Published in final edited form as:

Clin Pharmacol Ther. 2011 March ; 89(3): 355–365. doi:10.1038/clpt.2010.314.

Pharmacogenomics of the RNA World: Structural RNA Polymorphisms in Drug Therapy

W Sadee¹, D Wang¹, AC Papp¹, JK Pinsonneault¹, RM Smith¹, RA Moyer¹, and AD Johnson²

¹Program in Pharmacogenomics, Department of Pharmacology, The Ohio State University College of Medicine, Columbus, Ohio, USA

²The Framingham Heart Study of the National Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA

Abstract

The use of pharmacogenomic biomarkers can enhance treatment outcomes. Regulatory polymorphisms are promising biomarkers that have proven difficult to uncover. They come in two flavors: those that affect transcription (regulatory single-nucleotide polymorphisms (rSNPs)) and those that affect RNA functions such as splicing, turnover, and translation (termed structural RNA SNPs (srSNPs)). This review focuses on the role of srSNPs in drug metabolism, transport, and response. An understanding of the nature and diversity of srSNPs and rSNPs enables clinical scientists to evaluate genetic biomarkers.

Genetic factors play a significant role in drug response and toxicity. A growing number of genetic variants are being shown to alter the metabolism of drugs and their interactions with target tissues, and this has led to the use of biomarkers to guide drug therapy. Yet only a few validated pharmacogenomic tests (<http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm>) are routinely used in clinical practice. Despite the significant role that genes often play in affecting the course of a disease and the treatment outcome, the nature and extent of genetic variability have been inadequately explored. As a consequence, we currently understand only a few of the relevant genetic factors. This makes it more difficult to address gene–gene interactions, a step that is critical for predicting treatment outcomes that typically involve more than a single gene. With large-scale genome-wide association studies (GWASs) and full-genome sequencing yielding an ever increasing number of candidate genes, it is now essential to unravel the molecular genetic pathways germane to clinical applications. This review addresses gene regulation and focuses on a class of underappreciated genetic polymorphisms in the transcribed regions of genes that affect RNA functions. We use the term “structural RNA single-nucleotide polymorphisms” (srSNPs)^{1,2} to describe these polymorphisms that are now emerging as critical factors in the genetic diversity seen in humans. A broad survey of trait/disease-associated SNPs derived from GWASs reveals that nonsynonymous SNPs account for only ~9% and SNPs in intergenic regions for ~43%, whereas presumed srSNPs account for ~49% (synonymous SNPs, 2%; those in 5'- and 3'-untranslated regions, 2%; intronic, 45%).³ The goal of this review is to enable clinical scientists to recognize the types of genetic variation, understand their relative

significance in the context of environment and target tissue, and assess the validity and strength of evidence supporting the roles of different variations in affecting specified clinical outcomes.

THE EVOLVING VIEW OF GENETIC VARIATION

SNPs are the most frequent contributors to genetic variation; insertions/deletions, copy number variants, and chromosomal rearrangements are other contributors. To simplify this discussion, we use the term “SNP” in relation to genetic variation in general, unless specified otherwise, and propose three main groups characterized by distinct mechanisms. Traditionally, researchers have focused on nonsynonymous SNPs that alter the amino acid sequence of encoded proteins (coding SNPs, cSNPs; Figure 1a). These are easily discovered after sequencing, and tools are available to study their effects on protein functions; their impact is manifested in virtually all tissues in which the protein is expressed. cSNP mutations tend to lead to physiological defects and are therefore negatively selected in evolution, thereby decreasing their frequency relative to that of other types of polymorphisms.

Over the past several years, it has become apparent that another class of polymorphisms is more prevalent than cSNPs, namely, regulatory SNPs (rSNPs; Figure 1) that alter transcription of protein-coding genes (residing mostly in intergenic regions). rSNPs can also affect the expression of noncoding genes, which have emerged as an important part of the cellular machinery;¹ however, their application in pharmacogenomics is still in its infancy. Given that gene regulation depends heavily on the nature of the tissue target, greater flexibility for selective evolutionary paths affecting tissue-specific events can lead to positive selection and high frequencies of certain alleles in the human population. Genome-wide technologies have opened a path for large-scale exploration of mRNA expression quantitative trait loci (eQTLs) that are identified by applying GWASs to mRNA profiles in target tissues (ref. 2 and references therein). The vast majority of SNPs responsible for driving these eQTLs remain unknown, owing to their widely scattered distribution across a gene locus. Researchers typically categorize eQTL variants as likely rSNPs that affect transcription, leading to changes in mRNA expression levels.

A third general type of genetic variant, here termed “structural RNA SNPs” (srSNPs; Figure 1a),^{1–4} may be at least as prevalent as rSNPs. srSNPs are variants that reside in the transcribed portion of a gene locus and are therefore present in each nascent RNA transcript as well as in mature mRNA if present in exons, thereby affecting RNA functions in numerous ways. Figure 1b illustrates the regions of pre-mRNA and mature mRNA in which srSNPs can alter functions. It is important to note that some promoter/enhancer variants reside within transcribed regions, and these are considered to be rSNPs that affect transcription; the classification is driven by the mechanism involved. Similarly, nonsynonymous SNPs may exert their effect at the RNA level, acting as srSNPs rather than (or in addition to) affecting protein activity directly. Therefore, one must be on the alert when assigning biological roles to any variant solely on the basis of location in the gene locus.

