1 and Bcrp1/Abcg2 gene expression in normal tissue regeneration and after cancer therapy. *J Theor Biol* 2005; **232**: 41–5.

47 Morales MM, Falkenstein D, Lopes AG. The cystic fibrosis transmembrane regulator (CFTR) in the kidney. *An Acad Bras Cienc* 2000; **72**: 399–406.

48 Dobrzanski P, Ryseck RP, Bravo R. Differential interactions of Rel-NF-kappa B complexes with I kappa B alpha determine pools of constitutive and inducible NF-kappa B activity. *EMBO J* 1994; **13**: 4608–16.

49 Brouillard F, Bouthier M, Leclerc T, Clement A, Baudouin-Legros M, Edelman A. NF-kappa B mediates up-regulation of CFTR gene expression in Calu-3 cells by interleukin-1beta. *J Biol Chem* 2001; **276**: 9486–91.

Chapter 3: Alternative Splicing of the ATP-Binding Cassette Transporter Family Regulates Transporter Expression in Human Tissues

3.1 Abstract

Alternative splicing across tissues and across individuals can regulate transporter expression. In this study, a systematic evaluation of ABC transporter splicing was conducted across individuals and tissues. Transcriptome sequencing data from multiple human tissues were analyzed, and both previously annotated and unannotated splicing events were identified. Among transporters that were variably spliced across individuals or tissues, several were associated with changes in gene expression in *in vitro* assays, including alternate 5'UTRs for the genes *ABCC5* and *ABCA8* that may regulate inter-individual expression. A novel exon in the gene *ABCC6* was shown to trigger transcript decay and may regulate tissue specific expression. These splicing events represent potential pharmacogenetic and human disease markers.

3.2 Introduction

ATP-binding cassette (ABC) transporters are a family of transport proteins expressed on cell and organelle membranes responsible for transport of a broad range of metabolites and xenobiotics across cell membranes. These transporters

play a key role in both drug disposition and homeostasis of endogenous compounds[1]. Variation in expression of ABC transporters can affect these essential functions in the body. The impact of single nucleotide polymorphisms (SNPs) on ABC transporter expression have been studied[2], and are summarized in Chapter 1. Changes in secondary structure, or alternative splicing, of the transporters can also modulate transporter expression. This splicing-mediated regulation of gene expression can occur through changes in the proportion of functional transcripts, changes in the efficiency of translation, or stability of transcripts or proteins. Such splicing may be variable both across tissues[3,4] and across individuals[5–8], leading to differences in tissue-specific or inter-individual gene expression. Thus, comprehensive studies of alternative splicing for a given gene or gene family must examine such events in multiple human tissue types and across multiple individuals.

Alternative transcripts have been identified for many ABC transporters, and the functional impact has been characterized for a small number of events and transporters [9–11]. In this study, the first systematic evaluation of ABC transporter splicing across multiple individuals was conducted in four human tissue types (heart, adipose, liver and kidney) and lymphoblastoid cell lines using transcriptome sequencing data. Putative “functional” splicing of transporters that produces alternate transcripts that, when translated, generate functional protein, as well as “nonfunctional” splicing events that do not produce functional protein were identified. A significant portion of these events are not included in existing

transcript annotations. Further, the impact of specific functional and nonfunctional splicing events on transporter expression was evaluated *in vitro.* These events represent potential pharmacogenetic and human disease markers.

3.3 Materials and Methods

3.3.1 Samples, Library Preparation and Sequencing

Total RNA was extracted using Trizol (Invitrogen; Carlsbad, CA) from 24 liver (hepatocytes), 20 kidney (renal cortex), 25 adipose (subcutaneous), and 25 heart (ventricle) tissue samples, along with 44 lymphoblastoid cell line samples. Further details regarding tissues and RNA isolation have been published[12–18]. All samples used for library preparation had a minimum RNA integrity number (RIN) of 6 or higher assayed by Bioanalyzer. cDNA library preparation and sequencing was conducted at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) for liver, kidney, adipose, and heart tissue samples. Library preparation was conducted as described by Zhong et al.[19] Briefly, mRNA was selected by oligo(dT) and then fragmented. First strand cDNA was generated using Superscript III reverse transcriptase and random hexamer primers, followed by second strand synthesis using DNA polymerase I. Illumina adaptors were ligated to double stranded cDNAs and strand specificity was achieved by incorporating dUTP during second strand synthesis and digestion with Uracil-DNA Glycosylase prior to PCR amplification. Libraries were sequenced on an Illumina HiSeq instrument with five samples per lane. cDNA library preparation for LCLs was conducted by Covance, Inc (Princeton, NJ) according to manufacturer's instructions using the TruSeq RNA Sample Prep Kit v2 (Illumina; San Diego CA) and sequencing was conducted at the University of Washington

Northwest Genomics Center. Sequencing generated 100 base pair (bp) reads for all samples.

3.3.2 Alignment and Alternative Splicing Analysis

Reads were aligned to the human genome using Tophat v2.1.1[20]. Because read depth was variable across samples and tissues, read depth was normalized to 20 million reads per sample by subsampling reads. Transcript assembly for each sample was conducted with Cufflinks v2.0.2[21,22] using the Gencode(v12)[23] transcript annotation to guide assembly. Additional options were included for upper-quartile normalization (--upper-quartile-norm), library type (--library-type fr-firststrand), and maximum bundle length (--max-bundle-length 7500000). Cuffmerge was used to merge all per-sample transcript assemblies to create a single study-wide transcript assembly annotation. This study wide transcriptome annotation was used in the JuncBASE(v0.6)[24] tool along with junction reads identified by Tophat to characterize splicing events in ABC transporters. A Shannon entropy score of 2 was used in JuncBASE as a filter for unannotated splicing events. In addition, Cuffdiff (v2.2.1) (using the option --library-type fr-firststrand) was used to calculate sample and tissue specific mRNA expression estimates (fragments per kilobase per million, FPKM).

3.3.3 Functional annotation of splicing events

Functional annotation was prioritized for splicing events with at least 10 reads per event that were detected in at least two individuals. The median percent spliced in (PSI) was required to be between 10% and 90% in any sample type or high in one tissue and low in another (>90% and <10% in at least two different sample types) for functional consideration. Intron retention events were excluded from further functional annotation. Annotated splicing events are defined as those existing in UCSC (hg19), Gencode (v19), RefSeq (release 59-63), or Ensembl (v75) gene annotations.

For alternate 5'UTRs, RNA regulatory motifs were analyzed using RegRNA[25] and UTRScan[26] with 5'UTR motifs from UTRSite[26]. Minimum free energy of secondary structure estimates from mFold[27](with version 3 free energies and default parameters) and UTR length were also considered.

To identify premature termination codons (PTC), the impact of selected splicing events on coding sequence was examined, assuming the structure of the transcript upstream of each event matches the reference gene annotation. Where multiple transcript structures were available in the annotation databases, the transcript containing the splicing event for annotated events, or the transcript matching the CCDS transcript for unannotated events were used.

Potential single nucleotide polymorphisms (SNPs) that may result in a given alternative splicing event were identified by searching for polymorphisms (dbSNP 138 with minor allele frequency > 1%) at splice junctions. SNPs in splicing motifs (exonic and intronic splicing enhancers and silencers) defined in AEdb (the manually curated component of the Alternative Splicing Database[28]) were identified using RegRNA 2.0[29].

3.3.4 Luciferase Assays

For alternative 5'UTRs, each complete 5'UTR region was amplified from pooled human heart cDNA purchased from Clontech (Mountain View, CA) using primers containing NcoI restriction site sequences (Table 3.1) and Phusion High Fidelity DNA Polymerase (Thermo Scientific; Waltham, MA). Primers were purchased from Integrated DNA Technologies (Coralville, IA). Each amplified UTR was then inserted into the NcoI site in the pGL3-promoter vector [*luc*+/SV40] (Promega; Madison, WI), located immediately 5' of the luciferase gene in the vector.

Table 3.1: ABC Transporter 5'UTR Amplification Primers

Primer Name	Primer Sequence[1] **(5' -> 3')**
ABCA8_5UTR_A_F	TAAGCA[CCATGG]ATAACCTCCACTCTGAAAGCAGTC
ABCA8_5UTR_A_R	TGATGT[CCATGG]CTTGTTTCTGGGAAAATGGAAG
ABCA8_5UTR_B_F	TAAGCA[CCATGG]ATAACCTCCACTCTGAAAGCAGTC
ABCA8_5UTR_B_R	TGATGT[CCATGG]CTTGTTTCTGGGAAAATGGAAG
ABCC5_5UTR_A_F	TAAGCA[CCATGG]GGCCGATGCCGCTATAAA
ABCC5_5UTR_A_R	TGATGT[CCATGG]CATCTTCTCTGAGTGGAGGTTCC
ABCC5_5UTR_B_F	TAAGCACCATGGTTAAAGATAAGTAACAGCTATATCAACTTAGGG
ABCC5_5UTR_B_R	TGATGTCCATGGCTTCTCTGAGTGGAGGTTCCA
ABCC5_5UTR_C_F	TAAGCA[CCATGG]GAATTCTGATGTGAAACTAACAGTC
ABCC5_5UTR_C_R	TGATGT[CCATGG]TTCTCTGAGTGGAGGTTCCA

[1]NcoI restriction site shown in brackets

PCR conditions were as follows: one cycle for 60 sec at 98°C, followed by 40 cycles of 10 sec at 98°C, 30 sec at 60°C, and 30 sec at 72°C, and a final extension for 10 min at 72°C. Amplified UTRs and pGL3-promoter vector was digested with NcoI restriction enzyme (New England Biolabs; Ipswich, MA), ligated using T4 Ligase (New England Biolabs; Ipswich, MA), and transformed into DH5alpha competent cells. pGL3-promoter-UTR plasmids were purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific; Waltham, MA).

Each plasmid or empty pGL3-promoter vector was transfected into HEK293 cells using 0.5 μl Lipofectamine 2000 (Life Technologies, Carlsbad CA) per well, along with pGL4.74 [*hRluc*/TK] (Promega; Madison, WI) control vector in a 96 well plate at a 20:1 pGL3-promoter-UTR:pGL4.74 ratio. Luciferase and *Renilla*

bioluminescence were measured using the *Dual-Luciferase*® Reporter *Assay* System (Promega; Madison, WI), on a GloMax 96 microplate Dual Injector Luminometer (Promega, Madison, WI) 24 hours after transfection. For all luciferase assays, HEK293 cells were cultured in High-glucose Dulbecco's modified Eagle's medium (DMEM) (University of California San Francisco Cell Culture Facility) with 10% fetal bovine serum (Axenia BioLogix, Dixon CA) Opti-Minimal Essential (OptiMEM®) reduced-serum medium (Life Technologies; Carlsbad, CA) was used during transfections.

Relative luciferase activity for each UTR plasmid was calculated by first normalizing luciferase bioluminescence (in relative light units, RLU) to *Renilla* bioluminescence for each well. Relative luciferase activity for all plasmids was then expressed to the mean relative luciferase activity for six replicates of the plasmid containing the reference gene annotation (ABCC5_5UTR_A and ABCA8_5UTR_A). To test statistical significance of the difference in relative luciferase activity between un-annotated and reference UTR constructs was tested using a nonparametric two-sided Wilcoxon rank sum (Mann-Whitney) test.

3.3.5 Nonsense Mediated Decay Assays

Splicing events that created a PTC > 50 bases upstream of an exon-exon junction were evaluated *in vitro* to determine if the putative-PTC junction creates a transcript that is a target of the nonsense mediated decay (NMD) process. Cells were treated with the translation inhibitor puromycin to evaluate whether

splicing events were targets of NMD. Primers that overlap exon-exon junctions for both the putative PTC and the non-PTC event were used for qPCR. HepG2 or HEK293 cells were treated with 100 μg/ml puromycin (Life Technologies; Carlsbad, CA) for 6 hours, and RNA was extracted from cells with and without puromycin treatment with the Qiagen RNeasy MiniKit (Valencia, CA). cDNA was prepared using the iScript Select or iScript cDNA synthesis kits (Bio-Rad Laboratories Inc; Hercules, CA) using 1 μg RNA. qPCR assays were run using event specific primers (Table 3.2) (Integrated DNA Technologies; Coralville, IA) with the Maxima SYBR Green/ROX master mix (Thermo Scientific; Waltham, MA) using standard cycling protocols on an ABI7900 Real Time PCR System (Life Technologies; Carlsbad, CA). For all puromycin treatment assays, HepG2 and HEK293 cells were cultured in DMEM or RPMI-1640 (University of California San Francisco Cell Culture Facility) with 10% fetal bovine serum (Axenia BioLogix, Dixon CA).

Table 3.2: ABC Transporter Nonsense Mediated Decay Primers

Primer Name	Primer Sequence **(5' -> 3')**
ABCC3_Ext_PTC_F	CAGCGCTTCATGGACCTT
ABCC3_Ext_PTC_R	TACCAAAGTCCTGCCAGAGG
ABCC3_Ext_F	CAGCGCTTCATGGACCTT
ABCC3_Ext_R	CAAAGTCCTGCCAGAGGAAG
ABCC2_Trunc_PTC_F	GCCGGTGGTCAGATTATCA
ABCC2_Trunc_PTC_R	GAACAGGATGGGGTCTGG
ABCC2_Trunc_F	GATTATCATTGATGGAGTAGATATTGC
ABCC2_Trunc_R	GGGTCCTGGGGGATGAT
ABCA6_PTC_F	CCAAGATTCAGCATTCCAAA
ABCA6_PTC_R	GTGGCTCCCAAGGAATTGT
ABCA6_Trunc_F	AGATTCAGCATTCGCTCTCC
ABCA6_Trunc_R	TCATGACTATAATTTGGGTGAATG
ABCC2_PTC_F	GCCAACTTGTGGCTGTGATA
ABCC2_PTC_R	ATGATCTTGATGGTGATGTGC
ABCC2_F	GCCAACTTGTGGCTGTGATA
ABCC2 _R	GGTGCCTTGATGGTGATGT
ABCC3_PTC_F	GCTCTTTGCTGCACTATTTGC
ABCC3_PTC_R	AGGTCTACCTGCAAGGAGTAGG
ABCC3 _F	GCTGCACTATTTGCCGTCAT
ABCC3 _R	GCAAATGTCACCCTGCAA
ABCC6_PTC_F	GCCAAGATGAGAAACCAAAA
ABCC6_PTC_R	TCAATCCCTGCTATGAAATCAA
ABCC6_ F	GGGTCCTTGGTCCCATCTAC
ABCC6_ R	GAATCCAAGCACCCATCTTG
ABCA8_PTC_F	AGGAAGGAAGTTACTGTACATGAGTG

ABCA8_PTC_R	GTGACCTTCCTGTGCCCATC
ABCA8_ F	ATCTTCGAAAGTGATATTCCATTTA
ABCA8_R	GAGAAGATGAGAAGATAAGAACAAACA

Relative mRNA expression for each junction was represented by the average of three replicate cycle threshold (Ct) values (excluding any obvious outliers), which denote the cycle number at which *SYBR® Green fluorescence crosses a threshold value. Each Ct represents a two-fold decrease in expression, and higher overall Ct values represent relatively lower mRNA expression.* The relative mRNA expression of putative PTC junctions in each condition was calculated by normalizing average PTC junction Ct to the average non-PTC junction Ct values in each condition.

3.4 Results

3.4.1 Extensive Novel and Variable Splicing of the ATP-Binding Cassette Transporters

Paired-end RNA-seq data for 139 samples from four human tissues – adipose, liver, kidney and heart – and lymphoblastoid cell lines were used to explore alternative splicing of ABC transporters with the splice event identification tool JuncBASE[24]. JuncBASE identifies individual splicing events using junction reads (those that cross exon-exon boundaries) from transcriptome sequencing experiments. ABC transporter events identified in two or more individuals in any given tissue type and those with median PSI between 10 and 90% or >90% and <10% in at least two different sample types were defined as high confidence splicing events. This high confidence splicing set consists of 22 unannotated events (events that are not included in UCSC hg19, Gencode v19, RefSeq (release 59-63),, or Ensembl v75 gene annotations) and 28 annotated events across the four tissues and lymphoblastoid cell lines (Table 3.3). Unannotated events include both putative novel exons, cassette exons, and truncated or extended exons.

