Minireview

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, and Patrick L. Iversen

Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon

Received September 2, 2016; accepted February 6, 2017

ABSTRACT

The human genome encodes 57 cytochrome P450 genes, whose enzyme products metabolize hundreds of drugs, thousands of xenobiotics, and unknown numbers of endogenous compounds, including steroids, retinoids, and eicosanoids. Indeed, P450 genes are the first line of defense against daily environmental chemical challenges in a manner that parallels the immune system. Several National Institutes of Health databases, including PubMed, AceView, and Ensembl, were queried to establish a comprehensive analysis of the full human P450 transcriptome. This review describes a remarkable diversification of the 57 human P450 genes, which may be alternatively processed into nearly 1000 distinct mRNA transcripts to shape an individual's P450 proteome. Important P450 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 P450 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 P450 expression is influenced also 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 P450-nuclear receptor gene interactions, resembles a primitive immune-like system that can rapidly monitor, respond, and diversify to accommodate 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

Mapping the human genome sequence was completed in 2001 (Lander et al., 2001), helping to usher in the “postgenomic 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 precursor or unprocessed messenger RNA (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 physiologic and pathologic 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 previously described (Blencowe, 2006; Roy et al., 2013). The molecular basis for these mechanisms is 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 Fig. 1). Here we will focus primarily on the spectrum of phenotypic outcomes induced by alternative transcript splicing, in order to highlight the array of splice-sensitive features operating in the cytochrome P450 (P450) superfamily, a collection of 57 human genes that coordinate the metabolism of both drugs and endoxenobiotics.

This work focuses new attention on the biologic impact of alternative P450 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

ABBREVIATIONS: BP, benzo(a)pyrene; ER, endoplasmic reticulum; GWAS, genome-wide association studies; NMD, nonsense-mediated decay; NR, nuclear receptor; P450, cytochrome P450; PGx, pharmacogenomics; PPAR, peroxisome proliferator-activated receptors; pre-mRNA, precursor or unprocessed messenger RNA; ROS, reactive oxygen species; SNP, single-nucleotide polymorphism; snRNA, small nuclear RNA; VDR, vitamin D receptor.
Hapmap Project, which enables single-nucleotide polymorphism (SNP) arrays for \(>100,000\) SNPs; the 1000 Genomes Project; the ENCODE project for noncoding RNAs; and the National Human Genome Research Institute (NHGRI) genome-association studies (GWAS) catalog. GWAS have now identified two P450s that represent a biomarker for disease or a drug-response phenotype, including: 1) a CYP2C8 association with bisphosphonate-related osteonecrosis of the jaw in multiple myeloma (Sarasquete et al., 2008); and 2) a CYP2C19 association with clopidogrel interindividual 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. In this regard, we feel a new appreciation is needed for the complexity of alternative splicing. For not only is the world of both coding and noncoding 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 (exonic-splicing enhancers, exonic-splicing silencers, intronic splicing silencers, and intronic-splicing enhancers) with tissue-specific, trans-acting, splicing regulatory factors. 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 referencing 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 P450 splice variants, and their relationship to human health and disease, remains poorly understood. A summary of known P450 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 whose sole basis is 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 physiologic impact of genetic variations among developmentally regulated P450 genes (Hines and McCarver, 2002). To expand on this concept, and explore the array of alternative splicing products that serve as an improved platform for predicting disease whose basis is the functional genome, we annotated all of the known human P450 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. From this computational meta-analysis, we found that the 57 human P450 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 P450

