Pharmacogenetics of Drug Metabolizing Enzymes Involved in Cardiovascular Drug Treatment

DISSENTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jonathan Christian Sanford

Graduate Program in Integrated Biomedical Science Program

The Ohio State University

2014

Dissertation Committee:

Wolfgang Sadee (Advisor)

Mitch Phelps

Amanda Toland

Daren Knoell
Abstract

Regulatory variants, less well characterized than coding polymorphisms, have the potential to serve as clinically relevant biomarkers for determining drug efficacy or toxicity. This study focuses on the identification of functional regulatory single nucleotide polymorphisms (SNPs) in the candidate genes *cytochrome P450 2C19* (*CYP2C19*), *carboxylesterase 1A1* (*CES1A1*) and *angiotensin-I converting enzyme* (*ACE*), which are involved in the metabolism of many cardiovascular drugs. Using allelic mRNA ratios (an indicator of functional regulatory variants), classic molecular genetics and next-generation sequencing technologies, I demonstrate that *cis*-acting regulatory polymorphisms in these genes are present and may have utility as pharmacogenetic biomarkers.

The drug metabolizing enzyme cytochrome P450 2C19 plays a major role in activation of clopidogrel (Plavix). Ultra-rapid metabolizer phenotype status has been associated with enhanced transcription caused by promoter allele *CYP2C19*/*17*. Here, I characterize the effect of *CYP2C19*/*17* on gene expression in human liver. *CYP2C19*/*17* heterozygotes and homozygotes show an increase in total *CYP2C19* mRNA expression of 1.8-fold (p=0.028) and 2.9-fold (p=0.006), respectively compared to homozygous reference allele livers. Allelic mRNA ratios of ~1.8 fold (SD±0.6, p<0.005), also significantly associate with *CYP2C19*/*17*. Potential novel regulatory variants were identified in an individual of African descent whose allelic mRNA ratio (~2-fold) is not
accounted for by *CYP2C19*17. This indicates that additional regulatory SNPs may affect *CYP2C19* expression in non-Caucasian populations.

Carboxylesterase 1A1 metabolizes clopidogrel, ACE inhibitors and many other compounds. Coding polymorphisms have been shown to affect carboxylesterase 1A1 function and decrease clopidogrel efficacy. The previously described *CES1A1VAR*, a highly linked block of 11 SNPs in the 5′ region of carboxylesterase 1A1, shows a strong association with allelic mRNA ratios (~1.35-fold). *CES1A1VAR* carriers also show a decrease in total gene expression of 2.65-fold (P = 0.003), and a 1.33-fold decrease in protein quantity (P= 0.008). Additional study of the carboxylesterase gene locus has identified a 5 SNP linkage block in the 5′ UTR (termed *CES1A1SVAR*). Utility as a biomarker for these variants will require further study.

Previously, our group had shown *angiotensin converting enzyme-1* (a target of ACE inhibitors) harbors enhancer SNPs prevalent in African populations with function in heart tissue. Performed here, a survey of other tissues indicates that enhancer SNP function is limited to the heart. Further, allelic mRNA analysis provides evidence of regulatory variants in Caucasian liver tissue (~2-fold), putamen (~1.7-fold, Hispanic and African descent) and adipose (RNA-sequencing, ~1.35-fold).

These targeted studies demonstrate the impact genetic variability can have on gene expression. *CES1A1VAR* in particular has potential as a pharmacogenetic biomarker, especially in predicting clopidogrel response. *CYP2C19* and *ACE* studies also indicate that these genes harbor functional variants active in certain tissues or present in certain populations. These enzymes affect the metabolism of numerous drugs
(cardiovascular or otherwise) making pharmacogenetic characterization crucial to effective prediction of drug efficacy or toxicity.
This work is dedicated to my mother, father, sister, brother-in law, extended family and friends whose constant love and support allowed me to weather the exciting journey that is graduate school. I dedicate this also to Stephanie, your love, confidence, and enthusiasm was essential to my success.
Acknowledgments

Considerable thanks are given to my advisor Wolfgang Sadee who gave me the wonderful opportunity to work with others in a cutting edge research area that can rapidly affect human health. The time and energy he spent mentoring me has developed my scientific acumen and have motivated me in areas outside the lab. I must also thank my fellow Sadee lab members and collaborators. All of you acted as excellent sources of information and discussion for my scientific inquiries. Lastly, I would thank my committee members Dr. Mitch Phelps, Dr. Amanda Toland and Dr. Daren Knoell for their participation and advisement in my scientific education.
Vita

May 14th, 1987 ............................................Born, Rochester NY

2005 .................................................................Brockport Central High School

2009 .................................................................B.S. Molecular Genetics, SUNY Fredonia

2009 to present ........................................Graduate Research Assistant, Department of Pharmacology, The Ohio State University

Publications


Fields of Study

Major Field: Integrated Biomedical Science Program
# Table of Contents

Abstract................................................................................................................................. ii

Vita................................................................................................................................. vii

Table of Contents ............................................................................................................... viii

List of Tables ..................................................................................................................... x

List of Figures ..................................................................................................................... xi

Chapter 1: Fundamentals of Pharmacogenetics ................................................................. 1

1.1 Introduction .................................................................................................................. 1

1.2 A Brief History of Pharmacogenetics ......................................................................... 2

1.3 Basics of Modern Pharmacogenetics ......................................................................... 11

1.4 Pharmacogenetics of Cardiovascular Drug Metabolizing Enzymes ....................... 18

Chapter 2: Regulatory Polymorphisms in CYP2C19 .................................................... 22

2.1 Introduction .................................................................................................................. 22

2.2 Materials and Methods ............................................................................................... 28

2.3 Results .......................................................................................................................... 32

2.4 Discussion .................................................................................................................... 44

Chapter 3: Regulation of gene expression at the CES1A1 gene locus ............................ 49
3.1 Introduction ........................................................................................................... 49
3.2 Materials and Methods ....................................................................................... 53
3.3 Results .................................................................................................................. 62
3.2 Discussion ............................................................................................................ 74

Chapter 4. Evidence of regulatory polymorphisms at the ACE gene locus ............. 81
4.1 Introduction .......................................................................................................... 81
4.2 Materials and Methods ....................................................................................... 85
4.3 Results .................................................................................................................. 87
4.4 Discussion ............................................................................................................ 92

Chapter 5. Conclusions and Final Thoughts ............................................................. 96

References .................................................................................................................. 103

Appendix A: Tables .................................................................................................... 123
Appendix B: Figures .................................................................................................... 135
Appendix C: Abbreviations ....................................................................................... 139
List of Tables

Table 1. Allele frequency and metabolizer status of common CYP2C19 alleles. ........ 25
Table 2. CES1A1 5’ Region SNP #s................................................................. 56
Table 3. Primer Sequences and PCR conditions.............................................. 123
Table 4. CYP2C19 SNP genotypes and Hardy-Weinberg Equilibrium............... 130
Table 5. CES1A1 SNPs genotypes and Hardy-Weinberg Equilibrium............... 131
Table 6. ACE. Marker SNP and Hardy-Weinberg Equilibrium.......................... 132
Table 7. CES1A1-pGL3-Basic construct variations............................................ 133
Table 8. Average ACE-fold AEI in RNA-sequencing........................................ 134
List of Figures

Figure 1. Common types of single nucleotide polymorphisms. .................................................. 13
Figure 2. Allelic Expression Imbalance .................................................................................... 17
Figure 3. Cardiovascular Drug/Gene Network ........................................................................ 19
Figure 4. Schematic of CYP2C19 gene locus and relevant SNPs. ............................................. 23
Figure 5. Association between CYP2C19*17 genotype and total CYP2C19 expression. ..33
Figure 6. Allelic RNA expression imbalance of CYP2C19 at marker SNP rs17885098 (T>C) in nine individual livers. .......................................................................................... 35
Figure 7. GATA4 gene locus and allelic ratios at rs3729856 .................................................... 38
Figure 8. Allelic expression imbalance in the 3’ UTR of GATA4. ............................................ 39
Figure 9. HNF4A gene locus and allelic expression imbalance in at rs6130615................. 41
Figure 10. Allelic expression imbalance in the gene AHR. .................................................... 42
Figure 11. Association between CYP2C19*17 and CYP2C19 enzyme activity by CYP2C19 diplotypes .................................................................................................................. 46
Figure 12. CES1A1 Gene Structure. ......................................................................................... 54
Figure 13. Polymorphisms of CES1A1 and known variants .................................................... 55
Figure 14. CES1A1 5’ RACE Indicates UTR lengths ............................................................... 65
Figure 15. CES1A1VAR associates with decreased CES1 gene expression ....................... 66
Figure 16. Allelic expression at rs12149370 ............................................................................ 69
Figure 17. Correlation between CES1 protein and RNA, CES1 protein quantity association with CES1A1VAR. ................................................................. 70
Figure 18. CES1A1VAR does not associate with metabolic ability. .......................... 72
Figure 19. Luciferase activity of CES1A1 constructs in HepG2 cells. ....................... 73
Figure 20. 3D Structures of CES1A1 5’ UTR Isoforms. ........................................ 78
Figure 21. Schematic of ACE gene locus and relevant SNPs. .................................... 83
Figure 22. Allelic mRNA ratios in Prefrontal Cortex.............................................. 88
Figure 23. Allelic mRNA ratios in Prefrontal Cortex............................................... 90
Figure 24. Allelic mRNA ratios in human liver. ...................................................... 91
Figure 25. Allelic mRNA ratios determined by RNA-Sequencing Data..................... 95
Figure 26. Linkage Disequilibrium of CYP2C19 alleles. ....................................... 135
Figure 27. Linkage of relevant polymorphisms in CES1A1. ................................. 136
Figure 28. CES1 protein expression in human liver tissue.................................... 137
Figure 29. Relative Gene Expression of ACE in Prefrontal Cortex and Putamen. ...... 138
Chapter 1: Fundamentals of Pharmacogenetics

1.1 Introduction

Pharmacogenetics—the use of genetic biomarkers to predict an individual’s response to drug treatment—is a discipline of genetics that has grown alongside our understanding of the human genome. The field developed at the intersection of four areas, inborn errors of metabolism, inheritance, molecular genetics, and the characterization of new pharmacological drug entities. Seminal studies in these areas build on one another to create the framework for modern pharmacogenetics. Previously, pharmacogenetics existed alongside treatment dogma (treating patients according to the drug). However, understanding the role of genetic variation in drug response has shifted this focus to the individual patient and their predicted response.

Cardiovascular drugs are a highly prescribed group of compounds with broad function. These drugs are activated or metabolized by key drug metabolizing enzymes (DMEs) including cytochrome p450 2C19 (CYP2C19), carboxylesterase-1 (CES1) and angiotensin-I converting enzyme (ACE), among others. As such, these genes are strong candidates for pharmacogenetic study. Here, I will explore the development of modern pharmacogenetics from a historical perspective and define the important components of a pharmacogenetic study. This will provide context and background relevant to the chief aim of this thesis, the identification and characterization of pharmacogenetic biomarkers in DMEs involved in cardiovascular drug treatments.
1.2 A Brief History of Pharmacogenetics

**Evolution and Heredity**

Understanding evolution and heredity is critical when defining a population’s ability to metabolize xenobiotics or drugs. Importantly, these mechanisms generate and maintain the diversity that grants populations the robust metabolic phenotypes pharmacogenetics seeks to exploit. Charles Darwin’s “On the Origin of the Species” was the first well researched and highly popularized publication on the theory of evolution. The public both derided it for its anti-theist message and praised it for its well documented findings. The sum of his worldly travels, Darwin presented evidence for significant biodiversity among plant and animal life (Darwin, 1859). He ultimately concluded that this biodiversity was the product of an organism’s adaptations to its surroundings. One well known example of this is Darwin’s finches, which were identified during his visit to the Galapagos Islands. Darwin and others have documented significant variation in beak shape, function, and even song structure (Lack, 1947; Podos et. al., 2004). In subsequent years, we have learned that these finches represent a specific type of evolution known as adaptive radiation (Schluter, 2000). Here, a single ancestral species rapidly speciates to fill many niches (in this case feeding opportunities). In the case of the Galapagos Islands, there were many available feeding niches that the finches adapted to as evidenced by their varied beak shape and function. This example, along with others, (see Darwin’s thoughts on domestic pigeons) made it clear that adaptation of ancestor organisms drives speciation ultimately producing biodiversity. Darwin’s proposed mechanism driving adaptation was gemmules, which function as hereditary
particles (Darwin, 1890). Though these are not what we know as genes, they are an early hypothesis for the transmission of traits, a concept important to pharmacogenetics.

Diverse environmental compound exposure allows a population to adapt heterogeneous drug metabolizing ability. This qualifies DME genes as strong candidates to contain genetic biomarkers that predict enzyme activity and therefore drug efficacy or toxicity.

Shortly following Darwin’s work, the friar Gregor Mendel defined the laws of segregation and dominance (Mendel et. al., 1901). Mendel developed these laws through crossing various “races” of common pea plants. Though there were many traits, Mendel focused on simple characteristics such as pea shape or color. Through careful planning and artificial cross pollination between plants, Mendel was able to observe how traits of the parent plants would manifest in the offspring. From this he developed the law of dominance, stating that in hybrids (offspring) the dominant trait would be visible while the recessive trait would not be. The second, the law of segregation, states that when first generation hybrids are crossed amongst themselves they will produce offspring which express all ancestral characteristics. These second generation hybrids will also adhere to the law of dominance, though some offspring will exhibit no visible dominant traits allowing recessive traits to be observed (Weldon, 1902). This work pioneered the concept of discrete, dominant and recessive traits. These advances would define heritability, which is essential to early pharmacogenetic studies.

Heritability of Inborn Errors of Metabolism

In the early 1900s, the physician and scientist Sir Archibald Garrod observed variable pigment in the urine of individuals with alkaptonuria (Garrod, 1996). William
Bateson – the man who would coin the term genetics – informed Garrod that his observations indicated that alkaptonuria followed basic Mendelian inheritance. Following extensive study of alkaptonuria and other variations in metabolism, Garrod published “Inborn Errors of Metabolism”. Here he stated, “It is to be regretted that so many observers of these anomalies have paid little or no attention to their heredity… but when scraps of information available are pieced together there is revealed an underlying resemblance between them, which is specially evident in their sex-incidence, in their tendency to occur in several members of one generation of a family, and in the rarity of their direct transmission from parent to child. If the lack of a special enzyme be in each instance the underlying factor, it is to be expected that they should behave as Mendelian recessive characters” (Garrod, 1923). From these experiences Garrod developed the hypothesis of “Chemical Individuality”, that is that individual chemical variation was present in all individuals, often with little consequence. Garrod punctuated this hypothesis throughout his career with exhaustive characterization of many inborn errors in metabolism and their inheritance patterns. Garrod’s work would act as the basis for early pharmacogenetic familial studies and solidify him as the father of biochemical genetics (Harper, 2005).

The first pharmacogenetic study was far from what we imagine today. It started when A.L Fox at Dupont spilled a sugar substitute he was developing (phenylthiocarbamide, PTC) sending it into the air. Fox was not bothered by the airborne powder, but a coworker remarked on the bitter flavor of the substitute. They soon found that they had significant differences in their taste perception of PTC (Fox,
Another researcher, L.H. Snyder was interested in Fox’s observation and collected 800 families to study the differences in PTC taste perception and its inheritance. From this study he found that “taste blindness” was inherited in an autosomal-recessive manner (similar to many inborn errors of metabolism as studied by Garrod) and that taste blind individuals were of diverse ethnic origin (Snyder, 1931a; Snyder, 1931b). Considered the first large study of a genetic biomarker this would become a model for future pharmacogenetic studies. Further, it was noted for its importance in describing the racial heterogeneity of metabolic processes.

The structure of DNA

By the early 1950s it was clear deoxyribose nucleic acid (DNA) was the heritable molecule through which genetic information was communicated within an organism (Avery et. al., 1944; Hershey et. al., 1952). However, the obstacle to understanding the finer points of genetic heritability was the unknown structure of DNA. By 1953, X-ray crystallography suggested that nucleic acid molecules were helical, though it was unclear in what way (Franklin et. al., 1953). It was hypothesized that nucleic acid structure must be stable, replicable and capable of containing a massive amount of information. Early in 1953, Linus Pauling – who defined the protein α-helix structure – and Robert B. Corey postulated that DNA was a tri-helical molecule with a phosphate tetrahedral at the center axis and the nucleotides positioned on the outside of the helix (Pauling et. al., 1953). However, as Watson and Crick would point out in their seminal paper, this conclusion was likely the result of misinterpretation of x-ray crystallography results. Instead, their proposed structure was a two chain helical structure, with complementary bases in the
center (Watson et al., 1953). In large part their findings would not have been possible without the research of others including but not limited to Erwin Chargaff, Rosalind Franklin and Maurice Wilkins who provided critical information to discerning the structure of DNA.

Understanding the structure of DNA is a crucial component of pharmacogenetics. In order to define heritable traits that predict drug response, the organization of genetic information needed to be understood. Prior to defining the structure of DNA, chromosome mapping and genes had been identified. However, it wasn’t until the genetic code had been cracked that pharmacogenetics would truly have the tools to identify genetic biomarkers. Classical molecular genetics techniques and ultimately DNA sequencing would drive new pharmacogenetic studies.

**Genetics as a driver of drug metabolizing enzyme variability**

During the years following, scientists and physicians continued to observe heritable metabolic variation as described by Garrod. Arno Motulsky collected many of these studies to examine “drug allergy” and the role of genetics in determining drug response. The studies collected observed, an increased incidence of hemolytic events in African Soldiers when compared to Caucasian soldiers receiving anti-malarial drugs (primaquine), prolonged apnea in select individuals following exposure to succinylcholine, and variable efficacy of isoniazid therapy (Motulsky, 1957). Though it would be some time before these differences in drug response would be associated with specific genetic biomarkers, it was clear that variation in DNA was not only relevant in inborn errors of metabolism but was also a driving factor in drug response variability.
Friedrich Vogel would recognize this trend and coin the term pharmacogenetics by 1959 (Vogel, 1959).

Familial studies have been an important component of human molecular genetics studies, when studying human disease or otherwise. One example of a familial pharmacogenetic study was the characterization of phenylbutazone metabolism in monozygotic and dizygotic twins (Vesell et. al., 1968). Here they observed that identical twins (monozygotic) had similar half-lives for phenylbutazone while drug half-lives in fraternal twins (dizygotic) varied to a much larger degree. This study provided more evidence that interindividual variation in drug metabolizing ability was driven by genetic factors.

**Early applications of DNA sequencing to pharmacogenetics**

A major contribution to the development of modern pharmacogenetics was the invention of DNA sequencing technologies (Sanger et. al., 1975; Sanger et. al., 1977). With evidence of heritable drug metabolizing ability, knowledge of DNA structure, and DNA sequencing, pharmacogenetics shifts focus to candidate gene studies. Candidate genes typically encode enzymes or proteins that are directly related to the metabolism or action of a compound. Here, non-synonymous polymorphisms that affect an enzyme’s protein sequence have been easily defined—especially when the mutation is deleterious. The function of these variants can be characterized, defining them as pharmacogenetic biomarkers which can then be used to guide drug treatments.