Table 3.3: Summary of ABC Transporter Splicing Events

Event ID [1]	Annotated [2]	Gene	Event Type	Median Percent Spliced In (PSI) [3]					# Detected [4]				
				Adipose	Heart	Liver	Kidney	LCLs	Adipose	Heart	Liver	Kidney	LCLs
1	Y	*ABCA1*	Alt Last Exon	0.0	66.7	81.3	0.0		25	22	19	18	0
	N	*ABCA1*	Alt Last Exon	100.0	33.3	18.8	100.0		25	22	19	18	0
2	N	*ABCA1*	Alt Acceptor		0.0	18.2	100.0		0	9	11	1	0
	Y	*ABCA1*	Alt Acceptor		100.0	81.8	0.0		0	9	11	1	0
3	Y	*ABCA1*	Alt Acceptor	0.0	100.0	78.2			16	13	18	0	0
	N	*ABCA1*	Alt Acceptor	100.0	0.0	21.8			16	13	18	0	0
4a	Y	*ABCA10*	Cassette	100.0	0.0	100.0		0.0	13	13	1	0	8
	Y	*ABCA10*	Cassette	0.0	100.0	0.0		100.0	13	13	1	0	8
4b	Y	*ABCA10*	Cassette	100.0	0.0	100.0		0.0	13	13	1	0	8
	Y	*ABCA10*	Cassette	0.0	100.0	0.0		100.0	13	13	1	0	8
5	Y	*ABCA2*	Alt Acceptor	0.0	100.0	0.0	0.0	0.0	22	23	8	10	1
	Y	*ABCA2*	Alt Acceptor	100.0	0.0	100.0	100.0	100.0	22	23	8	10	1
6	Y	*ABCA4*	Cassette				21.4		0	0	0	7	0
	Y	*ABCA4*	Cassette				78.6		0	0	0	7	0
7a	Y	*ABCA5*	Cassette	26.7	0.0	8.3	10.0	0.0	25	25	19	18	45
	Y	*ABCA5*	Cassette	73.3	100.0	91.7	90.0	100.0	25	25	19	18	45
7b	Y	*ABCA5*	Cassette	26.7	0.0	8.3	10.0	0.0	25	25	19	18	45
	Y	*ABCA5*	Cassette	73.3	100.0	91.7	90.0	100.0	25	25	19	18	45
8	N	*ABCA6*	Alt Donor	0.0	50.0	14.3	0.0	0.0	25	24	19	2	42
	Y	*ABCA6*	Alt Donor	100.0	50.0	85.7	100.0	100.0	25	24	19	2	42
9	N	*ABCA6*	Cassette	0.0	0.0	0.0	0.0	30.0	15	25	18	2	45
	Y	*ABCA6*	Cassette	100.0	100.0	100.0	100.0	70.0	15	25	18	2	45
10	N	*ABCA6*	Alt Acceptor	16.7	0.0	8.3	0.0	0.0	25	24	19	7	45
	Y	*ABCA6*	Alt Acceptor	83.3	100.0	91.7	100.0	100.0	25	24	19	7	45

11	N	*ABCA8*	Alt Donor	0.0	0.0	35.7	0.0		25	25	19	11	0
	Y	*ABCA8*	Alt Donor	100.0	100.0	64.3	100.0		25	25	19	11	0
12	Y	*ABCA8*	Cassette	0.0	17.2	20.0	0.0		23	25	11	1	0
	Y	*ABCA8*	Cassette	100.0	82.8	80.0	100.0		23	25	11	1	0
13	N	*ABCA8*	Alt Acceptor	4.4	12.0	2.8	0.0		25	25	16	11	0
	Y	*ABCA8*	Alt Acceptor	95.7	88.0	97.2	100.0		25	25	16	11	0
14	Y	*ABCA8*	Alt First Exon	50.0	58.3	100.0			2	24	1	0	0
	Y	*ABCA8*	Alt First Exon	50.0	41.7	0.0			2	24	1	0	0
15	Y	*ABCA9*	Cassette	26.7	0.0	0.0	0.0		25	25	14	3	0
	Y	*ABCA9*	Cassette	73.3	100.0	100.0	100.0		25	25	14	3	0
16	N	*ABCA9*	Alt Acceptor	22.2	25.0	45.8	0.0		25	25	18	8	0
	Y	*ABCA9*	Alt Acceptor	77.8	75.0	54.2	100.0		25	25	18	8	0
17	Y	*ABCB4*	Cassette	0.0	0.0	16.7		46.4	2	15	19	0	18
	Y	*ABCB4*	Cassette	100.0	100.0	83.3		53.6	2	15	19	0	18
18	Y	*ABCB4*	Alt Acceptor		0.0	10.7		0.0	0	12	19	0	17
	Y	*ABCB4*	Alt Acceptor		100.0	89.3		100.0	0	12	19	0	17
19	Y	*ABCB4*	Mutually Exclusive	50.0	0.0	0.0		0.0	2	10	19	0	15
	Y	*ABCB4*	Mutually Exclusive	50.0	100.0	100.0		100.0	2	10	19	0	15
20	Y	*ABCB6*	Mutually Exclusive	60.0	58.3	100.0	100.0	50.0	7	2	1	3	2
	Y	*ABCB6*	Mutually Exclusive	40.0	41.7	0.0	0.0	50.0	7	2	1	3	2
21	Y	*ABCB7*	Alt Acceptor	63.3	72.7	70.8	66.7	66.7	22	25	14	11	45
	Y	*ABCB7*	Alt Acceptor	36.7	27.3	29.2	33.3	33.3	22	25	14	11	45
22	Y	*ABCB8*	Alt Donor	0.0	11.8	0.0	0.0	0.0	25	24	19	18	45
	Y	*ABCB8*	Alt Donor	100.0	88.2	100.0	100.0	100.0	25	24	19	18	45
23	Y	*ABCB8*	Alt Last Exon	20.0	25.0	28.6	16.7	17.4	25	25	19	17	45
	Y	*ABCB8*	Alt Last Exon	80.0	75.0	71.4	83.3	82.6	25	25	19	17	45
24	Y	*ABCC1*	Mutually Exclusive	75.0	100.0			0.0	2	4	0	0	9
	Y	*ABCC1*	Mutually Exclusive	25.0	0.0			100.0	2	4	0	0	9

25	N	*ABCC10*	Alt Donor	33.3	14.3	33.3	0.0	0.0	17	21	9	13	40
	Y	*ABCC10*	Alt Donor	66.7	85.7	66.7	100.0	100.0	17	21	9	13	40
26	Y	*ABCC2*	Alt Donor	0.0		78.3	33.3		1	0	19	16	0
	N	*ABCC2*	Alt Donor	100.0		21.7	66.7		1	0	19	16	0
27	N	*ABCC2*	Alt Acceptor	0.0		15.0	0.0	0.0	7	0	19	12	1
	Y	*ABCC2*	Alt Acceptor	100.0		85.0	100.0	100.0	7	0	19	12	1
28	Y	*ABCC2*	Alt Acceptor			37.5	100.0		0	0	19	7	0
	N	*ABCC2*	Alt Acceptor			62.5	0.0		0	0	19	7	0
29	N	*ABCC3*	Alt Donor	33.3	100.0	50.0	50.0	0.0	25	2	19	17	1
	Y	*ABCC3*	Alt Donor	66.7	0.0	50.0	50.0	100.0	25	2	19	17	1
30	N	*ABCC3*	Alt Acceptor	0.0		10.5	6.7		23	0	19	11	0
	Y	*ABCC3*	Alt Acceptor	100.0		89.5	93.3		23	0	19	11	0
31	N	*ABCC3*	Alt Acceptor	5.9	50.0	21.6	13.4		25	2	19	14	0
	Y	*ABCC3*	Alt Acceptor	94.1	50.0	78.4	86.6		25	2	19	14	0
32	Y	*ABCC4*	Cassette				71.4	0.0	0	0	0	6	41
	Y	*ABCC4*	Cassette				28.6	100.0	0	0	0	6	41
33	N	*ABCC5*	Alt Donor	0.0	16.7	0.0	0.0	0.0	3	24	1	12	29
	Y	*ABCC5*	Alt Donor	100.0	83.3	100.0	100.0	100.0	3	24	1	12	29
34	Y	*ABCC5*	Alt Acceptor	0.0	7.1	0.0	12.5	0.0	10	25	9	17	15
	Y	*ABCC5*	Alt Acceptor	100.0	92.9	100.0	87.5	100.0	10	25	9	17	15
35	N	*ABCC5*	Alt Acceptor	0.0	25.0	0.0	0.0	0.0	8	25	2	14	35
	Y	*ABCC5*	Alt Acceptor	100.0	75.0	100.0	100.0	100.0	8	25	2	14	35
36	Y	*ABCC6*	Alt Donor	50.0	0.0	61.9	22.2		25	3	19	18	0
	Y	*ABCC6*	Alt Donor	50.0	100.0	38.1	77.8		25	3	19	18	0
37	N	*ABCC6*	Cassette	28.6		4.6	0.0		25	0	19	14	0
	Y	*ABCC6*	Cassette	71.4		95.5	100.0		25	0	19	14	0
38	Y	*ABCC6*	Alt Acceptor	50.0		100.0	100.0		2	0	15	1	0
39	Y	*ABCC8*	Alt Acceptor		100.0	0.0			0	5	1	0	0

40	Y	*ABCC9*	Cassette	100.0	18.2	100.0	0.0		1	25	9	1	0
	Y	*ABCC9*	Cassette	0.0	81.8	0.0	100.0		1	25	9	1	0
41	Y	*ABCC9*	Alt Acceptor	0.0	10.5	5.1	0.0		25	25	18	9	0
	Y	*ABCC9*	Alt Acceptor	100.0	89.5	94.9	100.0		25	25	18	9	0
42	Y	*ABCC9*	Alt Acceptor	66.7	81.3	90.0	100.0		1	25	2	4	0
	Y	*ABCC9*	Alt Acceptor	33.3	9.5	10.0	0.0		1	25	2	4	0
43	N	*ABCD3*	Alt Donor	0.0	6.7	16.7	5.8	0.0	22	25	19	18	45
	Y	*ABCD3*	Alt Donor	100.0	93.3	83.3	94.2	100.0	22	25	19	18	45
44	Y	*ABCD4*	Alt Last Exon	42.9	33.3	33.3	28.6	33.3	25	25	19	18	45
	Y	*ABCD4*	Alt Last Exon	57.1	66.7	66.7	71.4	66.7	25	25	19	18	45
45	Y	*ABCD4*	Cassette	22.2	0.0	0.0	18.3	0.0	25	25	19	18	45
	Y	*ABCD4*	Cassette	77.8	100.0	100.0	81.7	100.0	25	25	19	18	45
46	Y	*ABCE1*	Alt Acceptor	0.0	0.0	0.0	0.0	13.6	23	24	13	12	45
	Y	*ABCE1*	Alt Acceptor	100.0	100.0	100.0	100.0	86.4	23	24	13	12	45
47	Y	*ABCF1*	Alt Acceptor	40.0	37.5	37.5	44.4	38.5	23	25	19	17	45
	Y	*ABCF1*	Alt Acceptor	60.0	62.5	62.5	55.6	61.5	23	25	19	17	45
48	Y	*ABCG2*	Alt First Exon	100.0	0.0	100.0	0.0	100.0	5	6	5	1	1
	Y	*ABCG2*	Alt First Exon	0.0	100.0	0.0	100.0	0.0	5	6	5	1	1
49	N	*ABCG5*	Alt Donor			13.8			0	0	19	0	0
	Y	*ABCG5*	Alt Donor			86.2			0	0	19	0	0
50	N	*ABCG5*	Alt Acceptor			19.4			0	0	19	0	0
	Y	*ABCG5*	Alt Acceptor			80.7			0	0	19	0	0
51	Y	*ABCB2*	Alt Donor	85.7	92.3	84.6	100.0	98.0	25	25	19	18	45
	N	*ABCB2*	Alt Donor	14.3	7.7	15.4	0.0	2.0	25	25	19	18	45
52	N	*ABCB2*	Alt Donor	37.0	22.2	33.3	33.3	7.2	25	25	19	18	45
	Y	*ABCB2*	Alt Donor	63.0	77.8	66.7	66.7	92.8	25	25	19	18	45
53	Y	*ABCB2*	Alt Acceptor	29.6	18.5	20.0	20.4	9.4	25	25	19	18	45
	N	*ABCB2*	Alt Acceptor	70.4	81.5	80.0	79.6	90.6	25	25	19	18	45

54	Y	*ABCB2*	Alt Acceptor	14.3	7.1	7.7	0.0	0.9	25	25	19	18	45
	Y	*ABCB2*	Alt Acceptor	85.7	92.9	92.3	100.0	99.1	25	25	19	18	45
55	Y	*ABCB3*	Alt Acceptor	65.0	54.6	57.1	72.1	89.8	25	25	19	18	45
	Y	*ABCB3*	Alt Acceptor	35.0	45.5	42.9	27.9	10.2	25	25	19	18	45

[1] Where two events are shown per event ID these represent two alternate junctions or junction sets for a given event, e.g. inclusion and exclusion of a cassette exon.
[2] Annotated events are those in UCSC (hg19), Gencode (v19), RefSeq (release 59-63), or Ensembl (v75) gene annotations
[3] Blank cells represent events with insufficient reads to estimate PSI. Where PSI was detectable, median value is shown. Range of PSIs across samples varies widely depending on the event.
[4] Number of samples event was detected in (PSI ≥ 0) in 25 adipose, 25 heart, 24 liver, 20 kidney and 44 LCLs

3.4.2 "Functional" splicing events modulate transporter expression

"Functional" splicing events were defined as those that result in a predicted functional transcript – that is, the splicing event does not change the function of the protein itself, but rather alters the untranslated region or changes the coding sequence in a region of the protein that will not disrupt function. Events that alter the coding sequence but retain both ATP-binding domains and all transmembrane domains were also included in the set of putative functional events. Events that met these criteria for generating potential functional protein and had variable PSI either between individuals or between tissues were included in the set of putative functional events (Table 3.4); these events are more likely to be involved in regulation of gene expression. Two functional splicing events in the 5' UTR were characterized further; these events occur in two genes expressed in the human heart - *ABCA8* and *ABCC5*.

Table 3.4: Putative Functional ABC Transporter Events

Gene	Event ID[1]	Annotated[2]	Event	Effect of splicing event
ABAC8	13	N	Last Exon	Early termination event occurs after all ATP-binding and transmembrane domains
ABCA5	7	N	Cassette Exon	Deletion of 38 amino acids after all ATP-binding and all transmembrane domains
ABCA6	10	N	Last Exon	Early termination event occurs after all ATP-binding and transmembrane domains
ABCA8	12	Y	Cassette Exon	Changes structure of 5'UTR
ABCA9	15	Y	Cassette Exon	Creates an early termination event after all ATP-binding and transmembrane domains
ABCB4	17	Y	Cassette Exon	Deletion of 47 amino acids does not change frame and occurs outside of ATP-binding and transmembrane domains
ABCC2	27	N	Truncated Exon	Creates an early termination event after all ATP-binding and transmembrane domains
ABCC5	33	N	First Exons	Changes structure of 5'UTR
ABCC9	41,42	Y	Alternative Last Exons	Early termination events occur after all ATP-binding and transmembrane domains

[1] See Table 3.3 for event descriptions
[2] Annotated events are those in UCSC (hg19), Gencode (v19), RefSeq (release 59-63), or Ensembl (v75) gene annotations

Inclusion of a cassette exon in the 5'UTR of ABCA8 results in two alternate 5'UTRs - ABCA8_5UTR_A and ABCA8_5UTR _B (Figure 3.1A). The percent of transcripts including the cassette exon varies between 8 to 40% (Figure 3.1B) across individuals in the heart.

