Alternative Splicing in the Cytochrome P450 Superfamily Expands Protein Diversity to Augment Gene Function and Redirect Human Drug Metabolism.

Andrew J. Annalora, Craig B. Marcus, Patrick L. Iversen

Department of Environmental and Molecular Toxicology, Oregon State University, 1007 Agriculture & Life Sciences Building, Corvallis, OR 97331; USA – AJA, CBM & PLI
Running Title: Alternative CYP Splicing Redirects Human Drug Metabolism.

Corresponding author and person to whom reprint requests should be addressed:
Andrew J. Annalora, Ph.D
Department of Environmental and Molecular Toxicology
Oregon State University
1007 Agriculture & Life Sciences Building
Corvallis, OR 97331; USA
Phone: 541-737-2706
Fax: 541-737-0497
e-mail: Andrew.Annalora@oregonstate.edu

Text Pages: 29
Number of Tables: 2
Figures: 4
References: 195

Word Count:
Abstract: 247
Introduction: 562
Discussion: 8,859

Abbreviations:
AR - androgen receptor
CYP - cytochrome P450
ESS - exonic splicing silencers
ESE - exonic splicing enhancers
GWAS - genome wide association studies
hnRNPs - heterologous nuclear ribonuclear proteins
hnRNPC - heterogeneous nuclear ribonucleoprotein C
ISS - intronic splicing silencers
ISE - intronic splicing enhancers
kDa - kiloDaltons
mRNA - messenger RNA
NIH - National Institutes of Health
NMD - nonsense mediated decay
P450 - cytochrome P450
PGx - pharmacogenomics
PPAR - peroxisome proliferator-activated receptors
pre-mRNA - precursor or unprocessed messenger RNA
PTC - premature termination codon
RNA - ribonucleic acid
RNAseq - RNA sequencing
RAR - retinoic acid receptor
RXR - retinoid X receptor
SNP - single nucleotide polymorphism
snRNA - small nuclear RNA
snRNP - small nucleolar ribonucleoprotein
SRF - splicing regulatory factor
VDR - vitamin D receptor
ABSTRACT

The human genome encodes 57 cytochrome P450 (CYP) genes whose enzyme products metabolize hundreds of drugs, thousands of xenobiotics and unknown numbers of endogenous compounds including steroids, retinoids and eicosanoids. Indeed, CYP genes are the first line of defense against daily environmental chemical challenges in a manner that parallels the immune system. Several NIH databases, including PubMed, AceView, and Ensembl, were queried to establish a comprehensive analysis of the full human CYP transcriptome. This review describes a remarkable diversification of the 57 human CYP genes, which may be alternatively processed into nearly 1000 distinct mRNA transcripts to shape an individual's CYP proteome. Important CYP splice variants from families: 1A, 1B, 2C, 2D, 3A, 4F, 19A and 24A have now been documented, with some displaying alternative subcellular distribution or catalytic function directly linked to a disease pathology. The expansion of CYP transcript diversity involves tissue-specific splicing factors, transformation-sensitive alternate splicing, trans-splicing between gene transcripts, single-nucleotide polymorphisms, and epigenetic regulation of alternate splicing. Homeostatic regulation of variant CYP expression also is influenced by nuclear receptor signaling, suppression of nonsense mediated decay or premature termination codons, mitochondrial dysfunction or host infection. This review focuses on emergent aspects of the adaptive gene splicing process, which when viewed through the lens of CYP-NR gene interactions, resembles a primitive immune-like system that can rapidly monitor, respond and diversify to acclimate to fluctuations in endo-xenobiotic exposure. Insights gained from this review should aid future drug discovery and improve therapeutic management of personalized drug regimens.
INTRODUCTION

The human genome sequence was completed in 2001 (Lander et al., 2001), helping to usher in the “post-genomic era” of personalized medicine. Over a decade later, however, the promise of pharmacogenomics (PGx) has yet to fully materialize, and the composition of the human genome remains enigmatic, as numerous questions linger concerning the complexity of its content and organization. For example, how can a genome with only 22,000 genes produce a proteome with over 200,000 distinct proteins? While transcription is a relatively well understood phenomenon, the mechanisms involved in regulating alternative pre-mRNA splicing, the driving force behind both transcriptome and proteome expansion, remains less appreciated. The ability to link a single gene to its full suite of RNA transcripts and polypeptide products will improve our ability to assess the functional nature of its role in both physiological and pathological conditions (La Cognata et al., 2014). The spectrum of alternate splicing mechanisms underlying the expansion of the proteome involves the use of seven main splicing types, as described (Blencowe, 2006; Roy et al., 2013). The molecular basis for these mechanisms are complex, and beyond the scope of this review, however a brief description of the phenomenon and its key players is provided in the Supplemental Materials (see Supplemental Figure 1). Here we will focus primarily on the spectrum of phenotypic outcomes induced by alternative transcript splicing, to highlight the array of splice-sensitive features operating in the cytochrome P450 (CYP or P450) superfamily, a collection of 57 human genes that coordinate the metabolism of both drugs and endo-xenobiotics.

This work focuses new attention on the biological impact of alternative CYP gene splicing, an underappreciated component of phase I drug metabolism that may complicate or disrupt personalized approaches to medicine. Precision medicine, therefore, through the universal application of genetics, still faces many challenges beyond cost, ethical considerations, and the need for additional, whole genome sequences. Multiple resources have now been developed, in addition to the Human Genome Project, to address these challenges, including: the Hapmap
Project, which enables single nucleotide polymorphism (SNP) arrays for >100,000 SNPs; the 1000 Genomes Project; the ENCODE project for non-coding RNAs; and the National Human Genome Research Institute (NHGRI) genome-wide association studies (GWAS) catalogue.

GWAS have now identified two CYPs that represent a biomarker for disease or a drug response phenotype, including: (i) A CYP2C8 association with bisphosphonate-related osteonecrosis of the jaw in multiple myeloma (Sarasquete et al., 2008); and (ii) A CYP2C19 association with clopidogrel inter-individual variation as an antiplatelet drug (Shuldiner et al., 2009). Aside from showing strong associations, few pharmacogenetic biomarkers of this sort have successfully transitioned from discovery to clinical practice (Carr et al., 2014). Many GWAS studies have failed to provide prognostic value because they are not designed to evaluate how variant gene expression is influenced by both cellular and environmental splicing factors. It is in this regard, that we feel a new appreciation for the complexity of alternative splicing is needed. For not only is the world of both coding and non-coding RNA more diverse and more complex than we could have imagined even a decade ago, it is teeming with untold amounts of RNA “dark matter” and other relics of the RNA world, whose structure and function must be fully appreciated and classified, before a new guiding orthodoxy for “individualized medicine” can be safely conceived (St. Laurent et al., 2014; Cowie et al., 2015).

META-ANALYSIS OF ALTERNATIVE GENE SPLICING IN THE CYTOCHROME P450 SUPERFAMILY

Alternative splicing regulated by tissue specific factors is a means of adapting to physiologic demands. Tissue-specific splice variants may arise from tissue-specific promoter elements (Wiemann et al., 2005) or tissue-specific splicing regulatory elements (Black, 2000). Exon and intron definition requires interaction of discrete cis-elements (ESE, ESS, ISS, and ISE) with tissue-specific, trans-acting, splicing regulatory factors (SRF). In this way the same pre-mRNA transcript can be processed into tissue-specific, alternately spliced forms (Wang and...
Burge, 2008). The cytochrome P450 superfamily of genes is well studied from many perspectives (Guengerich, 2013; Johnson and Stout, 2013; Pikuleva and Waterman, 2013). As of late 2015, Pubmed listed over 23,790 citations regarding human P450 genes (according to Genecards.org), with much of this work being focused on tissue-specific expression or catalytic function of individual, reference forms. Alternative splicing of P450s has also been well-studied for more than two decades, however, the structural and functional diversity of CYP splice variants, and their relationship to human health and disease, remains poorly understood. A summary of known CYP splice variants linked to human disease is presented in Table 1. Reports from groups attempting to synthesize a global view of alternative P450 splicing, and its relevance to our understanding of human metabolism in the context of PGx or personalized medicine, are exceedingly rare (Nelson et al., 2004; Turman et al., 2006).

Our review has led us to conclude that personalized approaches to medicine based solely on GWAS may be misleading, in that they fail to account for variability in compensatory splicing mechanisms, which can modulate the expression of some highly-penetrant mutations. This phenomenon is exemplified by ontogenetic regulatory factors that can mask the physiological impact of genetic variations among developmentally-regulated CYP genes (Hines and McCarver, 2002). To expand upon this concept, and explore the array of alternative splicing products that serve as an improved platform for predicting disease based on the functional genome, we annotated all of the known human CYP splice variants currently listed in the NCBI’s Pubmed, Aceview and Ensembl databases (Thierry-Mieg and Thierry-Mieg, 2006; Cunningham et al., 2015) as of late 2015. Based on this computational meta-analysis, we found the 57 human CYP genes produce a P450 “transcriptome cloud” composed of at least 965 unique mRNA transcripts (Table 2). This number represents all known reference (or wild-type) and variant CYP mRNAs including some with retained introns, premature termination codons (PTCs), and those subject to nonsense mediated decay (NMD); we also included several experimentally-derived transcripts described in Pubmed-listed citations (identified by searching CYP* or P450 splice variant or
alternative CYP* or P450 splicing), that were not represented in the Aceview or Ensembl databases.