One example of developing technologies shaping pharmacogenetic research is in the stories of debrisoquine and sparteine metabolism. Debrisoquine, an anti-hypertensive
drug was the subject of study in 1975 at St. Mary’s Hospital Medical School. To explore adverse drug effects, Robert L. Smith and four other laboratory members received 32 mg of the drug. Within a few hours, Smith experienced symptoms of orthostatic hypertension that lasted for days following drug administration (Goldstein et al., 2003; Meyer, 2004). Notably, these symptoms were not experienced by the others who had taken the drug. Examination of the urine of these initial participants indicated that individuals with sensitive responses (such as Smith) did not metabolize debrisoquine to 4-hydroxydebrisoquine. A follow-up study in a group of medical students revealed that debrisoquine metabolism was bimorphically distributed to poor and extensive metabolizers (Mahgoub et al., 1977).

Simultaneously, two researchers were studying the pharmacokinetics of sparteine, an antiarrhythmic drug. When Michel Eichelbaum and Hans Dengler administered the drug to a study population two participants reported side effects including nausea, blurred vision and headaches (Eichelbaum et al., 1975). These symptoms were indicative of an overdose and in these individuals the plasma levels of sparteine were 3-4 times higher than normal (Meyer, 2004). Continued studies showed that sparteine metabolism was also divided into two distinct phenotypic groups as seen in debrisoquine (Eichelbaum et al., 1979).

The observations from both sets of studies confirmed that an individual’s metabolic ability for debrisoquine and sparteine was following Mendelian inheritance patterns. It was also believed that the same genetic mutation was responsible for the polymorphic drug responses. Soon after, the responsible cytochrome P450 enzyme was
isolated in human liver microsomes in the lab of F.P. Guengerich (Distlerath et. al., 1985). Then referred to as P-450DB, this enzyme metabolized debrisoquine, sparteine, and many other compounds. Following this work, the location of the gene was determined to be on chromosome 22 (Eichelbaum et. al., 1987). This cytochrome P450 would be named CYP2D6 and was cloned and sequenced in 1988 (Gonzalez et. al., 1988). Upon sequencing, three variant mRNA isoforms were identified and attributed to the common defective enzyme phenotypes observed for CYP2D6 metabolism. Additional characterization of the gene locus occurred in following years uncovering over 100 CYP2D6 alleles, including hybrid genes, copy number variations and single nucleotide polymorphisms (SNPs) (Marez et. al., 1997; Panserat et. al., 1995; Soyama et. al., 2004). The story of CYP2D6 demonstrates how many disciplines have intersected to identify pharmacogenetic biomarkers.

The Genomic Era and Pharmacogenetics

The completion of the Human Genome Project (HGP) was another critical advance in the field of pharmacogenetics. The HGP aimed to map the entire human genome and was completed in collaboration of 20 groups spanning the globe. Planned as early as 1984, the project formally began in 1990 and was declared complete in 2003 (Lander et. al., 2001). The first released draft sequence (2001) was predicted to cover 94% of the genome. Upon completion, a predicted 30-40,000 protein coding genes with complex gene splicing were identified. By 2001, an estimated 1.4 million single nucleotide polymorphisms (SNPs) had already been identified. The sequence was relatively high quality, and predicted to have 9 incorrect bases per 1 MB with a
significant error every 6 MB (Schmutz et. al., 2004). In the spirit of competition the Human Genome Project was not the only genome released in 2001. A second genome was published by Celera Cooperation. Using shotgun sequencing and public data already obtained by the HGP a privately funded genome of 2.91 billion bases was assembled in a period of 3 years (Venter et. al., 2001).

With the human genome mapped by two groups, it became clear the extreme variation of an individual’s genome. An estimated 3-million SNPs exist in any one genome with an average of 1 SNP per 1000 bases (Cargill et. al., 1999). Of these, an estimated 1% of SNPs would affect protein function (Venter, 2001). As SNPs occurring in >1% of the population have been identified, microarray technologies (using targeted oligonucleotides) have been used to genotype thousands (and sometimes millions) of SNPs. A major step forward following sequencing of the human genome was genome wide association studies (GWAS). Here, a large number of SNPs are genotyped in a clinical population and then statistically associated to a phenotypic endpoint.

**Genome Wide Association Studies**

One classic genome wide study published in 2008, focused on the drug simvastatin and its relationship to statin-induced myopathy (Link et. al., 2008). Here, using a HumanHap300 chip (318,237 SNPs), over 300k SNPs were genotyped and associated to incidence of statin induced myopathy. Genome wide association indicated SNP rs4363657 in the gene *SLCO1B1* (encoding the transporter OATP1B1), was associated with myopathy risk ($P = 4 \times 10^{-9}$). The identified SNP was in high linkage with rs4149056 ($R^2 = 0.97$), a nonsynonymous coding SNP which has been associated with
statin metabolism (Kameyama et. al., 2005; Pasanen et. al., 2007; Pasanen et. al., 2006). Presence of the rs4149056 minor allele increases an individual’s odds ratio of developing statin induced myopathy to 4.5 and accounts for ~60% of all observed cases. This study which utilized many post genome techniques is a culmination of advances that have furthered pharmacogenetics. One of the best examples of genome wide association, this study accurately identified a biomarker associated with an adverse drug reaction.

The historical advances outlined here are not an exhaustive overview of all the projects that have contributed to the development of pharmacogenetic research. The identification of the DNA molecule (prior to its structure), genetic codons for protein synthesis, mapping of the first disease gene (Huntington’s), invention of polymerase chain reaction (PCR), and other advances in DNA sequencing have been critical for continued advancement of the pharmacogenetics.

1.3 Basics of Modern Pharmacogenetics

Drug metabolizing ability is the sum of the environment, age, gender, diet and genetics of an individual. With the development of genetic and genomic technologies, pharmacogenetic studies have now been applied to countless genes. As a result, the Food and Drug Administration has drug label indications for over 150 genetic variants known to affect drug metabolism or response. Many of these variants are coding polymorphisms with dramatic effect on an enzymes function. Even common drug treatments such as Warfarin have seen pharmacogenetic biomarkers in the genes VKORC1 and CYP2C9.

**Types of single nucleotide polymorphisms**
The Human Genome Project revealed the extent of individual variability in our genomes. With the entire genome sequenced and polymorphisms mapped we are able to better interrogate potential functional polymorphisms. Single nucleotide polymorphisms come in three major varieties (Figure 1). Coding SNPs (cSNPs) are located in mature mRNA and have the potential to change the protein sequence in addition to the mRNA sequence. Regulatory SNPs (rSNPs) are found in the regulatory regions of a gene, typically a promoter or enhancer. These variants often change gene transcription through creation or ablation of transcription factor or enhancer binding sites. Here, the effects of rSNPs can easily be measured through changes in gene expression. Finally, there are structural SNPs (srSNPs). These variants are found sometimes in 5′ and 3′ UTRs, exons (near splice points), or introns. They are typically present in the pre-processed heteronuclear RNA, and will change mRNA processing, namely splicing, mRNA folding and more.

*Cytochrome P450 2C19 (CYP2C19)* is a gene which contains all three types of SNPs. Here, the common loss of function polymorphism, *CYP2C19*<sup>*2*</sup> (srSNP), has been shown to create an aberrant splice site that creates a stop codon downstream resulting in a non-functional protein (de Morais et. al., 1994b). The cSNP *CYP2C19*<sup>*10*</sup>, has been shown to decrease mephenytoin metabolism (Blaisdell et. al., 2002). Finally, promoter rSNP, *CYP2C19*<sup>*17*</sup>, which is also a subject of study in Chapter 2, has been shown to increase gene transcription (Sim et. al., 2006). These SNPs represent the wide array of SNP functions that may be present in a gene.
Figure 1. Common types of single nucleotide polymorphisms.

Indicates variant name, location in RNA and functional effects typically observed. Some polymorphisms may fall into multiple categories.
Identification of regulatory variants

Many studies have searched for regulatory polymorphisms through expression Quantitative Trait Loci (eQTL) studies. Here, gene expression from a microarray is combined with SNP genotype information to identify polymorphisms associating with increases or decreases in gene expression. In human liver, Schadt et al used gene expression data and a SNP panel of over 300,000 variants to identify 6000 potential eQTLs (Schadt et. al., 2008). The corresponding SNPs were characterized as cis- or trans- depending on their distance to the gene locus itself (trans> 1MB). Similar to the study on SLCO1B1, both eQTL and GWAS studies can indicate genes experiencing cis-regulatory action. The associated SNP may or may not be the functional polymorphism (they may be in linkage). However, regulatory elements can exist at a large distance from a gene locus (enhancer region). These studies allow regulatory effects that are likely present and frequent in the genome to be identified on a genome wide scale.

To identify SNPs with regulatory function, the focus of these studies, the Sadee lab has traditionally measured allelic mRNA expression in combination with molecular genetics studies (Wang et. al., 2005;Zhang et. al., 2007;Zhang et. al., 2005). This method (termed SNaPshot), unlike an eQTL study, interrogates gene expression at an allelic level allowing the action of a cis-acting polymorphism to be observed on an individual allele basis (Figure 2). Here, coding SNPs with high frequency are selected as marker SNPs for analysis. Polymerase chain reactions (PCR) are developed (of similar size) for both gDNA and mRNA surrounding the selected marker SNP, and then amplified. An additional amplification (SNaPshot) combines a primer extension primer which (5’>3’)
stops adjacent to the marker SNP, and a mix containing fluorescent dideoxy nucleotide triphosphates (ddNTPs). A single fluorescent base is added for either allele at the marker SNP and transcription is aborted. In individuals heterozygote for the marker SNP this results in two populations of PCR product, one for each allele. The gDNA acts as a control (often with a ratio of 1:1 between alleles), and deviation from the gDNA ratio in the mRNA is indication that one of the alleles is experiencing the effect of a cis-acting regulatory polymorphism. Notably, the effect of trans-acting polymorphisms—which by definition would act upstream of gene expression—cannot be seen because they do not discriminate on an allelic basis. Still, deviation in the mRNA ratio indicates there is a functional SNP present and influencing gene expression. These polymorphisms may be affecting splicing, transcription, mRNA folding and more. Often, candidate genes can be selected that have seen study previously because regulatory SNP effects may not have been assessed, particularly at an allelic level. Evidence of cis-acting regulatory polymorphisms is followed by molecular genetic studies, and SNP screening to identify and characterize the causative variant.

A natural continuation of the SNaPshot assay is seen in DNA/RNA sequencing. When sequencing a samples RNA, the obtained “transcriptome” contains the allelic ratio of heterozygote SNPs at all coding SNPs in all expressed genes. This allows rapid identification of candidate genes experiencing the effects of regulatory SNPs on total gene expression, allelic mRNA ratios, splicing and more. From here, detailed molecular genetic and allelic studies can validate RNA-sequencing findings. Transcriptome sequencing will drive future discoveries in the field of pharmacogenetics.
SNPs mainly exist in haplotypes, which is a group of SNPs that are inherited together. SNPs existing within the same haplotype block are considered in terms of their linkage disequilibrium with one-another. Here the terms D’ and R^2 are used to measure the likelihood that two SNPs will appear in the same individual, and on the same allele, respectively. Haplotype blocks can exist over large regions with SNPs in varying degrees of linkage. Using these measures; highly, moderately, and poorly linked SNPs are identified. This is information is critical in pharmacogenetic studies. When a SNP is associated to changes in gene expression, allelic expression imbalance, or a complex phenotype (like statin induced myopathy), it may not be the causative SNP but rather highly linked to the functional variant.

 Currently there is some disagreement on what type of polymorphisms should be focused on when studying disease and pharmacogenetic phenotypes. For human disease, rare variants can be highly relevant, and they certainly can affect drug response. Conversely, common variants typically confer a smaller effect and are not often disease causing. However, pharmacogenetic variants often are only relevant given exposure to a certain drug. Further, small or large effects of these variants may only be significant given certain temporal and environmental conditions. This is unlike disease causing variants where severe phenotypes are easily identified. Because coding polymorphisms have been well-studied, it is likely that unexplained variability in drug metabolism is driven by common regulatory variants. These variants will likely have a wide array of functions (with small to large effect sizes) and potential utility as a biomarkers for predicting drug metabolizing enzyme expression or activity.
Paternal and Maternal genomic DNA is inherited in a 1:1 ratio. Allelic mRNA ratios under action of a cis-acting regulatory polymorphism will deviate from this ratio. Allelic ratios are visualized using the fluorescent output obtained from SNaPshot product run using ABI 3730
This dissertation focuses on 3 candidate genes (\textit{CYP2C19}, \textit{CES1}, \textit{ACE}) related to one another through the cardiovascular drugs they metabolize (Figure 3). Here, I make use of DNA-sequencing technologies, SNaPshot and other molecular genetics techniques in search of \textit{cis}-acting functional polymorphisms. Due to the frequent administration of drug metabolized by my enzymes of interest, pharmacogenetic biomarkers in one or more genes may be relevant in determining drug response.

1.4 Pharmacogenetics of Cardiovascular Drug Metabolizing Enzymes

The focus of this dissertation is the pharmacogenetics of cardiovascular drug metabolism. Three enzymes which are directly related to the metabolism and action of multiple cardiovascular drugs are interrogated for presence of \textit{cis}-acting regulatory biomarkers (Figure 3). The first, Cytochrome P450 2C19, is a drug metabolizing enzyme responsible for the metabolism of a large number of clinically prescribed drugs. Its primary site of action is the liver, and it is involved in the metabolism of proton pump inhibitors, anti-epileptic drugs, and the anti-platelet prodrug clopidogrel. Clopidogrel, a prodrug prescribed to over 25 million people, requires activation by CYP2C19 for drug action. CYP2C19 competes with carboxylesterase 1A1, an enzyme which is the primary inactivator of clopidogrel. Ultimately, 15\% of the administered prodrug is activated, with the remaining 85\% inactivated. The 15\% is in part determined by the activity of CYP2C19. Previously described polymorphisms affecting the coding region and gene splicing have been associated with poor drug activation and response. Regulator
Cardiovascular drug treatments and their relationship to the candidate enzymes (genes) of interest. These genes are good candidates for pharmacogenetic study with relevance in multiple treatment environments. Green arrows: Drug activating, red arrows: Drug inactivating.
polymorphisms affecting CYP2C19 gene expression have already been identified. However, the most common, CYP2C19*17 has been poorly characterized in human liver tissue, and additional polymorphisms in non-Caucasian populations may be present.

The second gene included in this study is carboxylesterase 1a1. Though this enzyme is involved in the inactivation of the prodrug clopidogrel, it is also involved in the metabolism of cardiovascular drugs such as ACE inhibitors (enalapril, imidapril, trandolapril, ramipril) and the anti-clotting drug dabigatran(Geshi et. al., 2005;Pare et. al., 2013;Thomsen et. al., 2014). Non-cardiovascular compounds metabolized include chemotherapeutics (irinotecan, capecitabine), anti-virals (oseltamivir), heroin, and cocaine (Guichard et. al.;Pindel et. al., 1997;Sai et. al., 2010;Shi et. al., 2006). In the context of clopidogrel, variability in the speed of inactivation can increase compound bioavailability for activation by CYP2C19. Previous studies have shown that moderately frequent polymorphisms which ablate enzyme function can significantly affect clopidogrel efficacy (Lewis et. al., 2013). It is not currently clear to what extent cis-regulatory polymorphisms affect CES1A1 gene expression and protein activity.

CES1A1, which is involved in the activation of some ACE inhibitors, implicates angiotensin converting enzyme-I as a potential gene to harbor pharmacogenetic variants. ACE is a ubiquitously expressed gene with a broad range of functions, particularly in the heart. A well studied insertion/ deletion (I/D) polymorphism has been associated to many disease phenotypes, though it’s effect as a regulatory polymorphism is unclear(Al-Rubeaan et. al., 2013;Das et. al., 2013;Dhangadamajhi et. al., 2010). Previous work in our lab, led by Andrew Johnson, identified three enhancer polymorphisms specific to
African ancestry populations that carry a significant risk for non-fatal myocardial infarction in hypertensive patients (Johnson et al., 2009). These SNPs and their function in alternative tissues is still unclear. There has also been limited screening for potential regulatory SNPs in non-cardiovascular tissues and non-African populations. Addressed here, functional SNPs in these tissues may relevant in the treatment or development of Alzheimer’s, diabetes or other disease phenotypes.

Here, I further the field of pharmacogenetics through study of these inter-related genes. I define \textit{CYP2C19*17}, a promoter polymorphism, as having a 2-fold increase on gene expression in the human liver. I also present functional rationale for a \textit{CES1A1} genetic variant which decreases gene expression and protein quantity. Finally, I indicate that \textit{ACE} likely harbors multiple functional polymorphisms with effects in multiple tissues while the enhancer SNPs show heart specific function. This work adds to our understanding of enzymes closely involved with cardiovascular drug treatments and further defines variants that have potential as pharmacogenetic biomarkers. Moreover, diverse functional genetic features of these genes indicate that consideration of SNP-SNP interactions should be a feature of future studies where relevant drugs are prescribed.
Chapter 2: Regulatory Polymorphisms in \textit{CYP2C19}

2.1 Introduction

CYP2C19 is a member of the cytochrome P450 enzyme family. Hepatic CYP enzymes are main contributors to the metabolism of many endogenous and exogenous compounds (Guengerich, 2008). The CYP2C19 isoform metabolizes 8\%–10\% of prescribed drugs, including clopidogrel, omeprazole and citalopram (Karam et. al., 1996; Kazui et. al., 2010; Ohlsson Rosenborg et. al., 2008). \textit{CYP2C19} consists of nine coding exons spanning 90,209 bases with a coding region of 1473 bases (Figure 4). Similar to other CYP genes, \textit{CYP2C19} carries genetic variants with strong effects on substrate metabolism. Two frequent alleles lacking metabolic ability in vivo and in vitro are \textit{CYP2C19}*2 (rs4244285, cryptic splice site) and \textit{CYP2C19}*3 (rs4986893, premature stop codon), causing intermediate (heterozygous) or poor metabolizer phenotypes (homozygous) (de Morais et. al., 1994a; de Morais, 1994b). Minor allele frequencies vary by race, with \textit{CYP2C19}*2 frequent in Caucasian and African populations, whereas *3 is common in Asians and Africans (Table 1A).

The anti-platelet prodrug clopidogrel (Plavix) requires enzymatic activation mediated mainly through CYP2C19 (Kazui, 2010). Clinical evidence indicates that *2 and *3 carriers fail to benefit fully from clopidogrel therapy, thereby increasing the risk for reoccurring cardiac events such as myocardial infarction, stroke and stent
Figure 4. Schematic of CYP2C19 gene locus and relevant SNPs.

Gray boxes indicate coding exons.
thrombosis (Hwang et. al., 2011; Mega et. al., 2009; Mega et. al., 2010; Shuldiner et. al., 2009). With an FDA-issued black box warning regarding clopidogrel treatment in CYP2C19*2 or*3 carriers, these have emerged as clinical biomarkers to predict the metabolizer status of CYP2C19 (Table 1B).