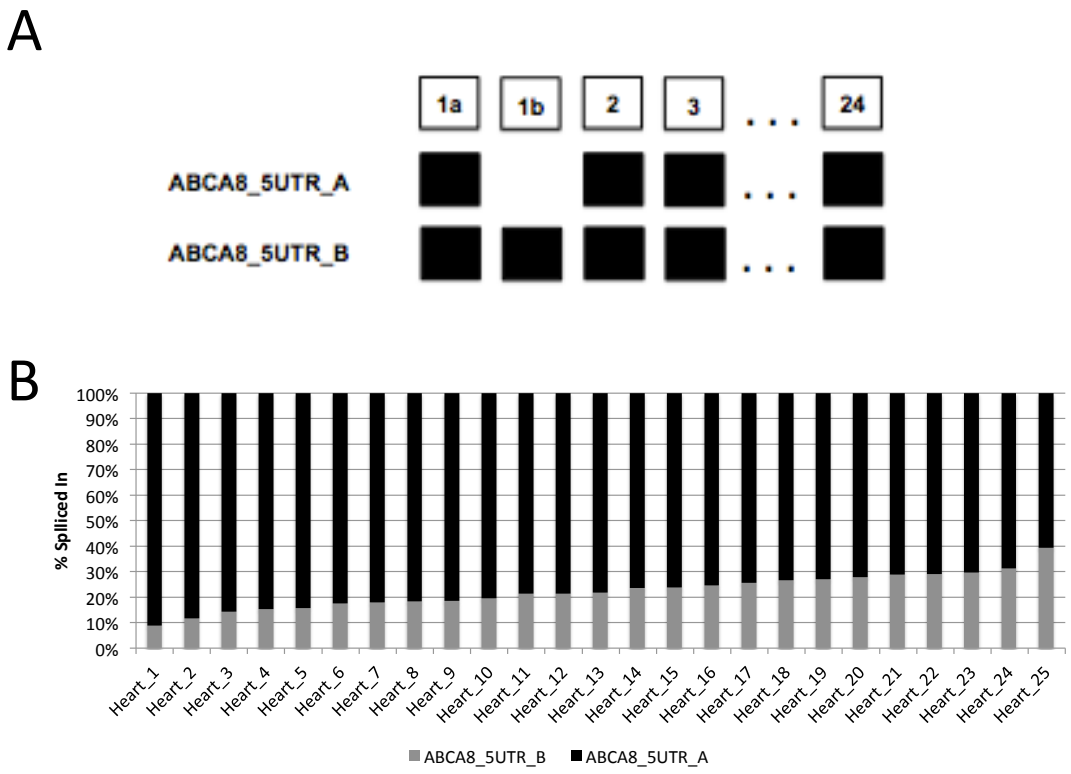


Figure 3.1: ABCA8 alternate 5' UTRs. A: Structure of alternate ABCA8 5' UTRs. Exon 1b is a cassette exon excluded in ABCA8_5UTR_A. B: Percent of reads representing inclusion of each splicing event by individual in the heart.

The two alternate 5'UTRs for ABCA8 were evaluated *in vitro* by inserting each UTR into a luciferase vector upstream of the luciferase gene. These plasmids were transfected into cells to look for a difference in luciferase activity between the two sequences. The shorter ABCA8_5UTR_A (without cassette exon) showed significantly lower luciferase activity relative to the longer ABCA8_5UTR_B (inclusion of the cassette exon) (Figure 3.2). UTR motif analysis of the alternate UTRs predicted the introduction of an Internal Ribosomal Entry Site (IRES) in ABCA8-5UTR-B (Table 3.5).

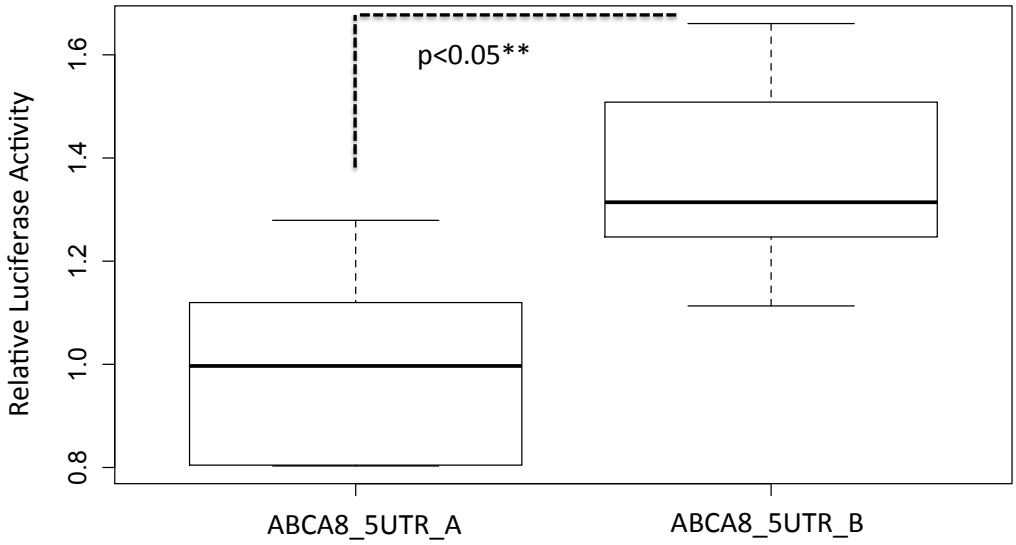


Figure 3.2: Transcriptional activity of ABCA8 5'UTR sequences. Luciferase activity is expressed relative to *Renilla* and the ABCA8_5UTR_A plasmid. ABCA8_5UTR_B has significantly higher luciferase activity relative to ABCC5_5UTR_A ($p<0.05$). Data shown from six replicates in one experiment and are representative of at least three independent experiments.

Table 3.5: Analysis of *ABCA8* 5'UTRs

	Length (bp)	Min – ΔG[1]	Motifs[2]
ABCA8_5UTR_A	179	-28.6	TOP
ABCA8_5UTR_B	340	-84.6	IRES, TOP, uORF

[1] Minimum free energy of 5'UTR folding predicted using mFold
[2] 5'UTR RNA binding motifs predicted using RegRNA: uORF: Upstream Open Reading Frame; IRES: Internal Ribosomal Entry Site; TOP: Terminal Oligopyramidine Tract

Two previously unannotated first exons for *ABCC5* in the heart result in a novel 5'UTR– ABCC5_5UTR_B (Fig 3.3A), in addition to the annotated 5'UTR, ABCC5_5UTR_A. The two alternate 5'UTRs for ABCC5 are expressed variably across individuals in the heart (Fig 3.3B). The first 58 bases of the *ABCC5* 5'UTR (ABCC5_5UTR_C), which exists as an independent 5'UTR for *ABCC5* in gene annotations, was not detected in this sample set, but was included in functional assays. As with the ABCA8 UTRs, the alternate UTRs were evaluated in an *in vitro* luciferase assay; ABCC5_5UTR_C, was also included in *in vitro* assays. ABCC5_5UTR_B resulted in a significant decrease in luciferase activity relative to the reference ABCC5-5UTR_A, and ABCC5-5UTR-C resulted in a significant increase in luciferase activity relative to the reference ABCC5-5UTR-A (Figure 3.4). Predicted 5'UTR motifs and estimates of folding free energy change for ABCC5 5'UTRs are included in Table 3.6. A predicted IRES site in ABCC5_5UTR_A and ABCC5_5UTR_B is not found in the shorter transcript.

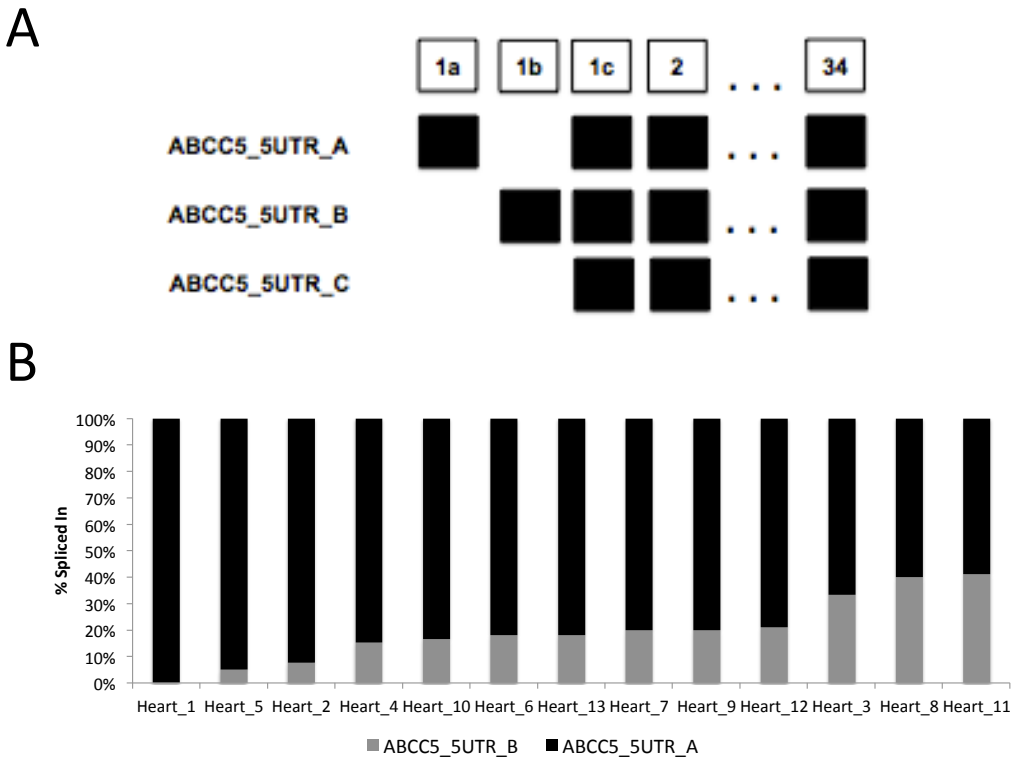


Figure 3.3: ABCC5 alternate 5' UTRs. A: Structure of alternate ABCC5 5' UTRs. Exons 1a and 1b represent two alternate first exons for ABCC5. ABCC5_5UTR_C represents the constitutive portion of the ABCC5 5'UTR, but does not exist as an independent event in this dataset. B: Percent of reads representing inclusion of each splicing event by sample in the heart. Only samples with sufficient read count in the 5'UTR to estimate a PSI are shown.

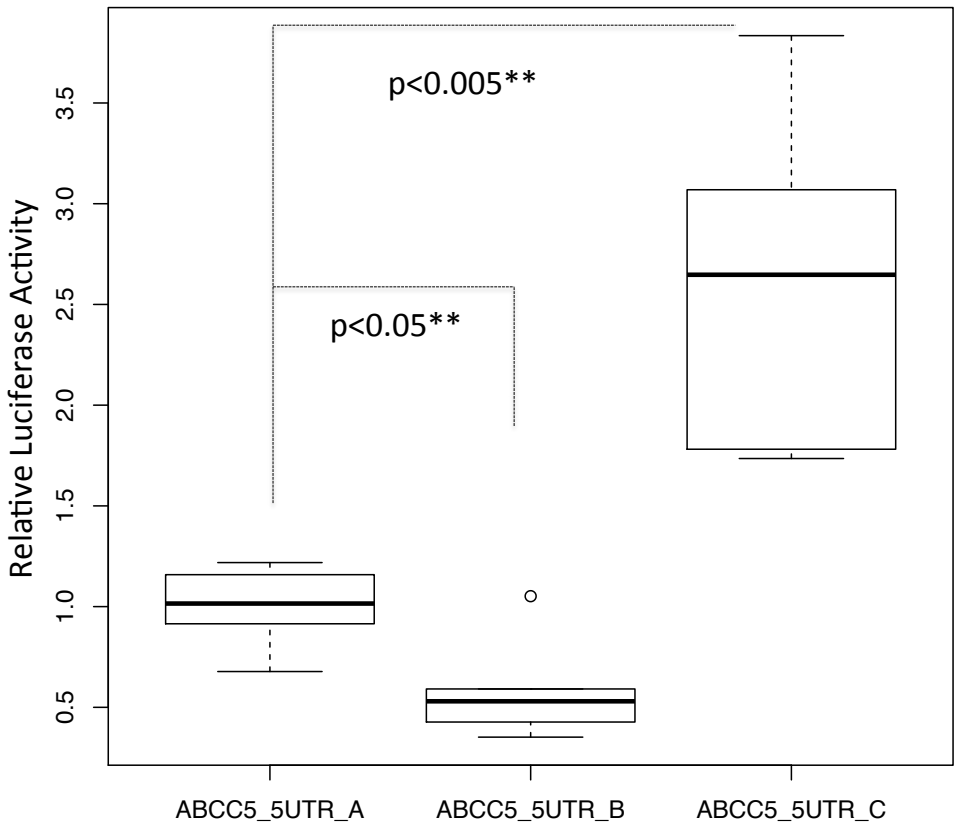


Figure 3.4: Transcriptional activity of ABCC5 5'UTR sequences. Luciferase activity is expressed relative to *Renilla* and the ABCC5_5UTR_A (annotated 5'UTR) plasmid. ABCC5_5UTR_B has significantly lower luciferase activity relative to ABCC5_5UTR_A ($p<0.05$), and ABCC5_5UTR_C has significantly higher luciferase activity relative to ABCC5_5UTR_A ($p<0.005$). Data shown from six replicates in one experiment and are representative of at least three independent experiments.

Table 3.6: Analysis of *ABCC5* 5'UTRs

	Length	Min -ΔG[1]	Motifs[2]
ABCC5_5UTR_A	153	-53.3	IRES, TOP, uORF
ABCC5_5UTR_B	190	-46.4	IRES, TOP
ABCC5_5UTR_C	58	-10.7	TOP

[1] Minimum free energy of 5'UTR folding predicted using mFold
[2] 5'UTR RNA binding motifs predicted using RegRNA: uORF: Upstream Open Reading Frame; IRES: Internal Ribosomal Entry Site; TOP: Terminal Oligopyramidine Tract

3.4.3 "Nonfunctional" splicing of ABCC6 regulates transporter expression

Splicing events in ABC transporters that are predicted to produce a non-functional protein, for example because of a premature termination codon (PTC) that would truncate the translated protein, were also identified. Many transcripts with a PTC will be targets of the nonsense mediated decay (NMD) process, which serves to eliminate erroneously spliced transcripts. Selected ABC transporter splicing events that may yield transcripts that are targets of the NMD process were identified by looking for events that create a premature termination codon >50-55 bases upstream of an exon-exon junction (Table 3.7) [30].

Table 3.7: Putative Nonsense Mediated Decay Events

Gene	Event ID[1]	Annotated[2]	Event
ABCC3	30	N	Extended Exon
ABCC2	27	N	Truncated Exon
ABCA6	10	N	Truncated Exon
ABCC2	28	N	Exon
ABCC3	29	N	Exon
ABCC6	36	N	Exon
ABCA8	13	N	Exon

[1] See Table 3.3 for event descriptions
[2] Annotated events are those in UCSC (hg19), Gencode (v19), RefSeq(release 59-63), or Ensembl (v75) gene annotations

To determine whether the identified events are targets of the NMD process, changes in relative transcript mRNA levels were measured by qPCR in HepG2 and HEK293 cells before and after treatment with puromycin, a translation inhibitor. A transcript targeted by the NMD process should proportionally increase

in expression after translation inhibition and decrease in expression after removal of the translation inhibitor. Of the seven events tested, five did not meet criteria for nonsense mediated decay under the conditions used, and one (*ABCA6* truncated exon, event ID 10) was not detectable at high enough levels in either cell line tested. One event, a novel exon in *ABCC6* (Figure 3.5A), was a target of the NMD process (Figure 3.6). Further, in these samples, tissue level PSI of the *ABCC6* NMD targeted splicing event is inversely correlated with tissue level mRNA expression (FPKM) of ABCC6 *(*Figure 3.5 C), suggesting that usage of this alternate event is a mechanism for tissue-specific regulation of ABCC6 expression.