Our results approximate that the average CYP gene encodes nearly 20 unique mRNA transcripts, capable of yielding both reference and splice variant proteins, and an ensemble of non-coding RNA molecules. Over 53% of these transcripts (515) are predicted to express viable proteins, based on the most stringent expression criteria utilized by both the Aceview and Ensembl databases (Table 2). Subsequent analysis of the GeneCards.org website revealed that the 57 human CYPs are associated with over 48,384 SNPs, approximating to roughly 850 mutations per CYP gene (Stelzer et al., 2011). While most CYPs fall within this general range of genetic variation, some, including CYP 2C19, 5A1, 19A1 and 39A1 express more than twice as many SNPs than average, suggesting the accumulation process is non-random (Table 2). The total number of diseases associated with each of the 57 human genes (1057) was also calculated using the Malacards.org website (Rappaport et al., 2014). CYP families 1 (203) and 2 (413) account for over half of the total disease associations, while orphan P450s CYP20A1 (1) and CYP39A1 (2) are associated with the fewest.

A side-by-side, radar plots comparison of the total number of alternative CYP splice variants and SNPs, for each of the 57 human CYP genes, is shown in Figure 1. Pharmacologically-relevant drug targets, like CYP2C9, CYP2D6, CYP3A5, CYP11A1 and CYP19A1, are among the most highly-spliced P450 genes, the majority of which are also subject to trans-splicing events. While differences in gene organization or cell-specific, selection pressure may play a role in this phenomenon, our observations suggest that drug-dosing guidelines derived from knowledge of a patient’s individual CYP transcriptome may be far superior to those based on genotyping alone, as SNP-based approaches may tend to overstate (or oversimplify) the biological influence of a point mutation on gene function. While computational studies indicate that 85% of all human mRNA comes from a single major transcript (Gonzales-Porta et al., 2013), the CYP superfamily of genes is highly diverse at the genomic level, making it uniquely-sensitive
to alternate splicing events that alter the biological output of both genetic and epigenetic signaling cascades. Nuclear receptors are now recognized for their role in mediating global gene splicing events (Auboeuf et al., 2005), and an improved framework for how this process functions to regulate drug metabolism appears to be in order, one that can potentially re-contextualize the utility and complexity of the convoluted metabolic cross-talk network that underlies the 48 NR signaling pathways operating in humans (Meyer, 2007; Tralau and Luch, 2013).

Furthermore, drug metabolizing P450s of CYP families 1-4, have traditionally been considered more polymorphic than CYPs that metabolize endogenous, cholesterol-based substrates, and are assumed to be more sensitive to SNP-based modulation of alternative splicing (Lewińska et al., 2013). Based on our analysis, it does appear that that the majority of steroid-metabolizing P450s, including CYP17A1 (91 SNPs) and CYP21A2 (173 SNPs), are among the least polymorphic P450 genes. As described above, mutations in these CYP enzymes are closely linked to congenital defects such as pseudohermaphroditism and adrenal hyperplasia (Yamaguchi et al., 1998; Doleschall et al., 2014), and SNPs may be less well tolerated in these systems due to their central role in the production of sex hormones, progestins and corticoids. In contrast, the 35 drug metabolizing P450s that comprise CYPs family 1-4 in humans, average over 700 SNPs per isoform, and the 25,659 total polymorphisms in this group represent over 53% of the total P450 SNP population. While CYP2C9 (1422 SNPs), CYP2C18 (1230 SNPs) and CYP2C19 (2467) are the most polymorphic drug-metabolizing P450s, they are considerably less polymorphic than genes like CYP5A1 (5274 SNPs) which metabolizes prostaglandins, and CYP7B1 (3400 SNPs), which converts cholesterol to bile acids. CYP5A1 is also known as thromboxane synthase (TBXAS1), a functionally-distinct P450 that catalyzes molecular rearrangements of its substrates (Hecker and Ullrich, 1989). CYP5A1-based SNPs represent more than 10% of all CYP polymorphisms in humans and are recognized for altering tissue-specific splicing in blood and lung cells (Wang et al., 1994), and for promoting alternative exon 12 inclusion to promote cerebrovascular disease (Kimouli et al., 2009). CYP2E1 displayed the
lowest number of SNPs (33), and the highest alternative transcript-to-SNP ratio (0.73) among human CYP genes (see Figure 1). CYPs 21A2 (0.33), 27B1 (0.15) and CYP2D6 (0.11) also skew above normal CYP superfamily genes with respect to the average alternative transcript-to-SNP ratio of 0.05. Improved understanding of how key SNPs interact with the environment to alter individual CYP gene splicing outcomes, therefore, could be very helpful in improving the predictive power of personalized medicine based on genetic screening.

Therefore, while some SNPs alter gene function by modulating alternative splicing events, the total number in a population may be irrelevant from an evolutionary perspective, if they do not manifest in alternative phenotypes. Furthermore, it is intriguing that several steroid hormone-metabolizing P450s, including CYP7B1 (3400 SNPs), CYP19A1 (2178 SNPs), and CYP39A1 (2129 SNPs) display disproportionate amounts of polymorphisms compared to all other human CYPs, reiterating the possibility that SNP expansion among P450 genes is neither random, nor predictable, based on the tissue-specific expression or substrate-selectivity of an individual CYP. What may be more important with respect to the accumulation of SNPs may be their ability to be silenced or masked via alternative- or trans-splicing paradigms that render them invisible or inconsequential to the natural selection process. The defective nature of a CYP-related SNP may depend on its ability to manipulate the CYP transcriptome, therefore improved understanding of how key mutations alter CYP gene splicing outcomes may be critical to addressing the goal of precision medicine in the post-genomic era.

**TISSUE-SPECIFIC ALTERNATIVE SPLICING OF CYTOCHROME P450 GENES: A HISTORICAL PERSPECTIVE**

During the late 1980’s and early 1990’s, several groups began to appreciate the possibility that differential RNA splicing might serve as an additional mode for regulating CYP function. In 1987, the first P450 splice variant for hepatic P450-1, later renamed CYP2C8, was reported (Okino et al., 1987). Over the next decade, several examples of alternate splicing were
documented for several human CYPs including lung CYP4F2 (Nhamburo et al., 1990), liver CYP2B6 (Yamano et al., 1989; Miles et al., 1989) and CYP2A7 (Ding et al., 1995). In the latter case, a tissue-specific, alternative splicing effect was documented, where a 44 kDa CYP2A7-AS splice variant (compared to the 49 kDa wild-type protein) found only in 20% of liver samples was the dominant protein expressed in human skin fibroblasts. Just two years later, a similar pattern of alternative splicing behavior was documented for CYP2C18, which selectively encodes splice variants skipping different combinations of exons 4, 5, 6 and 7 in the epidermis (Zaphiropoulos, 1997). Circular RNA transcripts synthesized from the donor and acceptor sites from the skipped segments of exons 4-7 were also documented, and although their potential epigenetic functions remains unknown, they were easily detected based on their increased stability (Zaphiropoulos, 1997). Important studies focused on environmentally-responsive, alternative CYP splicing among xenobiotic-metabolizing P450s of the rat 2A family were also conducted during this era (Desrochers et al. 1996).

By 1999, the CYP2C18 alternative splicing story became even more complex when it was discovered that CYP2C18 and CYP2C19, which cluster on chromosome 10q24 with CYP2C9, also participate in tissue-specific, trans-splicing events, among the 3 genes (Zaphiropoulos, 1999). Over the next decade, cell-specific splicing factors responsible for directing the complex, post-transcriptional gene assembly of P450s (and other genes) were identified (Wang and Burge, 2008). However, the physiological importance of the more unusual CYP splice variant and chimeric forms remains unclear, particularly with respect to how their variable, tissue-specific expression alters inter-individual sensitivity to hormones, pharmaceuticals and xenobiotics.

As discussed above, a single CYP gene can be translated into multiple, variant protein sequences, based only on tissue-specific differences in the expression of key, gene splicing machinery (e.g. snRNAs). Several alternately-spliced CYP1A1 transcripts identified in human brain tissues exemplify this phenomenon, as well as the concept of tissue-specific spliceosome activity (Bauer et al., 2007). Inducible CYP1A1 is constitutively expressed in the brain, where it
is localized to neurons in the cortex, cerebellum and hippocampus. An 87 bp deletion in exon 6 is observed in the brain-specific CYP1A1 form, but not in liver (Chinta et al., 2005). Human brain tissues expressing the Δexon6-CYP1A1 variant do not express wild-type CYP1A1 and they metabolize known substrates (e.g. benzoxy-resorufin and benzo(a)pyrene (BP)) at different rates than wild-type enzyme (Kommaddi et al., 2007). Furthermore, cells expressing wild-type CYP1A1 generate DNA adducts via the formation of reactive 3-OH-BP products, while cells expressing the Δexon6-CYP1A1 variant in brain, do not (Kommaddi et al., 2007).

Based on the structural analysis of CYP1A1 in Figure 2 (Derived from PDB: 4I8V; Walsh et al., 2013), tissue-specific, alternative splicing of exon 6 may re-shape the size and polarity of the CYP1A1 substrate access channel, to alter substrate recruitment and recognition processes and redox-partner binding interactions. Exon 6 of CYP1A1 appears to be organized for “plug and play” utilization in the brain, where targeted skipping does not compromise the global P450 fold, but alters substrate specificity for a spectrum of endogenous substrates, including testosterone. This tissue-specific, exon 6 usage of CYP1A1 is complimented by a similar pattern of alternative exon 2 usage found in human endothelial cells, leukocytes and tumor cells (Leung et al., 2005, Bauer et al., 2007). As shown in Figure 2, targeted-skipping of an 84 bp cryptic intron in exon 2 of CYP1A1, produces catalytically-unique, splice variant with diminished estradiol metabolism that localizes to the nucleus and mitochondria rather than the ER (Leung et al., 2005). The added diversity in both structure/function and subcellular distribution observed for exon 2-modified, CYP1A1 splice variants found both in normal and tumor tissues remains enigmatic and denotes a higher level of structural plasticity with respect to tertiary structure, than might be currently appreciated or expected from classical structural studies of wild-type CYP forms (Johnson and Stout, 2013).