Not yet commonly represented in CYP2C19 allele biomarker panels, is the gain of function allele CYP2C19*17 which has been reported to increase transcription, resulting in ultra-rapid metabolism of CYP2C19 substrates (Sim, 2006). Clinical association studies have shown CYP2C19*17 increases clopidogrel efficacy in reducing reoccurring cardiac events, risk of stent thrombosis and residual platelet aggregation, while enhancing the risk of a major bleeding event (Pare et. al., 2010; Sibbing et. al., 2010a; Sibbing et. al., 2010b; Tiroch et. al., 2010). The Clinical Pharmacogenetics Implementation Consortium, which intends to define how genetic biomarkers should be used clinically, has classified CYP2C19*17 carriers as ultra-rapid metabolizers in the context of clopidogrel treatment (Table 1B) (Scott et. al., 2013). However, this classification may not stand for other CYP2C19 substrates. The effect of CYP2C19*17 has been measured for several drugs, including omeprazole, amitriptyline, voriconazole, mephenytoin, pantoprazole and escitalopram (Baldwin et. al., 2008; de Vos et. al., 2011; Dolton et. al., 2011; Gawronska-Szklarz et. al., 2012; Goldstein et. al., 1994; Ohlsson Rosenborg, 2008; Sim, 2006; Wang et. al., 2009). In many cases, CYP2C19*17 does increase enzyme activity, but when comparing the effect on multiple drugs, CYP2C19*17 is not always a clinically relevant contributor to drug metabolism. Alternate routes of clearance between drugs may be responsible for some of these discrepancies, but this does draw into question the use of
Table 1. Allele frequency and metabolizer status of common CYP2C19 alleles.

(A) Minor allele frequencies of common CYP2C19 biomarkers and (B) associated metabolic phenotypes defined by Clinical Pharmacogenetics Implementation Consortium guidelines for individuals receiving clopidogrel (Scott, 2013)

A.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Caucasian</td>
</tr>
<tr>
<td>CYP2C19*2</td>
<td>0.133</td>
</tr>
<tr>
<td>CYP2C19*3</td>
<td>0.00</td>
</tr>
<tr>
<td>CYP2C19*17</td>
<td>0.208</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th>SNP</th>
<th>CYP2C19*1</th>
<th>CYP2C19*2</th>
<th>CYP2C19*3</th>
<th>CYP2C19*17</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19*1</td>
<td>*1/*1(wt)</td>
<td>*1/*2</td>
<td>*1/*3</td>
<td>*1/*17</td>
</tr>
<tr>
<td></td>
<td>Extensive</td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Ultra-rapid</td>
</tr>
<tr>
<td></td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
</tr>
<tr>
<td>CYP2C19*2</td>
<td>*2/*2</td>
<td>*2/*3</td>
<td>*2/*3</td>
<td>*2/*17</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
</tr>
<tr>
<td>CYP2C19*3</td>
<td>*3/*3</td>
<td>*3/*3</td>
<td>*3/*3</td>
<td>*3/*17</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
<td>metabolizer</td>
</tr>
<tr>
<td>CYP2C19*17</td>
<td></td>
<td></td>
<td></td>
<td>*17/*17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ultra-rapid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>metabolizer</td>
</tr>
</tbody>
</table>
the ultra-rapid metabolizer classification outside of clopidogrel treatment. Recent meta-analyses have questioned the role of CYP2C19*17 (and to some extent *2) as a biomarker (Bauer et. al., 2011; Jang et. al., 2012; Li et. al., 2012), and others propose that CYP2C19*17 homozygotes should be considered extensive metabolizers, rather than ultra-rapid (Li-Wan-Po et. al., 2010).

Whereas the impact of *2 and *3 alleles on CYP2C19 function is well documented, the regulation of gene expression by *17 and possibly other polymorphisms is less well established. In human liver, CYP2C19*17 carrier status has not been explored in the context of gene expression, which fails to support the ultra-rapid metabolizer classification of CYP2C19*17 carriers. However, evidence supporting a gain-of-function for *17 includes electrophoretic mobility shift assays showing enhanced binding of the CYP2C19*17 allele (-806T) to nuclear receptor cellular fractions and increased luciferase reporter gene activity in mice (Sim, 2006). Another study in transfected cells failed to show that CYP2C19*17 is active, instead implicating variants such as rs4986894, part of a haplotype including two other promoter single-nucleotide polymorphisms (SNPs), with increased expression. In the case of rs4986894, high linkage disequilibrium (LD) with *2 (D'=1.00, R²=1.00, 1000 Genomes Project), denoted as CYP2C19*2C or *21, renders this variant of little clinical consequence (Fukushima-Uesaka et. al., 2005; Satyanarayana Chakradhara Rao et. al., 2011).

Absent from these analyses is an accurate measure of the effect size of CYP2C19*17 and its interindividual variability on CYP2C19 expression in human liver. This is crucial information if *17 were to be used as a predictor of ultra-rapid metabolizer
status for a wide range of CYP2C19 substrates. A further confounding factor is the estimated low linkage disequilibrium of *17 with *2 (D'=1.0, r2 =0.04, 1000 Genomes Project) (Pedersen et. al., 2010). The high D', low R^2 and different allele frequencies indicate that *17 resides mainly on the opposite allele compared to *2 (namely, on the wild-type *1 allele). Therefore, the enhanced transcriptional activity of *17 is not annulled by residing on the inactive *2 allele in compound heterozygotes, although the latter case cannot be excluded in some subjects. An additional source of variation in gene expression may be driven by variable gene expression for transcription factors that initiate expression of CYP2C19 and other cytochrome p450 enzymes. Functional polymorphisms within transcription factors may be particularly relevant in the case of CYP2C19*17 where the functional polymorphism lies in the promoter region and is predicted to create a binding site for the transcription factor GATA4 (Sim, 2006).

In this study, I determine whether CYP2C19*17 enhances CYP2C19 expression in target tissues (liver) and to what extent. Further, I discern whether additional promoter polymorphisms are contributing to variability in CYP2C19 expression and I reexamine the relationships between *17 and *2 by measuring CYP2C19 total mRNA expression and enzyme activity in human livers. In addition, I apply allelic mRNA expression analysis, a more accurate and precise measure of the effect of cis-acting regulatory variants on mRNA levels. When determined in livers heterozygous for a marker SNP in the mRNA, a significant deviation of allelic mRNA ratios from the gDNA ratios indicates the presence of allelic expression imbalance (AEI), a measure of cis-acting regulatory polymorphisms (Lim et. al., 2007). Whereas the scarcity of frequent exonic CYP2C19
marker SNPs limits the scope of AEI analysis, I show that \textit{CYP2C19*17} accounts for 2-fold enhanced allelic and 1.6-fold increased mRNA expression, with some degree of interindividual variability. Allelic expression analysis in transcription factor GATA4 and other select transcription factors showed no indication of strong functional polymorphisms that may be acting upstream of CYP2C19 expression. Further, AEI not accounted for by \textit{CYP2C19*17} indicates the presence of a candidate regulatory variant(s) in a sample of African descent. These results indicate that \textit{CYP2C19*17} may compensate for the presence of loss of function alleles such as *2 in compound heterozygous samples, but this requires further confirmation in a larger number of subjects.

2.2 Materials and Methods

\textbf{Tissue samples}

A total of 125 biopsy or autopsy human liver samples were obtained from the Cooperative Human Tissue Network (Midwest and Western Division) under approval of the Ohio State Institutional Review Board. Fifty liver samples were obtained from other sources that included the Medical College of Wisconsin (Milwaukee, WI, USA), Medical College of Virginia (Richmond, VA, USA), Indiana University School of Medicine (Indianapolis, IN, USA) or University of Pittsburgh (Pittsburgh, PA, USA). These livers were under protocols approved by the appropriate committees for the conduct of human research. Samples were primarily from Caucasians, with ~15% of African descent. Liver microsomes were prepared by differential centrifugation using standard procedures (van
der Hoeven et. al., 1974) and characterized for protein content by the Lowry method (Lowry et. al., 1951).

**DNA and RNA isolation and genotyping**

Genomic DNA and RNA were prepared from liver tissue samples as previously described (Wang et. al., 2011). DNase I was used for RNA isolation to prevent contamination of genomic DNA in cDNA synthesis. cDNA was prepared using RNA and Reverse Transcriptase SSIII (Invitrogen, San Francisco, CA, USA), with controls lacking reverse transcriptase to test for residual gDNA. SNPs in CYP2C19 were genotyped in liver samples using a primer extension assay (SNaPshot, Life Technologies) or fluorescently labeled PCR-restriction fragment length polymorphism (RFLP) analysis as previously described (Pinsonneault et. al., 2004; Wang, 2005). PCR conditions and primers for all PCR assays (including genotyping) are given in Appendix A.1. For quality control, the Hardy-Weinberg equilibrium was assessed for each SNP, and allele frequencies were compared to existing population genotype data (Appendix A, Table 4).

**Measurements of gDNA and mRNA allelic ratios using a primer extension assay (SNaPshot)**

The protocol for SNaPshot has been previously described (Pinsonneault, 2004; Wang, 2011). It uses PCR amplification surrounding a marker SNP in an exonic region, from a heterozygous sample. This is followed by the addition of fluorescent ddNTPs at the marker SNP. Next, the SNaPshot product is read out as peaks on an ABI3730 sequencer (Life Technologies) to provide the relative amounts of each allele. The average of allelic ratios of gDNA is used to normalize allelic mRNA ratios in each
tissue. A finding of significant AEI in each tissue was established with a cutoff of three standard deviations from the mean of the genomic DNA allele ratios calculated from all tissues assayed. For the marker SNP in CYP2C19, normalized allelic mRNA ratios, measured at rs17885098 (T > C) in exon 1, > 1.26-fold in either direction (> 1 or < 1) were considered an indication of AEI. SNaPshot measurements for CYP2C19 were completed in duplicate for gDNA and at minimum triplicate for mRNA. To achieve sufficient accuracy, CYP2C19 mRNA expression needed to be robust (Ct of < 26 measured via qRT-PCR) for a liver to be considered for AEI analysis. Only nine livers were heterozygous for SNP rs17885098 and passed quality control for AEI analysis. PCR conditions and primers are available in Appendix A, Table 3. Additional allelic ratios were surveyed in selected transcription factors. Here, ratios were measured by sample with duplicates in mRNA and a single measurement in gDNA. Analytically significant deviation from gDNA was considered 3-standard deviations.

**SNP scanning and sequencing of CYP2C19**

The CYP2C19 promoter region (4kb) and exons were sequenced to scan for regulatory variants (for PCR conditions and primer sequences, see Appendix A.1, (Blaisdell, 2002)). PCR products were quantitated by QUBIT Broad Range DNA spectroscopy (Invitrogen). Fragments were mixed in equi-molar ratios to generate libraries that were analyzed by an IonTorrent PGM (Life Technologies) and barcoded for multiple samples per sequencing run. A total of four runs were performed on 15 samples. Sequences were analyzed in CLC Bio’s Genomics Workbench (CLC Bio, Katrinebjerg, Denmark). SNP calling required, at minimum, 20 reads at a given base with 30% reads
per allele. Select regions and samples were also sequenced using Sanger sequencing at OSU’s sequencing facility.

**Analysis of mRNA levels with quantitative real-time PCR**

mRNA levels of *CYP2C19* and transcription factors *PXR*, *RXRa*, *HNF4a*, *CAR* and *GATA4* were measured in human liver samples using quantitative real-time PCR (qRT-PCR) (in duplicate) on a 7500 Fast Real Time PCR System (Applied Biosystems, Carlsbad, CA, USA). *GAPDH* served as a housekeeping gene yielding expression representative of overall RNA quality in human liver with low interindividual variability (for primers, see Appendix A, Table 3)(Barber et. al., 2005). Standard curves were made using serial dilutions of cDNA for each gene and used for mRNA quantitation.

**Measurement of CYP2C19 activity in human liver microsomes**

CYP2C19 enzyme activity was measured in 40 liver microsomes using S-mephenytoin as the substrate as described previously (Baker JAR, 2010). The formation of 4’-hydroxy mephenytoin (nmol/min/mg) at a single substrate concentration (≤Km), represented by metabolic velocity (V), was used to measure the metabolic activity of CYP2C19. The data represent duplicate assays, transformed into log scale for analysis by linear regression and independent t-test.

**Data analysis**

Linkage disequilibrium (LD) and Hardy-Weinberg equilibrium were analyzed using Haploview, and Helix Tree SNP and Variation Suite (Golden Helix, Bozeman, MT, USA), respectively (Appendix A, Table 4; Appendix B, Figure 26). Significant differences in allelic mRNA ratios from mean allelic gDNA ratios (presence of AEI)
were confirmed by t-test. Linear regression was used to test the effects of SNPs and
transcription factor expression on CYP2C19 mRNA expression and protein activity using
SPSS (IBM, Armonk, NY) and Minitab (Minitab, State College, PA), respectively. P-
Values <0.05 were considered significant for all associations.

2.3 Results

**CYP2C19*17 is associated with increased CYP2C19 total mRNA expression**

Total mRNA expression for CYP2C19 was measured by qRT-PCR in 83 liver samples.
On average, *CYP2C19*17 heterozygotes showed a 1.8-fold increase in mRNA
expression (p=0.028, 95% CI 1.08-fold to 3.00-fold, t-test) over the tissue of *17 non-
carriers. *CYP2C19*17 homozygotes showed 2.9-fold increased expression (p=0.006,
95% CI 1.5-fold to 5.8-fold, t-test, equal variances not assumed) with large variability
(Figure 5). Known regulators of CYP2C19 expression (transcription factors) including
the pregnane X receptor (PXR), constitutive androgen receptor (CAR) and GATA4
(Chen et. al., 2003;Mwinyi et. al., 2010) were measured by qRT-PCR to determine
whether they significantly contribute to *CYP2C19* expression. Hepatic nuclear factor-α
(HNF4α) and retinoid x receptor-α (RXRα), additional factors that do not directly interact
with the *CYP2C19* promoter but may affect expression of other *CYP2C19* regulators
were also included. CYP2C19 expression was positively correlated with the expression of
*GATA4* and negatively correlated with *CAR* and *RXRa* mRNA (p<0.05). After adjusting
for the expression of *GATA4*, *CAR* and *RXRa*, *CYP2C19*17 carriers remain significantly
associated with increased expression of CYP2C19 mRNA (1.6-fold, p=0.014).
Figure 5. Association between \textit{CYP2C19*17} genotype and total CYP2C19 expression. \textit{CYP2C19*17} heterozygotes and homozygotes were associated with 1.8-fold (p=0.028) and 2.9-fold (p=0.006) increases in CYP2C19 mRNA, respectively. Data are presented as a box plot and whisker plot (box shows median with the 25th and 75th percentiles), and minimum and maximum values are shown by whiskers. Outliers that do not fall within the bounds of the plot and that are greater than three box-heights are indicated with o and *, respectively.
**CYP2C19*17 is associated with increased allelic expression of CYP2C19 in human liver tissues**

*CYP2C19* allelic RNA expression was measured to determine whether and to what extent *CYP2C19*17 increases gene transcription and whether any additional functional polymorphisms are present. Allelic expression was measured using marker SNP rs17885098, located in the first exon of *CYP2C19* (minor allele frequency=0.083, 1000 Genomes Project). Lack of frequent marker SNPs in CYP2C19 exons limited the number of livers accessible for allelic RNA expression analysis. Of 175 livers, nine identified heterozygotes passed the quality controls outlined in Materials and Methods. Of these nine samples, seven were Caucasian and two were of African ancestry (LL34, L123). Five samples showed significant AEI using criteria described in the Material and Methods (Figure 6, AEI >1.26-fold).

To test the association of AEI with *CYP2C19*17, I genotyped *CYP2C19*17 and other known functional SNPs *CYP2C19*2, *CYP2C19*3 and promoter SNP rs4986894 (Satyanarayana Chakradhara Rao, 2011). Functional SNPs must be heterozygous in all samples showing AEI and absent or homozygous in all samples not showing AEI. Four samples lacking AEI were also lacking *CYP2C19*17. Of the five with AEI, all were *17 heterozygotes except one (L123), which is one of two samples of African descent. Significant allelic mRNA ratios (AEI-positive) within the *CYP2C19*17 carriers ranged from 1.4- to 2.8-fold, with an average of 1.8-fold (SD±0.68-fold) (Figure 3). These results support the hypothesis that *CYP2C19*17 is functional, increasing RNA expression ~2-fold and also indicates additional regulatory polymorphisms may be
Figure 6. Allelic RNA expression imbalance of *CYP2C19* at marker SNP rs17885098 (T>C) in nine individual livers.

The allelic RNA ratio is expressed as a ratio of C/T averaged from a minimum of three measurements. Three standard deviations of the genomic DNA control (>1.26-fold AEI) indicate a biologically significant allelic expression imbalance (indicated by arrows), showing significant difference from the gDNA control at p<0.005. L123 is the only sample showing AEI and not carrying *I17*. Data are mean±SD.
present in African Americans.

Sample L123 did not contain the functional SNPs for *2 and *3 haplotypes. To identify any additional polymorphisms causing AEI in L123, *CYP2C19* coding exons and 4 kb upstream of the first exon (promoter) were sequenced using the Ion Torrent PGM. Four additional livers were also sequenced for comparison: L114 (*1/*2, no AEI), LL16 and LL19 (*1/*17, with AEI), and L90, an additional African liver (no AEI data). *CYP2C19*17 was the only allele uniquely associated with AEI in LL16 and LL19, which is consistent with *CYP2C19*17 causing increased gene expression. Comparing sequencing results of L123 with genotypes of LL34 (an AEI− sample of African descent) as control, I found a group of multiple SNPs as being potentially causative (rs11568730, rs7919698, rs710258 and rs56043006); however, additional studies will be needed to establish the function of these regulatory variants. These observations indicate that there are additional polymorphisms that may affect the transcription of *CYP2C19* in African American populations.

**Allelic expression in transcription factors interacting with CYP2C19 and other CYP 450s**

To determine the affect transcription factors may have in trans- on the expression of *CYP2C19* or other drug metabolizing enzyme gene loci, allelic mRNA ratios were measured. Allelic mRNA ratios were measured at *GATA4*, *AHR* and *HNF4a*. *GATA4* has been shown to associate with an eQTL (rs809204) and also induce expression of *CYP2C19* (Mwinyi, 2010; Schadt, 2008). Transcription factors, due to their upstream
(trans-) action on a gene locus will require a robust allelic expression imbalance to significantly affect gene expression and ultimately drug metabolizing ability.

Allelic mRNA expression was measured at 3 SNPs in the coding region of GATA4. Marker SNPs were selected based on their relative allele frequency and their linkage to a GATA4 eQTL (rs809204). Marker SNPs selected for analysis are rs3729856, rs904018 (eQTL LD, R² = 0.9414) and rs884662 (eQTL LD, R² = 0.400). Marker SNP rs3729856 located in exon 5 exhibited low allelic mRNA ratios ranging from 0.9 to 1.09-fold, with significant allelic expression considered > 1.38-fold (Figure 7). Though there is slight variation in allelic mRNA ratios it is not analytically significant. Further it is not of an adequate magnitude to have a significant effect on a downstream target such as CYP2C19.

Previous studies have implicated rs809204 as a GATA4 eQTL (Schadt, 2008). As described, rs884662 and rs904018 were selected due to their linkage with the eQTL which may increase likelihood of identifying significant allelic expression imbalance. Marker SNP rs884662 showed analytically significant allelic mRNA ratios with an average 1.18-fold change in 8 of 18 samples (Figure 8A). When allelic mRNA levels were measured at rs904018, 12 of 24 samples showed analytically significant allelic expression mRNA ratios of averaging at ~1.23 fold (Figure 8B).

HNF4α was selected due to its role as a central regulator of drug metabolism genes (Watt et. al., 2003). As previously described, allelic mRNA ratios were assessed in HNF4α at marker SNP rs6130615 to determine if any cis-acting functional polymorphisms were present (Figure 9A). Analytically significant allelic mRNA ratios
Figure 7. *GATA4* gene locus and allelic ratios at rs3729856.