### TABLE 1

<table>
<thead>
<tr>
<th>P450 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)</td>
</tr>
<tr>
<td>CYP1B1</td>
<td>Partial ΔExon 2</td>
<td>Glaucoma</td>
<td>Stoelov et al. (1998)</td>
</tr>
<tr>
<td>CYP2A6</td>
<td>Deletions</td>
<td>Lung cancer</td>
<td>Liu et al. (2013b)</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>ΔExons 4–6</td>
<td>Liver/colon cancer</td>
<td>Miles et al. (1998)</td>
</tr>
<tr>
<td>CYP8A1</td>
<td>ΔExons 2 and 9</td>
<td>Essential hypertension</td>
<td>Nakayama et al. (2002b)</td>
</tr>
<tr>
<td>CYP11B2</td>
<td>Trans-splicing to CYP11B1 ΔExons 2, 4, 6, or 8</td>
<td>Congenital adrenal hyperplasia, 17-α Hydroxylase deficiency, congenital adrenal hyperplasia</td>
<td>Hamps et al. (2001)</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>ΔExons 2, 4, 6, or 8</td>
<td>Congenital adrenal hyperplasia</td>
<td>Yamaguchi et al. (1998)</td>
</tr>
<tr>
<td>CYP19A1</td>
<td>ΔExon 5</td>
<td>Aromatase deficiency</td>
<td>Costa-Santos et al. (2004)</td>
</tr>
<tr>
<td>CYP24A1</td>
<td>ΔExons 9–11</td>
<td>Breast, colon, and prostate cancer</td>
<td>Muindi et al. (2007)</td>
</tr>
<tr>
<td>CYP26B1</td>
<td>ΔExon 2</td>
<td>Atherosclerosis</td>
<td>Elmabsout et al. (2012)</td>
</tr>
<tr>
<td>CYP27A1</td>
<td>ΔExons 4, 6, 7</td>
<td>Oral cancer</td>
<td>Chen et al. (2014)</td>
</tr>
<tr>
<td>CYP27B1</td>
<td>ΔExons 4–7</td>
<td>Cerebrotendinous xanthomatosis</td>
<td>Garuti et al. (1996)</td>
</tr>
<tr>
<td>CYP6A1</td>
<td>Intron 2</td>
<td>Glomblasta, leukemia, melanoma, cervical cancer, ovarian cancer</td>
<td>Verrips et al. (1997)</td>
</tr>
<tr>
<td>CYP27B1</td>
<td>ΔExons 4–7</td>
<td>Glioblastoma, leukemia, melanoma, cervical cancer, ovarian cancer</td>
<td>Chen et al. (1998)</td>
</tr>
<tr>
<td>CYP27B1</td>
<td>ΔExon 2</td>
<td>Alzheimer disease</td>
<td>Li et al. (2013a)</td>
</tr>
<tr>
<td>CYP6A1</td>
<td>Intron 2</td>
<td>Alzheimer disease</td>
<td>Li et al. (2013b)</td>
</tr>
</tbody>
</table>
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 P450 gene encodes nearly 20 unique mRNA transcripts, capable of yielding both reference and splice-variant proteins, and an ensemble of noncoding RNA molecules. Over 53% of these transcripts (515) are predicted to express viable proteins, on the basis of the most stringent expression criteria used by both the AceView and Ensembl databases (Table 2). Subsequent analysis of the GeneCards.org website revealed that the 57 human P450s are associated with over 48,384 SNPs, approximating to roughly 850 mutations per P450 gene (Stelzer et al., 2011). Although most P450s fall within this general range of genetic variation, some, including CYPs 2C19, 5A1, 19A1, and 39A1, express more than twice as many SNPs as the average, suggesting the accumulation process is nonrandom (Table 2). The total number of diseases associated with each of the 57 human genes (1057) was also calculated using the MalaCards Human Disease Database (Rappaport et al., 2014). P450 families 1 (203) and 2 (413) account for over half of the total disease associations, whereas 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 P450 splice variants and SNPs for each of the 57 human P450 genes is shown in Fig. 1. Pharmacologically relevant drug targets, such as CYP2C9, CYP2D6, CYP3A5, CYP11A1, and CYP19A1, are some of the most highly spliced P450 genes, the majority of which are also subject to trans-splicing events. Although 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 P450 transcriptome may be far superior to those resulting from genotyping alone, as SNP approaches may tend to overstate (or oversimplify) the biologic influence of a point mutation on gene function. Although computational studies indicate that 85% of all human mRNA comes from a single major transcript (González-Porta et al., 2013), the P450 superfamily of genes is highly diverse at the genomic level, making it uniquely sensitive to alternate splicing events that alter the biologic 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 recontextualize the utility and complexity of the convoluted metabolic crosstalk network that underlies the 48 nuclear receptor (NR) signaling pathways operating in humans (Meyer, 2007; Tralau and Luch, 2013).