Perhaps the most dramatic, and mutually exclusive example of tissue-specific, alternate exon usage in P450s was documented for CYP4F3, which can be expressed as two, distinct splice variant forms (CYP4F3A and CYP4F3B). These two CYP4F3 isoforms differ only by their
variable usage of exons 3 and 4, which encode distinct portions of the active site and substrate access channel, allowing them to fine-tune substrate specificity (Christmas et al., 1999; Christmas et al., 2001). CYP4F3A is expressed in blood and bone marrow myeloid cells, selectively expresses cassette exon 4 (and not 3), and has low affinity for leukotriene B(4) (LTB4). In contrast, CYP4F3B, expressed exclusively in the liver and kidney, encodes exon 3 (and not 4) and has high affinity for LTB4, in addition to other functional differences. The highly concerted, tissue-specific splicing of CYP4F3 exemplifies the functional utility of the modular CYP gene platform, one that allows cells to sample a continuum of a gene products, as dictated by tissue-specific, physiological stress conditions.

However, it should also be clearly noted, that not all CYP genes are divided into an equal number of coding and non-coding regions, which suggests that the accumulation of introns may be an adaptive and ongoing process. For example, the CYP1B1 gene, which encodes a P450 enzyme that metabolizes both xenobiotics and endogenous substrates, contains only 2 coding exons (Supplemental Figure 2), and therefore is not subject to the same patterns of alternative splicing as other CYP1 family members, 1A1 & 1A2, which have 6 coding exons (Stoilov et al., 1998). In this regard, the most well-studied splice variants of CYP1B1 all represent severe truncations or deletion mutants, not exon skipping (Tanwar et al., 2009), and CYP1B1 retains an unusual, elongated C-terminus, further distinguishing it from other human CYP forms. The CYP1B1 gene is also uniquely located on chromosome 2 at locus 2p21-22, which in humans, is derived from an ancient fusion of chimpanzee chromosomes 2a and 2b (Faiq et al., 2014). It is difficult to speculate why the CYP1B1 gene is organized so differently from closely related CYP1A genes located on chromosome 15. However, observations linking a common set of nuclear proteins, to both the alternative splicing process and NMD, seem to imply that expansion of gene complexity may be more related to the rate at which a given mRNA transcripts are processed by the spliceosome and ribosome, rather than other evolutionary factors (Lejeune and Maquat, 2005; McGlincey and Smith, 2008). Ultimately, the adaptive forces driving the modular organization of
genes into variable numbers of introns and exons remain a point of speculation. However, a gene's intronic complexity clearly influences its sensitivity to alternate splicing events and the domain swapping paradigm. The tissue-specific use of alternative promoters and alternative (AUG) translational-start-sites also is facilitated by this type of hidden gene complexity that allows discrimination among a set of well-defined exons and poorly-defined pseudoexons.

In 2005, a truncated, CYP24A1 splice variant was identified in the human macrophage that is incapable of metabolizing hormonal forms of vitamin D (Ren et al., 2005). CYP24A1 is the prototypical mitochondrial P450 responsible for the side-chain cleavage of the vitamin D hormone and it possesses a mitochondrial localization sequence in the first 30 residues of its N-terminus (Annalora et al.; 2004). The CYP24A1 splice variant found in activated macrophage (CYP24A1-SV1) lacks 153 N-terminal residues encoded by exons 1 and 2, and gains 8 residues preceding the content from exon 3, via a partial insertion of intron 2. CYP24A1-SV1 is thought to function as a soluble variant that disrupts normal vitamin D hormone metabolism via the sequestration of key intermediates from both wild type CYP27B1 and CYP24A1 (Ren et al., 2005). In this regard, the CYP24A1-SV1 variant may promote a discrete pattern of vitamin D hormone accumulation in the macrophage, and potentially other tissues, such as the intestine, where vitamin D hormonal leakage from irritated cells is thought to contribute to the burst of extracellular hormone observed in inflammatory disorders like Crohn's disease (Abreu et al., 2004; Mangin et al., 2014). A structural analysis of the CYP24A1-SV1 variant is shown in Figure 3 (based on PDB: 3K9V; Annalora et al., 2010), and it highlights how the targeted loss of N-terminal residues from exons 1 and 2 could reconfigure the membrane-binding surface and substrate access channel to alter the catalytic properties of wild-type CYP24A1.

Interestingly, the orphan P450 CYP27C1, a close relative of CYP24A1, encodes a 372 residue wild-type protein that resembles the CYP24A1-SV1. CYP27C1 lacks a prototypical N-terminus and initiates coding in the middle of canonical helix C, at a position analogous to the alternative start site utilized by CYP24A1-SV1 (Ren et al., 2005). CYP27C1 is expressed in liver,
kidney, pancreas and several other tissues, and remained an orphan for over a decade (Wu et al., 2006), until recently, when it was shown to mediate the conversion of vitamin A1 (all-trans retinol) into vitamin A2 (all-trans 3,4-dehydroretinal) in cell culture (Kramlinger et al., 2016). It is intriguing that CYP27C1 may represent a codified version of the more soluble CYP isotype recapitulated by CYP24-SV1, and that this transition may mediate a general switch between vitamin A and vitamin D substrate specificity. Improved knowledge of CYP27C1 structure/function could help to unravel some of the complexity associated with the variant CYP regulation, particularly for variants subject to atypical sub-cellular trafficking events.

Therefore, while increasingly common in the literature, the functional significance of many tissue-specific CYP variants remain elusive. CYP2C8 for example, which is constitutively expressed in the liver at 5-7% of the total tissue P450, has more than 20 known alternately-spliced variants, and is a breast cancer biomarker due to its high expression in mammary tumors (Knüpfer et al., 2004). Bimodal targeting to both mitochondria and ER have been reported for CYP2C8 and several other CYPs, including CYP1A1, 1B1, 2B1, 2E1 and 2D6. While it is clear that alternative splicing directs this process in some tissues, proteolytic cleavage of N-terminal targeting sequences can also produce this effect (Bajpai et al., 2014). One 2C8 variant (variant 3; ~44 kDa) with an alternative N-terminus is targeted to the mitochondria where it may contribute to oxidative stress (Bajpai et al., 2014). The catalytic competency of these untethered variants remains a topic of intense debate, as it is often difficult to assign an orphan function to a variant protein that has lost selectivity to the gene’s prototypical substrates. A review of CYP splice variants, subject to alternate subcellular trafficking is shown in Supplemental Table 1. Computational analysis from the Aceview database (Thierry-Mieg and Thierry-Mieg, 2006) indicate that some CYP variants may be targeted to the nucleus, peroxisome, plasma membrane or vacuole compartments, in addition to classical targets in the ER, mitochondria or cytoplasm. This analysis hints that CYP splice variants, including tumor-specific forms, may be performing
an array of unrecognized functions, including diverse roles at the cell surface and beyond, which are rarely considered (Eliasson et al., 1996; Stenstedt et al., 2012).

TRANSFORMATION-SENSITIVE ALTERNATIVE SPICING IN THE CYTOCHROME P450 SUPERFAMILY

This transformation-specific, CYP splicing phenomenon was first documented in 1996 for the prototype P450, CYP19A1 (or aromatase), which facilitates estrogen biosynthesis via three successive hydroxylations of the androgen A ring and displays a unique pattern of exon 1 (and cryptic intron) usage in human breast tumor cells and tissues (Zhou et al., 1996). A similar pattern of CYP19A1 exon 1 variant expression also has been documented in sheep, cattle and rabbit, suggesting this paradigm is not exclusive to humans or transformed-tissue types (Vanselow et al., 1999; Bouraïma et al., 2001). More recently, it was determined that CYP19A1’s exon 5 also is defined poorly and subject to both splicing mutations and physiological alternative splicing events (Pepe et al., 2007). Skipping of exon 5, which encodes complete α-helices D and E in CYP19A1, eliminates prototypical P450 aromatase activity in both steroidogenic and non-steroidogenic tissues (Lin et al., 2007; Pepe et al., 2007). These observations suggest a tissue-specific, regulatory mechanism for aromatization based on alternative exon 5 inclusion in humans. A structural analysis of this CYP19A1 splice variant is depicted in Supplemental Figure 3 (Based on PDB: 3S79; Ghosh et al., 2012). In aromatase, exon 5 skipping naturally suppresses androgen metabolism by re-defining the determinants of substrate recognition. While transfection studies suggest CYP19A1 variant proteins are encoded and not subject to NMD, their ability to alter the safety and efficacy of several CYP19A1 (aromatase) inhibitor drugs (e.g. AIs; including Arimidex (anastrozole); Aromasin (exemestane); Femara (letrozole) and Teslac (testolactone) remains unclear (Hadfield and Newman, 2012) and an area of active investigation (Liu et al., 2016).