(A) Exons indicated by grey boxes, UTRs indicated by hatched regions, and introns indicated by white space. (B) The allelic RNA ratio is expressed as a ratio of Major/Minor alleles surveyed from two measurements. Three standard deviations of the genomic DNA control (>1.38-fold AEI) indicate significant allelic expression imbalance. Data are mean±SD.
Figure 8. Allelic expression imbalance in the 3' UTR of GATA4.

(A & B) The allelic RNA ratio is expressed as a ratio of Major/Minor averaged from a minimum of two measurements at marker SNPs in GATA4. Three standard deviations of the genomic DNA control (>1.11 and 1.14-fold AEI, respectively) indicate significant allelic expression. Data are mean±SD.
were considered greater than 3 standard deviations of the gDNA control (1.17-fold). Though there was some minor variability in allelic mRNA ratios, no robust functional regulatory polymorphisms are indicated (Figure 9B).

Allelic mRNA ratios were interrogated in *aryl hydrocarbon receptor* (*AHR*) due to this transcription factor’s involvement in the initiation of gene expression for *CYP1A1* and *CYP1A2* (Nebert et. al., 2004). Further, *AHR* contains polymorphisms implicated with variation in caffeine metabolism (Sulem et. al., 2011). Here, a single exonic polymorphism was selected for allelic mRNA analysis (rs2066853, Figure 10A). Analytically significant AEI was observed as high as 1.34-fold (average 1.24-fold), in 5 of 15 samples screened indicating that there is likely a *cis*-acting regulatory polymorphism acting at the gene loci (Figure 10B). Of the transcription factors (TFs) studied here no large effects were observed, making it unlikely that a variant would exhibit any effect in *trans*- on CYP2C19 or other DME gene expression.

**CYP2C19*17 association with CYP2C19 enzyme activity**

To confirm the function of *CYP2C19*/*17*, I tested whether *17* was associated with enzyme activity. CYP2C19 enzyme activity (V) was measured in 40 human liver microsomes using S-mephenytoin as substrate. To determine the gene dosage effect of *CYP2C19*/*17* on enzyme activity, CYP2C19 enzyme activity was compared between diplotypes using a t-test (Figure 4). In *1/*2 carriers (intermediate metabolizers, n = 9), the mean activity was 2.3-fold lower (V=0.0091 nmol/min/mg, SD±0.0086) than the average of wild-type samples (V=0.021 nmol/min/mg, SD±0.022,*1/*1, extensive metabolizers, n=11) (p=0.07, Figure 11). Although not significant at p=0.05, this
Figure 9. HNF4A gene locus and allelic expression imbalance in at rs6130615.

(A) Exons indicated by black boxes, UTRs indicated by grey boxes with introns indicated by white space. (B) The allelic RNA ratio is expressed as a ratio of C/T averaged from a minimum of three measurements. Three standard deviations of the genomic DNA control (>1.17-fold) indicate a biologically significant allelic expression imbalance. Data are mean±SD.
Figure 10. Allelic expression imbalance in the gene AHR.

(A) Gene schematic of the AHR gene locus. UTR and coding regions indicated according to schematic legend. Marker SNP rs2066853 indicated in exon 10. (B) Allelic mRNA ratios measured in human liver. Analytically significant measurements were considered > 1.11-fold. Data are mean±SD.

The average activity for *2/*17 (V=0.025 nmol/min/mg, SD±0.018, n=6) carriers did not
significantly differ from wild-type samples, with a 2.7-fold increase in the average activity over *1/*2 carriers (V=0.0091 nmol/min/mg, SD±0.0086, n=9) (p=0.075, Figure 11). Comparing the average enzyme activity between CYP2C19*17 carriers (both heterozygote and homozygote) and non-carriers showed no significant difference. However, this comparison does indicate a borderline (p=0.06) 2.3-fold increase in carriers over non-carriers of CYP2C19*17. This result suggests that *17, which increases allelic expression 2-fold may compensate for the CYP2C19*2 nonfunctional allele in individuals heterozygous for both *17 and *2. This implies that *2 and *17 are likely to be on different alleles, consistent with an inverse LD between *17 and *2 (Appendix C.1). Comparisons between CYP2C19*1/*1 (wild-type) and CYP2C19*17 heterozygotes or homozygotes show these groups are not significantly different (p=0.47 and p=0.512, respectively). However, the small sample size limits generalizations that could be drawn from this analysis.

One liver (LL44, *2/*17) displayed extremely high enzyme activity, exhibiting a velocity of 0.41 nmol/min/mg, more than 12-fold higher than the average level in our cohort. I hypothesized that LL44 would contain a rare variant responsible for the high activity. To test this, I sequenced the promoter (4 kb) and coding regions of sample LL44 and nine control samples using Ion Torrent PGM. Sequencing confirmed the presence of both *2 and *17 alleles in LL44. Sample LL44 showed 14 polymorphisms in the regions sequenced. Compared to nine control samples, no SNPs were unique to LL44 that could explain high enzyme activity, which indicates non-genetic factors or SNPs located in proximal regulatory regions may be responsible. A larger cohort of well-
phenotyped liver tissues will be needed to investigate variants associated with extremely high CYP2C19 enzyme activity.

2.4 Discussion

The results of this study show that *CYP2C19*17 is associated with a 2-fold increase in mRNA expression in human liver. Sequencing and SNP scanning further support the conclusion that *CYP2C19*17 is a functional SNP causing enhanced *CYP2C19* transcription. However, the presence of an additional regulatory variant was indicated by an allelic mRNA expression imbalance in a liver from a sample of African descent. Further regulatory variants may exist, as indicated by exceptionally high CYP2C19 metabolic activity not accounted for by *17 alone in a Caucasian liver. Although the regulation of CYP enzyme expression is complex, use of allelic mRNA expression analysis in human liver tissues proved a valuable tool for assessing the effect size of *CYP2C19*17. It also prompted the continued search for additional regulatory variants at the *CYP2C19* gene locus. Because of the absence of frequent marker SNPs for AEI analysis, large collections of liver tissues will be needed to complete this analysis. Individual CYP mRNA expression and drug metabolizing ability are highly variable (Furukawa et. al., 2004). Our initial analysis of the association between *17 and total CYP2C19 mRNA revealed a significant 2-fold increase for *17 heterozygotes and a 3-fold increase for *17 homozygotes. To account for confounding effects of a *trans*-acting source in CYP2C19 expression, I also measured the expression of five transcription factors. A positive correlation of CYP2C19 mRNA expression with GATA4 mRNA and
negative correlations with CAR and RXRα confirm trans-regulation by diverse processes. After controlling for the expression of these transcription factors, CYP2C19*17 carriers remained significantly associated with a 1.6-fold increase in CYP2C19 mRNA expression, which indicates that CYP2C19*17 increases mRNA transcription independent of these transcription factors. CYP2C19 allelic RNA expression and enzyme activity results further support CYP2C19*17 as a gain-of-function allele. It is possible that there is an interaction between the expression of transcription factors and *17, accounting for the lower estimated effect size when transcription factor expression is taken into account, but the sample size was too small to test this hypothesis further.

CYP2C19*17 effect size when considering the variability in CYP2C19 allelic mRNA levels, may be the result of variation upstream in transcription factors. Screening GATA4, the most likely candidate for interacting with the CYP2C19*17 loci, for allelic expression imbalance indicated low to no significant variation. In fact, the other survey TFs (AHR, and HNF4α), showed similar results with either low or no biologically relevant variation detected. Trans- effects exerted by transcription factors would require a substantial change in gene expression, allelic or otherwise, for marked differences in target gene expression. Lack of robust AEI may be due to the fact that TFs are tightly regulated due to their role in the expression of many genes. It is likely that TFs are subject to some regulatory polymorphisms, but their study is outside of the scope of this work.

The finding that CYP2C19*17 results in a 2-fold increase in CYP2C19 expression
Figure 11. Association between CYP2C19*17 and CYP2C19 enzyme activity by CYP2C19 diplotypes.

CYP2C19 enzyme activity was measured by conversion of S-mephenytoin to 4-hydroxymephenytoin in human liver microsomes. Data are presented as a box plot and whisker plot (box shows median with the 25th and 75th percentiles), and minimum and maximum values are shown by whiskers.
should be considered in the context of clinical CYP2C19 biomarker tests. Currently, the primary utility of CYP2C19*17 is in the context of monitoring treatment and managing risk (clopidogrel treatment). Any variation of CYP2C19*17 effect size between individuals poses a challenge if *17 is to be used as a biomarker. Shown here, the 2-fold variability in expression (in human liver) with allelic mRNA ratios of 1.4- to 2.8-fold is within a sufficiently narrow range for predicting metabolizer status. To fully define the effect size of *17, more samples are needed to quantitate differences in expression. Recently, improved metabolic phenotyping supports the hypothesis that *2/*17 compound-heterozygote metabolizer status is equivalent to an intermediate-metabolizer phenotype, suggesting that CYP2C19*17 may not completely compensate for a loss of function allele (Scott, 2013). For certain CYP2C19 substrates, where ultra-rapid metabolizer status prediction by CYP2C19*17 is questionable, alternate phenotype/genotype tables may be required. Further, this study shows that CYP2C19*17 does not explain all instances of AEI or increased enzyme activity, requiring a continued search for regulatory variants affecting CYP2C19. These additional marker SNPs will be needed to assure adequate prediction of the CYP2C19 metabolic phenotype in any single individual.

In conclusion, I have shown that CYP2C19*17 increases CYP2C19 expression approximately 2-fold in human liver. The defined effect size measured here supports the notion that CYP2C19*17 can serve as a biomarker, refining treatment decisions based on genotype, and solidifies clinical use in the context of risk management for bleeding events during clopidogrel treatment regimens. Further, in two samples with evidence of
enhanced transcription or increased enzyme activity, respectively, the CYP2C19*17 allele cannot account for this finding or CYP2C19*17 alone is insufficient. Therefore, additional regulatory polymorphisms may alter a patient’s response to drug therapy and their identification requires further study.
Chapter 3: Regulation of gene expression at the CES1A1 gene locus

3.1 Introduction

Carboxylesterase 1A1 (CES1A1) is a member of the CES1 family and is active primarily in the liver. CES1A1 plays a significant role in the metabolism of drugs, esters and amides (fatty acyls, cholesterol esters). The drugs metabolized by CES1 include the anti-platelet prodrug clopidogrel (Lins et. al., 1999), anti-virals (oseltamivir) (Shi, 2006; Suzaki et. al., 2013), ADHD medications (methylphenidate) (Sun et. al., 2004), chemotherapeutic agents (irinotecan) (Humerickhouse et. al., 2000), ACE inhibitors (imidapril, enalapril, trandolapril, ramipril) (Geshi, 2005; Song et. al., 2002; Thomsen, 2014) and others. Therefore, CES1 has a significant role in the metabolism and response to these drugs. More specifically, CES1A1 is of interest because of its high expression relative to the pseudogene CES1P1, which expresses at less than 2% of total CES1 gene expression (Fukami et. al., 2008). Spanning 30,312 bases, CES1A1 has not typically been considered in a pharmacogenetic context. However, recent studies have shown that polymorphisms within CES1A1 can have significant effect on enzyme function and drug efficacy (clopidogrel) (Lewis, 2013; Zhu et. al., 2008). Due to CES1A1’s major role in compound metabolism in the human liver, it is imperative to understand the genetics behind gene expression and enzyme function.

The role of CES1A1 in treatment response is drug dependent. In the case of clopidogrel – a prodrug which requires activation by CYP2C19 for therapeutic effect –
CES1A1 inactivates the drug, thereby lowering prodrug bioavailability and decreasing drug efficacy. I aim to determine if CES1A1 exhibits evidence of regulatory variation that affects gene expression or enzyme activity. These variants could have potential utility as pharmacogenetic biomarkers in predicting CES1A1 substrate response.

Previous studies have characterized the structure of the CES1 gene family locus (Fukami, 2008). The main features are two genes which are defined as CES1A1 and CES1P1 (Figure 12). The CES1P1 gene locus is considered a polymorphic pseudogene. It can be either CES1A2, a low expressing functional isoform, or CES1A3 a truncated isoform. CES1P1 is important due to high homology with CES1A1, differing only at bases in the 5’ UTR, exon 1 and intron 1 regions between the two (Figure 13, Table 2). The CES1A1 gene locus is also polymorphic with at least two alternative isoforms identified (Tanimoto et. al., 2007). The major isoform at this locus is CES1A1VAR. Here, the CES1A1 5’ UTR, exon 1 and intron 1 sequence matches the sequence of CES1P1 (Figure 12, Figure 13, Table 2). This variation (CES1A1VAR) is marked by 11 highly linked SNPs which cause CES1A1VAR mRNA to be identical to CES1P1 full length mRNA (the CES1A2 isoform). A major difference in this case is that CES1A1VAR (which matches CES1P1) is being driven by the CES1A1 promoter region instead of the pseudogene promoter. A previous study indicates a similar variant besides CES1A1VAR where CES1P1 sequence is present at exon 1 of CES1A1, though this is not observed in this study (Tanimoto, 2007). Previous studies have suggested that CES1A1VAR affects gene expression (Fukami, 2008). However, it not well characterized and has been applied
to pharmacogenetic studies in only a handful of substrates (irinotecan, isoniazid) (Sai, 2010; Yamada et. al., 2010).

Multiple association studies implicate CES1 polymorphisms in human disease and drug response. Previous work has utilized genome wide association studies to identify potential variants and searched for association between known variants and complex phenotypes. A CESIP1 SNP (rs3785161) has been shown to have increased promoter activity as well as increased clopidogrel and imidapril response (Geshi, 2005; Xie et. al., 2014; Zou et. al., 2014). However, haplotype analysis has shown this variant is in high linkage with the non-functional CES1A3 isoform (of the CESIP1 locus) suggesting an associated functional variant may be responsible (Sai, 2010). Additionally, CES1A1 SNP rs3815583 shows association with appetite reduction during methylphenidate treatment (Bruxel et. al., 2013). Further, CES1A1 SNP rs2244613 shows association with increased efficacy of the anticoagulation drug dabigatran (Pare, 2013) as well as the incidence of sadness during methylphenidate treatment (Johnson et. al., 2013).

Molecular genetic studies testing variants within CES1A1 for function or association with changes in gene and protein expression are lacking. A previous study on CPT-11 toxicity seems to indicate that CES1A2 gene expression (as a function of CES1A1VAR and CESIP1) is relevant to overall toxicity, but this association is unclear (Tanimoto, 2007). In fact, other studies testing association of major CES1 variants (such as CES1A1VAR), have not found significant associations (irinotecan, isoniazid) to drug response or toxicity (Sai, 2010; Yamada, 2010). This may be due to the numerous other enzymes involved in their metabolism, and in some cases lack of analytical stringency.
Previous work by a collaborator identified a coding polymorphism within 
*CES1A1* that significantly affects enzyme activity (Zhu, 2008). This is achieved through an amino acid change in the catalytic binding site of CES1A1. The polymorphism (rs71647871, G143E) was identified through sequencing and shown to ablate CES1 metabolic activity for the majority of substrates (Zhu et. al., 2012). Use as a biomarker was assessed in a clinical cohort of individuals receiving clopidogrel anti-platelet therapy (Lewis, 2013). In the case of clopidogrel, CES1 is the primary inactivator of all forms of the drug (including the active metabolite). Typically, 85% of the drug is inactivated (Lins, 1999). However, decreased CES1A1 activity increased clopidogrel efficacy. This indicates that *CES1A1* functional SNPs can have a significant affect on clopidogrel treatment.

Regulatory polymorphisms can be a significant source of variation in the expression, splicing, or translation efficiency of a gene. If regulatory variants are affecting *CES1A1* gene expression, they may have an effect on protein levels and therefore enzyme activity. In the case of clopidogrel, variants affecting enzyme activity in this way may have utility as clinical biomarkers similar to the G143E coding polymorphism. Even a moderate effect on gene expression could have a significant effect on active metabolite levels, due to the compounding of CES1A1 enzyme activity on all metabolites (prodrug, intermediate and active metabolites). Further, the associations of variants in the *CES1A1* gene locus to drug response indicate that *CES1A1* is a strong candidate gene for harboring functional regulatory polymorphisms.
CESIA1 has been screen for regulatory SNPs using a comprehensive approach developed by our lab. Here we show that the variant CESIA1VAR, associates with decreased overall CES1 mRNA expression, the presence of allelic expression imbalance (1.3-1.5 fold), and decreased protein quantity in human liver. Traditional Sanger sequencing further define the CESIA1VAR into a “short” and “long” subtype, where the “long” is the traditional CESIA1VAR allele while the “short” does not contain the exon 1 differences (CESIA1SVAR, Figure 13, Table 2). Screening for additional polymorphisms in the promoter of CESIA1 did not identify SNPs that better associated with changes in gene expression. No significant association of CESIA1VAR to metabolic capability was seen for enalapril or clopidogrel. Luciferase assay demonstrates CESIA1VAR is associated with a decrease in protein activity in HepG2 cells. Further, mRNA structure analysis of the various 5’ regions of CESIA1 in the luciferase construct indicate that the WT construct has significantly lower free energy with a more defined 3D folding structure when compared to variant alleles. This evidence indicates CESIA1VAR has function as a biomarker, though it may only be relevant during clopidogrel therapy where CES1 is a primary inactivator at all stages of metabolism.

3.2 Materials and Methods

Tissue samples

A cohort of 125 biopsy or autopsy human liver samples were obtained from the Cooperative Human Tissue Network (Midwest and Western Division) under approval o
Figure 12. *CES1A1* Gene Structure.

(A) *CES1* gene loci on chromosome 16. *CES1P1* can exist as a truncated isoform *CES1A3* (truncated at //), or a full length isoform (*CES1A2*). The **green** region in the 5′ end of *CES1P1* represents sequence that can appear as a group of 11 highly linked SNPs at the *CES1A1* gene locus creating the *CES1A1VAR* isoform. (B) A closer schematic of *CES1A1*. The 5′ region which can match *CES1P1* sequence creating *CES1A1VAR* is indicated by a **green bracket**. Additional sequence information can be viewed in Figure 13 and Table 2.
Figure 13. Polymorphisms of CES1A1 and known variants.

5'UTR and exon 1 of CES1A1 compared to CES1A1VAR, CES1A1SVAR and CES1P1. Yellow bases indicate polymorphic sites and blue bases indicate the start of exon 1. Annotated 5'UTR and an intermediate 5' UTR are indicated in text. Numbers correspond to SNP rs#s from Table 2.
Table 2. *CES1A1* 5’ Region SNP #s.

*CES1A1VAR* and *CES1A1SVAR* are comprised of highly linked SNPs which match the sequence of *CES1P1*. In the case of *CES1P1*, these variants would not be SNPs, but rather the major alleles.