Furthermore, drug metabolizing P450s of families 1–4 have traditionally been considered more polymorphic than P450s that metabolize endogenous, cholesterol-based substrates and are assumed to be more sensitive to SNP-based modulation of alternative splicing (Lewińska et al., 2013). On the basis of our analysis, it does appear 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 P450 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 owing to their central role in the production of sex hormones, progestins, and corticoids. In contrast, the 35 drug-metabolizing P450s that compose P450 families 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. Although 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

### Table 2

**Summary of human P450 splice variants**

<table>
<thead>
<tr>
<th>Enzyme Family</th>
<th>Transcript Size</th>
<th>Exons</th>
<th>Amino Acids</th>
<th>Mol. Weight</th>
<th>Variant RNAs</th>
<th>SNPs</th>
<th>Disease Links</th>
<th>Ratio of Viable Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1 (3)</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 (16)</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 (4)</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 (12)</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 (1)</td>
<td>2093</td>
<td>13</td>
<td>594</td>
<td>66.9</td>
<td>30</td>
<td>5274</td>
<td>48</td>
<td>0.48 (14)</td>
</tr>
<tr>
<td>CYP7 (2)</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 (2)</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 (3)</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 (1)</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 (1)</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 (1)</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 (1)</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>CYP24 (1)</td>
<td>3266</td>
<td>12</td>
<td>514</td>
<td>58.9</td>
<td>17</td>
<td>711</td>
<td>19</td>
<td>0.46 (8)</td>
</tr>
<tr>
<td>CYP26 (3)</td>
<td>1569–4536</td>
<td>6–7</td>
<td>497–522</td>
<td>56.2–57.5</td>
<td>23</td>
<td>858</td>
<td>22</td>
<td>0.68 (16)</td>
</tr>
<tr>
<td>CYP27 (3)</td>
<td>2579–5117</td>
<td>8–9</td>
<td>372–531</td>
<td>42.6–60.2</td>
<td>55</td>
<td>1582</td>
<td>58</td>
<td>0.38 (32)</td>
</tr>
<tr>
<td>CYP39 (1)</td>
<td>2623</td>
<td>12</td>
<td>469</td>
<td>54.1</td>
<td>15</td>
<td>2179</td>
<td>2</td>
<td>0.50 (8)</td>
</tr>
<tr>
<td>CYP46 (1)</td>
<td>2186</td>
<td>15</td>
<td>500</td>
<td>56.8</td>
<td>11</td>
<td>952</td>
<td>7</td>
<td>0.57 (6)</td>
</tr>
<tr>
<td>CYP51 (1)</td>
<td>2817</td>
<td>6</td>
<td>860</td>
<td>98.0</td>
<td>18</td>
<td>388</td>
<td>2</td>
<td>0.65 (12)</td>
</tr>
<tr>
<td>Total 57</td>
<td>1569–5632</td>
<td>1–16</td>
<td>372–860</td>
<td>42.6–98.0</td>
<td>965</td>
<td>48384</td>
<td>1097</td>
<td>0.534 (515)</td>
</tr>
</tbody>
</table>

^Ratio of viable protein represents the fraction of human P450 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.

^Total number of variant RNA transcripts represents the total of all nonredundant transcripts identified for human P450 forms in the AceView, Ensembl, and PubMed databases. Transcripts with the same predicted protein molecular weight (in kDa) with a fewer than four base-pair differences 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 database.
its substrates (Hecker and Ullrich, 1989). CYP5A1-based SNPs represent more than 10% of all P450 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 P450 genes (see Fig. 1). CYPs 21A2 (0.33), 27B1 (0.15), and CYP2D6 (0.11) also skew above normal P450 superfamily genes with respect to the average alternative transcript-to-SNP ratio of 0.05. Improved understanding of how key mutations alter P450 gene splicing may be critical to addressing the goal of precision medicine in the post-genomic era.