The tissue-specific, alternative start site usage of CYP24A1, discussed above, appears to be exacerbated in several human transcript variants, where an array of truncated splice variants...
of approximately 40, 42, and 44 kDa have been identified with similar N-terminal modifications (Fischer et al., 2009; Horváth et al., 2010; Scheible et al., 2014). An additional, cancer-related, splice variant of CYP24A1 that cleanly skips exon 10 has also been documented (Horvath et al., 2010; Muindi et al., 2007). In CYP24A1, exon 10 encodes much of the protein's proximal surface, including the CYS loop, and a portion of helix L involved in heme-thiolate bond formation and adrenodoxin recognition (shown in Figure 3). While the function of this cancer-specific variant remains unknown, it may have reduced ability to bind heme and redox partners compared to truncated variants (skipping exons 1 and 2), and thus may represent an alternative, dominant negative variant also capable of modulating vitamin D hormone metabolism and metabolite trafficking at the plasma membrane level. As discussed previously, the cellular mechanisms regulating CYP24A1 alternative splicing in the macrophage remains unclear, although cell-specific variations in splicing factors (e.g. heterogeneous ribonuclear protein A1) are thought to be responsible (Ren et al., 2005). A more complex pattern of “transformation-sensitive” alternative CYP24A1 gene splicing also has been documented in human tumors of the prostate (Muindi et al., 2007), breast (Fischer et al., 2009; Scheible et al., 2014) and colon (Horváth et al., 2010; Peng et al., 2012). It is thought that while intronic SNPs may facilitate the alternate splicing pattern seen for CYP24A1 in prostate cancer cells, vitamin D hormone exposure also putatively alters splicing via discrete interactions among the vitamin D receptor (VDR) and nuclear splicing factors like the heterogeneous nuclear ribonucleoprotein C (hnRNPC; Zhou et al; 2014). Tumor cells overexpressing CYP24A1, therefore, may not retain sufficient hormone to activate the VDR properly, disrupting normal gene splicing events needed to properly transduce the hormone's immunogenic, pro-differentiation properties. CYP24A1 splice variants have now been documented among multiple human breast tumor cell lines (e.g. MCF-7 and MCF-10) and in both healthy and malignant breast tumor tissues. In benign tissue, wild-type CYP24A1 (56 kDa) was the only isoform present. In malignant tissues, 3 splice variants (40, 42 and 44 kDa) were expressed, similar to cultured MCF-7 breast tumor cells that express the 42 and 44 kDa variants.
(Fischer et al., 2009; Scheible et al., 2014). A similar pattern of transformation-sensitive splicing is seen in human colon cancer tissues, where the histological grade and the gender of the patient modulates the formation of the 3 detectable CYP24A1 splice variants (Horváth et al., 2010). Peng and co-authors further refined our understanding of this process in the colon, by demonstrating that cell- or tissue-specific splicing patterns were suppressed by parathyroid hormone signaling, but only in the absence of the vitamin D hormone, which may be the master regulator of both CYP24A1 gene expression and splicing (Peng et al., 2012).

The phenotypic instability exemplified by the CYP24A1 gene in human tumors complicates the therapeutic landscape from which rational drugs for cancer can be conceived (Trump et al., 2006). Epigenetic modification of the CYP24A1 gene and its promoter also may contribute to tissue-specific differences in P450 expression (Johnson et al., 2010). Frontiers in this area of research will likely address the coordinated role that nuclear hormone receptors and epigenetic splicing factors (including non-coding forms of RNA) play in managing both the transcriptome and the epigenome.

It should be noted that the CYP27B1 gene, which is highly related to CYP24A1, was also recognized early as being sensitive to transformation-specific alternative splicing in malignant human glioma cells (Maas et al., 2001). CYP27B1 variant expression has now been documented in tumors of the breast, skin, cervix, kidney and ovaries (Fischer et al., 2007; Wu et al., 2007; Diesel et al., 2004; Fischer et al., 2009). Tissue-specific CYP27B1 splicing also is common in healthy skin, where a 59 kDa variant with a 3 kDa insertion between exons 2 and 3, exclusive to skin, was observed, and linked to the activity of CRE-binding elements (Flanagan et al., 2003). Fluctuations in skin cell density, calcium concentrations, and UV-B light exposure were also shown to effect the splicing process (Seifert et al., 2009). Collectively, these findings help to validate our core hypothesis that alternative gene splicing, dictated by both cellular and environmental factors such as UV light exposure and nutritional status, can alter a patient's
metabolic profile in tissue-, age-, sex-, exposure- and disease-specific manners that are difficult to predict.

While research focused on CYPs 19A1, 24A1 and 27B1 are among the most comprehensive studies on transformation-sensitive alternative splicing in human P450s, the literature is replete with anecdotal reports of alternative CYP splicing among various transformed cell-types and tissues. For example, CYP2E1 transcripts that selectively skip exons (2, 2-3, and 2,4,5-5′,6) were identified in lung carcinoma cell lines, but not in corresponding lung tumor tissue or liver extracts, where splice variants were found to be non-inducible by ethanol exposure (Bauer et al., 2005). The authors noted at the time that none of these splice variants were accounted for in the NIH’s Aceview database, highlighting the longstanding challenge associated with cataloging splice variant forms derived from cell–specific abnormalities versus those generated from normal, tissue-specific splicing paradigms. Furthermore, several orphan P450s, including CYP1B1, CYP2S1 and CYP2W1 are highly associated with specific, human tumor types, making them important pro-drug targets for chemoprevention strategies (Stenstedt et al., 2012). Unfortunately, as discussed above, the alternative splicing behavior of orphan CYPs is often as poorly understood, as their cryptic function.

Interestingly, orphan CYP2W1, which is expressed human tumor types of the colon, is associated almost exclusively with aberrant P450 behavior, including an inverted orientation in the ER membrane, and glycosylation-sensitive trafficking to the plasma membrane (Gomez et al., 2010; Stenstedt et al., 2012). The increased expression of CYP2W1 in tumors is associated with demethylation of a CpG island in the exon 1-intron 1 junction of the gene. How this mechanism relates to the paucity of normal CYP2W1 expression in healthy tissues, remains unclear, but it implies that the expression of some specialized CYPs may only be invoked under extreme cellular stress, where an alternative CYP activity may be critical for balancing cellular homeostasis. The abnormal expression and trafficking of CYP2W1 to the plasma membrane could promote an autoimmune response as part of its mechanism. A similar autoimmune reaction is associated
with aberrant CYP2E1 trafficking in animal tissues treated with halothane (Eliasson et al., 1996). The NIH’s Aceview and Ensembl (Cunningham et al., 2015) databases indicate CYP2W1 encodes at least 11 transcript variants, three of which lack one or more exons and are associated with hepatocellular carcinomas, adenocarcinomas and neuroblastomas. Nearly 200 of the 965 alternatively-spliced CYP transcripts we identified in public database searches were associated with one of 40 human tumor types. Improved understanding of their tumor-specific cellular roles, and the mechanisms regulating their induction, should aid in the identification of improved disease biomarkers and CYP variant-specific therapeutic agents.

TRANS-SPlicing of Cytochrome P450 Gene Chimeras

It is well understood that both transformation-sensitive cellular changes and CYP polymorphisms can alter protein expression in multiple ways that are often difficult to predict from primary sequence analysis alone. This challenge is exacerbated in situations where chimeric CYP genes are created via trans-splicing events among extended (or multiple) pre-mRNA transcripts. Therefore, while mutations in a gene like CYP2C9 are linked to poor metabolism of several drugs, including phenytoin and tolbutamide, the role that alternative- or trans- gene splicing plays in manifesting a deleterious CYP2C9-related metabolic phenotype are rarely considered. For example, several CYP2C9 splice variants (CYP2C9sv) have been identified (Ohgiya et al., 1992), including a liver-specific form that skips exon 2 (Ariyoshi et al., 2007). The internally-truncated CYP2C9 variant is spectrally-active, but does not metabolize prototype CYP2C9 substrates. A structural analysis of the CYP2C9 crystal structure (PDB: 4NZ2; Brändén et al., 2014) shown in Supplemental Figure 4A, hint that removal of exon 2 would reshape portions of the membrane-binding surface, substrate access channel, active site and redox partner binding surface relative to the reference protein. CYP2C9 is located within a cluster of CYP genes on chromosome 10q24 and is subject to non-random, trans-splicing events with neighboring CYPs 2C8, 2C18 and 2C19 in liver and skin, where CYP2C18 exon 1-like sequences were found
splice into different combinations of introns and exons from the CYP2C9 genes (Warner et al., 2001). Common 2C9 related SNPs linked to various metabolic disorders, therefore, may only function to modulate the innate splicing process, indirectly. This is particularly true if their coding regions are spliced out of chimeric 2C forms being expressed on an inter-individual basis, highlighting another major flaw in GWAS targeting the components of gene clusters.