<table>
<thead>
<tr>
<th>Figure #</th>
<th>rs#</th>
<th>Figure #</th>
<th>rs#</th>
<th>Figure #</th>
<th>rs#</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>rs3815583</td>
<td>4</td>
<td>rs12149370</td>
<td>8</td>
<td>rs201577108</td>
</tr>
<tr>
<td>1</td>
<td>rs12149373</td>
<td>5</td>
<td>rs12149368</td>
<td>9</td>
<td>rs114788146</td>
</tr>
<tr>
<td>2</td>
<td>rs12149371</td>
<td>6</td>
<td>rs111604615</td>
<td>10</td>
<td>rs12149366</td>
</tr>
<tr>
<td>3</td>
<td>rs12149322</td>
<td>7</td>
<td>Unknown</td>
<td>11</td>
<td>rs12149359</td>
</tr>
</tbody>
</table>
the Ohio State Institutional Review Board. An additional fifty liver samples were obtained from other sources that included the Medical College of Virginia (Richmond, VA, USA), Medical College of Wisconsin (Milwaukee, WI, USA), Indiana University School of Medicine (Indianapolis, IN, USA) or University of Pittsburgh (Pittsburgh, PA, USA). All livers were used under protocols approved by the appropriate committees for the conduct of human research. Cohort composition is primarily Caucasian, with ~15% of African ancestry.

**DNA and RNA isolation and genotyping**

Genomic DNA and RNA were prepared from liver tissue samples using either nuclei lysis buffer or Trizol respectively (Life Technologies, Carlsbad, CA). Nucleic acid isolation has been previously described (Wang, 2011). Briefly, following tissue lysis genomic DNA samples were phenol chloroform purified, ethanol washed and resuspended in TE. RNA isolation used SpinSmart Columns (Denville, Metuchen, NJ) with DNaseI to prevent contamination of genomic DNA in cDNA synthesis. cDNA was prepared using RNA and Reverse Transcriptase SSIII (Invitrogen, San Francisco, CA, USA), with controls lacking reverse transcriptase or template to test for residual gDNA contamination. 5’RACE product was synthesized using the FirstChoice RLM-RACE RNA Ligase Mediate RACE kit in a CES1A1WT/WT individual (Life Technologies). SNPs in CES1A1 were genotyped in liver samples using a primer extension assay (SNaPshot, Life Technologies), GC-clamp real time-PCR assay or fluorescently labeled PCR-restriction fragment length polymorphism (RFLP) analysis as previously described.
(Pinsonneault, 2004; Wang, 2005). PCR conditions and primers for all PCR assays (including genotyping) are given in Appendix A, Table 3. Allele frequencies were checked for Hardy-Weinberg equilibrium (Appendix A, Table 5).

**Measurements of gDNA and mRNA allelic ratios by primer extension (SNaPshot)**

The SNaPshot protocol has been previously described (Pinsonneault, 2004; Wang, 2011). In short, PCR amplification surrounding a marker SNP in an exonic region, from a heterozygous sample is visualized using a primer extension primer and fluorescent ddNTPs. SNaPshot is quantitated by peaks on an ABI3730 sequencer (Life Technologies) that provide the relative amounts of each allele. The average of the allelic ratios of gDNA is used to normalize allelic mRNA ratios in each tissue. The primary marker SNP for CES1A1 was rs1714370 a feature of CES1A1VAR. Due to CES1A1’s homology to the CES1P1 gene locus, the genomic DNA control PCR product was larger than the mRNA product. This allowed specific allelic gDNA ratios, though the product size difference prevents a definitive statistical analysis or 3-standard deviation cutoff for significance as previously described in Chapter 2. Despite this limitation, One-Way ANOVA indicates 13 of 28 samples show significant allelic mRNA ratios when compared to gDNA allelic ratios (P <0.05, *data not pictured*). A more defined cutoff for significant allelic mRNA ratios was set at 1.2-fold for which SNaPshot is sufficient in measuring. Note, mRNA measurements targeted both CES1A1 and CES1A1VAR isoforms. This was acceptable because it has been shown that CES1P1, which is identical to CES1A1VAR, gene expression is <2% of total CES1 gene expression.
(Fukami, 2008). All SNaPshot measurements were completed in duplicate for gDNA. Allelic mRNA ratios were measured in at minimum in triplicate for 26 of 28 samples, while the remaining were measured in duplicate (L36 and L120). Primers for this assay can be viewed in Appendix A, Table 3.

**Determination of CES1 mRNA by quantitative real-time PCR**

*CES1* mRNA expression was measured using quantitative real time PCR (qRT-PCR) on a 7500 Fast Real Time PCR System (Applied Biosystems, Carlsbad, CA, USA). Amplification of *CES1* mRNA captured *CES1A1* and *CES1P1* (<2% total mRNA) locus mRNA but also allowed *CES1A1VAR* mRNA to be amplified. Relative quantitation used *GAPDH* as a housekeeping gene where expression is representative of overall RNA quality in human liver and has low interindividual variability (Barber, 2005).

**Measurement of CES1 protein quantity and metabolic ability**

CES1 protein was isolated from human liver tissue and western blot was performed using CES1 antibodies (Bascom, Cambridge, MA). Using equivalent input quantities, total CES1 protein expression was determined using UVP VisionWorksLS (Upland, CA) and normalization to a representative sample. All CES1 protein quantities were measured in duplicate. CES1 enzyme activity was measured using isolated S9 fractions. Protein content was measured by Pierce BCA Assay Kit (Pierce, Rockford, IL). Equivalent protein S9 protein quantities were incubated with clopidogrel or enalapril for 90 minutes at 37°C. Following incubation, LC MS/MS was used to measure the
concentration of the hydroxyl metabolites for each sample. All metabolic activities were performed in triplicate.

**Sequencing of the CES1A1 Promoter Region**

The *CES1A1* promoter region (5kb) was sequenced to scan for regulatory variants. PCR products were visualized on agarose gel to determine relative quantity. Products were mixed and fragmented on a Covaris S220 (Woburn, MA). NEBNext Fast DNA Library Prep Set for Ion Torrent (New England Biosciences, Ipswich, MA) was used to generate barcoded libraries that were run on an IonTorrent PGM (Life Technologies). Sequencing results were analyzed in CLC Bio’s Genomics Workbench (CLC Bio, Katrinebjerg, Denmark), and SNP calling required, at minimum, 20 reads at a given base with 20% reads per allele.

**Infusion cloning and construct preparation**

Luciferase assay vectors were prepared using pGL3-Basic (Promega, Madison WI) and a *CES1A1* promoter region-5’UTR-exon 1 region insert (Figure 18). The Infusion Cloning System (Clontech, Mountain View, CA) was used to insert the fragment at the HindIII multiple cloning site of pGL3-Basic. Completed reactions were transformed into XL-2 (Agilent, Santa Clara, CA) cells and plated on LB agar with carbenicillin. Individual clones were screened and plasmids isolated by the Zyppy Mini Prep kit (Zymo Research, Irvine, CA). Plasmids were screened for inserts by HindIII digestion and sequenced for confirmation of the insert present in frame with the luciferase gene. Positive clones with haplotypes of interest were subcloned into DH5α
cells and screened for insert presence. Three positive subclones per haplotype were then combined and cultured via a Qiagen MidiPrep Kit (Qiagen, Valencia, CA). Stock cultures were frozen in a 1:1 ratio of bacteria to glycerol for future use.

**Luciferase Assay**

Luciferase assay was performed using pGL3-Basic vector prepared as described. HepG2 cells were selected due to their high endogenous expression of CES1 (www.biogps.org)(Wu et. al., 2009). Prior to transfection, cells were plated in 12-well plates and grown to ~60% confluence. Transfection plasmid mixes of CES1A1-pGL3B and pRL (Promega), a background Renilla fluorescence vector, were prepared in a ratio of 1ug to 0.2 ug in OptiMem Media (Life Technologies). Plasmid mix was added to an equivalent OptiMem mix of transfection reagent FuGene HD (Promega) at 1ug pDNA/ 3 ul reagent. Six hours post transfection media containing antibiotics was added to transfected wells (Penicillin/Streptomycin, Life Technologies). Following incubation for 24 hours luciferase activity was measured by Dual-Glo Luciferase Assay System (Promega), and read out on a Packard Fusion plate reader (PerkinElmer Life and Analytical Sciences, Shelton, CT). Luciferase activity was determined as a ratio of luciferase fluorescence over renilla fluorescence. Luciferase was measured in 3 experiments, with 3-replicates and two measurements per replicate.

**Data Analysis**

Significant association of CES1A1VAR to gene expression and protein quantity was performed in SPSS (IBM, Armonk, NY). Stepwise linear regression including
covariates, gender, race, rs3815583 carrier status, rs2244613 carrier status, and 
*CESIA1VAR* carrier status were included in these analyses. Correlation statistics for 
RNA-Protein relationship, One-way ANOVA for AEI and luciferase activity were 
performed in GraphPad Prism. Figures were generated in Graph Pad, or designed in 
Photoshop/PowerPoint. P-values < 0.05 were considered significant for all associations. 
Hardy-Weinberg equilibrium and linkage for *CESIA1VAR* and other variants of interest 
was assessed using Helix Tree SNP and Variation Suite (Golden Helix, Bozeman, MT, 
USA) (Appendix A, Table 5; Appendix B, Figure 27).

**Generation of CESIA1 mRNA folding structure**

Structure (2D) and free energy was assessed using mFOLD, with sequence based 
on the intermediate length 5’ UTR, CES1A1 exon, and the first 18 bases of pGL3-Basic 
downstream of the insertion site. Folding structures and free energy was recorded and 
compared between the various constructs by One-way ANOVA (GraphPad Prism). To 
best visualize the overall difference in the predicted structures, three dimensional models 
were generated using RNAFold and RNAComposer. Based on the medium 5’ UTR 
length, the rs3815583 construct was identical to the WT and was omitted from analysis.

3.3 Results

**An alternative 5’ variant of CES1A1**

To understand the makeup of *CESIA1* and its relationship to *CESIP1*, Sanger 
sequencing of the *CESIA1* gene locus was performed. Previous studies characterized the
gene loci for CES1A1 and CES1P1 identifying CES1A1VAR (where CES1P1 “replaces” CES1A1). Multiple CES1A1 isoforms have been previously described, though the number of variants in unclear. Sequencing samples of CES1A1 homozygotes, CES1A1VAR heterozygotes and CES1A1VAR homozygotes defined the major regions affected by CES1A1VAR. The 5’ boundary begins between rs3815583 (unaffected) and rs12149373, and ends in intron 1 (Figure 13). Screening for CES1A1VAR SNPs, while using CES1A1 specific primers identified 3 samples containing only CES1A1VAR 5’UTR associated SNPs. This shorter variant, referred here as CES1A1SVAR was confirmed by sequencing and represents 5’ sequence variation seen at the CES1A1 gene locus.

CES1A1 5’ UTR Usage

To measure the most abundant mRNA isoforms for allelic mRNA analysis, CES1A1 5’ UTR length was tested by 5’ RACE. In a CES1A1 homozygote, fluorescent CES1A1-RACE PCR product indicates that the primary length of CES1A1 mRNA in human liver is ~ 50 bases shorter than mRNA annotated by RefSeq and the UCSC Genome Browser (Figure 14). qPCR confirms abundance of the intermediate 5’ UTR length mRNA and indicates the annotated 5’ UTR measures at roughly 1-2% of total CES1 gene expression (data not pictured). A shorter, exon 1 coding region is also likely present based on RACE analysis, though in small quantity. Based on these observations, the medium 5’ UTR is an abundant mRNA isoform for SNaPshot analysis.

CES1A1 associates with decreased total mRNA expression
Previous studies have shown that CES1A1 gene expression is highly variable, though it is not clear to what extent it associates with any CES1A1 variants. Total CES1 mRNA was measured in 57 human liver samples. Analysis was performed by stepwise linear regression including gender, race, rs3815583, rs224816 and CES1A1VAR carrier statuses. For individuals with at least one copy of CES1A1VAR (heterozygotes and homozygotes) total CES1 expression was 2.65-fold lower (p =0.003) than in non-carriers (Figure 15). No other covariates were significantly associated with gene expression.

*CES1A1VAR is moderately associated with allelic expression imbalance.*

Allelic mRNA ratios were measured determine if the CES1A1 gene locus is being affected by CES1A1VAR or a cis- acting functional regulatory polymorphism(s). Again, due to the minor contribution of CES1P1 to total gene expression, allelic mRNA ratios have been measured using CES1 primers. Allelic expression was measured at rs12149370 in the 5′UTR of CES1A1 (included in the intermediate length mRNA transcript). This SNP is included in both CES1A1VAR (minor allele frequency = 0.17) and CES1A1SVAR (MAF < 2%) isoforms. Allelic mRNA ratios were screened in 28 human liver samples. Statistical analyses indicate 13 of these show significantly different allelic mRNA ratios at P < 0.05. However, due to the differences in gDNA and mRNA PCR amplification, a cutoff for potential allelic expression imbalance was set at >1.2-fold. Based on this, 18 samples showed allelic expression imbalance greater than 1.2-fold (1.23 to 1.4-fold, with a 2.23-fold outlier, Figure 16). It is clear that there is some degree of allelic expression imbalance in the majority of the samples screened.
Figure 14. *CES1A1* 5′ RACE Indicates UTR lengths.

5′UTR length was assessed by 5′RACE. Medium 5′ UTR transcript is indicated by arrows consists of 2 similarly sized transcripts flanking the defined “---“labeled *Med 5′ UTR* in Figure 13. Data also suggests a small population of *CES1A1* transcript with no 5′ UTR, and no annotated 5′ UTR transcript product was observed.
Figure 15. *CES1A1VAR* associates with decreased CES1 gene expression.

*CES1A1VAR* carrier status was the only significant predictor of gene expression when analyzed with gender, race, rs2244613 and rs3815583, decreasing gene expression 2.65-fold when compared to non-carriers (P = 0.003; WT, n = 29; VAR, n = 31). Data are presented as a box plot and whisker plot (box shows median with the 25th and 75th percentiles), and minimum and maximum values are shown by whiskers.
CES1A1SVAR carriers, L50, L51, and LL10 all show some degree of AEI, though it varies as in CES1A1VAR carriers. Because the majority of samples screened show some degree of AEI, this indicates that CES1A1VAR (and CES1A1SVAR) may be a functional variant, or is highly linked to the functional variant(s). Notably, allelic mRNA ratios in one case measured at ~2-fold, and in other case at ~5-fold (data not pictured), suggesting uncommon variants may also be present.

**Sequencing of the CES1A1 Promoter Region**

One possible conclusion of the allelic expression assay is that CES1A1VAR is in linkage with another functional polymorphism. To test this, 5kb of the promoter region was screened for polymorphisms that associate with instances of allelic expression imbalance. No single SNP associated with allelic mRNA ratio presence, though rs38135583 was present in all AEI – samples (n = 4), and only present in 1 of 6 AEI+ samples. This SNP is located in the 5’UTR of CES1A1, is not associated with CES1A1VAR and is associated to appetite reduction during methylphenidate therapy. However, genotyping revealed no association with AEI status (or other measured phenotypes).

**CES1A1VAR association with Protein Quantity**

Based on the association of CES1A1VAR with decreased total CES1 gene expression, I sought to determine if these effects were present at the protein level. CES1 protein quantity was measured by western blot and quantitated using protein input and a sample for normalization (arbitrary unit) (Appendix B, Figure 28). To determine the
effect of CES1A1VAR on protein quantity, carriers (heterozygotes n=15, and homozygotes n=4) were compared as a single group against non-carriers. CES1A1VAR carrier protein quantity was 1.33-fold lower (1.018 units, SD 0.07277 N= 19) when compared to non-carriers (1.34 units, SD 0.08228 N = 24). When analyzed by stepwise linear regression including gender, race, rs3815583 carrier status and rs224816 carrier status, CES1A1VAR was the only significant contributor to decreased protein quantity (p = 0.008) (Figure 17). To strengthen the association between CES1A1VAR carrier status and CES1 protein quantity, correlation between CES1 RNA and protein was analyzed. In this cohort, gene expression and protein quantity were not significantly associated where both data were available (Figure 17, p = 0.065, N = 46).

**CES1A1VAR association with protein activity**

The effect of CES1A1VAR on CES1 protein levels was further interrogated using human liver S9 fractions. The velocity of hydrolysis of two CES1 substrates, clopidogrel and enalapril were measured and tested for association with CES1A1VAR carrier status by linear regression. Hydrolysis of clopidogrel in CES1A1VAR individuals was not significantly decreased when compared to wild-type (WT: 329.7 +41.7 pmole/min/mg N = 24 vs. VAR: 277.7 + 32.07 pmole/min/mg N = 19) (Figure 18). Similar results were obtained when comparing enalapril hydrolysis between CES1A1WT and CES1A1VAR individuals (WT: 65.91+ 10.3 pmole/min/mg, N =24 vs VAR: 59.8 +7.342 pmole/min/mg, N = 19) (Figure 18). In both cases, no covariates (gender, race, rs3815583, rs224613 and rs71647871 (G143E) were significantly associated with protein
Figure 16. Allelic expression at rs12149370

The allelic RNA ratio is expressed as a ratio of major/minor alleles. Allelic expression is considered significant when >1.2-fold. AEI+ samples have an average of ~1.35-fold allelic mRNA ratio (18 of 28 samples). An outlier showing ~5-fold allelic expression imbalance was also identified (data not pictured). Data are mean±SD.
Figure 17. Correlation between CES1 protein and RNA, CES1 protein quantity association with *CES1A1VAR*.

(A) CES1 mRNA and protein level do not significantly correlate with one another ($R^2 = 0.05134$, $P = 0.065$, $N = 46$). (B) CES1 protein levels associate with a 1.33-fold decrease in protein expression when an individual has at least one copy of *CES1A1VAR* ($P = 0.008$). Horizontal lines indicate population mean.
activity. Notably, including both \( CYP2C19^*2 \) and \( CYP2C19^*17 \) (due to the role of \( CYP2C19 \)) with the above covariates results in rs3815583 carrier status as a significant predictor of clopidogrel metabolic ability (though only when analyzed in a non-stepwise manner, *data not pictured*).

**\( CES1A1 \) 5′ UTR activity by luciferase in HepG2**

To determine the function of the various \( CES1A1 \) 5′ UTRs, luciferase assay was performed in HepG2 cells. Constructs contained 5′UTR regions corresponding to \( CES1A1, CES1A1VAR, CES1A1SVAR \) and \( CES1A1WT + \) rs3815583 (a complete list of construct structure and variation can be viewed in Figure 19A and Appendix A, Table 7, respectively). \( CES1A1VAR \) showed a significant 35% decrease in luciferase activity when compared to \( CES1A1WT \) (\( P < 0.01 \)) (Figure 19B).

Interestingly, \( CES1A1SVAR \) showed a 36% increase in luciferase activity when compared to \( CES1A1WT \) (\( P < 0.01 \)), though there was considerable variation in activity between replicates. \( CES1A1 + \) rs3815583 showed no significant difference in luciferase activity. Deviation of \( CES1A1VAR \) and \( CES1A1SVAR \) suggest that these two variants may be functional. Notable significant differences are indicated by * (\( WT \) vs. \( VAR \)) or ** (\( WT \) vs. \( SVAR \)) (Figure 19).

**RNA folding in the 5′ UTR of \( CES1A1 \)**

To explore the potential mechanism for which 5′ UTRs have exhibited functions in this study, they were put into various RNA folding programs. The \( CES1A1WT \) construct showed 2 potential folding patterns, with an average final \( \Delta G = -27.74 \). The \( CES1A1VAR \) construct yielded 4 different structures with an average \( \Delta G = -40.58 \). The
Figure 18. CES1A1VAR does not associate with metabolic ability.

Gender, race, rs3815583, rs2244613 and CES1A1VAR do not significantly associate with metabolic ability for (A) clopidogrel or (B) enalapril. Horizontal lines indicate population mean.
Figure 19. Luciferase activity of *CES1A1* constructs in HepG2 cells.