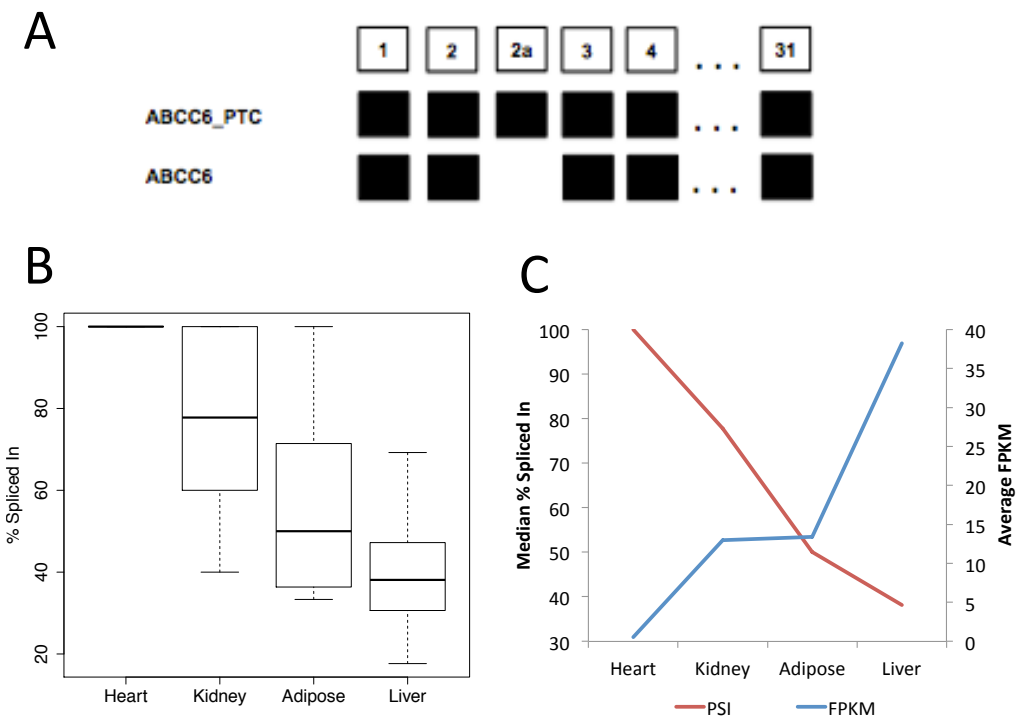


Figure 3.5: *ABCC6* novel exon. A: Structure of alternate ABCC6 transcript. Exon 2a is a putative novel exon. B: The *ABCC6* novel exon is expressed at variable levels between individuals and between tissues. C: Median PSI for the event is inversely correlated with average FPKM across tissues, suggesting that inclusion of the novel event may be one mechanism for regulating tissue-specific expression of ABCC6.

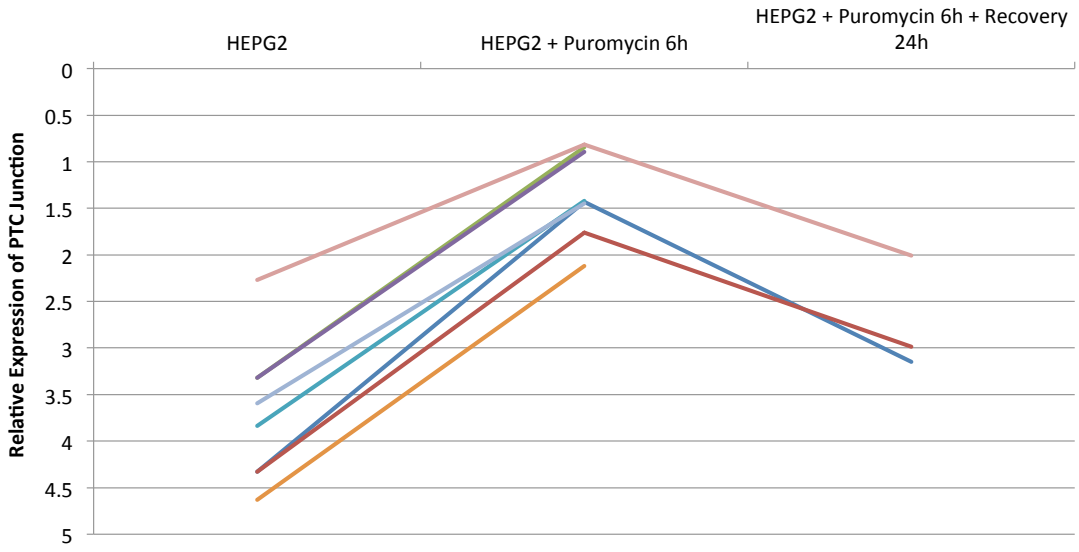


Figure 3.6: Relative expression of *ABCC6* premature termination codon (PTC) junction. Relative expression of PTC junctions represents the average Ct of three replicates for the putative PTC junction normalized to the average of three replicates for the non-PTC junction in each condition. HEPG2 human liver cells were exposed to translation inhibitor puromycin for 6 hours. Recovery from translation inhibition was evaluated by replacing puromycin containing media and allowing 24 hours recovery. Each colored line represents a single set of experiments. Relative expression of the PTC junction decreases with inhibition of translation (and inhibition of nonsense mediated decay) and increases after re-initiation of translation and the nonsense mediated decay process, suggesting that this transcript is a target of the nonsense mediated decay process.

3.5 Discussion

Splicing events for ABC transporters were identified in four human tissues and lymphoblastoid cell lines, including a number of previously unannotated splicing events. The functional outcome of selected splicing events in ABC transporters were evaluated, focusing on events that may modulate gene expression.

While alternative splicing at any position in a transcript may alter gene expression, splicing in regulatory elements of the gene, particularly the 5' UTR , are most likely to modulate gene expression[31]. Alternative 5' UTRs in *ABCC5* and *ABCA8*, key ABC transporters in the human heart, were shown to modulate luciferase activity *in vitro*, suggesting that these alternate events may regulate gene expression *in vivo*. In the heart, ABCC5 has been localized to cardiac myocytes and cardiac endothelial cells[32]. ABCC5 encodes the multidrug resistance associated protein 5 (MRP5) that transports cyclic nucleotides, including cyclic GMP; cGMP acts as a second messenger of nitric oxide levels, and plays a role in regulating cardiac contractility and protecting cardiomyocytes during cardiac ischemia[32]. ABCA8 is expressed at high levels in the heart and has been shown to transport drugs and endogenous compounds, including glucuronide conjugates, leukotriene C_4, and estrone sulfate [33]. Expression of alternate 5'UTRs in these transporters varies across individuals in the heart, suggesting that alternative splicing in these transporters may regulate transporter expression between individuals.

To identify potential mechanisms driving differences in expression, 5'UTR motifs were identified in these alternate UTRs. An IRES site was identified in the 5'UTR of the longer ABCA8 5'UTR, associated with increased luciferase activity *in vivo*. In mammalian mRNAs, IRES sites have been associated with increased translation under particular cellular conditions[34]. Thus it is possible that inclusion of the cassette exon allows for finer control of ABCA8 expression in certain individuals. However, computational predictions of RNA binding elements, such as IRES, are still being refined and the results of such analyses must be interpreted with caution[35].

Secondary structure of each 5'UTR was also examined. More stable secondary structures, typically associated with longer UTRs, are generally associated with less efficient translation.[36,37] As expected, more efficient translation (as indicated by higher luciferase activity) was found for the shortest *ABCC5* 5'UTR with lowest minimum free energy (ABCC5_5UTR_C)[38]. A decrease in luciferase activity was associated with the longest 5'UTR, however no 5'UTR motifs or significant changes in stability of the secondary structure were identified unique to this UTR that could explain this change in translation.

Variability in alternative splicing across individuals, as detected for *ABCA8* and *ABCC5* in the heart, may be driven by polymorphisms at canonical splicing acceptor and donor sites[39] or in the binding sites of proteins that regulate splicing

in exons and introns[40–42]. These splicing factors can enhance or suppress splicing, and their binding sites are termed exonic and intron splicing enhancers or silencers. [43] No common (minor allele frequency > 1%) polymorphisms were identified that occur within splicing acceptors or donors or predicted splicing silencers or enhancers in the *ABCA8* and *ABCC5* UTRs. As with the identification of RNA binding elements in UTRs, methods to predict splicing factor binding sites are still being developed, and it is possible that novel splicing enhancer or silencer sites occur around or within identified events.

The NMD process can also regulate expression of a wide range of genes[44]. Putative NMD inducing events were studied *in vitro* and one event targeted by the nonsense mediated decay process was identified for *ABCC6*. ABCC6 is expressed in multiple tissues, with highest expression in the liver; mutations in *ABCC6* cause the hereditary metabolic disorder pseudoxanthoma elasticum[45] and other disorders involving mineralization of tissues. Further, there is some evidence in humans of association between more common *ABCC6* polymorphisms and cholesterol and cardiac phenotypes, such as plasma high density lipoprotein (HDL) and triglyceride levels and coronary heart and artery disease [46–48]. Very little is known about the precise function of ABCC6 in the body, but it is possible that its role in lipid levels is due to either some direct involvement in lipid transport or indirectly due to mineralization and damage of renal tissues, as renal failure has been associated with dyslipidemia[49,50].

Inclusion of the NMD targeted splicing event is inversely correlated with tissue level expression of ABCC6, and may be involved in regulating tissue-specific expression of this important transporter. Further, inclusion of the identified event varies across individuals, and may contribute to inter-individual variability in expression of ABCC6 and potentially the activity of its encoded transporter, MRP6. In mice, *Abcc6* gene dose is associated with response to cardiac injury, including infarct size and level of calcification, with more severe phenotypes with lower Abcc6 levels[51,52]. Thus variation in expression of ABCC6 across individuals may increase susceptibility to or modulate severity of any phenotypes associated with this ABC transporter.

One limitation of RNA-seq data is the lack of complete transcript sequence for alternative events. As a result, an underlying assumption of the reported functional analyses is that each event that was detected in the RNA-seq analysis is the only event in the transcript. It is possible that some of the predicted premature termination codons are mis-annotated where the transcript structure upstream of the event does not match the reference gene annotation. While it is known that some true premature termination codons can evade the nonsense mediated decay process[53], it is likely some of the putative NMD events identified do not in fact meet criteria for NMD. Likewise, because the NMD process eliminates transcripts, there are likely additional splicing events that could not be identified in this dataset that are targets of NMD.

Identification of alternative splicing events is dependent on both read depth and gene expression. As membrane proteins, ABC transporters can play significant roles in cell function without high expression levels. In this study, only common splicing events were captured for transporters with higher expression levels. Transporters without identified splicing events tend to be expressed at lower levels, and may require greater read depth to capture alternative splicing events. For example, no splicing events were identified in the transporter *ABCG2* encoding the breast cancer resistance protein, BCRP. BCRP is expressed at low levels even in tissues in which it serves a key physiologic function, such as the liver and kidney. It is possible that splicing events for *ABCG2* do exist in these tissues but did not meet read count thresholds in the current analysis. Further, intron retention events were not considered in this study. While intron retention events do occur in the human transcriptome, and have been associated with biological outcomes[54–56], it is difficult to separate biologically meaningful intron retention from incompletely or un-processed transcripts or genomic DNA contamination.

Rare ABC transporter splicing events, events in transporters expressed at low levels in all studied tissues, and biologically meaningful intron retention events, might not have been detected in this analysis. However, a number of previously unannotated splicing events in ABC transporters were identified. Further, by including multiple individuals per tissue type, a valuable picture of variability in splicing of ABC transporters was developed. Alternative splicing of transporters

was found to modulate inter-individual and inter-tissue gene expression, and may represent a source of variation in susceptibility to drug toxicity or common disorders, variability in drug response, or tissue specificity of drug disposition or disease phenotype.

3.6 References

1 Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 2001; **11**: 1156–66.

2 Matsson P, Yee SW, Markova S, Morrissey K, Jenkins G, Xuan J, Jorgenson E, Kroetz DL, Giacomini KM. Discovery of regulatory elements in human ATP-binding cassette transporters through expression quantitative trait mapping. *Pharmacogenomics J* 2012; **12**: 214–26.

3 Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008; **456**: 470–6.

4 Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol* 2004; **5**: R74.

5 Coulombe-Huntington J, Lam KCL, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS Genet* 2009; **5**: e1000766.

6 Zhang W, Duan S, Bleibel WK, Wisel S a, Huang RS, Wu X, He L, Clark T a, Chen TX, Schweitzer AC, Blume JE, Dolan ME, Cox NJ. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* 2009; **125**: 81–93.

7 Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 2008; **40**: 225–31.

8 Hull J, Campino S, Rowlands K, Chan M-S, Copley RR, Taylor MS, Rockett K, Elvidge G, Keating B, Knight J, Kwiatkowski D. Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* 2007; **3**: e99.

9 Stojic J, Stöhr H, Weber BHF. Three novel ABCC5 splice variants in human retina and their role as regulators of ABCC5 gene expression. *BMC Mol Biol* 2007; **8**: 42.

10 Lamba JK. Nonsense mediated decay downregulates conserved alternatively spliced ABCC4 transcripts bearing nonsense codons. *Hum Mol Genet* 2003; **12**: 99–109.

11 Yabuuchi H, Shimizu H, Takayanagi S, Ishikawa T. Multiple splicing variants of two new human ATP-binding cassette transporters, ABCC11 and ABCC12. *Biochem Biophys Res Commun* 2001; **288**: 933–9.

12 Chaudhry AS, Thirumaran RK, Yasuda K, Yang X, Fan Y, Strom SC, Schuetz EG. Genetic variation in aldo-keto reductase 1D1 (AKR1D1) affects the expression and activity of multiple cytochrome P450s. *Drug Metab Dispos* 2013; **41**: 1538–47.

13 Dahlin A, Geier E, Stocker SL, Cropp CD, Grigorenko E, Bloomer M, Siegenthaler J, Xu L, Basile AS, Tang-Liu DD-S, Giacomini KM. Gene expression profiling of transporters in the solute carrier and ATP-binding cassette superfamilies in human eye substructures. *Mol Pharm* 2013; **10**: 650–63.

14 Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson D a, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD. Identification, replication, and functional fine-mapping of expression quantitative trait Loci in primary human liver tissue. *PLoS Genet* 2011; **7**: e1002078.

15 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107.

16 Mangravite LM, Dawson K, Davis RR, Gregg JP, Krauss RM. Fatty acid desaturase regulation in adipose tissue by dietary composition is independent of weight loss and is correlated with the plasma triacylglycerol response. *Am J Clin Nutr* 2007; **86**: 759–67.

17 Lundquist AL, Manderfield LJ, Vanoye CG, Rogers CS, Donahue BS, Chang PA, Drinkwater DC, Murray KT, George AL. Expression of multiple KCNE genes in human heart may enable variable modulation of I(Ks). *J Mol Cell Cardiol* 2005; **38**: 277–87.

18 Simon JA, Lin F, Hulley SB, Blanche PJ, Waters D, Shiboski S, Rotter JI, Nickerson DA, Yang H, Saad M, Krauss RM. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. *Am J Cardiol* 2006; **97**: 843–50.

19 Zhong S, Joung J-G, Zheng Y, Chen Y, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ. High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* 2011; **2011**: 940–9.

20 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–11.

21 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; **28**: 511–5.

22 Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; : 1–5.

23 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012; **22**: 1760–74.

24 Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res* 2011; **21**: 193–202.

25 Huang H-Y, Chien C-H, Jen K-H, Huang H-D. RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* 2006; **34**: W429–34.

26 Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 2010; **38**: D75–80.

27 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; **31**: 3406–15.

28 Stamm S, Riethoven J-J, Le Texier V, Gopalakrishnan C, Kumanduri V, Tang Y, Barbosa-Morais NL, Thanaraj TA. ASD: a bioinformatics resource on alternative splicing. *Nucleic Acids Res* 2006; **34**: D46–55.