The CYP2D gene family also is complex, clustered and subject to trans-splicing events; it is comprised of CYP2D6 and 4 pseudogenes (CYP2D7P1 and 2 and CYP2D8P1 and 2). Alternative splicing of the CYP2D family pre-mRNAs has been detected in human liver, breast and lung tissue (Huang et al., 1996; Huang et al., 1997), including a highly-expressed splice variant which completely skips exon 6. As shown in Supplemental Figure 4B, analysis of the CYP2D6 crystal structure (PDB: 4WNU: Wang et al., 2015) reveals that exon 6 encodes the complete coding sequence for helix I, and that skipping of this important region would dramatically alter protein structure/function, by removing key catalytic residues (e.g. Thr-309) and remodeling the distal pocket of the active site. Alternate CYP2D6 exon usage is tissue-selective and variant forms skipping exon 3 and portions of exon 4 have been documented that alter enzyme function and subcellular localization (Sangar et al., 2010). CYP2D6 polymorphisms are linked to modified enzyme activity and may serve as biomarkers for poor metabolizing phenotypes sensitive to dose-dependent drug toxicity. Recently it was reported that intronic polymorphisms in CYP2D6*41 individuals (considered intermediate metabolizers) could dramatically increase the expression levels of CYP2D6 variants lacking exon 6 (Toscano et al., 2006). These findings, coupled with the complex organization of the CYP2D gene cluster, indicate that both alternative and trans-splicing mechanisms may underlie the complex genotype-phenotype relationships associated with variable CYP2D6 metabolism in humans, and that CYP2D6 is well-organized structurally to exploit both tissue- and transformation-selective splicing mechanisms targeting helices C and D (exon 3) and helix I (exon 6).
The CYP 3A gene family (comprised of CYP 3A4, 3A5, 3A7 and 3A43) is also highly complex in humans, with each gene consisting of 13 exons (with ~71-88% amino acid identity) that cluster on chromosome 7q21-22.1, where the CYP3A43 gene is in a head-to-head orientation with the other three genes (Finta and Zaphiropoulos, 2000). Several chimeric forms of CYP3A mRNAs have been described where CYP3A43 exon 1 is joined at CYP3A4 and CYP3A5 canonical splice sites, implying a trans-splicing mechanism that bypasses classical transcriptional control paradigms. In Supplemental Figure 5A, a structural analysis of CYP3A4 (Based on PDB: 4K9T; Sevrioukova et al., 2013) reveals the highly-complex segmentation of the CYP3A4 gene into 13 exonic domains, which, as described above, may potentiate rapid translation upon induction (Lejeune and Maquat, 2005; McGlincy and Smith, 2008). Primary amino acid sequence alignments of key CYP3A genes, suggest exon 1 may encode distinct transmembrane anchor helices and membrane binding architecture (Supplemental Figure 5B), hinting that CYP3A chimeras may encode a spectrum phenotypes with unique tissue-specific distribution and functionalities. A CYP3A4 trans-splicing variant containing CYP3A43 exon 1, followed by CYP3A4 exons 4-13, has been described, which retains detectable, testosterone 6-β-hydroxylase activity despite losing membrane binding features encoded by exons 2 and 3, comprising helices A’ and A, and portions of the β1-1 sheet (Supplemental Figure 5C). A CYP3A4 trans-splicing variant containing CYP3A43 exon 1, followed by CYP3A4 exons 7-13 also has been studied, and it displays minimal 6-β-hydroxylase activity (Supplemental Figure 5D). This story is further complicated by the presence of additional CYP3A pseudogenes (CYP3AP1 and CYP3AP2) in the intergenic regions between CYP3A4, 3A7 and 3A5, where discrete exons from the pseudogenes also have been found trans-spliced into CYP3A7 transcripts (Finta and Zaphiropoulos, 2002). Ultimately, CYP3A4 and CYP3A5 splicing complexity may help to explain how intronic SNPs in CYP3A4*22 patients alter tacrolimus metabolism, without introducing a non-synonymous mutation to an exon (Elen et al., 2011). The predictive power of GWAS may therefore need to be reconsidered, particularly with respect to the most complex CYP gene
families defined by the presence of pseudogenes (e.g. CYP 2C, 2D, 3A, 21A and 51A families) and trans-splicing mechanisms.

SNP-SENSITIVE ALTERNATIVE SPLICING IN THE CYTOCHROME P450 SUPERFAMILY

As discussed above, tissue-specific, alternative- and trans-splicing behavior has now been documented for several, important CYP gene families, however single nucleotide polymorphisms (SNPs) can destabilize this process in complex ways that are often difficult to predict. SNPs that target cryptic intronic recognition elements or discrete intron/exon splice junction boundaries can alter translation start and stop site usage and the expression ratios among wild-type and variant CYP transcripts. Some CYP SNPs are exceedingly rare, occurring in less than 1% of the population. These rare mutations are linked to congenital defects, including glaucoma (CYP1B1; Stoilov et al., 1998; Tanwar et al., 2009; Sheikh et al., 2014), 17-alpha hydroxylase deficiency (CYP17A1; Yamaguchi et al., 1998, Costa-Santos et al., 2004; Hwang et al., 2011; Qiao et al 2011), congenital adrenal hyperplasia (CYP21A2; Robins et al., 2006; Szabó et al., 2013, Lee, 2013; Sharaf et al., 2014), spina bifida (CYP26A1; Rat et al., 2006), focal facial dermal dysplasia (CYP26C1; Slavotinek et al., 2013), and cerebrotendinous xanthomatosis (CYP27A1; Garuti et al., 1996; Verrips et al., 1997; Chen et al., 1998; Tian et al., 2011). Other CYP polymorphisms that alter splicing are associated with neurological and metabolic diseases, including Parkinson's disease (CYP2D6; Denson, 2005; CYP2J2; Searles et al., 2013), hypertension (CYP4A11; Zhang et al., 2013; CYP17A1; Wang et al., 2011), breast cancer (CYP2D6; Huang et al., 1996; CYP19A1; Kristensen et al., 2000;) colon cancer (CYP2W1; Stenstedt et al., 2012) and lung cancer (CYP2D6; Huang et al., 1997; CYP2F1; Tournel et al., 2007). However, the complex etiology of these disorders can make it difficult to interpret the exact role that a P450 SNP might play in the actual onset or progression of a disease. Multiple studies of various intergenic, intronic and exonic polymorphisms that alter human drug and xenobiotic metabolism have also been reported (CYP1A1; Allorge et al., 2003; CYP2C19; deMorais et al., 1994, Ibeanu et al., 1999,
Satyanarayana et al., 2009, Sun et al., 2014; CYP2D6; Toscano et al., 2006, Lu et al., 2013, Wang et al., 2014; CYP3A4; Elens et al., 2011; and CYP3A5; Kuehl et al., 2001, Busi and Cresteil, 2005, Lee et al., 2007, García-Roca et al., 2012). In many cases, the most pharmacologically-relevant SNPs are those affecting the innate splicing behavior of a given gene. This observation highlights why knowledge of a patient’s discrete CYP transcriptome is so important when selecting a personalized medicine regime.

It is now clear that numerous CYP-related SNPs are found between genes and exons, or in the 5’ or 3’ untranslated regions of RNA transcripts. Polymorphisms in this class have now been identified for virtually every human CYP isoform, with each having the potential to alter splicing in unique ways in different individuals. For example, the CYP3A5*3 SNP (rs776746, also known as 6986A>G) introduces a mutation in intron 3 that alters normal splicing of exon 4, producing a nonfunctional, truncated CYP3A5 protein that predominates in many Caucasians (Kuehl et al., 2001). Expression of the wild-type CYP3A5*1 protein is highly-variable among populations, and is linked to an increased sensitivity to salt-induced hypertension. Interestingly, some individuals that are homozygous CYP3A5*3/*3 can express both the truncated variant and properly-spliced, wild-type CYP3A5 (Lin et al., 2002). Because the CYP3A5*3 mutation occurs in intron 3, over 100 base pairs upstream from the splice donor site of exon 4, it is unclear how this SNP alters normal recognition of the intron 3-exon 4 junction. While variations in the expression of ubiquitous splicing factors (e.g. U1snRNP) may account for inter-individual variability in recognizing the CYP3A5*3 SNP, there are currently no models to explain how this complex, salt-sensitive, splicing phenomenon is being regulated in the kidney. The alternative splicing of CYP3A5 variant exemplifies why SNP-based approaches to personalized medicine are so challenging to implement, when the physiological impact of a given SNP, may be the opposite of what is expected. Computational approaches that consider a patient’s age, genotype, nutrition and disease status, may ultimately be needed to make good predictions concerning the impact of a given SNP on an individual’s transcriptome.
The challenge of addressing SNPs for personalized medicine is further complicated by observations that some xenobiotics, including aminoglycosides and cyclohexamides (Busi and Cresteil, 2005), can alter the proofreading capabilities of the ribosome, allowing inhibition of PTCs and translational read-thru of alternate transcripts that may or may not remain in-frame with respect to the wild-type transcript. Improved knowledge of a patient’s chemical exposure history and epigenetic status, therefore, may become increasingly important when making predictions about food and drug safety. Furthermore, personalized approaches to medicine, based solely on SNPs identified via cohort studies or GWAS, may be fundamentally-flawed, and potentially misleading, in that they fail to account for the milieu of compensatory cellular mechanisms that are adept at masking the most deleterious effects of even the most highly-penetrant mutations.

ADDRESSING NON-TRADITIONAL CYP FUNCTION AND TRAFFICKING EVENTS

Our meta-analysis and review of splice variability within the cytochrome P450 superfamily has revealed an under-appreciated level of alternative transcript expansion and variant protein structural complexity that is poorly addressed by conventional CYP structure/function paradigms. Well-studied splice variants from the CYP1A family reveal a sophisticated domain swapping mechanism that allows for tissue-specific sampling of alternative protein conformational space. Discrete arrays of tissue-specific snRNAs (U1-U9) and ancillary splicing factors can alter the biological syntax of a gene, thus providing a granular mechanism for this natural phenomenon, which is highly-responsive to environmental factors, including diet and chemical exposure. In human brain, the alternative splicing of CYP1A1 appears linked to a tissue-specific sensitivity to a reactive CYP-product (e.g. 3-OH BP; Kommaddi et al., 2007). How this gene-specific sensitivity is transduced and imprinted at the genetic level remains enigmatic, but the cellular microenvironment clearly dictates what hidden structural complexity of a CYP gene may be exploited. This observation helps explain why some CYP variants, like CYP24A1-SV1, which are normally expressed in the immune system, can become major players in tumor cells, where they...
may elaborate a common emergency function, such as the suppression of vitamin D hormone metabolism. This cellular phenomenon corresponds well with related, vitamin D hormone-scavenging strategies, such as the 3-epimerization pathway, which can be induced to modify the vitamin D hormone’s chemical ring structure to limit its efficient catabolism via CYP24A1 (Rhieu et al., 2013). To date the enzyme responsible for the 3-epimerization of the vitamin D hormone remains unknown (Bailey et al., 2013).