![Fig. 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 P450 transcript variants and (B) SNPs for each of the 57 human cytochrome P450 genes are shown, on the basis of a meta-analysis of information of the NCBI’s Ensembl, PubMed, and AceView databases and the GeneCards.org website. Total transcript variants for individual P450 genes identified in this study are 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 interindividual variability in polymorphic P450 gene splicing, predicated on existing transcript databases. Radar plot comparison of the total human SNPs associated (as reported by GeneCards.org) revealed CYP5A1 (or thromboxane synthase 1) as the most polymorphic P450 gene (5274), with CYPs 2C19 (2467), 7B1 (3400), 19A1 (2178), and 39A1 (2129) showing higher rates of polymorphism than the average P450 gene (~850 mutations per gene). (C) Bar graph representation of alternative P450 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 P450 genes. CYPs 21A2 (0.33), 27B1 (0.15), and CYP2D6 (0.11) also skew above normal for this ratio (~0.05 alternative transcripts per SNP). Improved understanding of the SNP-based mechanisms that alter P450 gene splicing will facilitate the identification of novel SNP biomarkers that are predictive of drug interactions and human disease.]

Tissue-Specific Alternative Splicing of Cytochrome P450 Genes: Historical Perspective

During the late 1980s and early 1990s, several groups began to appreciate the possibility that differential RNA splicing might serve as an additional mode for regulating P450 function. In 1987, the first P450 splice variant for hepatic P450-1, later renamed CYP2C8, was reported.
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 three 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 physiologic importance of the more unusual P450 splice variant and chimeric forms remains unclear, particularly with respect to how their variable, tissue-specific expression alters interindividual sensitivity to hormones, pharmaceuticals, and xenobiotics. As discussed above, a single P450 gene can be translated into multiple, variant protein sequences, solely on the basis of tissue-specific differences in the expression of key gene-splicing machinery ([e.g., small nuclear RNA (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 (Kommeddah et al., 2007). Furthermore, cells expressing wild-type CYP1A1 generate DNA adducts via the formation of reactive 3-OH-BP products, whereas cells expressing the Δexon6-CYP1A1 variant in brain do not (Kommeddah et al., 2007).

On the basis of the structural analysis of CYP1A1 in Fig. 2 (derived from PDB: 4I8V; Walsh et al., 2013), tissue-specific, alternative splicing of exon 6 may reshape 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 complemented 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 Fig. 2, targeted skipping of an 84-bp cryptic intron in exon 2 of CYP1A1 produces a catalytically unique splice variant with diminished estradiol metabolism that localizes to the nucleus and mitochondria rather than the endoplasmic reticulum (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 classic structural studies of wild-type P450 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, 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 P450 gene platform, one that allows cells to sample a continuum of gene products as dictated by tissue-specific physiologic stress conditions. However, it should also be clearly noted that not all P450 genes are divided into an equal number of coding and noncoding 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 two coding exons (Supplemental Fig. 2) and, therefore, is
not subject to the same patterns of alternative splicing as other CYP1 family members 1A1 and 1A2, which have six 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 P450 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 (Faaij 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 related more to the rate at which a given mRNA transcript is processed by the spliceosome and ribosome, rather than other evolutionary factors (Lejeune and Maquat, 2005; McGlincey and Smith, 2009). 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 intrinsic complexity clearly influences its sensitivity to alternate-splicing events and the domain swapping paradigm. The tissue-specific use of alternative promoters and non-AUG translation start-sites also is facilitated by this type of hidden gene complexity, which allows discrimination among a set of well defined exons and poorly defined pseudoexons.

In 2005, a truncated, CYP24A1 splice variant incapable of metabolizing hormonal forms of vitamin D was identified in the human macrophage (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 eight 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 Fig. 3 (on the basis of 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 used 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 retinal) into vitamin A2 (all-trans 3,4-dehydroretinal) in cell culture (Kramlinger et al., 2016). It is intriguing that CYP27C1 may represent a modified version of the more soluble P450 isoform recapitulated by CYP24A1-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 P450 regulation, particularly for variants subject to atypical subcellular trafficking events.

Therefore, although increasingly common in the literature, the functional significance of many tissue-specific P450 variants remains elusive. CYP2C8, for example, is constitutively expressed in the liver at 5–7% of the total tissue P450, has more than 20 known alternatively spliced variants, and is a breast cancer biomarker owing to its high expression in mammary tumors (Knüppel et al., 2004). Bimodal targeting to both mitochondria and ER have been reported for CYP2C8 and several other P450s, including CYP1A1, 1B1, 2B1, 2E1, and 2D6. Although 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 unaltered 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 P450 splice variants subject to alternate subcellular trafficking is shown in Supplemental Table 1. Computational analysis from the AceView
Transformation-Sensitive Alternative Splicing in the Cytochrome P450 Superfamily