29 Chang T-H, Huang H-Y, Hsu JB-K, Weng S-L, Horng J-T, Huang H-D. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* 2013; **14**: S4.

30 Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 1999; **96**: 307–10.

31 Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 2009; **136**: 731–45.

32 Dazert P, Meissner K, Vogelgesang S, Heydrich B, Eckel L, Böhm M, Warzok R, Kerb R, Brinkmann U, Schaeffeler E, Schwab M, Cascorbi I, Jedlitschky G, Kroemer HK. Expression and localization of the multidrug resistance protein 5 (MRP5/ABCC5), a cellular export pump for cyclic nucleotides, in human heart. *Am J Pathol* 2003; **163**: 1567–77.

33 Tsuruoka S, Ishibashi K, Yamamoto H, Wakaumi M, Suzuki M, Schwartz GJ, Imai M, Fujimura A. Functional analysis of ABCA8, a new drug transporter. *Biochem Biophys Res Commun* 2002; **298**: 41–5.

34 Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell* 2010; **40**: 228–37.

35 Baird SD, Turcotte M, Korneluk RG, Holcik M. Searching for IRES. *RNA* 2006; **12**: 1755–85.

36 Pelletier J, Sonenberg N. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 1985; **40**: 515–26.

37 Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 1989; **9**: 5134–42.

38 Davuluri R V, Suzuki Y, Sugano S, Zhang MQ. CART classification of human 5' UTR sequences. *Genome Res* 2000; **10**: 1807–16.

39 Kurmangaliyev YZ, Sutormin RA, Naumenko SA, Bazykin GA, Gelfand MS. Functional implications of splicing polymorphisms in the human genome. *Hum Mol Genet* 2013; **22**: 3449–59.

40 Woolfe A, Mullikin JC, Elnitski L. Genomic features defining exonic variants that modulate splicing. *Genome Biol* 2010.

41 Fairbrother WG, Holste D, Burge CB, Sharp PA. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol* 2004; **2**: E268.

42 Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 2004; **5**: 389–96.

43 Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 2008; **14**: 802–13.

44 Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 2004; **36**: 1073–8.

45 Bergen A, Plomp A, Schuurman E. Mutations in ABCC6 cause pseudoxanthoma elasticum. *Nat Genet* 2000; **25**: 228–31.

46 Wang J, Near S, Young K, Connelly PW, Hegele RA. ABCC6 gene polymorphism associated with variation in plasma lipoproteins. *J Hum Genet* 2001; **46**: 699–705.

47 Peloso GM, Demissie S, Collins D, Mirel DB, Gabriel SB, Cupples LA, Robins SJ, Schaefer EJ, Brousseau ME. Common genetic variation in multiple metabolic pathways influences susceptibility to low HDL-cholesterol and coronary heart disease. *J Lipid Res* 2010; **51**: 3524–32.

48 Trip MD, Smulders YM, Wegman JJ, Hu X, Boer JMA, ten Brink JB, Zwinderman AH, Kastelein JJP, Feskens EJM, Bergen AAB. Frequent Mutation in the ABCC6 Gene (R1141X) Is Associated With a Strong Increase in the Prevalence of Coronary Artery Disease. *Circulation* 2002; **106**: 773–5.

49 Kaysen GA. The Role of Oxidative Stress-Altered Lipoprotein Structure and Function and Microinflammation on Cardiovascular Risk in Patients with Minor Renal Dysfunction. *J Am Soc Nephrol* 2004; **15**: 538–48.

50 Prinsen BHCMT, De Sain-Van Der Velden MGM, De Koning EJP, Koomans HA, Berger R, Rabelink TJ. Hypertriglyceridemia in patients with chronic renal failure: Possible mechanisms. *Kidney Int* 2003; **63**: 121–4.

51 Christopher Brampton ZA. The level of hepatic ABCC6 expression determines the severity of calcification after cardiac injury. *Am J Pathol* 2014; **184**: 159 – 70.

52 Mungrue IN, Zhao P, Yao Y, Meng H, Rau C, Havel J V, Gorgels TGMF, Bergen AAB, MacLellan WR, Drake TA, Boström KI, Lusis AJ. Abcc6

deficiency causes increased infarct size and apoptosis in a mouse cardiac ischemia-reperfusion model. *Arterioscler Thromb Vasc Biol* 2011; **31**: 2806–12.

53 Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* 2003; **100**: 189–92.

54 Hir H Le, Charlet-Berguerand N. 5′-End RET splicing: absence of variants in normal tissues and intron retention in pheochromocytomas. *Oncology* 2002; **63**: 84–91.

55 Wong JJ-L, Ritchie W, Ebner OA, Selbach M, Wong JWH, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, Thoeng A, Khoo T-L, Bailey CG, Holst J, Rasko JEJ. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 2013; **154**: 583–95.

56 Buckley PT, Lee MT, Sul J-Y, Miyashiro KY, Bell TJ, Fisher SA, Kim J, Eberwine J. Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron* 2011; **69**: 877–84.

Chapter 4: Identification of Transcriptional Regulators of ABC Transporter Expression

4.1 Abstract

Transcription factors play a key role in regulation of ABC transporter expression, and transcription factor mediated regulation of transporters has been characterized for specific genes and transcription factor families. As genes with similar co-expression patterns in a given tissue may be regulated by the same transcription factors, sets of genes co-expressed with ABC transporters in several healthy human tissue types were extracted and transcription factor binding motifs enriched in the 1000 bp region upstream of these genes were identified. Known and novel transcription factor regulators of transporter expression were identified in each tissue type. Further, enrichment of gene ontology biological processes in genes co-expressed with ABC transporters was evaluated to generate hypotheses about transporter function.

4.2 Introduction

Transcription factors play a key role in regulation of ABC transporter expression, and transcription factor mediated regulation of transporters has been characterized for specific genes and transcription factor families[1]. Discovery of new transcription factor regulators of ABC transporter expression typically occurs through accidental discovery during study of a disease or drug related

phenotype[2] or targeted evaluation of a specific transcription factor[3]. An unbiased approach to identification of transcription factor regulators of transporter expression may reveal novel associations. On the assumption that genes with similar co-expression patterns in a given tissue are regulated by the same transcription factors, sets of genes co-expressed with ABC transporters were extracted and then transcription factor binding motifs enriched in the 1000 bp region upstream of these genes were identified.

Several approaches to co-expression analysis have been applied to gene expression data. At the simplest level, correlations in expression patterns between two genes across multiple samples can be calculated. This pairwise analysis can provide valuable information about the relationships between any two genes, but doesn't reflect true biological interactions between multiple genes. Clustering methods have also been applied to gene expression data[4]. Such methods do allow for identification of sets of genes whose expression patterns are similar across individuals, but do not reveal the relationship between genes in a cluster. Co-expression network analyses address these issues; in these methods, a graph is constructed with links (edges) between genes (nodes) representing co-expression between pairs of genes. Edges can be dependent on some threshold of co-expression[5–7], over which pairs of genes are linked, or can be weighted by the correlation between pairs of genes[8,9]. Clustering methods applied to the graph allow for the detection of co-expression network modules or clusters, while still preserving the relationship between pairs of genes[10,11]. In this study, a weighted co-expression network method and associated R package,

Weighted Gene Coexpression Network Analysis (WGCNA)[12] was used to identify co-expression modules containing ABC transporters in four human tissues (liver, adipose tissue, heart and kidney) and lymphoblastoid cell lines.

Many methods and approaches have been developed for motif enrichment in a set of sequences. Broadly, such methods look for overrepresentation of a specific sequence motif in a set of test sequences relative to a set of background sequences. Motifs may be pre-defined (such as a set of selected transcription factor motifs) or may be identified *ab initio* from the test sequence itself[13]. The latter approach allows for identification of potential novel transcription factors, but is a less effective method for detecting overrepresentation of known transcription factor motifs. Transcription factor binding motif enrichment, even of known motifs, is complicated by the fact that these sequences are short, 7-10 bp, and can be highly degenerate. Consensus sequences describe the probability of seeing each nucleotide (A,T,C,G) at each position in the motif. These sequences can be used to estimate the affinity of a given transcription factor to a particular DNA sequence. Most published methods for estimation of motif enrichment rely on a contingency table approach to estimate enrichment, counting the number of times a motif reaches some defined affinity threshold in test and background sequences[14–16], with some methods taking into account the fact that a given motif can be identified multiple times in a single sequence[17]. However, multiple binding sites with varying degrees of affinity for a single transcription factor can exist within a given regulatory region[18,19]. In this study, a motif enrichment method, Clover[20], that takes into account motif affinity scores in the enrichment analysis,

was used rather than setting a hard threshold on affinity, and allows for multiple matches for a given motif in a single sequence.

Finally, because co-expressed genes may share functional relationships in addition to transcriptional regulators[4,21], an evaluation of genes co-expressed with ABC transporters may generate novel hypotheses about the function of ABC transporters; to explore function, annotated biological pathways enriched in sets of genes co-expressed with transporters are identfied. Together, the co-expression network analysis and transcription factor motif enrichment will generate novel hypotheses about the function and transcriptional regulation of ABC transporters in human tissues. An overview of the methodology used in this study is included in Figure 4.1.

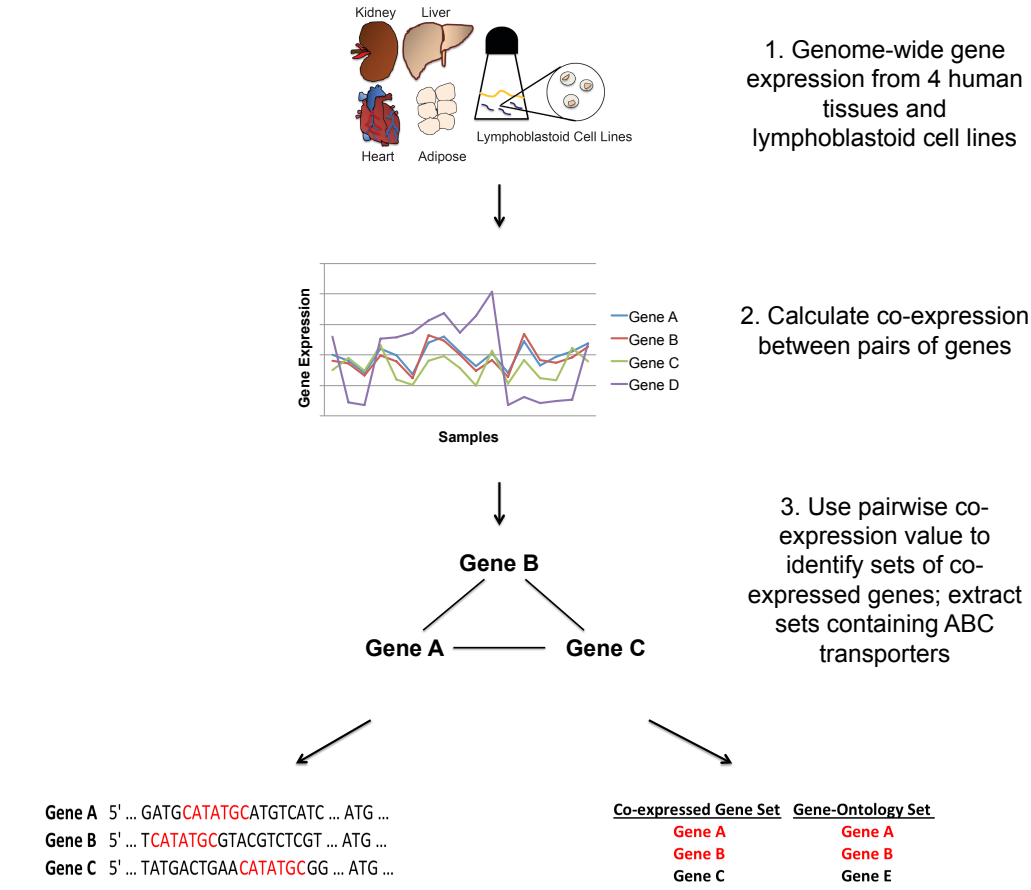


Figure 4.1. Overview of methodology. Steps 1-3 are conducted for each tissue type studied. Steps 4 and 5 are conducted for each ABC transporter co-expression module identified in a given tissue type.

4.3 Methods

4.3.1 Co-expression Network Analysis

All samples for co-expression analyses come from healthy human heart, liver, adipose, and kidney tissue and lymphoblastoid cell lines. Gene expression (FPKM, Fragments Per Kilobase Of Exon Per Million Fragments Mapped) estimates come from analysis of RNA-Seq data with Tophat v2.1.1[22], Cufflinks v2.0.2[23,24], and Cuffdiff (v2.2.1). Further details regarding sample preparation, alignment, and analysis are included in Chapter 2. The analyses in this section were restricted to 18 randomly selected samples per tissue type for consistency across tissues. Gene expression for each individual was calculated by summing FPKM values for all isoforms for a given gene for each individual from the isoforms.read_group_tracking file generated by Cuffdiff. Genes with any isoform with a status of "HIDATA" were excluded from subsequent analyses.

Data preprocessing steps included removal of genes with zero variance, setting minimum FPKM values to 1, and transforming FPKM values to log2 scale. Co-expression network construction was conducted with the Weighted Gene Co-expression Analysis (WGCNA) package[12]. Soft thresholding was used to transform coexpression (pairwise Pearson correlations) values by a power, beta; these values were selected for each tissue type using the pickSoftThreshold() function. The following powers (β**)** were used for each tissue: kidney 10, heart 12, adipose 10, liver 10, LCLs 8. Scale free topology R^2 for each selected β was >0.8, with the exception of heart tissue (R^2 = 0.77 for β = 12). Weighted

unsigned modules were generated using parameters minModuleSize=5, deepSplit=4, and mergeCutHeight=0.1.

4.3.2 Gene ontology enrichment analysis

To generate hypotheses about the function of ABC transporters that have not been characterized, biological process ontologies in the Gene Ontology[25] (as defined in the bioconductor package GO.db release 2.14) overrepresented in sets of genes co-expressed with ABC transporters were identified using the conditional hypergeometric test as implemented in the bioconductor package GO.stats[26]. Genes expressed in each tissue (FPKM ≥ 1） were used as background gene sets.

4.3.3 Transcription factor motif enrichment analysis

For each module in a given tissue type containing an ABC transporter, the 1000 bases upstream of the transcription start site (TSS) of each gene in the module was extracted using the R package biomaRt with Ensembl gene definitions[27]. This sequence set was analyzed for enrichment of transcription factor binding sites using Clover[20] with default parameters. A background sequence set was generated containing the 1000 bases upstream of the TSS of all human genes. The human chromosome 20 sequence and human CpG islands (from UCSC genome browser, 14-Apr-2003) were also used as additional background sequence sets.

Transcription factor motifs for enrichment analysis come from analysis of the 1000 bases upstream of ABC transporter TSSs using TRANSFAC® Match with the 2014.4 matrix library, high quality matrices, and the minimize false positives (minFP) setting (Table 4.1). Motifs identified in the ABC transporter upstream region were used in Clover motif enrichment search.

Table 4.1: Transcription factor binding sites evaluated in enrichment analysis

AHR	CRX	AP1	LRH1	POU2F1	TATA
BEN	CTCF	GFI1	MAF	POU6F1	TBX5
TTF1	DEAF1	GKLF	ARID5a	RBPJkappa	TEF1
XVENT1	deltaEF1	GLI	MAFA	RELA	
YY1	DLX3	GRE	MAZ	REST	
ZFP105	DMRT4	HES1	MEF2	NR1D1	
ZF5	LXR	HIC1	MEIS1	RFX1	
ZFX	DRI1	HIF1aplpha	MUSCLE	RFX	
ZNF333	E2A	HMGIY	MYOGENIN	RNF96	
ZSCAN4	Ebox	HNF1	MZF1	RREB1	
BRCA1	AML1	HNF3beta	NANOG	RUSH1alpha	
CDP	EGR1	AP2alpha	NF1	SF1	
CDX2	ERalpha	HNF4	BBX	BCL6	
CEBPA	EVI1	HSF1	NF-AT1	SMAD4	
CHCH	FAC1	ING4	NFY	SOX10	
cMYB	FPM315	IRF1	NKX2	SP100	
COE1	FREAC3	ISLET1	p53	SP1	
CPBP	GATA	KAISO	PBX	SREBP	
ATF2	GCM2	LBP9	PIT1	SRY	
AIRE	INITIATOR	LEF1	PLZF	STAT1	

4.4 Results

Using gene expression data from multiple individuals in multiple human tissue types, co-expression networks were constructed for ABC transporters. A representative example of these networks is shown in Figure 4.2.