New appreciation for the increasingly diverse and complex roles that CYP splice variants play in normal biology should help reinvigorate discussions of CYP gene structure and evolution, particularly with respect to the selective forces that regulate CYP gene duplication and mutation events. While the origin and expansion of spliceosomal introns remains highly-debated (Rogozin et al., 2012; Qu et al., 2014), the intron-exon boundaries of 17 of 18 mammalian CYP families appear to be well-conserved across almost 420 million years of evolution, as evidenced by the fugu (pufferfish) genome (Nelson, 2003); only the CYP39 gene is missing from fugu and all known fish. The bulk of P450 structural diversity encoded by the human genome, therefore, likely predates the divergence of tetrapods from ray-finned fish over 420 million years ago. In this regard, brain-selective expression of CYP1A1 variants lacking exon 6 in humans, may be no more evolved than tissue-specific expression of CYP19A1 splice variants found in fugu. In each case, the induction of CYP splice variants appears to provide a novel mechanism for modulating steroid hormone pleiotropism among a diverse set of target tissues.

The selective pressures that created the modular framework of the CYP gene, therefore likely predate the Cambrian explosion 540 million years ago, and lurk somewhere in the history of deuterostome evolution (Nelson, 2003). However, contemporary analysis of CYP splicing mechanisms are needed to help elaborate the relationship between alternative subcellular trafficking and alternative CYP variant function in the mitochondria, nucleus, peroxisome, cytoplasm or plasma membrane (Supplemental Table 2). Because alternative splicing and alternative start site usage can generate even greater CYP structural diversity than those
predicted to arise from post-translational modifications of intracellular targeting motifs, the mechanisms underlying their emergence and regulation deserve greater attention. For example, if subtle differences in the membrane binding features of CYP1A1 and CYP1A2 can modulate redox partner recruitment and protein interactions within ordered and disordered regions of the lipid bilayer (Park et al; 2015), metamorphic structural changes associated with some CYP variants may underlie an even greater array of uncharacterized (and potentially moonlighting) P450 functions (Zhao and Waterman 2011, Lamb and Waterman, 2013).

While the landscape of alternate CYP gene functions continues to expand, the cellular basis for alternatively-targeting microsomal CYP forms to the mitochondria (e.g. CYP1A1, CYP1A2, CYP1B1 CYP2B1, CYP2B2, CYP2E1, CYP3A1 and CYP3A2) is still being investigated (Anandatheerthavarada et al, 1997). However, a putative role in modulating destructive, subcellular ROS production has recently been proposed (Dong et al., 2013). This theory emerges from evidence that the alternative trafficking of the related NADPH:quinone oxidase-1 (NQO1) gene, which limits ROS by preventing semiquinone redox cycling, can protect the mitochondria from stress-induced oxidative damage. CYP variants may potentially play a similar role by redirecting the metabolism of reactive subcellular compounds. However, because CYPs can also become uncoupled and directly produce ROS via the release of reactive products (i.e. via uncoupled Fenton-type reactions), further attention to the role of alternative CYP splicing in the modulation of mitochondrial stress and dysfunction seem warranted. In this regard, it is intriguing that many CYP splice variants that undergo alternative trafficking often lack a significant portion of the membrane binding domain and/or substrate access architecture (typically encoded by exons 1-3), which may expand heme accessibility or the rate of uncoupled ROS production. When paired with observations that alternative splicing is a primary mechanism by which mitochondria alter cellular phenotypes to mitigate environmental stress (Guantes et al., 2015), new, non-classical roles for CYP variants linked to disease, seem likely to emerge.
In this regard, it is interesting that the consensus sequence (GU-AG) of most RNA splice junctions contains a minimum of two guanine nucleotides, which are highly sensitive to ROS-mediated oxidation. In the presence of radical oxygen, guanine is rapidly converted to 8-oxo-guanine and related metabolites that disrupt normal base pairing and local secondary structure. Formation of 8-oxo-guanine at splice junctions can impede target site recognition by splicing factors (e.g. snRNAs and SRF proteins) and alter RNA Pol II processivity. Furthermore, because some introns are defined by AU-AC junctions rather than GU-AG junctions, the splicing of these introns may be less sensitive to ROS-mediated events, and therefore may be subject to a different suite of regulatory splicing mechanisms. For example, the induction of an environmentally-responsive genes, like adenosine-deaminase could alter the recognition of a prototypical AU-AC junction in a similar way, via adenosine-to-inosine (A-to-I) RNA editing (Solomon et al., 2013). Even mitochondrial genes lacking spliceosomal introns could be affected by this phenomenon, as the discrete dinucleotide sequences that coordinate their autocatalytic, self-splicing events, could also be potentially altered by CYP-mediated ROS modulation (Zimmerly and Semper, 2015).

**SHAPING AN IMPROVED ROADMAP TOWARDS PRECISION MEDICINE**

Because changes in mitochondrial function can alter global alternative splicing events, it is not surprising that both phenomenon are linked to the induction of cellular heterogeneity and human disease pathology progression (Raj and van Oudenaarden, 2008; Hanahan and Weinberg, 2012; Pagliarini et al., 2015). Despite a growing appreciation for the role that alternative splicing plays in promoting phenotypic variability, the identification of genetic polymorphisms linked to disease, which may or may not alter gene splicing events, remain a core focus of Pharmacogenomics. Mutations in single genes have now been identified for over 4000 human diseases, of which, 5 to 15 % are the result of SNPs resulting in nonsense mutations (Mort et al., 2008). However, SNPs occur in coding regions of genes with a frequency of approximately
1.5 %, and of those, only one third are expected to result in non-synonymous mutations and a much smaller number alter pre-mRNA processing, leading to a frequency in change of phenotype only ~0.5 % of the time. When one considers humans express almost 50,000 SNPs across the 57 human CYP genes, fewer than 250 may capable of significantly altering a patient’s metabolic phenotype substantially, an insight we contend may help streamline PGx approaches to personalized medicine (Zhou et al., 2009; Zanger and Schwab, 2013). For example, while CYP2D6 genotyping is no longer recommended in the clinical setting for Tamoxifen treatments (Abraham et al., 2010; Wu, 2011, Lum et al., 2013), the FDA still considers CYP2D6 clinically actionable from a PGx perspective for codeine and other drugs (Crews et al., 2014). Unfortunately, over 74 allelic variants of CYP2D6 have already been described, which greatly complicates genetic testing and the clinician’s ability to select the appropriate pharmacotherapy and dose (Zhou, 2009). Fortunately, organizations such as the Clinical Pharmacogenetics Implementation Consortium (CPIC) the Pharmacogenomics Knowledgebase (PharmGKB) are working to standardize methodologies for PGx data analysis and clinical guidance for specific gene-drug pairings (Caudle et al., 2016). However, CPIC concedes that some rare CYP variants may not be included in genetic tests, and that patients may be assigned “wild-type” genetics by default (Crews et al. 2014). When paired with additional uncertainties regarding CYP-specific genotype/phenotype associations and the untold numbers of unclassified SNPs, it becomes clear why accurate genetic screening remains challenging, and only one aspect of the therapeutic decision-making process.

Ultimately, the falling cost and broader availability of pyrosequencing technologies support our call for improved RNAseq strategies for personalized medicine, as accurate identification of the patient’s functional genome is a crucial component of precision medicine. While transcriptomic analysis offers superior guidance in the design of personalized therapeutic options, its broad implementation will require technical improvements to sample collection and processing that are also problematic for genomic testing. In this regard, complimentary metabolomics approaches
directed at variant-specific metabolism may provide more feasible, short-term improvements to PGx screening and precision-based approaches to medicine (Beger et al., 2016). Innovative, gene-directed therapeutic technologies such as splice-altering antisense oligonucleotides and CRISPR/Cas9 genome editing systems may also become feasible tools for manipulating a patient’s transcriptome to optimize therapeutic outcomes. Key examples of splice-switching technology already being investigated to treat human disease are listed in Supplemental Table 2. In this regard, our group has participated in the development of Eteplirsen, the new antisense oligonucleotide drug which received accelerated approval from the FDA for the treatment of Duchenne muscular dystrophy (DMD; Syed, 2016; Niks and Aartsma-Rus, 2017). Eteplirsen’s development evolved from early studies of exon skipping in the murine dystrophin model (Fall et al., 2006; Fletcher et al., 2007; Adams et al., 2007), a canine model of DMD (McClorey et al., 2006), and in human muscle explants (McClorey et al., 2006). Our group has also employed exon skipping oligomers to refine the immune response-mediated gene expression of IL-10 in an Ebola virus lethal challenge mouse model (Mourich and Iversen, 2009; Panchal et al., 2014) and of CTLA-4 in a murine model of autoimmune diabetes (Mourich et al., 2014). We hypothesize that similar splice-altering technology may be useful in re-directing the function of drug metabolizing CYPs, like CYP3A4 (Arora and Iversen, 2001), whose metabolism of drugs like tamoxifen, is linked to genotoxicity (Mahadevan et al., 2006). As our appreciation for transcriptome expansion and the mechanisms of alternative CYP gene splicing evolve, new therapeutic gene editing options are likely to emerge that could scarcely be predicted using genetic testing alone.