(A) Construct schematic, specific construct differences can be viewed in Appendix A, Table 7. (B) *CES1A1VAR* luciferase activity was significantly lower than *CES1A1WT* (35%, $P < 0.01$). *CES1A1SVAR* was significantly higher than CES1A1WT activity (36%, $P < 0.01$). Data are presented as the average of 3 transfection experiments. Significant differences when compared to WT are indicated by * (*CES1A1VAR*) and ** (*CES1A1SVAR*).
CES1A1SVAR construct yielded 3 structures with an average $\Delta G = -40.68$. When compared to the WT, both VAR and SVAR constructs had significantly lower free energy (P <0.001, One-way ANOVA). Though there is some overlap in the predicted structures, the difference in $\Delta G$ indicates that CES1A1VAR and CES1A1SVAR, which also showed significant differences in luciferase activity in cell culture, are more stable in their predicted structures when compared to the WT. RNAFold outputs confirmed mFOLD free energy predictions. The RNAFold 2-D structure was visualized in 3-D using RNAComposer. Here, drastically different 3-D structures can be seen between the WT, VAR and SVAR constructs (Figure 20). These indicate overall structural differences that may affect transcription or translation.

3.2 Discussion

The results of this study indicate that CES1A1 contains regulatory polymorphisms that decrease mRNA and protein expression in human liver. This conclusion is supported by total gene and allelic mRNA expression data, as well as protein quantitation and luciferase activity. The primary variant associating with a change in mRNA and protein expression is CES1A1VAR, which consists of 11 SNPs in the 5’ UTR and Exon 1 of CES1A1 (Figure 13, Table 2). Present at a minor allele frequency of ~17%, this variant results in CES1P1 (CES1A2) mRNA expression from the CES1A1 gene locus. In addition to this potential pharmacogenetic biomarker, I also better define CES1A1 5’ UTRs and their role in gene expression.
Measurement of the average 5’ UTR length by 5’RACE (in a CES1A1/CES1A1 individual) indicates that the most frequent UTR length is shorter than the RefSeq annotation. Interestingly, this intermediate length 5’ UTR is considered an exonic coding region at the CES1P1 gene locus (where no 5’ UTR is annotated). Targeted sequencing also identified an uncommon (MAF < 2%) alternative mRNA isoform referred to as CES1AISVAR, where only the 5’ UTR region at CES1A1 contains CES1P1 variants (Figure 13). Interestingly, CES1AISVAR and a previous study suggest that there are likely numerous, uncommon CES1A1 variations in addition to CES1A1VAR (Tanimoto, 2007).

Decreases in CESI mRNA transcription are supported by our human liver protein data. These results are further recapitulated as CES1A1VAR decreases protein expression through luciferase activity in HepG2 cells (while CES1AISVAR may increase protein activity). The three dimensional structures based on the 5’ UTRs of CES1A1 indicate that significantly different free energy and folding conformations which may affect transcription or translation. If translation is also negatively affected by CES1A1VAR this would indicate two mechanisms both acting in the same phenotypic direction.

Due to the complex nature of the CESI gene loci, gene expression is difficult to measure. Therefore, the majority of this study has focused on the promoter and 5’ region of the gene where the few differences between the two isoforms exist. In this study, in addition to CES1A1VAR carrier status other relevant SNPs associated with CES1A1 phenotypes, such as rs2244613 (associated to dabigatran metabolism and methylphenidate side effects), and the 5’ UTR SNP rs3815583 (associated to
methylphenidate side effects) were included. Notably, these polymorphisms were not considered significant contributors to mRNA expression or CES1 protein quantity. Further, inclusion of rs71641730 (G143E) showed no significant associates with protein activity, though this may be due to small sample size.

The \textit{CES1P1} variant rs3785161 has not yet been assessed in relationship to the work here. Though function has been shown in cell culture, the true functional effect of this polymorphism is unclear. If it does truly increase gene expression it would have to be have a large effect size to increase gene expression from <2\% of total \textit{CES1} mRNA to a significant amount. This could potentially interfere with our allelic or total \textit{CES1} mRNA measurements, but would indicate a ~15+ fold increase in gene expression.

Moreover, an increase in gene expression would be counter intuitive to the decreases in gene and protein expression observed. The original association to imidapril response for this variant is what is expected –increased promoter activity, CES1 activity, drug activation and efficacy. However, the two recent studies testing for association with clopidogrel efficacy take opposing views. Increases in CES1 activity as a result of this variant should decrease clopidogrel efficacy, as more of the drug is inactivated as seen in Xie, et. al. (Xie, 2014). However, the publication by Zou et. al. suggests that variant presence improves clopidogrel efficacy, which is unexpected given the role of CES1 in the inactivation of clopidogrel (Zou, 2014). Notably, this takes into account supports the defined linkage of rs3785161 with the non-functional \textit{CES1A3} (at the \textit{CES1P1} locus) (Sai, 2010) and suggests that rs3785161 is likely a variant in high linkage with a functional SNP. The relationship between rs3785161 (or associated functional SNP),
*CESIA1VAR*, and any potential effects on mRNA or protein expression warrant further exploration.

CES1 gene expression is highly variable as evidenced by our quantitative PCR and protein quantity data. Carboxylesterase, having a diverse role in the liver is undoubtedly under the influence of both *trans*- genetic and environmental factors, which contribute to variability. This variability may have a significant effect on observing significant associations of *CESIA1VAR* to metabolic activity. Clopidogrel and enalapril metabolism was not significantly associated with *CESIA1VAR* carrier status. It has previously been shown that *CESIA1VAR* does not associate with irinotecan and isoniazid metabolic ability, though CES1 is not the only involved DME (Sai, 2010; Whirl-Carrillo et. al., 2012; Yamada, 2010). One explanation for the lack of association in the protein activity experiments is additional enzymes interacting within the S9 fraction with either clopidogrel or enalapril. Due to well-defined *CYP2C19* genetic variant alleles, linear regression was performed with the other available covariates, *CYP2C19*<sup>∗</sup>2 and *CYP2C19*<sup>∗</sup>17 to determine if any of the genotyped variants affected clopidogrel metabolism. Interestingly, rs3815583 carrier status was associated only in a non-stepwise linear regression. However, no significant association is seen with mRNA or protein expression and this variant. One explanation is that this variant is marking a functional SNP in the coding region of CES1A1. This requires further study to determine if rs3815583 has any effect on enzyme function.

*CESIA1VAR* is not completely associated with the presence of allelic mRNA ratios (average ~1.35-fold). Still, *CESIA1VAR* associates with decreases in total mRNA
Figure 20. 3D Structures of*CES1A1* 5′ UTR Isoforms.

Three dimensional structures are based on the intermediate *CES1A1* 5′UTR and a portion of the following pGL3-Basic vector. Structures for *WT* (A), *VAR* (B) and *SVAR* (C) include pink dNTPs which represent positions where variation occurs between isoforms. In all cases *CES1A1* exonic region differences are located in the upper helix, while 5′ region variation is in the lower region.
expression, protein quantity, and lower luciferase activity. One explanation is that AEI may be driven by enzyme induction, in some cases producing larger allelic mRNA ratios such as the ~2-fold AEI as seen in L18. An outlier sample was also identified in this study, with allelic mRNA ratios at ~5-fold (data not pictured). Though this sample is a carrier of CES1A1VAR it is likely that either gene duplication or an unrelated SNP is responsible. To determine if any candidate SNPs were associating with typical AEI or changes in gene or protein expression, the promoter region was sequenced. In 5kb of the promoter region, no SNPs strongly associating with AEI + samples were observed, and a moderate association was seen with rs3815583 and AEI – samples (present in only a single AEI + sample). This is interesting considering association of the polymorphism with variation in methylphenidate treatment (Bruxel, 2013). However, genotyping and subsequent analyses indicate this SNP is not clearly associated with CES1A1 AEI status, total mRNA or protein expression and does not alter luciferase activity. Functional polymorphisms might exist outside the screened region, but the density of variation in the 5'UTR of CES1A1, and the subsequent findings demonstrate rationale for CES1A1VAR as a functional variant.

CES1A1VAR’s utility as a pharmacogenetic biomarker is not clear, though here I have gathered evidence for a function in human liver tissue. Interestingly, CES1A1 shows a higher free energy when compared to both variant alleles, implying it may be more receptive to translation. It has already been shown that rs71641730 (G143E) can have a significant effect on clopidogrel efficacy. With this loss of function allele, heterozygote carriers have their metabolic ability halved. Though CES1A1VAR does not
have as strong an effect in decreasing gene expression or protein quantity it may be relevant in clopidogrel therapy. Clopidogrel isoforms, which consists of a prodrug, intermediate and active metabolite are always inactivated by CES1. These three points of entry for CES1 may allow the modest effect of CES1A1VAR to have a measurable effect on clopidogrel efficacy. Distinct from hydroxylation as measured in this study, a clinical population with endpoints such as platelet aggregation, incidence of bleeding or coronary stent thrombosis would be a more informative measure of clinical utility.

Here I show that the previously described CES1A1VAR decreases gene and protein expression in the human liver by at least ~30%. I also indicate the primary 5′ UTR length for CES1A1, and identify an uncommon alternative 5′ UTR sequence (CES1A1SVAR). It remains to be seen if CES1A1VAR will have utility as a clinical biomarker, as it has already been shown to not associate with drug response in select compounds. However, given a substrate where CES1 is the primary metabolizer (clopidogrel) and relevant clinical endpoints are available, there may be a significant association. It is worth considering that, CES1A1 expression is variable and may harbor additional regulatory variants (as evidenced by outliers), though these are likely uncommon.
Chapter 4. Evidence of regulatory polymorphisms at the *ACE* gene locus

4.1 Introduction

Angiotensin I-converting enzyme (ACE) regulates blood pressure through the renin-angiotensin-system (RAS), among other physiological functions. Renin facilitates the formation of angiotensin-I from the peptide angiotensinogen. ACE hydrolyzes angiotensin I to angiotensin II (vasopressor) and inactivates bradykinin (vasodilator) (Soffer, 1976). In turn this increases blood pressure. Hypertension (high blood pressure) increases risk of stroke and heart failure (Wolf et al., 1991). Due to its role in the RAS, ACE is an attractive drug target for managing blood pressure. ACE inhibitors are widely prescribed, lower blood pressure and decrease adverse cardiac events in individuals with hypertension and heart failure (Wing et al., 2003; Yusuf et al., 2000). A 25 exon gene spanning ~25 kb, *ACE* is ubiquitously expressed and implicated in numerous diseases in addition to hypertension (Figure 21). These include myocardial infarction, diabetic neuropathy, and amyloid plaque formation in Alzheimer’s disease (Cambien et al., 1992; Lewis et al., 1993; Yang et al., 2008). The aim of this study is to survey *ACE* for evidence of *cis*-regulatory polymorphisms affecting *ACE* gene expression in non-cardiac tissues while assessing the function of previously identified variants on gene expression in these tissues. Both classic molecular genetics and next-generation sequencing methods will be applied to assess these questions.
Early familial studies have determined ACE activity is highly heritable (Cambien et. al., 1988). The vast majority of genetic studies have focused on an insertion/deletion (I/D, rs1799752, Figure 21) polymorphism located in intron 16. However, in vivo studies are unable to identify an effect of the I/D variant on gene expression or splicing (Lei et. al., 2005;Rosatto et. al., 1999). Further, meta-analyses have not consistently associated the I/D polymorphism to hypertension(Sayed-Tabatabaei et. al., 2006). However, significant associations have been reported for Alzheimer’s, ischemic stroke and diabetic nephropathy suggesting that the I/D polymorphism is a surrogate marker for a functional variant (Fujisawa et. al., 1998;Lehmann et. al., 2005;Ng et. al., 2005;Zhang et. al., 2012).

Our group has found an association between three enhancer SNPs located ~3kb upstream of ACE and a 4-fold decrease in cardiac ACE expression. These enhancer SNPs also increase risk for primary outcomes in hypertensive individuals with coronary disease (mostly myocardial infarction with an odds ratio of ~6) (Johnson, 2009). The polymorphisms (rs4290, rs7213516, rs7214530) are in high linkage disequilibrium with each other (but not with the I/D variant). They are common in African populations (MAF ~17%) and to a lesser degree in Hispanics and Caucasians (~4% and 0.2% MAF, respectively). Reporter gene assays fail to definitively identify which of the variants is active though it is clear these account for a significant decrease in ACE mRNA expression in the heart.

African populations are known to have high risk for cardiac events and adverse drug reactions (ADRs) from ACE inhibitors (angioedema, dry cough, and hypotension) (Gibbs et. al., 1999;Israíl et. al., 1992;Spinar et. al., 2000). These adverse reactions are
Figure 21. Schematic of $ACE$ gene locus and relevant SNPs.

Gray boxes indicate UTR and black boxes indicate coding exons. SNPs of interest are indicated with arrows. Enhancer SNPs rs4290, rs7214530 and rs7213516 are located ~3 kb upstream from exon 1 of $ACE$. 
not limited to African populations but are less frequent in Caucasians (Israeli et. al., 1995; Murray et. al., 1998). Notably, the I/D polymorphism is present in similar frequencies among populations, implying the variant responsible for increasing ADRs is unrelated and specific to African populations. These ADRs can be therapy limiting prompting alternative antihypertensive treatments. Further, while the enhancer SNPs may affect ADRs, additional variants may be active at the ACE gene locus and affect drug response and efficacy.

Although the I/D variant is poorly characterized and unlikely to have function per se, thousands of clinical studies continue to rely on it as the sole variant representing genetic factors in ACE. Lacking frequent nonsynonymous SNPs of proven effect, ACE is likely to harbor regulatory variants such as the enhancer SNPs (Johnson, 2009). Our group has developed a comprehensive approach to discovery of regulatory variants that I apply here in non-cardiovascular tissues. In short, allelic mRNA ratios are measured at a coding SNP in both mRNA and gDNA by SNaPshot. Deviation of the mRNA ratio from the gDNA ratio is termed allelic expression imbalance (AEI), and indicates a cis-acting functional polymorphism is present.

The work here expands on our previous study and aims to determine whether the enhancer SNPs have activity in other tissues, and if evidence of regulatory variants are present in alternative tissues. Here I determine that the enhancer SNPs show no notable effect on gene expression in liver and two brain tissues (putamen, prefrontal cortex). This is supported by their existence in a MEF2A cardiac transcription factor binding site (Johnson, 2009). A survey of different tissues (liver, prefrontal cortex, putamen, adipose
and heart) for indication of \textit{cis}-acting regulatory polymorphisms by allelic mRNA ratios, identifies liver (~2-fold AEI), putamen (~1.7-fold), and adipose (1.4-fold) as tissues where \textit{cis}-acting polymorphisms are active. This was achieved in two ways, by allelic mRNA ratios and RNA-sequencing transcriptomes which allow the capture of allelic ratios for all the SNPs in ACE coding regions.

4.2 Materials and Methods

**Tissues**

Allelic mRNA survey was performed on 50 heart tissues, obtained under an existing IRB protocol from heart transplant patients (collaboration with Dr. P. Binkley, OSU). Additional tissues were analyzed from IRB approved cohorts of 125 liver tissues, and over 200 brain tissues (prefrontal cortex and putamen).

**Genotyping**

SNPs of interest were genotyped using primer extension primers, GC-clamp real-time PCR, and fluorescent restriction fragment length polymorphism. This includes the three enhancer SNPs, and rs4343, the marker SNP used for allelic expression ratios. The enhancer SNP rs4290 was used as the primary indicator of enhancer SNP status in all tissues due to the high linkage of these polymorphisms. Primer sets and cycling conditions can be viewed in Appendix A, Table 3. Genotypes for rs4343 were assessed for Hardy-Weinberg Equilibrium using Helix Tree SNP and Variation Suite (Golden Helix, Bozeman, MT, USA) (Appendix A, Table 6).

**Allelic mRNA ratios**
SNaPshot as previously described (Pinsonneault, 2004; Wang, 2011), uses PCR amplification surrounding a marker SNP in an exonic region, from a heterozygous sample. The SNaPshot reaction outputs fluorescent peaks on an ABI3730 sequencer (Life Technologies) that provides the relative amounts of each allele. Average gDNA allelic ratios are used to normalize allelic mRNA ratios in each tissue. Deviation of allelic mRNA ratios from unity following normalization is termed allelic expression imbalance (AEI), and indicates a cis-acting regulatory SNP. In ACE, normalized allelic mRNA ratios are measured at rs4343. Significant mRNA AEI in each tissue is considered three standard deviations from the mean of the genomic DNA.

**Allelic ratios from RNA-sequencing data**

RNA-sequencing transcriptomes were obtained from human prefrontal cortex, putamen, heart, adipose and liver tissues. Transcriptome alignment was completed by either TopHat or LifeScope. The allelic ratios of all SNPs in the coding region of ACE were extracted from the aligned data. Allelic mRNA ratios were selected for analysis that had a minimum of 30 reads at the variant’s position. Variants that met the read depth criteria were all aligned to the genome using TopHat and an IUPAC reference genome (which counts SNPs as mismatched bases). To account for potential loss of minor allele reads introduced into allelic ratios by this method I employed the correction methods previously outlined by our group (Smith et. al., 2013). Briefly, all allelic ratios are converted to their absolute fold change (e.g. 60%VAR = 1.5-fold, 40%VAR = 0.667 with an overall change of 1.5-fold). Absolute fold-changes were adjusted for IUPAC bias
with a reduction of 0.3 per 1 fold (e.g. 2-fold increase would subtract 0.6-fold equaling 1.4-fold final ratio).

4.3 Results

**Allelic mRNA ratios in Prefrontal Cortex**

Our group previously identified enhancer SNPs that were shown to decrease ACE gene expression in human heart. To test the effects of these polymorphisms in prefrontal cortex, and determine if any additional cis-acting polymorphisms are present, allelic mRNA ratios were assessed. Significant allelic ratios were considered >1.3-fold as determined by gDNA ratios. Sample MB145, an individual of African descent and heterozygote carrier of the enhancer SNPs shows no significant allelic expression imbalance. In the other samples assayed, a single sample showed border-line allelic expression imbalance (~1.3-fold) (Figure 22). Significant standard deviation was observed in some samples. This is likely due to the low average expression of ACE in prefrontal cortex (Appendix B, Figure 29). Further, there is no observed association of allelic mRNA ratios to the I/D variant.

**Allelic mRNA ratios measured in human putamen.**

Due to potential function of ACE, and use of ACE inhibitors in Alzheimer’s disease, I surveyed putamen for evidence of cis-acting polymorphisms and enhancer SNP function (Hu et. al., 2001; Ohrui et. al., 2004). Putamen is a good candidate due to high expression of ACE when compared to prefrontal cortex expression (Appendix B, Figure
Figure 22. Allelic mRNA ratios in Prefrontal Cortex.

The allelic RNA ratio is expressed as a ratio of major/minor alleles averaged from a survey of two measurements. Three standard deviations of the genomic DNA control (>1.3-fold AEI) indicate a biologically significant allelic expression imbalance. Minor allelic ratios are present in at least one sample (MB33). ACE I/D carrier status is heterozygous unless otherwise indicated. No I/D association with AEI was observed. MB145, indicated by *, is a carrier of the enhancer SNPs and shows no significant AEI. Variability may be driven by low expression of ACE in the PFC. Data are mean±SD.
89. Significant allelic mRNA ratios were considered >1.2-fold, with some large standard deviation observed. Of samples assessed, 2 of 19 showed significant allelic expression averaging 1.73-fold, indicating that cis-acting regulatory polymorphisms are likely present in these individuals (Figure 23). These samples, MB72, and MB80 are of African and Hispanic ancestry, respectively. Sample MB55 is a heterozygote carrier of the enhancer SNPs, though no functional effect is observed (as seen in PFC). No association between allelic mRNA ratios and the I/D polymorphism was observed.