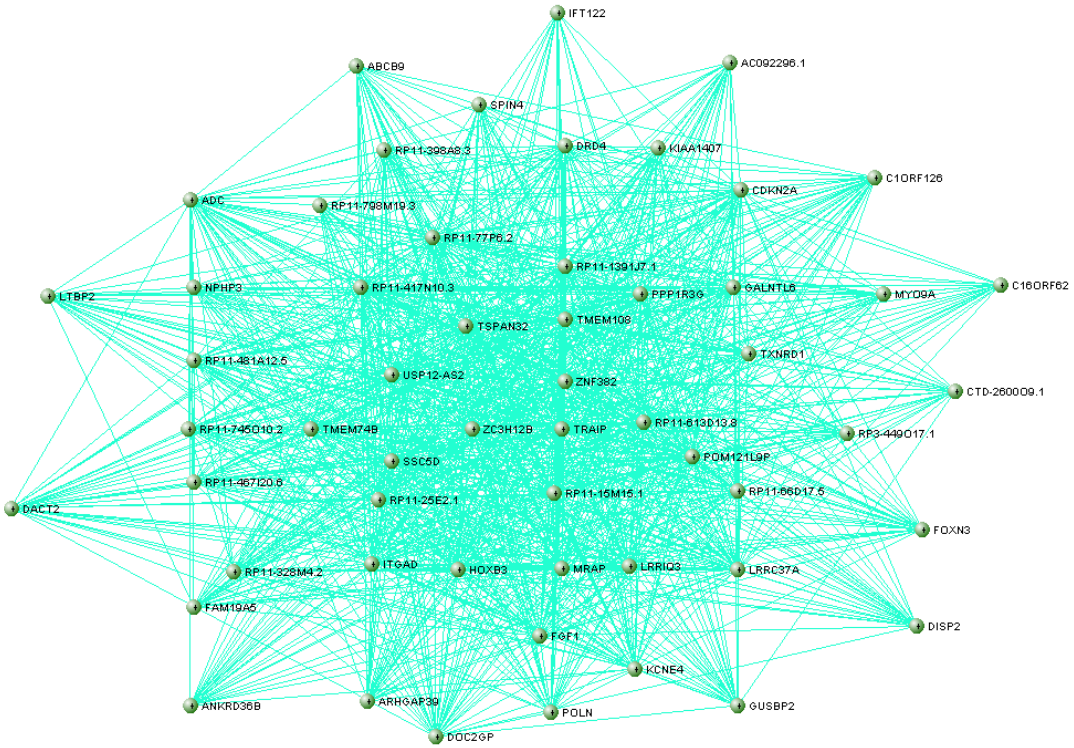


Figure 4.2. Unsigned co-expression network module for *ABCB9* in liver tissue. Only edges between genes with adjacency > 0.1 are shown**.**

Transcription factor motif enrichment analysis was conducted on the 1000 bp region upstream of genes within each module containing an ABC transporter. Only motifs identified in the 1000 bases upstream of the ABC transporter were tested. Enrichment of a specific motif was evaluated relative to its presence in the 1000 bases upstream of all human genes, the sequence of human chromosome 20, and in human CpG islands. Enriched motifs were defined as those with enrichment p-value < 0.01 for all three background sequence sets and a positive raw Clover enrichment score. Transcription factors with motifs enriched in genes co-expressed with an ABC transporter are included in Table 4.2, and enriched motifs that occur in two or more tissues for a single ABC transporter are listed in Table 4.3. Several known transcriptional regulators of specific ABC transporters were identified, along with many novel associations.

As co-expressed genes may also be functionally related, an evaluation of sets of genes co-expressed with transporters may generate novel hypotheses about the role of the transporter in each tissue type. Enrichment of gene ontologies representing biological processes was calculated using a conditional hyper-geometric test, which takes into account the tree-like structure of the ontologies in calculating enrichment. Ontologies enriched (enrichment $p<0.001$, >5 genes present in transporter gene set) in genes co-expressed with transporters are listed in Table 4.4. Up to three ontologies for each transporter are included in the table.

Table 4.2: Enriched transcription factor motifs in ABC transporter modules

LCLs		Kidney		Heart		Adipose		Liver	
ABCA3	BBX	*ABCA10*	ZFP161, ZF5	*ABCA1*	AP-2alpha	*ABCA1*	HNF4	*ABCA1*	IRF1
ABCB10	LBP9	*ABCB1*	MAZ	*ABCA1*	CDX2	*ABCA10*	STAT1	*ABCA1*	MUSCLE INITIATOR B
ABCB1	EVI1	*ABCB1*	MZF1	*ABCA1*	DRI1	*ABCA2*	EVI1	*ABCA1*	NF-AT1
ABCB8	FAC1	*ABCB1*	NANOG	*ABCA1*	EGR1	*ABCA3*	CDX2	*ABCA3*	E2A
ABCB8	ZFX	*ABCB1*	RREB1	*ABCA1*	HNF3B	*ABCA3*	HNF-3beta, FOXA2	*ABCA3*	HES1
ABCB8	EGR1	*ABCB1*	SP1	*ABCA1*	NFY	*ABCA3*	LBP9	*ABCA3*	Nkx-2.5
ABCB8	NFY	*ABCB10*	EGR1	*ABCA1*	ZFX	*ABCA3*	TBX5	*ABCA4*	GKLF
ABCB8	AHR	*ABCB10*	HNF-3beta, FOXA2	*ABCA3*	SP1	*ABCA4*	DR4	*ABCA7*	CHURCHILL
ABCC3	SP1	*ABCB10*	MUSCLE INITIATOR B	*ABCA7*	AP-2alpha	*ABCA4*	RNF96	*ABCA7*	DRI1
ABCC3	NANOG	*ABCB10*	NFAT1	*ABCA7*	CDX2	*ABCA7*	GKLF	*ABCA7*	EGR1
ABCC3	ZFX	*ABCB10*	RNF96	*ABCA7*	DRI1	*ABCA8*	AP-2alpha	*ABCA7*	HNF-3beta, FOXA2
ABCC3	EGR1	*ABCB10*	SP1	*ABCA7*	EGR1	*ABCA8*	DRI1	*ABCA7*	IRF1
ABCC3	DRI1	*ABCB10*	SRY	*ABCA7*	HNF3B	*ABCA8*	SRY	*ABCA7*	MUSCLE INITIATOR B
ABCC3	NFY	*ABCB10*	ZFX	*ABCA7*	NFY	*ABCA8*	ZFX	*ABCA7*	NF-AT1
ABCC4	IRF1	*ABCB8*	NFY	*ABCA7*	ZFX	*ABCA9*	KAISO	*ABCA7*	RNF96
ABCC5	MEF2	*ABCB8*	SP1	*ABCA8*	LBP9	*ABCA9*	ZFP161, ZF5	*ABCA7*	SRY
ABCC5	GRE	*ABCB9*	CHURCHILL	*ABCA8*	Rev-ErbALPHA	*ABCB10*	ZFX	*ABCA7*	ZFP161, ZF5
ABCD2	EGR1	*ABCB9*	EGR1	*ABCA9*	LBP9	*ABCB4*	CHURCHILL	*ABCA7*	ZFX
ABCD3	ZFP161, ZF5	*ABCB9*	EVI1	*ABCA9*	Rev-ErbALPHA	*ABCB4*	DMRT4	*ABCA8*	TEF-1
ABCD4	REST	*ABCB9*	MUSCLE INITIATOR B	*ABCB1*	RREB1	*ABCB4*	HNF-3beta, FOXA2	*ABCB1*	EGR1
ABCA3	BBX	*ABCB9*	SP1	*ABCB9*	AP-2alpha	*ABCB7*	AP-2alpha	*ABCB10*	AHR
ABCB10	LBP9	*ABCB9*	ZFX	*ABCB9*	CDX2	*ABCB7*	DRI1	*ABCB10*	EGR1
ABCB1	EVI1	*ABCC1*	MAZ	*ABCB9*	DRI1	*ABCB7*	SRY	*ABCB10*	IRF1
ABCB8	FAC1	*ABCC1*	MZF1	*ABCB9*	EGR1	*ABCB7*	ZFX	*ABCB10*	LBP9

ABCB8	ZFX	*ABCC1*	NANOG	*ABCB9*	HNF3B	*ABCB8*	NFY	*ABCB10*	MAZ
ABCB8	EGR1	*ABCC1*	RREB1	*ABCB9*	NFY	*ABCC1*	ZFX	*ABCB10*	SP1
ABCB8	NFY	*ABCC1*	SP1	*ABCB9*	ZFX	*ABCC10*	MAZ	*ABCB4*	RNF96
ABCB8	AHR	*ABCC10*	EGR1	*ABCC1*	CDX2	*ABCC3*	GRE	*ABCB8*	ZFP161, ZF5
ABCC3	SP1	*ABCC10*	RNF96	*ABCC1*	DRI1	*ABCC5*	COE1	*ABCB9*	EGR1
ABCC3	NANOG	*ABCC10*	ZFP161, ZF5	*ABCC1*	HMGIY	*ABCC5*	EGR1	*ABCB9*	ZFX
ABCC3	ZFX	*ABCC5*	CRX	*ABCC1*	HNF3B	*ABCC6*	REST	*ABCC4*	ZFP161, ZF5
ABCC3	EGR1	*ABCC5*	DRI1	*ABCC1*	SP1	*ABCC8*	DR4	*ABCC4*	ZFX
ABCC3	DRI1	*ABCC6*	HNF-3beta, FOXA2	*ABCC1*	SRY	*ABCC8*	RNF96	*ABCG2*	GRE
ABCC3	NFY	*ABCC8*	CPBP	*ABCC10*	DRI1	*ABCC9*	NFY		
ABCC4	IRF1	*ABCC8*	ZFP161, ZF5	*ABCC10*	NANOG	*ABCD1*	SP1		
ABCC5	MEF2	*ABCC8*	GKLF, KLF4	*ABCC10*	NFY	*ABCD2*	CHURCHILL		
ABCC5	GRE	*ABCC8*	MAZ	*ABCC10*	Rev-ErbALPHA	*ABCD2*	EGR1		
ABCD2	EGR1	*ABCD1*	EGR1	*ABCC10*	ZNF333	*ABCD2*	GKLF		
ABCD3	ZFP161, ZF5	*ABCD1*	PLZF	*ABCC10*	ZSCAN4	*ABCD3*	EGR1		
ABCD4	REST	*ABCD3*	ZFX	*ABCC3*	MAZ	*ABCD4*	EVI1		
		ABCD4	GRE	*ABCC8*	FPM315				
				ABCC8	GKLF				
				ABCC9	HNF4				
				ABCD1	HSF1				
				ABCD1	RBPJkappa				
				ABCD2	DRI1				
				ABCD2	EGR1				
				ABCD2	MUSCLE INITIATOR B				
				ABCD2	RNF96				
				ABCD2	SP1				
				ABCD2	ZFX				

ABCD3	DRI1
ABCD3	GKLF
ABCD3	HNF3B
ABCD3	LBP9
ABCD3	Rev-ErbALPHA
ABCD3	SP1
ABCD4	EVI1
CFTR	GFI1

Table 4.3: Enriched transcription factor motifs in ABC transporter modules in two or more tissues

Tissues	Transporter	Transcription Factor	Alias	Known Association?
Heart, Liver	*ABCA7*	Dead ringer homolog	DRI-1, ARID3A	
Heart, Liver	*ABCA7*	Early growth response protein 1	EGR1	
Heart, Liver	*ABCA7*	Hepatocyte nuclear factor 3,beta	HNF-3beta, FOXA2	
Heart, Liver	*ABCA7*	Zinc finger X-chromosomal protein	ZFX	
Kidney, Heart	*ABCB1*	Ras responsive element-binding protein 1	RREB1	
Kidney, Liver	*ABCB10*	Early growth response protein 1	EGR1	
LCLs, Liver	*ABCB10*	Fatty acid-binding protein homolog 9	LBP9	
Kidney, Liver	*ABCB10*	Specificity protein 1	SP1	
Kidney, Adipose	*ABCB10*	Zinc finger X-chromosomal protein	ZFX	
LCLs, Kidney, Adipose	*ABCB8*	Nuclear transcription factor Y	NFY	
Kidney, Heart, Liver	*ABCB9*	Early growth response protein 1	EGR1	
Kidney, Heart, Liver	*ABCB9*	Zinc finger X-chromosomal protein	ZFX	
Kidney, Heart	*ABCC1*	Specificity protein 1	SP1	[28,29]
Kidney, Heart	*ABCC8*	Zinc finger protein 263	ZNF263,FPM315	
Kidney, Heart	*ABCC8*	Gut enriched Krüppel-like factor	GKLF, KLF4	
LCLs, Heart, Adipose	*ABCD2*	Early growth response protein 1	EGR-1	
Heart, Adipose	*ABCD4*	Ecotropic virus integration site 1 protein homolog	EVI1	

Table 4.4: Biological process gene ontologies enriched in sets of genes co-expressed with ABC transporters

Gene	Tissue	Gene Ontology Biological Process
ABCA1	Liver	protein modification process
		cellular protein metabolic process
		protein modification by small
	Heart	response to retinoic acid
		cellular response to cAMP
		mesenchymal cell proliferation
ABCA10	Liver	inner ear morphogenesis
		neural tube closure
		homophilic cell adhesion
ABCA2	Heart	respiratory electron transport chain
		mRNA metabolic process
		protein polyubiquitination
ABCA3	Adipose	cellular macromolecule catabolic process
ABCA4	Adipose	ion transport
		metal ion transport
ABCA6	Liver	carboxylic acid metabolic process
		organic acid metabolic process
		cellular ketone metabolic process
	Adipose	macromolecule methylation
	Heart	positive regulation of signaling
		positive regulation of cell
		regulation of cell proliferation
ABCA7	Liver	protein modification process
		cellular protein metabolic process
		protein modification by small protein conjugation or

		removal
	Adipose	immune response-activating cell surface
		immune response
		immune response-regulating signaling pathway
	Heart	response to retinoic acid
		cellular response to cAMP
		mesenchymal cell proliferation
ABCA8	Adipose	cellular process
		response to ionizing radiation
ABCA9	Liver	carboxylic acid metabolic process
		organic acid metabolic process
		cellular ketone metabolic process
ABCB1	Liver	transport
ABCB11	Liver	carboxylic acid metabolic process
		organic acid metabolic process
		cellular ketone metabolic process
	Adipose	macromolecule methylation
ABCB4	Heart	respiratory electron transport chain
		mRNA metabolic process
		protein polyubiquitination
ABCB7	Adipose	cellular process
		response to ionizing radiation
	Heart	energy derivation by oxidation of organic compounds
		monocarboxylic acid catabolic process
		lipid oxidation
ABCB8	LCLs	energy derivation by oxidation of organic compounds
		small molecule metabolic process

		primary metabolic process
ABCB9	Heart	response to retinoic acid
		cellular response to cAMP
		mesenchymal cell proliferation
ABCC1	Heart	endosomal transport
		ribonucleoprotein complex subunit organization
		cellular macromolecular complex assembly
ABCC3	Liver	organic acid catabolic process
		carboxylic acid metabolic process
	LCLs	cellular component assembly
		protein complex assembly
ABCC4	Adipose	immune system process
		response to stress
		response to biotic stimulus
	LCLs	positive regulation of protein serine/threonine kinase activity
ABCC5	Liver	transport
ABCC6	Heart	sensory perception
ABCC8	Adipose	ion transport
		metal ion transport
	Heart	immune response
ABCC9	Adipose	lysosome organization
		sphingolipid catabolic process
		leukocyte migration
ABCD1	Adipose	nucleic acid metabolic process
		histone modification
		chromatin modification
ABCD2	Adipose	cellular response to vitamin

		cellular response to retinoic acid
		cellular response to external stimulus
	Heart	mitotic cell cycle
ABCD3	LCLs	RNA splicing, via transesterification
		RNA processing
		nuclear mRNA splicing, via spliceosome

4.5 Discussion

Using gene expression data from eighteen individuals each from lymphoblastoid cell lines and human liver, kidney, adipose tissue, and heart, we identified putative transcriptional regulators of ABC transporter expression by transcription factor motif enrichment in co-expressed gene modules containing ABC transporters.