In summary, the human cytochrome P450 family transcriptome contains over 965 different variant forms (Table 2), many with common structural features sensitive to alternative splicing events that expand CYP protein diversity. The transcription and processing of CYP gene transcripts is complex and coordinately-regulated within the nucleus by multiple factors, including nuclear receptor (NR) signaling via environmental sensors like the peroxisome proliferator-
activated receptors (PPARs), PPARγ and PPARα, which interact with the PGC-1α transcriptional coactivator to regulate oxidative metabolism and mitochondrial biogenesis (Wu et al., 1999, Monsalve et al., 2000). Multiple steroids, including products of CYP 1A, 1B, 2A, 2B, 2C, 2D, 3A, 7A, 17A, 19A, 24A1 and 51A metabolism, bind NRs leading to interactions with NR coregulators through LxxLL or FxxLF motifs that modulate the assembly of the spliceosome complex and pre-mRNA splicing (Auboeuf et al., 2005). The androgen receptor (AR), which binds multiple metabolites of CYPs 2A, 2C, 2D, 3A, 17A, 19A, and 21A, can also directly interact with nucleolar splicing factors (e.g. U5snRNPs), indicating a receptor-mediated role in transcription that is coupled to pre-mRNA splicing mechanisms (Zhao et al., 2002). Vitamin D receptor (VDR) activation mediated by metabolites of CYPs 2R, 2J, 3A, 11A, 27A, 24A, 27B can also alter CYP gene expression and splicing, through NR mediated-crosstalk (e.g. PPARs) transduced via interactions with the retinoid X receptor (RXR; Matsuda and Kitagishi, 2013), recruitment of the NCoA62/SKIP coactivator complex (Zhang et al., 2003) and discrete interactions with the heterogeneous nuclear ribonucleoprotein C splicing factor (hnRNPC; Zhou et al., 2014). Traditional VDR signaling in the nucleus is further refined by several non-traditional NR functions operating near the plasma membrane, which alter gene expression via modulation of key membrane-based paracrine signaling pathways, mediated by agents like Wnt and epidermal growth factor (Larriba et al., 2014). Vitamin A metabolites (or retinoids) of CYPs 1A, 2B, 2C and 26, signaling through the retinoic acid receptor (RAR) and RXR, are also known to guide the recruitment of SC35 coactivators to regulate the alternate splicing of protein kinase delta (PKCδ) among other pre-mRNAs (Apostolatos et al., 2010). Collectively, these data reveal a novel CYP-based mechanism for adaptive transcriptome remodeling, whereby xenobiotics and endogenous substrates, monitored by one of several tissue or disease-specific “CYP clouds”, are metabolized in a coordinated fashion that harmonize NR signaling cascades with alternative gene expression and splicing events that promote adaptive responses to cell stress or stimuli (Figure 4).
In conclusion, the human metabolome adapts to substrate burden through the induction of gene transcription, which helps to maintain homeostasis in a well-documented pathway guided by NR binding and signaling events. In this respect, the metabolic response to xenobiotics (via CYP induction) is adaptive in a manner reminiscent of the immune response to viral antigen; there is a recognition phase of the chemical by the CYP active site, an activation phase when the chemical (or CYP metabolite) interacts with the NR, and an effector phase where the coordinated transcription and splicing of CYP transcripts occurs to feedback-modulate NR signaling. The analogy to the immune response is appropriate here in that specific transcript variants are produced in response to a specific chemical stimulus. The ability to tightly control cellular homeostasis via NR-mediated gene expression and alternative splicing implies a high order of sophistication operating in what might be considered a primitive, chemical immune response. Molecules that engage this primitive CYP-based immune system transduce transcriptome-wide biological responses capable of reshaping both the phenotype and "epigenotype" of the cell (via the regulation of both coding and noncoding RNA), allowing for reversible, environmental adaptation, as well as imprinting, which in rare cases, may persist trans-generationally (Hochberg et al., 2011). Improved knowledge of both the adaptive and maladaptive epigenome remodeling processes induced by xenobiotics may ultimately help reconcile inter-individual variability in efficacy and toxicity that plague many FDA-approved drugs. These insights will provide new guidance for developing "individualized" therapeutic strategies more sensitive to a patient's adaptive transcriptome or functional genome, which as exemplified by CYP superfamily of genes, is the ultimate expression of the heritable epigenotype, which appears to remain environmentally-responsive throughout all phases of the human lifecycle.

ACKNOWLEDGMENTS

We would like to thank Dr. Ronald N. Hines (US-EPA) for his thoughtful comments and suggestions concerning this review.
AUTHORSHIP CONTRIBUTIONS

Participated in research design: Annalora, Iversen and Marcus

Performed data analysis: Annalora and Iversen

Wrote or contributed to the writing of the manuscript: Annalora, Iversen and Marcus
REFERENCES


FOOTNOTES

1.) This research was supported by start-up funds from the Office of Research and the College of Agricultural Sciences at Oregon State University (Corvallis, OR) awarded to Drs. Marcus and Iversen.
FIGURE LEGENDS

Figure 1. Comparative Analysis of the Complete Human Cytochrome P450 Transcriptome and the Total Number of Single Nucleotide Polymorphisms – Radar plot comparison of the total number of A.) Human CYP transcript variants and B.) SNPs for each of the 57 human Cytochrome P450 genes are shown, based on a meta-analysis of information of the NCBI’s Ensembl, Pubmed and Aceview databases and the GeneCards.org website. Total transcript variants for individual CYP genes identified in this study are shown superimposed (in pink) over the current number listed in the Aceview database alone (in purple). The expansion in variant numbers we identified highlights the challenge of predicting inter-individual variability in polymorphic CYP gene splicing, based on existing transcript databases. Radar plot comparison of the total human SNPs associated (as reported by Genecards.org), revealed CYP5A1 (or thromboxane synthase 1 (TBXAS1) as the most polymorphic CYP gene (5274), with CYPs 2C19 (2467), 7B1 (3400), 19A1 (2178) and 39A1 (2129) showing higher rates of polymorphism than the average CYP gene (~850 mutations per gene). C.) A bar graph representation of alternative CYP transcript-to-SNP ratios for each gene is shown at the bottom. CYP2E1, with 24 transcript variants and the lowest number of SNPs (33), displays the highest alternative transcript-to-SNP ratio (0.73) among human CYP genes. CYPs 21A2 (0.33), 27B1 (0.15) and CYP2D6 (0.11) also skew above normal for this ratio (~0.05 alt-transcripts per SNP). Improved understanding of the SNP-based mechanisms that alter CYP gene splicing will facilitate the identification of novel SNP biomarkers that are predictive of drug interactions and human disease.

Figure 2. Tissue-specific and Transformation-sensitive Alternative Splicing of CYP1A1 in Humans – CYP1A1 is inducible in virtually every tissue of the body, however a brain-specific, CYP1A1 splice variant has been identified that preferentially skips exon 6 (Kommadi et al., 2007). This shortened form of the enzyme is spectrally-active, but does not metabolize benzo(a)pyrene to toxic metabolites in brain. Structural analysis of the CYP1A1 crystal structure (PDB: 4I8V)
indicates removal of exon 6 (in yellow ribbon) would: 1) expand the opening of the pw2b substrate access channel; 2) reshape the ligand binding pocket by altering the β1-4 sheet; and 3) alter the redox partner binding surface via elimination of the K’ helix. A similar pattern of tissue-specific, alternative exon 2 usage in CYP1A1 has been reported in tumor cells (Leung et al., 2005). Ovarian cancer cells skip an 84 bp cryptic intron in exon 2 of CYP1A1 but remain in-frame, to produce a catalytically-unique splice variant with diminished estradiol metabolism that localizes to the nucleus and mitochondria, rather than the ER. Cryptic intron removal in exon 2 eliminates 28 amino acid residues among helices E, F, and the E-F loop (shown in transparent gold ribbon), which putatively expands the opening of the pw3 substrate access channels located above the heme center, and helix I.

**Figure 3. Alternative-Splicing of the Human CYP24A1 Gene: Tissue- and Tumor-Specific Exon 1, 2 and 10 Inclusion** – CYP24A1 is a mitochondrial P450 comprised of 11 coding exons that metabolizes the vitamin D hormone to regulate its role in calcium homeostasis, cell growth and immunomodulation. Alternatively spliced transcript variants have been identified for this gene, including a CYP24A1 splice variant (CYP24sv) specifically-expressed in macrophage (Ren et al., 2005), which skips exons 1 and 2. The 372 amino acid variant protein (~43 kDa) lacks the N-terminus, helices A’, A, B and B’ and portions of the β1 sheet, but likely retains the ability to bind heme and substrate in some capacity. Ren and co-authors concluded that CYP24sv represents a cytosolic variant with dominant negative function that quarantines 25-hydroxyvitamin D3 and other hormonal forms of vitamin D, slowing the rate of their metabolism in the mitochondria or ER. Structural analysis of CYP24A1 (PDB: 3K9V) suggests that CYP24sv’s unique N-terminus, derived from intron 2 (shown in green), reshapes the pw2a substrate access channel, and provides contacts for sealing the ligand binding pocket via interactions with helices F-G and the β1 and β4 sheets systems. It is notable that exon 1 and 2 exclusion in CYP24A1 is exacerbated in human tumors of the breast and colon, where N-terminally-truncated splice
variants of approximately 40, 42, and 44 kDa have been identified (Fischer et al., 2009; Horváth et al., 2010; Scheible et al., 2014). An additional, prostate cancer-related, splice variant of CYP24A1 that cleanly skips exon 10 has also been described (Horvath et al., 2010; Muindi et al., 2007). Exon 10 encodes much of the proteins proximal surface, including the β1-3 sheet, meander region, CYS loop, and portions of the L-helix involved in heme-thiolate bond formation. Variants lacking exon 10 may express defects in heme and redox partner binding, giving rise to an alternative dominant negative form that may retain membrane binding features; loss of exon 10 could also refine the alternative functions of cytosolic splice variants lacking exons 1 and 2.