This transformation-specific, P450 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; Bouraima et al., 2001). More recently, it was determined that CYP19A1’s exon 5 also is defined poorly and subject to both splicing mutations and physiologic 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 nonsteroidogenic tissues (Lin et al., 2007; Pepe et al., 2007). These observations suggest a tissue-specific, regulatory mechanism for aromatization on the basis of alternative exon 5 inclusion in humans. A structural analysis of this CYP19A1 splice variant is depicted in Supplemental Fig. 3 (on the basis of PDB: 3S79; Ghosh et al., 2012). In aromatase, exon 5 skipping naturally suppresses androgen metabolism by redefining 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., Arimidex (anastrozole), Aromasin ( exemestane), Femara ( letrozole), and Teslac (testolactone)] remains unclear (Hadfield and Newman, 2012) and an area of active investigation (Liu et al., 2013a; Liu et al., 2016).

The tissue-specific, alternative start-site usage of CYP24A1, discussed above, appears to be exacerbated in several human transcript variants in which an array of truncated splice variants of approximately 40, 42, and 44 kDa have been identified with similar N-terminal modifications (Fischer et al., 2009a; 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 (Muindi et al., 2007; Horvath et al., 2010). 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 Fig. 3). Although the function of this cancer-specific variant remains unknown, it may have reduced ability to bind heme and redox partners compared with 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., 2009a; Scheible et al., 2014), and colon (Horváth et al., 2010; Peng et al., 2012). It is thought that whereas 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., 2015). 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, prodifferentiation 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 isof orm present. In malignant tissues, three 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., 2009a; 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 three detectable CYP24A1 splice variants (Horváth et al., 2010). Peng and coworkers (2012) 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.

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 (Diesel et al., 2004; Fischer et al., 2007, 2009b; Wu et al., 2007). Tissue-specific CYP27B1 splicing also is seen in healthy skin, where a 59-kDa variant with a 3-kDa insertion between exons 2 and 3 and exclusive to skin was observed and linked to the activity of CAMP response element–binding elements (Flanagan et al., 2003). Fluctuations in skin cell density, calcium concentrations, and UV-B light exposure were also shown to affect 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 ways that are difficult to predict.

Research focused on CYPs 19A1, 24A1, and 27B1 is among the most comprehensive on transformation-sensitive alternative splicing in human P450s, and the literature is replete with anecdotal reports of alternative P450 splicing among various transformed cell types and tissues. For example, CYP2E1 transcripts that selectively skip exons (2, 2–3, and 2,4,5–7,6) were identified in lung carcinoma cell lines but not in corresponding lung tumor tissue or liver extracts, where splice variants were found to be noninducible by ethanol exposure (Bauer et al., 2005). The authors noted at the time that none of these splice variants were accounted for in the National Institutes of Health (NIH) 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 prodrug targets for chemoprevention strategies (Stenstedt et al., 2012). Unfortunately, as discussed above, the alternative splicing behavior of orphan P450s is often as poorly understood as are their cryptic functions.

Interestingly, orphan CYP2W1, which is expressed in 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 P450s may only be invoked under extreme cellular stress, when an alternative P450 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 and Kenna, 1996). The NIH’s AceView and Ensembl (Cunningham et al., 2015) databases indicate that 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 P450 transcripts we identified in public database searches were associated with a well-characterized human tumor type. Improved understanding of their tumor-specific cellular roles and the mechanisms regulating their induction should aid in the identification of improved disease biomarkers and P450 variant-specific therapeutic agents.

**Trans-Splicing of Cytochrome P450 Gene Chimeras**

It is well understood that both transformation-sensitive cellular changes and P450 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 P450 genes are created via trans-splicing events among extended (or multiple) pre-mRNA transcripts. Therefore, although mutations in a gene like CYP2C9 are linked to poor metabolism of several drugs, including phenytoin and tolbutamide, the role that alternative or trans-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 Fig. 4A) hint that removal of exon 2 would reshape portions of the reference protein relative to its membrane-binding surface, substrate access channel, active site, and redox partner binding surface. CYP2C9 is located within a cluster of P450 genes on chromosome 10q24 and is subject to nonrandom, trans-splicing events with neighboring CYPs 2C8, 2C18, and 2C19 in liver and skin, where CYP2C18 exon 1–like sequences were found spliced 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 interindividual basis, highlighting another major flaw in GWAS targeting of the components of gene clusters.