Among the transporter-transcription factor associations, a number have been previously reported in the literature including regulation of ABCA1 expression by activating enhancer binding protein 2 alpha (AP-2alpha) [30] and forkhead box protein A2 (FOXA2) [31] and regulation of ABCB8[32], ABCB1[33], and ABCC1[28,29] expression by specificity protein 1 (Sp1). A number of novel associations were identified as well; some of these are discussed in detail below.

For some novel transcription factor-transporter associations, while there is no direct evidence in the literature for transcriptional regulation of these genes by the transcription factor, there is indirect evidence based on disease associations. For example, Ras responsive element-binding protein 1 (RREB-1) was identified as a transcriptional regulator of several ABC transporters. RREB-1 is upregulated in prostate cancer[34], as is ABCC1 expression[35], suggesting RREB-1 may play a role in the transcriptional activation of *ABCC1*. Likewise, RREB has been identified in an Alzheimer's disease specific gene cluster[36], and *ABCB1* has been implicated in Alzheimer's disease[37–39]. In other cases, a transcription factor-transporter association was identified in other members of a transporter

subfamily where there is known regulation by the transcription factor. For example, nuclear factor of activated T-cells, cytoplasmic 2 (NFAT2C) has been previously shown to be a regulator of ABCA1 expression and binds in the *ABCA1* promoter region[40]. While an association between *ABCA7* and NFAT1 has not been published previously, given the expected functional similarity between ABCA1 and ABCA7, it is likely that NFAT1 is also a transcriptional regulator of *ABCA7.* Likewise, hepatocyte nuclear factors (HNF1,3,4) have been widely implicated in regulation of ABC transporter expression[41–44], and may be involved in regulating additional transporters identified here. Similarly, specificity protein 1 (SP-1) activates transcription of a number of ABC transporters[28,29,32,33,45–47] from multiple subfamilies, and likely activates transcription of additional transporters. Likewise, early growth response protein 1 (EGR-1) regulates transcription of several trasporters.[45,48] Indeed, there is evidence of co-operative transcriptional regulation between EGR-1 and SP-1[45]; transporters regulated by SP-1 may also be regulated by EGR-1 in particular tissues or cell conditions.

While ABC transporters are generally studied in the context of their function in differentiated cells, they are expressed in and play a key role in the function and differentiation of stem cells, including embryonic[49] and hematopoietic stem cells[50,51]. Several identified transcription factors are known to play a role in hematopoiesis, such as the aryl hydrocarbon receptor (AHR)[52], myeloid zinc finger protein 1 (MZF1) [53,54], acute myeloid leukemia 1 protein (AML-1), early growth response protein 1 (EGR-1)[55], dead ringer homolog (DRI-1, ARID3A)[56], and ecotropic virus integration site 1 protein homolog (EVI-1)[57]. Others play a key

role in maintenance and differentiation of embryonic stem cells, such as homeobox protein NANOG[58]. Given the importance of both the ABC transporters and these transcription factors in stem cells, it is possible that they play a role in regulating transporter expression.

Transcription factors were identified from unsigned co-expression network construction methods. Unsigned methods allow for both positive and negative correlations between genes; transcription factor regulators of gene sets identified from such networks may be dual-regulators, causing increase in expression of some genes and decrease in expression of others. Many transcription factors can act as both repressors or enhancers depending on the gene target and cell conditions[59,60], and it is possible this dual regulatory role is important for many transcription factors. For example, gut-enriched Krüppel-like factor (GKLF) is a transcription factor with known ability to both activate and suppress transcription[61]. Suppression of transcription can be mediated by GKLF's ability to disrupt Sp-1 mediated transcriptional activation and enhance Sp-3 mediated transcriptional suppression[62]. Activation of transcription by GKLF can be mediated by interactions with CREB-binding protein (CBP) as part of the CBP/p300 complex[61]. In this analysis, ABCC8 was associated with GKLF in the kidney. Given that there is evidence that Abcc8 is regulated by Sp-1 family in mice[63], GLKF may act through this transcription factor to modulate transcriptional activation or repression in the identified transporters. Likewise, while little is known about ZNF263, a gene expression analysis after knockdown of the transcription factor suggested that it can both activate and repress gene

expression[64] , as can nuclear factor Y (NFY)[65]. Even transcription factors that generally act as either repressors or activators can take on a dual-regulatory role in some promoters. For example, SP1, a transcriptional activator of many genes, can act as a repressor of certain genes[66,67]. Given the potential for so many transcription factors to activate or repress transcription, unsigned co-expression networks can be valuable when studying transcriptional regulation.

Any analysis of transcription factor binding motifs, including the current analysis, is limited by the quality and availability of binding motifs. While the quality and quantity of these motifs continue to improve, it is likely that numerous additional transcription factors and further improvements to the consensus sequence of existing motifs will generate additional results in the future. Further, transcription factors can bind to sequences much farther upstream than the first 1000 bases before the transcription start site, and can also bind to sequences within the gene itself, most commonly within the first intron[68,69]. However increasing the sequence space in the analysis would also increase noise, thus in this study we limit sequences to the 1000 bp region upstream of each gene. Limiting the analysis to this regulatory region minimizes the number of false positive motifs, albeit at the cost of missing some true transcription factor motifs.

In addition to exploring transcriptional regulation, co-expression analyses can also be used to explore gene function. Biological process gene ontologies enriched in genes co-expressed with ABC transporters were identified in the current analysis. A handful of identified processes have been previously associated with function of transporters. For example, *ABCC8* encodes a

sulfonylurea receptor (SUR1) that is involved in potassium ion transport[70]. In adipose tissue, genes co-expressed with *ABCC8* are enriched for the ion transport biological process. Likewise, *ABCB8* and *ABCB7* were associated with the gene ontology biological process for energy derivation by oxidation of organic compounds occurring in the mitochondria. Both genes are mitochodnrial membrane transporters that play a key role in mitochondrial function[71,72]. However, a large number of the identified processes do not have any apparent relationship with ABC transporter function. As others have found previously in co-expression analyses conducted in other organisms, annotating gene function based on co-expression information alone has limited sensitivity[73], particularly in healthy tissues and cells[74]. While the goal was to study transporter function in healthy cells, such functional annotation may be best conducted in tissues or cells with some disease or environmental exposure that perturbs a particular biological system. Further, there has been some success in functional annotation of genes using co-expression methods that take into account conservation between species[75–77]; such methods may prove more fruitful for examination of transporter function. Finally, as with any study involving pathway and gene set databases and ontologies, the value and accuracy of the results depend largely on the quality of the gene set annotation in the literature. As the function of many genes have not been fully characterized, functional annotation in co-expressed gene sets will only be successful in sets in which the majority of genes have complete gene annotation.

Both motif enrichment and gene ontology enrichment analyses are limited by the quality of the input sequences that are determined to be co-regulated. All genes that fall within a co-expressed network module with ABC transporters are included, including both direct and indirect associations. While direct associations are more likely to be co-regulated with ABC transporters, exclusion of indirect associations may result in a loss of power for enrichment analysis. In addition, the parameters applied during module detection in WGCNA were selected to provide the most conservative definition of a co-expression network module, while still retaining information about biological interactions between genes.

Further, co-expression analyses only detect groups of genes in a given dataset with similar expression patterns. While these patterns of expression may be similar due to shared transcription factor regulators, it is also possible that co-expression is driven by technical artifacts, functional similarities between genes without co-regulation, or other regulatory mechanisms aside from transcription factors. The use of expression data generated by RNA-Seq rather than microarray and use of identical study protocols across all samples and tissues minimizes artifacts in the data, but it is possible additional factors that could not be controlled for in the study contribute to co-expression results.

Using high-quality gene expression data, co-expression network modules containing ABC transporters in human heart, liver, kidney, adipose tissue and lymphoblastoid cell lines were identified. Transcription factor motifs were found in ABC transporter 5' regulatory regions enriched in these co-expression modules.

The transcription factors that bind these motifs may represent novel regulators of ABC transporter expression.

4.6 References

1 Scotto KW. Transcriptional regulation of ABC drug transporters. *Oncogene* 2003; **22**: 7496–511.

2 Osborn MT, Chambers TC. Role of the stress-activated/c-Jun NH2-terminal protein kinase pathway in the cellular response to adriamycin and other chemotherapeutic drugs. *J Biol Chem* 1996; **271**: 30950–5.

3 Zollner G, Fickert P, Fuchsbichler A, Silbert D, Wagner M, Arbeiter S, Gonzalez FJ, Marschall H-U, Zatloukal K, Denk H, Trauner M. Role of nuclear bile acid receptor, FXR, in adaptive ABC transporter regulation by cholic and ursodeoxycholic acid in mouse liver, kidney and intestine. *J Hepatol* 2003; **39**: 480–8.

4 Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998; **95**: 14863–8.

5 Aggarwal A, Guo DL, Hoshida Y, Yuen ST, Chu K-M, So S, Boussioutas A, Chen X, Bowtell D, Aburatani H, Leung SY, Tan P. Topological and functional discovery in a gene coexpression meta-network of gastric cancer. *Cancer Res* 2006; **66**: 232–41.

6 Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science (80-)* 2003; **302**: 249–55.

7 Ray M, Ruan J, Zhang W. Variations in the transcriptome of Alzheimer's disease reveal molecular networks involved in cardiovascular diseases. *Genome Biol* 2008; **9**: R148.

8 Zhou X, Kao M-CJ, Wong WH. Transitive functional annotation by shortest-path analysis of gene expression data. *Proc Natl Acad Sci U S A* 2002; **99**: 12783–8.

9 Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet* 2006; **2**: e130.

10 Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. *Genome Res* 2004; **14**: 1085–94.

11 Zhu D, Hero AO, Cheng H, Khanna R, Swaroop A. Network constrained clustering for gene microarray data. *Bioinformatics* 2005; **21**: 4014–20.

12 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559.

13 Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov A V, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov A a, Noble WS, Pavesi G, Pesole G, Régnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol* 2005; **23**: 137–44.

14 Aerts S. Toucan: deciphering the cis-regulatory logic of coregulated genes. *Nucleic Acids Res* 2003; **31**: 1753–64.

15 Liu R, McEachin RC, States DJ. Computationally identifying novel NF-kappa B-regulated immune genes in the human genome. *Genome Res* 2003; **13**: 654–61.

16 Zheng J. An approach to identify over-represented cis-elements in related sequences. *Nucleic Acids Res* 2003; **31**: 1995–2005.

17 Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res* 2003; **13**: 773–80.

18 Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome DEVELOPMENTAL. *Construction* 2002; **99**.

19 Papatsenko DA, Makeev VJ, Lifanov AP, Regnier M, Nazina AG, Desplan C. Extraction of Functional Binding Sites from Unique Regulatory Regions: The Drosophila Early Developmental Enhancers. *Genome Res* 2002; **12**: 470–81.

20 Frith MC, Fu Y, Yu L, Chen J-F, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* 2004; **32**: 1372–81.

21 Zhou XJ, Kao M-CJ, Huang H, Wong A, Nunez-Iglesias J, Primig M, Aparicio OM, Finch CE, Morgan TE, Wong WH. Functional annotation and network reconstruction through cross-platform integration of microarray data. *Nat Biotechnol* 2005; **23**: 238–43.

22 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–11.

23 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; **28**: 511–5.

24 Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; : 1–5.

25 Harris M, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin G, Blake J, Bult C, Dolan M, Drabkin H, Eppig J, Hill D, Ni L, Ringwald M, Balakrishnan R, Cherry J, Christie K, Costanzo M, Dwight S, Engel S, Fisk D, Hirschman J. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004; **32**: D258–61.

26 Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. *Bioinformatics* 2007; **23**: 257–8.

27 Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009; **4**: 1184–91.

28 Wang Q, Beck WT. Transcriptional suppression of multidrug resistance-associated protein (MRP) gene expression by wild-type p53. *Cancer Res* 1998; **58**: 5762–9.

29 Zhu Q, Center MS. Evidence that SP1 modulates transcriptional activity of the multidrug resistance-associated protein gene. *DNA Cell Biol* 1996; **15**: 105–11.

30 Iwamoto N, Abe-Dohmae S, Ayaori M, Tanaka N, Kusuhara M, Ohsuzu F, Yokoyama S. ATP-binding cassette transporter A1 gene transcription is downregulated by activator protein 2alpha. Doxazosin inhibits activator protein 2alpha and increases high-density lipoprotein biogenesis independent of alpha1-adrenoceptor blockade. *Circ Res* 2007; **101**: 156–65.

31 Thymiakou E, Kardassis D. Novel mechanism of transcriptional repression of the human ATP binding cassette transporter A1 gene in hepatic cells by the winged helix/forkhead box transcription factor A2. *Biochim Biophys Acta* 2014; **S1874-9399**: 00104–7.

32 Sachrajda I, Ratajewski M. Mithramycin A suppresses expression of the human melanoma-associated gene ABCB8. *Mol Genet Genomics* 2011; **285**: 57–65.

33 Gromnicova R, Romero I, Male D. Transcriptional control of the multi-drug transporter ABCB1 by transcription factor Sp3 in different human tissues. *PLoS One* 2012; **7**: e48189.

34 Zou J, Milon BC, Desouki MM, Costello LC, Franklin RB. hZIP1 zinc transporter down-regulation in prostate cancer involves the overexpression of ras responsive element binding protein-1 (RREB-1). *Prostate* 2011. doi:10.1002/pros.21368.

35 Munoz M, Henderson M, Haber M, Norris M. Role of the MRP1/ABCC1 multidrug transporter protein in cancer. *IUBMB Life* 2007; **59**: 752–7.

36 Pasluosta CF, Dua P, Lukiw WJ. Nearest hyperplane distance neighbor clustering algorithm applied to gene co-expression analysis in Alzheimer's disease. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc* 2011; **2011**: 5559–62.

37 Vogelgesang S, Cascorbi I, Schroeder E, Pahnke J, Kroemer HK, Siegmund W, Kunert-Keil C, Walker LC, Warzok RW. Deposition of Alzheimer's beta-amyloid is inversely correlated with P-glycoprotein expression in the brains of elderly non-demented humans. *Pharmacogenetics* 2002; **12**: 535–41.

38 Cirrito JR, Deane R, Fagan AM, Spinner ML, Parsadanian M, Finn MB, Jiang H, Prior JL, Sagare A, Bales KR, Paul SM, Zlokovic B V, Piwnica-Worms D, Holtzman DM. P-glycoprotein deficiency at the blood-brain barrier increases amyloid-beta deposition in an Alzheimer disease mouse model. *J Clin Invest* 2005; **115**: 3285–90.

39 Vogelgesang S, Warzok RW, Cascorbi I, Kunert-Keil C, Schroeder E, Kroemer HK, Siegmund W, Walker LC, Pahnke J. The role of P-glycoprotein in cerebral amyloid angiopathy; implications for the early pathogenesis of Alzheimer's disease. *Curr Alzheimer Res* 2004; **1**: 121–5.

40 Maitra U, Parks JS, Li L. An innate immunity signaling process suppresses macrophage ABCA1 expression through IRAK-1-mediated downregulation

of retinoic acid receptor alpha and NFATc2. *Mol Cell Biol* 2009; **29**: 5989–97.