Figure 4. Endo-xenobiotic Crosstalk among Cytochrome P450 and Nuclear Receptor Genes Coordinate Alternative Splicing and Resemble a Primitive Immune System—Human tissues are subject to exposure from over 400 FDA approved drugs, > 10,000 xenobiotics, and untold numbers of endogenous substrates and their metabolites (x > 100,000). Cytochrome P450 genes participate in phase I detoxification of many of these compounds, including model substrates Benzo[a]pyrene (via CYP3A4) and Calcifediol (via CYP24A1). CYP genes are classically induced to silence endo-xenobiotic signaling through cognate nuclear receptors, which modulate global gene expression and splicing events by “coloring” or modulating the composition of coregulatory factors that comprise both the transcription complex and the spliceosome complex, which ultimately alter the nature of both ribosome assembly and gene expression (Auboeuf et al., 2005). Model substrates are subject to metabolism by a finite population of CYPs in a given tissue, however, because each gene is sensitive to an infinite number of environmentally-sensitive, alternative splicing events, each individual may express a unique, tissue-specific “CYP gene cloud” comprised of both wild-type (WT) and splice-altered variant forms (e.g. SV1, SV2, etc.). CYP splice variants can: 1) display reduced ability to metabolize model substrates; 2) function as dominant negatives to sequester compounds from metabolism or potentiate basal NR-mediated signaling; or 3) function as a conformationally-distinct protein
with alternative metabolic function or cellular role. When coupled with existing paradigms of alternate CYP trafficking and membrane-associated cooperativity, an integrated network of cross-talk among 57 CYP and 48 NR genes begins to emerge, as novel CYP metabolites may engage NR signaling pathways in unique ways that re-program gene splicing and expression to promote cellular homeostasis in the face of endocrine disruption. NR signaling cascades can alter both the transcriptome and epigenome of an individual, providing an elegant feedback mechanisms for adapting to cellular stress created by one's unique personal history and disease-status.
TABLE 1. Pathologies Associated with CYP Splice Variants

<table>
<thead>
<tr>
<th>CYP Isoform</th>
<th>Variant Form</th>
<th>Related Disease</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A1</td>
<td>ΔExon 2 and 6</td>
<td>Ovarian Cancer</td>
<td>Chinta et al. (2005) <em>J Neurochem</em> 93(3):724-736</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Leung et al. (2005) <em>Cancer Res</em> 65:3726-3734</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Li et al. (2014) <em>J Proteome Res</em> 13(5):2679-2687</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>ΔExons 4-6</td>
<td>Liver/Colon Cancer</td>
<td>Miles et al. (1989) <em>Nucl Acids Res</em> 17(20):8241-8255</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nakayama et al. (2002) <em>Am Heart J</em> 143(5):797-801</td>
</tr>
<tr>
<td>CYP11B2</td>
<td>Trans-splicing to CYP11B1</td>
<td>Congenital Adrenal Hyperplasia</td>
<td>Hampf et al. (2001) <em>Clin Endocrinol Metab</em> 86(9):4445-4452</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>ΔExons 2, 4, 6 or 8</td>
<td>17-alpha Hydroxylase Deficiency, Congenital Adrenal Hyperplasia</td>
<td>Yamaguchi et al. (1998) <em>Hum Genet</em> 102(6):635-639</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chen et al. (2014) <em>Sci World J</em> 2014:810561</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>CYP27A1</td>
<td>∆Exons 4, 6, 7</td>
<td>Cererotendinous Xanthomatosis (CTX) Garuti et al. (1996) <em>J Lipid Res</em> 37(7):1459-1467</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Verrips et al. (1997) <em>Hu Genet</em> 100(2):284-286</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chen et al. (1998) <em>Biochemistry</em> 37(43):15050-15056</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wu et al. (2007) <em>Endocrinology</em> 148(7):3410-3418</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fischer et al. (2009) <em>Anticanc Res</em> 29(9):3627-3633</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Summary of Human CYP Splice Variants

<table>
<thead>
<tr>
<th>Enzyme Family</th>
<th>Transcript Size (nucleotides)</th>
<th>Exons</th>
<th>Amino Acids</th>
<th>Mol. Weight (kDa)</th>
<th>Total # of Variant RNAs</th>
<th>Total # of SNPs</th>
<th>Disease Links</th>
<th>Ratio of Viable Protein (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1</td>
<td>2683-5988</td>
<td>3-7</td>
<td>512-543</td>
<td>58.2-60.8</td>
<td>49†</td>
<td>1052</td>
<td>203</td>
<td>0.64† (31)</td>
</tr>
<tr>
<td>CYP2</td>
<td>1681-5024</td>
<td>5-9</td>
<td>491-596</td>
<td>55.7-67.6</td>
<td>281</td>
<td>11853</td>
<td>413</td>
<td>0.50 (141)</td>
</tr>
<tr>
<td>CYP3</td>
<td>1604-2167</td>
<td>13-14</td>
<td>502-534</td>
<td>57.1-61.3</td>
<td>120</td>
<td>3029</td>
<td>93</td>
<td>0.47 (56)</td>
</tr>
<tr>
<td>CYP4</td>
<td>1678-3395</td>
<td>5-16</td>
<td>505-638</td>
<td>59.1-71.3</td>
<td>170</td>
<td>9725</td>
<td>37</td>
<td>0.53 (90)</td>
</tr>
<tr>
<td>CYP5</td>
<td>2093</td>
<td>13</td>
<td>594</td>
<td>66.9</td>
<td>30</td>
<td>5274</td>
<td>6</td>
<td>0.48 (14)</td>
</tr>
<tr>
<td>CYP7</td>
<td>2765-2950</td>
<td>6</td>
<td>504-506</td>
<td>57.7-58.3</td>
<td>4</td>
<td>3688</td>
<td>35</td>
<td>0.80 (3)</td>
</tr>
<tr>
<td>CYP8</td>
<td>3974-5632</td>
<td>1-10</td>
<td>501-527</td>
<td>58.1-59.5</td>
<td>9</td>
<td>1752</td>
<td>16</td>
<td>0.77 (7)</td>
</tr>
<tr>
<td>CYP11</td>
<td>2311-3558</td>
<td>9</td>
<td>503-521</td>
<td>57.6-60.1</td>
<td>52</td>
<td>1807</td>
<td>74</td>
<td>0.57 (30)</td>
</tr>
<tr>
<td>CYP17</td>
<td>2001</td>
<td>8</td>
<td>508</td>
<td>57.4</td>
<td>13</td>
<td>179</td>
<td>31</td>
<td>0.57 (7)</td>
</tr>
<tr>
<td>CYP19</td>
<td>4479</td>
<td>10</td>
<td>503</td>
<td>57.9</td>
<td>47</td>
<td>2178</td>
<td>45</td>
<td>0.57 (27)</td>
</tr>
<tr>
<td>CYP20</td>
<td>1811</td>
<td>13</td>
<td>470</td>
<td>53.4</td>
<td>21</td>
<td>1308</td>
<td>1</td>
<td>0.54 (11)</td>
</tr>
<tr>
<td>CYP21</td>
<td>2153</td>
<td>10</td>
<td>495</td>
<td>56.0</td>
<td>30</td>
<td>91</td>
<td>35</td>
<td>0.52 (16)</td>
</tr>
<tr>
<td>CYP Form</td>
<td>Transcript ID</td>
<td>Base Pair Difference</td>
<td>Predicted Protein Molecular Weight (kDa)</td>
<td>Unique Transcript Identification</td>
<td>Total Number of Predicted Proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>----------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP24</td>
<td>3266</td>
<td>12</td>
<td>58.9</td>
<td>17</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP26</td>
<td>1569-4536</td>
<td>6-7</td>
<td>56.2-57.5</td>
<td>23</td>
<td>858</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td></td>
<td></td>
<td>(16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP27</td>
<td>2579-5117</td>
<td>8-9</td>
<td>42.6-60.2</td>
<td>55</td>
<td>1582</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td></td>
<td></td>
<td>(32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP39</td>
<td>2623</td>
<td>12</td>
<td>54.1</td>
<td>15</td>
<td>2179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP46</td>
<td>2186</td>
<td>15</td>
<td>56.8</td>
<td>11</td>
<td>952</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP51</td>
<td>2817</td>
<td>6</td>
<td>98.0</td>
<td>18</td>
<td>388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1569-5632</strong></td>
<td><strong>1-16</strong></td>
<td><strong>42.6-98.0</strong></td>
<td><strong>965</strong></td>
<td><strong>48,384</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>57</strong></td>
<td></td>
<td></td>
<td><strong>1097</strong></td>
<td><strong>0.534</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(515)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Total Number of Variant RNA Transcripts represents the total of all non-redundant transcripts identified for human CYP forms in the AceView, ENSEMBL and Pubmed database. Transcripts with the same predicted protein molecular weight (in kilodaltons (kDa)) with a less than 4 base pair difference in nucleotide transcript length were considered the same transcript. Data for SNPs was obtained from GeneCards.org, and Disease Links were obtained from the MalaCards.org.

‡ Ratio of Viable Protein represents the fraction of human CYP variant transcripts predicted to code functional proteins using the highest confidence scoring metrics for the Aceview (very good) and ENSEMBL (Protein coding/merged Ensembl/Havana). Total number of predicted variant proteins indicated in parenthesis.
Figure 3

Figure 4