**Allelic mRNA ratios in human liver.**

To continue to explore the incidence of allelic expression imbalance in human tissues, allelic mRNA ratios were measured in the liver. Here, allelic mRNA ratios were considered significant when >1.15-fold (3-standard deviations of gDNA). No African samples were observed that carried our marker SNPs and the enhancer SNPs. Here, 3 of 7 samples of Caucasian ancestry show allelic expression imbalance of ~1.83-fold, with at least one additional sample showing borderline significance as determined by 3-standard deviations (Figure 24). There is no observed association of allelic mRNA ratios to the I/D polymorphism.

**Allelic mRNA ratios from RNA-sequencing.**

To survey additional tissues for indication of cis-acting functional polymorphisms, RNA-sequencing transcriptomes were utilized to obtain allelic mRNA ratios in many tissues. Following application of the criteria for observing allelic ratios in RNA-sequencing, only heart and adipose tissue were left for analysis. This may be due to lower expression of ACE in the other tissues when compared to adipose and heart.
The allelic RNA ratio is expressed as a ratio of major/minor alleles averaged from a survey of two measurements. Three standard deviations of the genomic DNA control (>1.2-fold AEI) indicate a biologically significant allelic expression imbalance. Though there is significant variation in some samples, two non-Caucasian individuals (MB72 and MB80) show significant AEI of ~1.73-fold. ACE I/D carrier status is heterozygous unless otherwise indicated. No I/D association with AEI was observed. MB55, indicated by *, is a carrier of the enhancer SNPs and shows no significant AEI. Data are mean±SD.

Figure 23. Allelic mRNA ratios in Prefrontal Cortex.
Figure 24. Allelic mRNA ratios in human liver.

The allelic RNA ratio is expressed as a ratio of major/minor alleles averaged from a survey of two measurements. Three standard deviations of the genomic DNA control (>1.15-fold AEI) indicate a biologically significant allelic expression imbalance.

Caucasian samples L66, L76 and L89 show average AEI of 1.83-fold. ACE I/D carrier status is heterozygous unless otherwise indicated. No I/D association with AEI was observed. Data are mean±SD.
Allelic mRNA ratios were collected and analyzed as described in Materials and Methods. Corrected allelic mRNA ratios are grouped by sample in Figure 25. Averaging all absolute allelic mRNA ratios indicates select adipose tissues (3) with overall AEI of ~1.35-fold. The I/D polymorphism was not assessed in these samples due to lack of gDNA. Still, this indicates presence of a functional polymorphism acting on ACE expression in adipose.

4.4 Discussion

The results of this survey of allelic mRNA ratios at ACE in non-cardiac tissue indicate that the previously identified enhancer SNPs only have function in the heart. It also demonstrates that moderate allelic expression imbalance is present in putamen, liver and adipose tissue. Further, the commonly studied ACE I/D polymorphism has no association with instances of AEI. This demonstrates that ACE is under influence from additional regulatory polymorphisms with tissue specific function.

The tissues in this study were selected based primarily on their availability. Prefrontal cortex, which does show low AEI in a single sample, is a poor expresser of ACE. However, even at a low level of gene expression the lack of enhancer SNP function can still be observed. Interestingly, the putamen has ties to Alzheimer’s amyloid plaque associated memory loss (de Jong et. al., 2008). Due to the role of ACE in breaking down amyloid plaque protein, this is an excellent target tissue (though not a direct application to cardiovascular drug metabolism) (Hu, 2001; Ohrui, 2004). Here, non-
Caucasian (Hispanic and African ancestry) samples exhibit AEI of ~1.73-fold. It is not clear what a modest change in gene expression might do to ACE activity in the putamen, or if it would affect memory loss or Alzheimer’s progression.

Interestingly, $ACE$ also showed significant allelic mRNA expression in human liver. To some extent $ACE$ in the liver is affected by $ACE$ inhibitors, as they are activated here by carboxylesterases (Thomsen, 2014). Moreover, the AEI identified here is in Caucasian samples, a subset of individuals where AEI is not seen in the heart. Though functional polymorphisms may be acting in this population in the liver it is not clear their relevance to drug treatments or other disease phenotypes.

Using RNA-sequencing data to screen multiple allelic mRNA ratios at SNPs in $ACE$ allows rapid identification of AEI. $ACE$ plays a role in cardiovascular risk in adipose, where an RAS is active. It has also been shown that $ACE$ in adipose may regulate body weight (Engeli et. al., 2005; Fleming, 2006). Following, analysis 3 samples showed an average of ~1.35-fold AEI. Importantly, in these samples standard deviation is relatively high, which may be a product of the IUPAC alignment correction, or the modest 30-read depth coverage. Additional follow up including targeted allelic expression assays or RNA-sequencing could further confirm the presence of AEI in adipose tissue. Though some heart was included in this RNA-sequencing analysis no striking allelic mRNA ratios were observed.

This study presents multiple tissues with evidence of functional cis-acting regulatory polymorphisms acting on $ACE$. It is not clear the role this AEI and associated variants might have in drug treatment or human disease progress. I also show a lack of
function for the *ACE* enhancer SNPs, which is expected because these SNPs exist in a MEF2A cardiac transcription factor binding site. Notably, the *ACE* I/D variant does not show any association with allelic mRNA ratios, which provides evidence that it is not functional and may associate with a *trans*-factor (if any at all). These results indicate that targeted studies to identify functional genetic variants affecting *ACE* should be completed in alternative (non-cardiac) tissues and diverse populations.
Figure 25. Allelic mRNA ratios determined by RNA-Sequencing Data.

Allelic mRNA ratios that have been corrected for lack of IUPAC correction are presented here. Bars indicate observed allelic mRNA ratios at a SNP, while groupings indicate all the SNPs in a particular sample. In some cases significant variability between intra-sample SNPs is observed. Average allelic mRNA ratios between all SNPs indicate that at least 3-samples may exhibit an average of ~1.35-fold allelic expression balance in adipose tissue (Data summarized in Appendix B.4).
Chapter 5. Conclusions and Final Thoughts

The work herein illustrates the complexity of regulatory influence on the expression and activity of candidate genes involved in the metabolism of cardiovascular drug metabolism in the human liver (and other relevant tissues). This regulatory control is presented as a promoter variant increasing gene transcription in cytochrome p450 2C19 (CYP2C19). In carboxylesterase-1A1 (CES1A1), decreases in gene expression and protein quantity are regulated by a polymorphic 5′UTR which significantly alters mRNA folding. Additionally regulatory variation is seen in the ubiquitously expressed angiotensin-I converting enzyme (ACE) which exhibits moderate allelic mRNA ratios in a wide array of tissues. Further, a known group of enhancer SNPs show tissue specific function in the heart. Though, functional ACE SNPs have not been identified outside the heart, it is clear that they are present in both a population and tissue specific manner. In the same way 90% of gene loci undergo alternative splicing (Wang et. al., 2008), the majority likely experience some degree of regulatory variation. In fact, alternative splicing and regulatory variants are linked by srSNPs, which may drive the presence or abundance of alternative splice forms. Further demonstrated here, robust allelic phenotypes, though relevant (especially in drug metabolism) are less common than regulatory variants with moderate effects. This work presents evidence for wide-spread regulatory and structural variation in DME genes. Though this may not significantly
affect clinical drug efficacy, it does contribute to the cadre of genetic variation that drives the drug metabolism interindividual variability.

A person’s metabolic capabilities are often irrelevant until they are prescribed a drug. As a system, drug metabolism is redundant, with multiple enzymes that have overlapping substrate specificity. Additionally, throughout an individual’s life, their drug metabolizing ability varies due to environment, gender, age, race and more. As a result this creates a population with a diverse spectrum of drug metabolizing ability which contributes to a dynamic response to new environmental compounds. This diversity is strengthened by the polymorphic nature of drug metabolism enzyme gene loci, which contributes to metabolic variation. This abundant genetic variation and the need to predict individual drug response drives pharmacogenetic studies.

As described, drug metabolism is subject to influence from non-genetic and genetic factors. Transcription factors (TFs) such as the central regulator of liver gene expression *hepatocyte nuclear factor-alpha* (*HNF4α*) shows little evidence of regulatory variants by allelic expression imbalance (AEI). Nuclear receptors (NRs), a class of TFs which induce drug metabolizing enzyme (DME) expression when binding certain ligands, have been shown to experience extensive alternative splicing(Lamba et. al., 2005). Though *cis* -acting variants may be present in these genes, large effect sizes are needed to observe significant changes in DME ability. Taken together, these *trans*-elements, (plus age, race and gender), are mechanisms for exerting environmental and genetic variation on drug metabolism. Understanding these different factors is a challenge pharmacogeneticists needs to address to predict drug response. Unlike some
pharmacogenetic applications, where a drug/compound and its target are the subject of interest, the huge number of variables involved in drug metabolism can make identification of clinically actionable biomarkers difficult.

This brings the field to an impasse. As “common” pharmacogenetics biomarkers with strong effect sizes (regulatory or coding variant sites) have been discovered, new biomarkers with large effect become more difficult to identify. Certainly there is merit in identification of these variants as predictors of metabolic ability, but as demonstrated here, moderate regulatory effects are also common in critical drug metabolizing enzymes (DMEs). Does a modest effect size invalidate a cis-acting polymorphism as a pharmacogenetic biomarker? I would venture that it does not. It does indicate that targeted studies are required to determine clinical relevance of these variants, particularly in relationship to other known functional SNPs. This approach will be essential as more functional SNPs are identified. Pharmacogeneticists must examine our ability to assess the true contribution of a variant (particularly of modest effect) to drug metabolizing ability.

An important application of these biomarkers is the manner in which their contributions to a phenotype are considered. Typically, polygenic models “add” the effects of identified variants to generate risk predictions in development of a phenotype. A model polygenic trait is height, where multiple gene loci have been shown to contribute to an increase height in an additive method (Weedon et. al., 2008). Alternatively, epistatic interactions are beginning to be considered when studying the effects of multiple SNPs. Here, variants are assessed in the context of one another, and
may have differing function in the presence or absence of a variant(s). The elucidation of SNP-SNP epistatic interactions is a large computational and statistical task. However, focused studies may be able to utilize previously characterized SNPs in disease or metabolic pathways, and assess any epistatic interactions present. A previous study by our lab showed an epistatic interaction between SNPs in dopamine receptor D2 (DRD2) and dopamine transporter (DAT). Here, homozygotes for a non-functional DAT variable number tandem repeat (VNTR, rs3836790) augmented the odds ratio for risk of cocaine-related death in the presence of the functional DRD2 srSNP rs2283265 to an odds ratio of 7.5 (Sullivan et. al., 2013). This study demonstrates that targeted studies can identify epistatic interactions, and that even non-functional variants can carry significant risk in a particular genetic context.

The same can be predicted for drug metabolism, though these studies are limited. In yeast, metabolic gene deletion was used to broadly define epistatic interaction within yeast metabolism. Importantly, they conclude each phenotype contains its own specific set of epistatic interactions. They also highlight the complexity of metabolic networks, stating that apparently unrelated metabolic processes may affect a phenotype in an epistatic and nonlinear fashion (Snitkin et. al., 2011). As such this highlights limitations on current pharmacogenetic studies. If we accept that complex genetic interactions with possible nonlinear effects on drug metabolism are prevalent, then assessment of a few well-defined SNPs in the context of drug metabolism may be inadequate. If SNP-SNP interactions are pervasive, then all SNPs regardless of effect size may be clinically relevant in a genetic or substrate specific context.
Twenty genotypes were assessed for their ability to predict development of type 2 diabetes when compared to, or in addition to phenotype based risk models. Using cumulative allele counts or odds ratios, genetic prediction was considered inferior to phenotype based models. Further, when combined with these models, no significant increase in risk prediction was observed (Talmud et. al., 2010). Though perhaps genetic prediction is not effective here, any present SNP-SNP interactions would be represented to some extent in phenotype risk models while an additive genetic model would not. Importantly, fields such as cancer are taking an epistatic approach to both predict disease risk and drug response (Chu et. al., 2014; Weigelt et. al., 2014).

Epistatic interactions could also have a role in studies adjacent to this body of work. \textit{CYP2C19} has been the primary focus of study in clopidogrel metabolism because it is responsible for the majority of prodrug activation. Notably, clopidogrel’s main two enzymes; CYP2C19 (drug activation) and CES1A1 (drug inactivation) have numerous variants that have not been tested for epistatic interaction. Additionally, of the 8 enzymes involved in clopidogrel metabolism, 6 harbor variants that are moderately associated with a drug’s efficacy or toxicity (not necessarily clopidogrel). Here the large numbers of involved enzymes and variants have not been assessed for epistatic interaction. The same holds true for the enzymes (and SNPs) involved in many therapeutic interventions.

Eventually, the SNP-SNP approach will be refined; allowing accurate prediction of patient metabolic ability. Ultimately, polymorphisms with modest effect size may find clinical relevance in these studies. Still, these analyses will be limited by an inability to account for non-genetic sources of metabolic variation. However, a shift towards
epistatic interactions should address missing heritability–unaccounted genetic contributions to a phenotype. In the case of *CYP2C19*2, only ~12% of variation of clopidogrel response is accounted for by the loss of function allele (Shuldiner, 2009). If a clearly functional polymorphism has a small overall effect as seen here, epistatic interactions have potential to increase the predictive power of genetic biomarkers.

As more functional SNPs in drug metabolism are characterized, scientists and clinicians will seek to identify these interactions. Databases such as PharmGKB or The Human Cytochrome P450 Allele Nomenclature Database, indicate the large number of variants defined (cSNPs, srSNPs, and rSNPs), in addition to current evidence for clinical implementation (www.pharmgkb.org, www.cypalleles.ki.se) (Whirl-Carrillo, 2012). Further, considering these polymorphisms in addition to the SNPs with large enough effect sizes to prompt FDA black box warnings, SNP-SNP interactions predicting drug response seem likely. This will elevate pharmacogenetic studies, drug response prediction, and should be reflected in guidelines as defined by the Clinical Pharmacogenetics Implementation Consortium (CPIC) or other similar groups.

This study utilizes cardiovascular drug metabolism as a representation of how multiple genes that interact at the point of treatment each have their own complex molecular genetic mechanisms. Though studied individually here, they represent a group of genes where SNP-SNP interactions may exist given the proper substrate (clopidogrel, ACE inhibitors). Here, I define *CYP2C19*17 in human liver, and the functional effect of *CES1A1VAR*. I also present evidence for *cis*-acting regulatory SNPs acting on *ACE* in multiple races and tissues (with lack of ACE enhancer SNP function in heart).
CESIA1VAR has potential for utility as a pharmacogenetic biomarker, and the characterization of CYP2C19*17 should improve classification as a biomarker. The evidence for cis-acting functional regulatory variants in multiple drug metabolizing enzyme genes (as seen here), is not only relevant to cardiovascular drugs but also a wide array of therapeutic interventions. Though variants observed in this study have a modest effect, they may be relevant clinically in certain treatments, or when considered as part of a SNP-SNP interaction. In the future, as genomic data is gathered, functional molecular genetic studies will provide rationale in the inclusion of variants into accurate metabolism prediction algorithms.
References


Darwin C. The variation of animals and plants under domestication. 2d ed. New York,: D. Appleton and company; 1890.

Das M, Pal S, Ghosh A. Synergistic effects of ACE (I/D) and Apo E (Hha I) gene polymorphisms on obesity, fat mass, and blood glucose level among the adult Asian Indians: A population-based study from Calcutta, India. Indian J


Furukawa M, Nishimura M, Ogino D, Chiba R, Ikai I, Ueda N, Naito S, Kuribayashi S, Moustafa MA, Uchida T, Sawada H, Kamataki T, Funae Y, Fukumoto M. Cytochrome p450 gene expression levels in peripheral blood mononuclear cells in


Lamba J, Lamba V, Schuetz E. Genetic variants of PXR (NR1I2) and CAR (NR1I3) and their implications in drug metabolism and pharmacogenetics. Curr Drug Metab. 2005;6(4):369-83.


Snyder LH. Linkage in Man. 1931b.


Appendix A: Tables

Table 3. Primer Sequences and PCR conditions.

Sequencing, quantitation, genotyping and SNAPshot primers are included. Primers are presented with their use, name, sequence and run conditions. Run conditions are annotated as (Annealing temp)PCR(HS=Hotstart)(# of cycles if >30). Extension times are 1 minute or default for the Taq specified.