The CYP2D gene family also is complex, clustered, and subject to trans-splicing events; it is composed 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, 1997), including a highly expressed splice variant that completely skips exon 6. As shown in Supplemental Fig. 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 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 poorly 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 CYP3A gene family (comprising CYPs 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 mRNA have been described, in which CYP3A43 exon 1 is joined at CYP3A4 and CYP3A5 canonical splice sites, implying a trans-splicing mechanism that bypasses classic transcriptional control paradigms. In Supplemental Fig. 5A, a structural analysis of CYP3A4 (on the basis of PDB: 4K9T; Sevrioukova and Smith, 2008) 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; McGlynn 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 Fig. 5B), hinting that CYP3A3 chimeras may encode a spectrum of 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; it 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 B1-1 sheet (Supplemental Fig. 5C). A CYP3A4 trans-splicing variant containing CYP3A43 exon 1, followed by CYP3A4 exons 7–13, has been studied also, and it displays minimal 6-β-hydroxylase activity (Supplemental Fig. 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 nonsynonymous 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 P450 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 behaviors have now been documented for several important P450 gene families; however, 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 P450 transcripts. Some P450 SNPs are exceedingly rare, occurring in less than 1% of the population. These rare mutations are linked to congenital defects, including glaucoma (CYP1B1; Stoiilov et al., 1998; Tanwar et al., 2009; Sheikh et al., 2014), 17-α 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; Lee, 2013; Szabó et al., 2013; Sharaf et al., 2015), spinia bifida (CYP26A1; Rat et al., 2006), focal facial dermal dysplasia (CYP26C1; Slavotinek et al., 2013), and cerebrotendinous xanthomatosis (CYP27A1; Garuti et al., 1996; Verris et al., 1997; Chen et al., 1998; Tian and Zhang, 2011). Other P450 polymorphisms that alter splicing are associated with neurologic and metabolic diseases, including Parkinson’s disease (CYP2D6, Denson et al., 2005; CYP2J2, Searles Nielsen 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; CYP2FI, 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, de Morais et al., 1994; Ibeanu et al., 1999; Satyanarayana et al., 2009; Sun et al., 2015; CYP2D6, Toscano et al., 2006; Lu et al., 2013; Wang et al., 2014; CYP3A4, Elen et al., 2011; and CYP3A5, Kuehl et al., 2001; Busi and Creteil, 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 P450 transcriptome is so important when selecting a personalized medicine regimen.

It is now clear that numerous P450-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 P450 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 who 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. Although variations in the expression of ubiquitous splicing factors (e.g., U1 small nuclear ribonuclear proteins) may account for interindividual 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 physiologic 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.

Addressing Nontraditional P450 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 P450 structure/function paradigms. Well studied splice variants from the CYP1A1 family reveal a sophisticated domain-swapping mechanism that allows for tissue-specific sampling of alternative protein conformations. Discrete arrays of tissue-specific snRNAs (U1–U9) and ancillary splicing factors can alter the biologic syntax of a gene, thus providing a granular mechanism for this natural phenomenon that 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 P450-product (e.g., 3-hydroxybenzo[a]pyrene; 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 P450 gene may be exploited. This observation helps explain why some P450 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 (Rhee 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 P450 splice variants play in normal biology should help reinvigorate discussions of P450 gene structure and evolution, particularly with respect to the selective forces that regulate P450 gene duplication and mutation events. Although 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 P450 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 P450 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 P450 gene, therefore, likely predate the Cambrian explosion 540 million years ago, and lurk somewhere in the history of deuterostome evolution (Nelson, 2003). However, contemporary analyses of P450 splicing mechanisms are needed to help elaborate the relationship between alternative subcellular trafficking and alternative P450 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 P450 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 P450 variants may underlie an even greater array of uncharacterized (and potentially lighting) P450 functions (Zhao and Waterman 2011; Lamb and Waterman, 2013).

As the landscape of alternate P450 gene functions continues to expand, the cellular basis for alternatively targeting microsomal P450 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 reactive oxygen species (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. P450 variants may potentially play a similar role by redirecting the metabolism of reactive subcellular compounds. However, because P450s 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 P450 splicing in the modulation of mitochondrial stress and dysfunction seem warranted. In this regard, it is intriguing that many P450 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), the recognition of new, nonclassic roles for P450 variants linked to disease seems probable.