41 Moschetta A. Welcoming Foxa2 in the bile acid entourage. *Cell Metab* 2008; **8**: 276–8.

42 Ohoka N, Okuhira K, Cui H, Wu W, Sato R, Naito M, Nishimaki-Mogami T. HNF4α increases liver-specific human ATP-binding cassette transporter A1 expression and cholesterol efflux to apolipoprotein A-I in response to cholesterol depletion. *Arterioscler Thromb Vasc Biol* 2012; **32**: 1005–14.

43 Qadri I, Iwahashi M, Kullak-Ublick GA, Simon FR. Hepatocyte nuclear factor (HNF) 1 and HNF4 mediate hepatic multidrug resistance protein 2 up-regulation during hepatitis C virus gene expression. *Mol Pharmacol* 2006; **70**: 627–36.

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. Cooperative interaction between hepatocyte nuclear factor 4 alpha and GATA transcription factors regulates ATP-binding cassette sterol transporters ABCG5 and ABCG8. *Mol Cell Biol* 2007; **27**: 4248–60.

45 Davis W, Chen ZJ, Ile KE, Tew KD. Reciprocal regulation of expression of the human adenosine 5'-triphosphate binding cassette, sub-family A, transporter 2 (ABCA2) promoter by the early growth response-1 (EGR-1) and Sp-family transcription factors. *Nucleic Acids Res* 2003; **31**: 1097–107.

46 Dabrowska M, Sirotnak FM. Regulation of transcription of the human MRP7 gene. Characteristics of the basal promoter and identification of tumor-derived transcripts encoding additional 5' end heterogeneity. *Gene* 2004; **341**: 129–39.

47 Douet V, VanWart CM, Heller MB, Reinhard S, Le Saux O. HNF4alpha and NF-E2 are key transcriptional regulators of the murine Abcc6 gene expression. *Biochim Biophys Acta* 2006; **1759**: 426–36.

48 Cornwell MM, Smith DE. SP1 activates the MDR1 promoter through one of two distinct G-rich regions that modulate promoter activity. *J Biol Chem* 1993; **268**: 19505–11.

49 Barbet R, Peiffer I, Hutchins JRA, Hatzfeld A, Garrido E, Hatzfeld JA. Expression of the 49 human ATP binding cassette (ABC) genes in pluripotent embryonic stem cells and in early- and late-stage multipotent mesenchymal stem cells: possible role of ABC plasma membrane

transporters in maintaining human stem cell pluripotency. *Cell Cycle* 2012; **11**: 1611–20.

50 De Grouw EPLM, Raaijmakers MHGP, Boezeman JB, van der Reijden BA, van de Locht LTF, de Witte TJM, Jansen JH, Raymakers RAP. Preferential expression of a high number of ATP binding cassette transporters in both normal and leukemic CD34+CD38- cells. *Leukemia* 2006; **20**: 750–4.

51 Tang L, Bergevoet SM, Gilissen C, de Witte T, Jansen JH, van der Reijden BA, Raymakers RAP. Hematopoietic stem cells exhibit a specific ABC transporter gene expression profile clearly distinct from other stem cells. *BMC Pharmacol* 2010; **10**: 12.

52 Smith BW, Rozelle SS, Leung A, Ubellacker J, Parks A, Nah SK, French D, Gadue P, Monti S, Chui DHK, Steinberg MH, Frelinger AL, Michelson AD, Theberge R, McComb ME, Costello CE, Kotton DN, Mostoslavsky G, Sherr DH, Murphy GJ. The aryl hydrocarbon receptor directs hematopoietic progenitor cell expansion and differentiation. *Blood* 2013; **122**: 376–85.

53 Morris JF, Hromas R, Rauscher FJ. Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol Cell Biol* 1994; **14**: 1786–95.

54 Le Mée S, Fromigué O, Marie PJ. Sp1/Sp3 and the myeloid zinc finger gene MZF1 regulate the human N-cadherin promoter in osteoblasts. *Exp Cell Res* 2005; **302**: 129–42.

55 Min IM, Pietramaggiori G, Kim FS, Passegué E, Stevenson KE, Wagers AJ. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* 2008; **2**: 380–91.

56 Webb CF, Bryant J, Popowski M, Allred L, Kim D, Harriss J, Schmidt C, Miner CA, Rose K, Cheng H-L, Griffin C, Tucker PW. The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. *Mol Cell Biol* 2011; **31**: 1041–53.

57 Zhu J, Emerson SG. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene* 2002; **21**: 3295–313.

58 Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003; **113**: 631–42.

59 Valin A, Gill G. Regulation of the dual-function transcription factor Sp3 by SUMO. *Biochem Soc Trans* 2007; **35**: 1393–6.

60 Mo X, Kowenz-Leutz E, Xu H, Leutz A. Ras induces mediator complex exchange on C/EBP beta. *Mol Cell* 2004; **13**: 241–50.

61 Geiman DE, Ton-That H, Johnson JM, Yang VW. Transactivation and growth suppression by the gut-enriched Krüppel-like factor (Krüppel-like factor 4) are dependent on acidic amino acid residues and protein-protein interaction. *Nucleic Acids Res* 2000; **28**: 1106–13.

62 Zhang W, Shields JM, Sogawa K, Fujii-Kuriyama Y, Yang VW. The gut-enriched Krüppel-like factor suppresses the activity of the CYP1A1 promoter in an Sp1-dependent fashion. *J Biol Chem* 1998; **273**: 17917–25.

63 Woo SK, Kwon MS, Geng Z, Chen Z, Ivanov A, Bhatta S, Gerzanich V, Simard JM. Sequential activation of hypoxia-inducible factor 1 and specificity protein 1 is required for hypoxia-induced transcriptional stimulation of Abcc8. *J Cereb Blood Flow Metab* 2012; **32**: 525–36.

64 Frietze S, Lan X, Jin VX, Farnham PJ. Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J Biol Chem* 2010; **285**: 1393–403.

65 Peng Y. The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood* 2002; **99**: 2408–17.

66 Zaid A, Hodny Z, Li R, Nelson BD. Sp1 acts as a repressor of the human adenine nucleotide translocase-2 (ANT2) promoter. *Eur J Biochem* 2001; **268**: 5497–503.

67 Law AYS, Yeung BHY, Ching LY, Wong CKC. Sp1 is a transcription repressor to stanniocalcin-1 expression in TSA-treated human colon cancer cells, HT29. *J Cell Biochem* 2011; **112**: 2089–96.

68 Kim H-D, Choe HK, Chung S, Kim M, Seong JY, Son GH, Kim K. Class-C SOX transcription factors control GnRH gene expression via the intronic transcriptional enhancer. *Mol Endocrinol* 2011; **25**: 1184–96.

69 Bornstein P, McKay J, Morishima JK, Devarayalu S, Gelinas RE. Regulatory elements in the first intron contribute to transcriptional control of the human alpha 1(I) collagen gene. *Proc Natl Acad Sci U S A* 1987; **84**: 8869–73.

70 Nichols CG. KATP channels as molecular sensors of cellular metabolism. *Nature* 2006; **440**: 470–6.

71 Ichikawa Y, Bayeva M, Ghanefar M, Potini V, Sun L, Mutharasan RK, Wu R, Khechaduri A, Jairaj Naik T, Ardehali H. Disruption of ATP-binding cassette B8 in mice leads to cardiomyopathy through a decrease in mitochondrial iron export. *Proc Natl Acad Sci U S A* 2012; **109**: 4152–7.

72 Pondarré C, Antiochos BB, Campagna DR, Clarke SL, Greer EL, Deck KM, McDonald A, Han A-P, Medlock A, Kutok JL, Anderson SA, Eisenstein RS, Fleming MD. The mitochondrial ATP-binding cassette transporter Abcb7 is essential in mice and participates in cytosolic iron-sulfur cluster biogenesis. *Hum Mol Genet* 2006; **15**: 953–64.

73 Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y. Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol* 2009; **150**: 535–46.

74 Childs KL, Davidson RM, Buell CR. Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS One* 2011; **6**: e22196.

75 Teichmann SA, Babu MM. Conservation of gene co-regulation in prokaryotes and eukaryotes. *Trends Biotechnol* 2002; **20**: 407–10.

76 Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science (80-)* 2003; **302**: 249–55.

77 Van Noort V, Snel B, Huynen MA. Predicting gene function by conserved co-expression. *Trends Genet* 2003; **19**: 238–42.

Chapter 5: Summary and Perspectives

5.1 Summary

ATP binding cassette (ABC) transporters are a family of proteins whose activity is vital to cell detoxification, protection against xenobiotics and oxidative stress, and maintenance of homeostasis of endogenous compounds.[1–3] While rare and extreme disorders typically result from complete loss of function or expression of these essential transporters, more subtle changes in endogenous expression level can also have clinical implications. In addition, the role of the ABC transporters in movement of pharmaceutical agents across cellular membranes is of particular interest; inter-individual differences in expression of ABC transporters can result in changes in exposure to pharmaceutical agents or their metabolites, leading to altered efficacy or drug-induced toxicities[3]. Variation in gene expression between individuals and between tissues can be caused by a number of factors that modulate the “normal” activity during transcription or translation. While a number of large scale studies have been conducted examining various sources of expression regulation, most of these have been carried out in lymphoblastoid cell lines or a limited number of other human tissues or cell lines. The goal in this dissertation was to characterize mechanisms that regulate mRNA or protein expression of ATP-binding cassette transporters in human tissues, with a special emphasis on mechanisms regulating inter-individual variation in expression.

In the second chapter, variants in DNA sequence, or single nucleotide polymorphisms (SNPs), that are associated with mRNA expression levels of ABC transporters in the human kidney were identified. While a number of such expression quantitative trait loci (eQTL) analyses have been conducted in human tissues, no such studies have been performed in the human kidney. Six SNPs located within 500 kb of ABC transporters and associated with mRNA expression levels of the transporter in human kidney samples were identified, and the role of one of these variants (rs1471400) in regulating gene expression was validated in an *in vitro* reporter gene assay. Transcription factors containing variants that are associated with ABC transporter mRNA expression levels were also identified. These transcription factor-QTLs (tf-QTLs) also provide evidence for transcription factors involved in regulating ABC transporter expression.

In the third chapter alternative splicing events in ABC transporters were examined. While splicing of transporters has been studied on a gene by gene and tissue by tissue basis in a small number of samples, ideally an unbiased exploration of transporter splicing would be conducted using transcriptome sequencing data in multiple samples and tissue types. Splicing events in ABC transporters were cataloged using transcriptome-sequencing data from over twenty samples each for four human tissues and lymphoblastoid cell lines; these events include both previously annotated events, as well as novel events that are not a part of existing gene annotations. Further, selected events were identified that may be responsible for regulating inter-individual or inter-tissue variation in transporter expression. For regulation of inter-individual expression, I focused on

two sets of splicing events in the 5' untranslated regions of transporters *ABCA8* and *ABCC5* in the human heart, and demonstrated using *in vitro* reporter gene assays that usage of alternate 5'UTRs results in variation in reporter gene expression. For regulation of inter-tissue expression, I focused on splicing events containing a premature termination codon that may be targets of the nonsense mediated decay process, and demonstrated that one event in the gene *ABCC6* is a target of the decay process, and may be responsible for regulating inter-tissue expression of *ABCC6.* These examples suggest that alternative splicing of transporters likely plays a significant role in regulation of expression, and should be considered in future studies of transporter expression.

Finally, transcription factor regulators of ABC transporter expression were identified by looking for overrepresented transcription factor binding motifs in sets of co-expressed genes in multiple tissue types. Both known and novel transcription factor – transporter associations were identified. Co-expression networks were also used to explore further the function of ABC transporters. These analyses provide novel hypotheses about the regulation and function of ABC transporters.

5.2 Perspectives and Future Directions

Three key mechanisms of regulation of ABC transporter expression were examined in this dissertation – DNA sequence variation, alternative splicing, and transcription factor usage. Missing here is an additional important mechanism –

the role of epigenetic modifications in regulating gene expression. Two main forms of epigenetic changes include addition of a methyl group to cytosine nucleotides and various forms of chemical modifications to histone proteins that are involved in maintaining DNA conformation and tertiary structure[4,5]. Changes in both types of epigenetic modifications regulate the accessibility of a particular sequence to the transcription machinery, thereby regulating gene expression. The impact of epigenetic modifications to ABC transporter expression has been studied for a handful of transporters, particularly in the context of multidrug resistance and cancer therapeutics[6–8], but in general this area requires more study. In particular, epigenetic changes may be responsible for gene expression changes in response to environmental exposure, such as drug treatment[9], and thus is of particular interest in the regulation of expression of drug transporters.

Four tissues types are included in this body of research – kidney, liver, heart, and adipose tissues – along with one cell line. These sample types were selected to represent important tissues for the function of ABC transporters. However, two key tissues were missing from this research – the brain, specifically the blood-brain barrier (BBB), and the intestine. At the BBB, ABC transporters such as Pglycoprotein (Pgp,ABCB1) and breast cancer resistance protein (BCRP, ABCG2) protect the brain by effluxing toxins and xenobiotics[10]. Transporters at the BBB are of particular interest in development drugs whose site of action is in the brain. In the intestine, a number of ABC transporters are expressed on the apical membrane or intestine epithelial cells, where they may efflux compounds back into the intestinal lumen and modulate bioavailability of orally ingested

drugs[11]. Research on the regulation of transporter expression in these tissues would be a valuable addition to the results presented in this thesis.

In addition to the inter-individual variability that was studied explicitly, I also noticed significant tissue-specificity in both expression patterns and in regulators of gene expression. Table 1.1 indicates that the majority of identified eQTLs have been identified in a single tissue type. Most of the kidney eQTLs detected have not been previously identified, despite the sizable number of tissue types that have been studied. Splicing patterns were also quite variable across tissues, and may themselves be a mechanism for regulating tissue-specific expression, as I suggest is the case for one splicing event in the *ABCC6* gene. These observations emphasize the need to study transporters in a variety of tissue types, and to avoid the use of cell lines as proxies for human tissues as far as possible.

Finally, over the course of this research I noted that the function of a large number of ABC transporters still has not been well characterized. While selected transporters have been studied in great detail, particularly those that were initially associated with drug transport and multidrug resistance, the majority of transporters still have not been studied, and their presumed function in the body is typically assigned based on structural similarity with other transporters. Given the clinical and biological importance of these ABC transporters in the body, efforts to study these genes will prove valuable. Functional annotation of transporters based on co-expression was discussed briefly in Chapter 4, but more targeted functional evaluation is necessary to understand the role of

transporters in the body and the clinical significance of variation in their expression.

5.3 References

1 Schinkel AH, Jonker JW. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. *Adv Drug Deliv Rev* 2003; **55**: 3–29.

2 Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* 2001; **42**: 1007–17.

3 Glavinas H, Krajcsi P, Cserepes J, Sarkadi B. The role of ABC transporters in drug resistance, metabolism and toxicity. *Curr Drug Deliv* 2004; **1**: 27–42.

4 Bird A. Perceptions of epigenetics. *Nature* 2007; **447**: 396–8.

5 Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007; **128**: 635–8.

6 Bram EE, Stark M, Raz S, Assaraf YG. Chemotherapeutic drug-induced ABCG2 promoter demethylation as a novel mechanism of acquired multidrug resistance. *Neoplasia* 2009; **11**: 1359–70.

7 Baker EK, Johnstone RW, Zalcberg JR, El-Osta A. Epigenetic changes to the MDR1 locus in response to chemotherapeutic drugs. *Oncogene* 2005; **24**: 8061–75.

8 Baker EK, El-Osta A. MDR1, chemotherapy and chromatin remodeling. *Cancer Biol Ther* 2004; **3**: 819–24.

9 Lötsch J, Schneider G, Reker D, Parnham MJ, Schneider P, Geisslinger G, Doehring A. Common non-epigenetic drugs as epigenetic modulators. *Trends Mol Med* 2013; **19**: 742–53.

10 Begley DJ. ABC transporters and the blood-brain barrier. *Curr Pharm Des* 2004; **10**: 1295–312.

11 Chan LMS, Lowes S, Hirst BH. The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. *Eur J Pharm Sci* 2004; **21**: 25–51.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature 

Date 5/28/2014