<table>
<thead>
<tr>
<th>CYP2C19 Primer List</th>
</tr>
</thead>
</table>

1. CYP2C19 Promoter and Exonic regions for Ion Torrent Sequencing

Promoter ~-4KB to -2KB
63.5PCRHS, 2 min 30 sec extension
2C192kbTL1F1, GACCGGAGCTGTTCGTATTTGGC
2C19P2kbTL1R1, TTGACTAGATTGGGGTCAGAAGAGTTTG

Promoter -2KB to start of Exon 1
63.5PCRHS, 2 min 30 sec extension
2C19P2kbTL2F2, CATTGTGATACTTTGTCTCACTGAGTCA
2C19P2kbTL2R2, CTACATTGGTTAAGGATTTGCTGACA

***Promoter -3KB to Exon 1
63.5PCRHS_35
2C19_3kbP_to_Ex1_F, TTATTTGTGCTAGGGCTCGTG
2C19_3kbP_to_Ex1_R, CTACGTGTTACCCCTCAGCC

Continued
Table 3 continued

***Exon 2 to Exon 3
60PCRHS
2C19_Ex2_to_Ex3_F, AGGTAGACACAAGAGTGCTGA, 0.025, desalt
2C19_Ex2_to_Ex3_R, TTCTCTGGTGACATGTCTTGGA, 0.025, desalt

***Exon 4 to Exon 5
60PCRHS_35
2C19_Ex4_to_Ex5_F, CCATTATTTAACCAGCTAGGC, 0.025, desalt
2C19_Ex4_to_Ex5_R, TCCTATCCTGACATCCTTATTG, 0.025, desalt

Exon 6
60PCRHS
2C19 I5F1, GACAAACCCACAGCCAATATCATACT
2C19 I6R1, GGGACAGATTACAGCTGCGG

Exon 7
60PCRHS
2C19_Ex7_F, AATGCTGAAGTGGGTTGTTG, 0.025, desalt
2C19_Ex7_R, ACCCTGACAGAAATTCTAGCCC, 0.025, desalt

Exon 8
64.1PCRHS, requires gel purification of ~951 bp band
CYP2C19I7F1, GTGCTGCAACCATTAAACTCATC
CYP2C19I8R1, TTTCCAAACACAGAAGTGAGCCTC

Exon 9
59.3PCRHS
CYP2C19I8F1, CCATCCATTCAATCCATTAATCCT
CYP2C193UPR1, CATTATGTGGCACTCAATGTAACTATTAT

*** From Blaisdell J, et al.
### Table 3 continued

2. Quantitation Primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA4QF</td>
<td>AGCCTGGCCCTGTCCATCTCACTA</td>
<td>GATA4QR, GACATCGCACTGACTGAGAAGTG</td>
</tr>
<tr>
<td>CARquan-F</td>
<td>CACATGGGACACCATGTGGTTA</td>
<td>CARquan-R, AAGGGCTGTGGATGGATGAA</td>
</tr>
<tr>
<td>CARquan-R</td>
<td>GAGCCCGCAGCGTGGCAAGG</td>
<td>RXRA-R, GGCAATGTTGGTGGACAGG</td>
</tr>
<tr>
<td>RXRA-F</td>
<td>GAGCGGCAGCGTGGCAAGG</td>
<td>RXRA-R, GGCAATGTTGGTGGACAGG</td>
</tr>
<tr>
<td>RXRA-R</td>
<td>AAAGGGCTGTGGATGGATGAA</td>
<td>PXRquan-F, CAAACGGGAGAAAGTGAACG</td>
</tr>
<tr>
<td>PXRquan-R</td>
<td>CAAACGGGAGAAAGTGAACG</td>
<td>PXRquan-R, CACAGATCTTTCCGGACCTG</td>
</tr>
<tr>
<td>HNF4A1A-F</td>
<td>ACATGGACATGGCGACTAC</td>
<td>HNF4A1A-R, CTCGAGGCACCGTAGGTGTT</td>
</tr>
<tr>
<td>HNF4A1A-R</td>
<td>CTCGAGGCACCGTAGGTGTT1</td>
<td>GAPDHquan-F, ACTCCTTCACCTTTGACCGT</td>
</tr>
<tr>
<td>GAPDHquan-R1</td>
<td>GGTCCACCACCTGTTTC</td>
<td>GAPDHquan-R1, GGTCCACCACCTGTTTC</td>
</tr>
<tr>
<td>CYP2C19_E8F2</td>
<td>GTCACTTTTCGGATGAGGTGTA</td>
<td>CYP2C19_E9_R, AGGTCCTTTGGGATCAATCAGAG</td>
</tr>
<tr>
<td>CYP2C19_E9_R</td>
<td>AGGTCCTTTGGGATCAATCAGAG</td>
<td>CYP2C19_E9_R, AGGTCCTTTGGGATCAATCAGAG</td>
</tr>
</tbody>
</table>

3. SNaPshot/Genotype Primers

**rs12248560 (CYP2C19*17)**

<table>
<thead>
<tr>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>60PCRHS</td>
<td>2C19 Promoter F1, GCCCTTAGACCAACATTCTCTG</td>
</tr>
<tr>
<td>2C19 Promoter R1, AACACCTTTACCATTAAACCCC</td>
<td>CYP2C19_rs560_PEF, CAAATTTGTGCTTCTGTTCTCAAG</td>
</tr>
</tbody>
</table>

**rs4244285 (CYP2C19*2)**

<table>
<thead>
<tr>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>60PCR</td>
<td>CYP2C19int4F, ATCTTATATTTCAAGATTGTAGAGAAGAATTGTTTG</td>
</tr>
<tr>
<td>CYP2C19int5R, CATCCGAGFTAAACCAAAACTAGTCAATG</td>
<td>CYP2C19_rs285PEFT10, T(10)CCCCTATCATGTATTATTTCCC</td>
</tr>
</tbody>
</table>
Table 3 continued

<table>
<thead>
<tr>
<th>rs17885098 (marker SNP)</th>
<th>60PCRHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2C19 E1DNA-R1, CTTACATTGGTTAAGGATTTGCTGACA</td>
<td></td>
</tr>
<tr>
<td>2C19 E1RNA-R1, AGAGATTTGGTTAAGGATTTGCTGACA</td>
<td></td>
</tr>
<tr>
<td>2C10 rs098PER2, ATTTCCAATCACTGGGAGGGAGT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rs4986894(-98T&gt;C)</th>
<th>60PCRHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Primer, [6FAMTCATTGGAAACCTTGTTAAGGT]</td>
<td></td>
</tr>
<tr>
<td>R Primer, CCTACATTGGTTAAGGATTTGCTGACA</td>
<td></td>
</tr>
<tr>
<td>Cut with BtsCI at 50°C in Buffer 4 (New England Biosciences, Ipswich MA)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rs4986893(2C19*)</th>
<th>60PCRHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Primer, GAATGAAAAACATCAAGATTGTAACGT</td>
<td></td>
</tr>
<tr>
<td>R Primer, GGATTTGCCAGAAAAAAAGACTGT</td>
<td></td>
</tr>
<tr>
<td>Primer Extension R, CAAAAACTTGGCCTTACCTGG</td>
<td></td>
</tr>
<tr>
<td>Tm: 60</td>
<td></td>
</tr>
<tr>
<td>Tm: 58</td>
<td></td>
</tr>
<tr>
<td>Tm: 60.5</td>
<td></td>
</tr>
</tbody>
</table>

4. Sanger sequencing primers for promoter region

2C19P7F1, CGGAGCTGTTCCGATTTGGC
2C19P7F2, AATAAGATGACAAGACAGACACTGGGA
2C19P7F3, AGAAATGAAATCAAGGAAAGCATGAAAGA
2C19P7F4, CATATTGCTTTCGCTGAGTCA
2C19P7F5, CCTCATTCCTTGAGATGGGTCAAT
2C19P7F6, GTGGTCTATTTAATGGAAGCCCTGT
2C19P7F7, TTAGCTATTTTCATGTTAGGCTGCTGTAT

CES1 Primer List

1. Genotyping and allelic mRNA expression assays

60PCRHS

CES1A1 gDNA amplification

CES1A1_gDNA_F2, AGTAGTGAGTTCTCTGACACCCATG
CES1A1_gDNA_R2, GTAAGGGACTAATTAAGGGACACGC

Continued
Table 3 continued

**CES1A1 mRNA amplification for SNaPshot**

60PCRHS
CES1A1Qmedalt_F, CAGAGACCTCGCAGGCC
CES1A1Qshort_R, CAAGCCGCGGAAGCAG

*Primer extension primers for preceding allelic mRNA ratios and gDNA reactions*
CES1A1 rs12149370 PER, TGGAAAGGCAGCACGTTCC
CES1A1 rs3815583 PER, TGCTGTCCAGCCCTGG

**CES1 rs2244613 genotyping**
60PCRHS
CES1A1_rs2244613_FAM_F,(6FAM)AGACAGCCATGTCACTCCTGTG
CES1A1_rs2244613_R, GGACACCAAACATCACATCTGCT
Digest with SmaI in NEB CutSmart Buffer (Cuts at minor allele 'C').

**CES1A1VAR genotyping**
7500 Fast, SYBR Default Cycling
Fukami_1A1_FR_F, CAAGGGTGAACCCTTATGTA, 0.025, desalt and
CES1A1WT
Fukami_1A1_ex1AS, AGAGAGTGGCCAGGATAAAG
or
CES1A1VAR
Fukami_1A2_3_ex1AS, CGAGAGTGGCCAGGACAAGA
Genotype by presence or absence of product in parallel reactions.
From Fukami et al. 2008.

2. mRNA quantitation and notable sequencing primers

**CES1 mRNA quantitation**
7500 Fast, SYBR Default Cycling
CES1_RNA_Q_F, GTTCCGCGGCTTGG
CES1_RNA_Q_R, CGGCTTTGGCAAAAAGGGATT

*See GAPDH quant primers from CYP2C19*
**Table 3 continued**

5kb promoter region amplification

61.8_PrimeStarTaq

See CES1A1_gDNA_F2

CES1A1_5K_F, ATGACGGCGATCAAGGCT

**Infusion Cloning Primers for insertion into pGL3B**

59.8_ADVHD Taq

CES1A1_Inf2_HIII_F,

CGATCTAAAGTAAGCTTTGCTCTTTGTGTAACAAGCTTTTGTG

CES1A1_Inf3_HIII_R, CCGGAATGCAAGCTTTCAAGCGCGGAAGCA

**Promoter Sanger Sequencing Primer**

CES1A1Qlong_F, AGGCAGAGCGCCTAAGCT

---

**ACE Primer List**

1. Genotyping and allelic mRNA primers

*ACE (I/D, rs13447447)*

60PCR

ACE_ID_R_FAM, (6FAM) GTGGCCATCACATTCGTCAG

ACE_ID_common_F, CCCATCCTTTTCTCCCATTT

ACE_ID_specific_F, GACCTCGTGATCCGCCC

Insertion = peaks at 191 bp and 462

Deletion = 175 bp peak

rs4343

7500 Fast, SYBR Default Cycling

ACE_rs4343_gcG_F, CTGACGAATGTGATGGCAAGG

ACE_rs4343gc_gcA_F, CGCGCCCGCCGCAGCTGACGAATGTGATGGCAAGG

ACE_rs4343_gc_R, GATGGCTCTCCCCGCC

rs4343 allelic mRNA ratio

60PCR

ACEI_rs4343_DNA_F, CCCTTACAAGGAGAGCTAAG

ACEI_rs4343_RNA_F, ACCACCTACAGCTGAC

ACEI_rs4343_R, CATGCCCATAACAGGTCTTCTATT

ACEI_rs4343_PEF, GACGAATGTGATGGCC

---

Continued
Table 3 continued

| rs7214530          | 7500 Fast, SYBR Default Cycling |
|--------------------|---------------------------------
| ACE_rs7214530_F1   | CATTTCTGGACTCTCAATTCTATTCCA     |
| ACE_rs7214530_R_A  | CACCAGTTACCACAGGAGAGAAGAAA     |
| ACE_rs7214530gcR_C | GGCGCGGGCGCGCGGCAACCAGTTACCACAGGAGAGAC |

Use half concentration of allele A compared to allele C

<table>
<thead>
<tr>
<th>rs4290</th>
<th>60PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE_rs4290_RFLP_F</td>
<td>TCTTATATTATTTCCTGCAAAAGGCTG</td>
</tr>
<tr>
<td>ACE_rs4290_RFLP_FAM_R</td>
<td>(6FAM)CACCTCCCTCCTGTAACGCTT</td>
</tr>
</tbody>
</table>

Digest with Taq1α in NEB CutSmart Buffer, cuts with major ‘C’ allele

<table>
<thead>
<tr>
<th>rs7213516</th>
<th>60PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE_rs7213516_RFLP_FAM_F</td>
<td>(6FAM)TGCACCCAGCCCCAAAT</td>
</tr>
<tr>
<td>ACE_rs7213516_RFLP_R</td>
<td>CATCTCTCTCCAGGAAGTACAGGT</td>
</tr>
</tbody>
</table>

268 BP fragment, Digest with HinF1 in NEB CutSmart Buffer
A allele cuts at 163 bp
G allele cuts at 175 bp

2. Quantitation Primers

<table>
<thead>
<tr>
<th>PGK1F,CTGTGGTCTCTGAAAGCAGCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGK1R,TTAGCCCGAGTGACAGCCTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACE_Q_F,CCCCTTCCCGCTACAACCTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE_Q_R,TCCCTGATACTTTGGTTCGAA</td>
</tr>
</tbody>
</table>
Table 4. CYP2C19 SNP genotypes and Hardy-Weinberg Equilibrium.

Varying allele frequencies between populations may cause lack of rs17885098 Hardy-Weinberg Equilibrium. (MAF = 0.083 in Caucasians, MAF = 0.180 in Africans).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor Allele</th>
<th>Minor Allele Frequency</th>
<th>Major Allele</th>
<th>Major Allele Frequency</th>
<th>HWE-P</th>
<th>LOG10 HWE P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17885098</td>
<td>G</td>
<td>0.128</td>
<td>A</td>
<td>0.872</td>
<td>2.20E-08</td>
<td>7.658</td>
</tr>
<tr>
<td>rs4244285</td>
<td>A</td>
<td>0.174</td>
<td>G</td>
<td>0.826</td>
<td>0.576</td>
<td>0.239</td>
</tr>
<tr>
<td>rs4986894</td>
<td>C</td>
<td>0.149</td>
<td>T</td>
<td>0.851</td>
<td>0.72</td>
<td>0.143</td>
</tr>
<tr>
<td>rs12248560</td>
<td>T</td>
<td>0.26</td>
<td>C</td>
<td>0.74</td>
<td>0.856</td>
<td>0.068</td>
</tr>
</tbody>
</table>
Table 5. CES1A1 SNPs genotypes and Hardy-Weinberg Equilibrium

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor Allele</th>
<th>Minor Allele Frequency</th>
<th>Major Allele</th>
<th>Major Allele Frequency</th>
<th>HWE-P</th>
<th>LOG10 HWE P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3815583</td>
<td>C</td>
<td>0.345</td>
<td>A</td>
<td>0.655</td>
<td>0.795</td>
<td>0.010</td>
</tr>
<tr>
<td>rs2244613</td>
<td>B</td>
<td>0.176</td>
<td>A</td>
<td>0.824</td>
<td>0.215</td>
<td>0.670</td>
</tr>
<tr>
<td>rs71647871</td>
<td>C</td>
<td>0.221</td>
<td>A</td>
<td>0.779</td>
<td>0.184</td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>0.017</td>
<td>G</td>
<td>0.983</td>
<td>0.850</td>
<td>0.071</td>
</tr>
</tbody>
</table>
rs4343 was genotyped and enhancer (rs4290) and I/D variants were genotyped on a by sample basis in rs4343 carriers.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Minor Allele</th>
<th>Minor Allele Frequency</th>
<th>Major Allele</th>
<th>Major Allele Frequency</th>
<th>HWE P</th>
<th>-log10 HWE P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4343</td>
<td>A</td>
<td>0.424</td>
<td>G</td>
<td>0.576</td>
<td>0.289</td>
<td>0.539</td>
</tr>
</tbody>
</table>
Table 7. CES1A1-pGL3-Basic construct variations.

CES1A1VAR and CES1A1SVAR primary SNPs are represented in Table X in the text. Additional SNPs are present, with rs3859102 and rs334428341 minor alleles present in all constructs. Allele rs28759040 is present in CES1A1WT and CES1A1WT+rs3815583 which show no significant difference in luciferase assay, indicating that likely non-functional.

<table>
<thead>
<tr>
<th></th>
<th>SNP Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES1A1WT</td>
<td>rs3859102</td>
</tr>
<tr>
<td></td>
<td>rs334428341</td>
</tr>
<tr>
<td></td>
<td>rs28759040</td>
</tr>
<tr>
<td>CES1A1VAR</td>
<td>rs3859102</td>
</tr>
<tr>
<td></td>
<td>rs28779040</td>
</tr>
<tr>
<td></td>
<td>Table 2 SNPs:1-11</td>
</tr>
<tr>
<td>CES1A1SVAR</td>
<td>rs3859102</td>
</tr>
<tr>
<td></td>
<td>rs334428341</td>
</tr>
<tr>
<td></td>
<td>rs28759040</td>
</tr>
<tr>
<td></td>
<td>Table 2 SNPs:1-5</td>
</tr>
<tr>
<td>CES1A1WT+rs3815583</td>
<td>rs3859102</td>
</tr>
<tr>
<td></td>
<td>rs28779040</td>
</tr>
<tr>
<td></td>
<td>rs3815583</td>
</tr>
</tbody>
</table>
Table 8. Average ACE-fold AEI in RNA-sequencing.

Absolute fold-change is computed and corrected for the lack of IUPAC alignment, which when not performed causes an allelic bias in the number of minor alleles mapped to the genome. Average fold is presented along with standard deviation to represent the overall allelic mRNA ratio at n SNPs in ACE.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Fold-AEI</th>
<th>Standard Deviation</th>
<th># of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart 1</td>
<td>1.122</td>
<td>0.318</td>
<td>5</td>
</tr>
<tr>
<td>Heart 2</td>
<td>1.094</td>
<td>0.433</td>
<td>2</td>
</tr>
<tr>
<td>Heart 3</td>
<td>0.871</td>
<td>0.073</td>
<td>3</td>
</tr>
<tr>
<td>Heart 4</td>
<td>1.225</td>
<td>0.247</td>
<td>2</td>
</tr>
<tr>
<td>Heart 5</td>
<td>0.975</td>
<td>0.331</td>
<td>2</td>
</tr>
<tr>
<td>Heart 6</td>
<td>1.211</td>
<td>0.612</td>
<td>4</td>
</tr>
<tr>
<td>Heart 7</td>
<td>1.235</td>
<td>0.233</td>
<td>2</td>
</tr>
<tr>
<td>Adipose 1</td>
<td>0.925</td>
<td>0.175</td>
<td>3</td>
</tr>
<tr>
<td>Adipose 2</td>
<td>0.976</td>
<td>0.157</td>
<td>5</td>
</tr>
<tr>
<td>Adipose 3</td>
<td>0.914</td>
<td>0.057</td>
<td>3</td>
</tr>
<tr>
<td>Adipose 4</td>
<td>1.378</td>
<td>0.364</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 5</td>
<td>0.990</td>
<td>0.198</td>
<td>5</td>
</tr>
<tr>
<td>Adipose 6</td>
<td>0.969</td>
<td>0.076</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 7</td>
<td>0.933</td>
<td>0.245</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 8</td>
<td>0.862</td>
<td>0.079</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 9</td>
<td>1.324</td>
<td>0.273</td>
<td>2</td>
</tr>
<tr>
<td>Adipose 10</td>
<td>0.813</td>
<td>0.112</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 11</td>
<td>0.939</td>
<td>0.240</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 12</td>
<td>0.907</td>
<td>0.082</td>
<td>4</td>
</tr>
<tr>
<td>Adipose 13</td>
<td>1.342</td>
<td>0.303</td>
<td>3</td>
</tr>
<tr>
<td>Adipose 14</td>
<td>1.078</td>
<td>0.310</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix B: Figures

Figure 26. Linkage Disequilibrium of CYP2C19 alleles.

Included *17(rs12248560), *2(rs4244285), marker SNP (rs17885098), rs4986894 and *3 (rs4986893) where data was available. LD expressed as a measure of $r^2$. *2 and rs4986894 are in high LD as expected. Our marker SNP and *3 are also in high LD however no *3 carriers were detected in this study. Linkage plots generated in Haploview using data from 1K Genomes Project.
Figure 27. Linkage of relevant polymorphisms in CES1A1. Polymorphisms included are CES1A1VAR, rs3815583, rs71647871 and rs2244613. LD is expressed as a measure of $r^2$. Linkage plots are developed from genotyping data and generated in Helix Tree SNP and Variation Suite.
Figure 28. CES1 protein expression in human liver tissue.

CES1 protein quantity was measured by normalization to input of L100 (arbitrary unit, Sample #32). Quantitation was performed in duplicate and averages were used for analysis.
Figure 29. Relative Gene Expression of ACE in Prefrontal Cortex and Putamen. ACE mRNA expression normalized to PGK1 housekeeping gene in human prefrontal cortex and putamen. In many cases, prefrontal cortex samples showed no measurable gene expression (not pictured).
Appendix C: Abbreviations

ACE - angiotensin-I converting enzyme
ACE I/D - ACE insertion/deletion polymorphism
AEI - allelic expression imbalance
AHR - aryl hydrocarbon receptor
CAR - constitutive androstane receptor
cDNA- complimentary deoxyribonucleic acid
CES1A1- carboxylesterase-1A1
CES1P1 - carboxylesterase pseudogene 1 (also CES1A2/CES1A3)
CYP2C19 - cytochrome P450 2C19
DAT- dopamine transporter
DME - drug metabolizing enzyme(s)
DRD2- dopamine receptor D2
eQTL- expression quantitative trait loci
GATA4 - GATA binding protein 4
gDNA- genomic deoxyribonucleic acid
HGP- Human Genome Project