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 splicing regulatory-factor proteins) and alter RNA polymerase (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 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 P450-mediated ROS modulation (Zimmerly and Semper, 2015).

### Shaping an Improved Roadmap toward Precision Medicine

Because changes in mitochondrial function can alter global alternative splicing events, it is not surprising that both phenomena are linked to the induction of cellular heterogeneity and human disease pathology progression (Raj and van Oudenaarden, 2008; Hanahan and Weinberg, 2011; 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, remains a core focus of pharmacogenomics. Mutations in single genes have now been identified for over 4000 human diseases, of which 5–15% 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 nonsynonymous mutations and a much smaller number alter pre-mRNA processing, leading to a frequency in change of phenotype of only ~0.5%. When one considers that humans express almost 50,000 SNPs across the 57 human P450 genes, fewer than 25 may be capable of significantly altering a patient’s metabolic phenotype, an insight we contend may help streamline PGx approaches to personalized medicine (Zhou et al., 2009; Zanger and Schwab, 2013). For example, although 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) and 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 P450 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 P450-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 RNA sequence strategies for personalized medicine, as accurate identification of the patient’s functional genotype is a crucial component of precision medicine. Although 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, complementary metabolomics...
Endoxenobiotic 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 1 detoxification of many of these compounds, including model substrates benzo[a]pyrene (via CYP3A4) and calcifediol (via CYP24A1). P450 genes are classically induced to silence endoxenobiotic 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 P450s in a given tissue, however, and because each gene is sensitive to an infinite number of environmentally sensitive, alternative splicing events, each individual may express a unique, tissue-specific "P450 gene cloud" comprising both wild-type (WT) and splice-altered variant forms (e.g., SV1, SV2, etc.). P450 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 P450 trafficking and membrane-associated cooperation, an integrated network of crosstalk among 57 P450 and 48 NR genes begins to emerge, as novel P450 metabolites may engage NR signaling pathways in unique ways that reprogram 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 adaptation to cellular stress created by unique personal history and disease status.

![Diagram](https://example.com/diagram.png)
indicating a receptor-mediated role in transcription that is coupled to pre-mRNA splicing mechanisms (Zhao et al., 2002). Vitamin D receptor activation mediated by metabolites of CYP2R2, 2J, 3A, 11A, 27A, 24A, and 27B can also alter P450 gene expression and splicing through NR-mediated crosstalk (e.g., PPARs) transduced via interactions with the retinoid X receptor (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 (Zhou et al., 2015). Traditional VDR signaling in the nucleus is further refined by several nontraditional NR functions operating near the plasma membrane that 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 and retinoid X receptor, are also known to guide the recruitment of SC535 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 P450-based mechanism for adaptive transcriptional remodeling, whereby xenobiotics and endogenous substrates, monitored by one of several tissue- or disease-specific “P450 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 (Fig. 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 P450 induction) is adaptive in a manner reminiscent of the immune response to viral antigen; there is a recognition phase of the chemical by the P450 active site, an activation phase when the chemical (or P450 metabolite) interacts with the NR, and an effector phase in which the coordinated transcription and splicing of P450 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 P450-based immune system transduce transcriptome-wide biologic 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 biologic responses capable of reshaping both the phenotype and functional genome, which as exemplified by P450 superfAMILY of genes is the ultimate expression of the heritable epigenotype and appears to remain environmentally responsive throughout all phases of the human life cycle.

Acknowledgments

The authors thank Dr. Ronald N. Hines (US-EPA) for his thoughtful comments and suggestions concerning this review.

Authorship Contributions

Participated in research design: Annalora, Iversen, Marcus.
Performed data analysis: Annalora, Iversen.
Wrote or contributed to the writing of the manuscript: Annalora, Iversen, Marcus.

References


Alternative P450 Splicing Redirects Human Drug Metabolism

389


Address correspondence to: Dr. Andrew J. Annalora, Department of Environmental and Molecular Toxicology, Oregon State University, 1007 Agriculture and Life Sciences Building, Corvallis, OR 97331. E-mail: Andrew.Annalora@oregonstate.edu