As discussed with respect to rSNPs, the effects of srSNPs are often tissue dependent, allowing for evolutionary selection of targeted favorable traits that enhance fitness. As a consequence, we expect some srSNPs also to occur frequently at population levels. Two characteristics promote the potential of srSNPs to alter the functions of a gene. First, RNAs, particularly newly transcribed heterogeneous nuclear RNAs, interact with numerous proteins that recognize motifs within the RNA, thereby enabling RNA processing and various RNA functions.⁵ These motifs can be based on primary structure (sequence) or RNA folding

domains (secondary/tertiary). Any SNPs in these domains can potentially disrupt function. Second, being a single-stranded nucleic acid, RNA folds onto itself in a dynamic fashion that is highly sensitive to sequence alterations. As a result, >90% of single-nucleotide substitutions are predicted to change the folding pattern, as illustrated in a thought experiment depicted in Figure 2. Here we calculated the effect on mRNA folding upon replacing each nucleotide *in silico*, one base at a time, along the entire mRNA coding region of OPRM1, the μ -opioid receptor. The vast majority of substitutions resulted in significant changes in the computed folding pattern, with the potential to affect function, as demonstrated for A118G SNP in *OPRM1* (see Figure 2 and refs. 6,7). Although these thermodynamic calculations may not always be accurate, a physicochemical SNP assay known as single-stranded conformational polymorphism separates >90% of all SNPs on a nondenaturing capillary column on the basis of induced folding changes of the nucleic acid chain. Crucial RNA processes that can be altered by srSNPs are listed in Figure 1a. Among these, alternative or aberrant splicing plays a prominent role in a majority of genetic diseases, displaying high tissue selectivity.⁸ Therefore the study of srSNPs could shed light on the “missing heritability”⁹ in health, disease, and therapy. Before describing examples from the pharmacogenomics literature, we first discuss the discovery and evaluation of regulatory polymorphisms as they affect the interpretation of clinical genetic association studies.

HOW ARE REGULATORY srSNPs DISCOVERED?

Large-scale studies typically employ mRNA expression profiles in combination with genome-wide SNP analyses. Using the relative levels of each mRNA in multiple donors as a quantitative trait, one can scan the genome for any SNPs that associate with the expression of any given transcript (eQTLs).² For each mRNA, one then identifies eQTLs that reside either within the same gene locus expressing the mRNA studied (*cis*-eQTL) or in other regions, implying genetic regulation of *trans*-acting factors (such as transcription factors). Applied to *MAOA* as an example for a *cis*-eQTL (i.e., *MAOA* mRNA levels associate with SNPs in the *MAOA* locus), calculated from published monocyte mRNA expression,¹⁰ multiple SNPs in a large haplotype block of *MAOA* are significantly associated with mRNA levels (Figure 3; for more discussion, see below). *cis*-eQTLs are more abundant—or perhaps more easily detected because of a stronger signal—than *trans*-acting eQTLs (ref. 10 and references therein). GWAS results indicate that rSNPs and srSNPs may account for *cis*-eQTLs in similar proportions,³ although not all intergenic SNPs are rSNPs (affecting noncoding RNA functions in different ways). Traditional bias has favored the recognition of rSNPs over srSNPs, and the possibility of nonsynonymous coding SNPs acting as srSNPs is also often ignored (e.g., *OPRM1* A118G (N40D) (ref. 6)). Moreover, expression arrays do not resolve all splice variants, potentially leaving many splice polymorphisms undetected. Finally, srSNPs that affect translation require separate detection techniques, including measurements of mRNAs bound to ribosomes in the process of translation,¹¹ and, finally, protein levels. Clearly, srSNPs are more prevalent than previously realized.

A refinement of RNA analysis that facilitates detection of rSNPs and srSNPs is determining the allelic ratios of transcripts. Any deviation of the allelic RNA ratio from unity (expected in an autosomal gene), termed “allelic expression imbalance” (AEI), indicates the presence of *cis*-regulatory factors.⁴ AEI offers a more precise relative measure of transcript activity as compared with total mRNA levels, which are confounded by the influence of *trans*-acting factors. Using AEI, we have studied well over 100 genes and have detected frequent and substantial AEI in a number of important pharmacogenomic candidate genes.^{4,6,7,12,13} AEI analysis has also been performed on a genome-wide basis by applying SNP arrays capable of determining the relative allele abundances of heterozygous loci (e.g., Illumina platforms) in both gDNA and RNA transcripts.¹⁴ The resultant AEI ratios reveal *cis*-eQTLs and their

distribution in promoter/enhancer regions (rSNPs) or transcribed regions (mostly srSNPs). Yet many central nervous system (CNS) drug targets are expressed at very low levels; as a result, important genes such as the dopamine and serotonin receptors often fall below the detection threshold of quantitative SNP RNA arrays.

Given the much greater dynamic range of detection now available, we can now apply next-generation sequencing of the transcriptome to uncover more details by determining the species, amount, and allelic ratios of each gene's transcripts. Early results demonstrate unfathomed transcriptome complexity and rather frequent AEIs.¹⁵ An example of possible AEIs from our transcript analysis of human prefrontal cortex tissue is given in Figure 4. From this technology, we can expect a wave of new regulatory variants, filling a knowledge gap that has slowed progress in resolving the "missing heritability."

PITFALLS OF USING SURROGATE GENETIC MARKERS RATHER THAN VALIDATED POLYMORPHISMS WITH KNOWN TRAIT EFFECTS

Although clinical association studies—in particular, GWASs—have revealed numerous candidate SNPs, follow-up studies to identify causative functional variants are largely lacking. The top-scoring markers are open to the influence of several factors (e.g., demographics), and these may not be causative ones. As a result, the true genetic penetrance with respect to a complex trait such as drug response often remains uncertain. Significant SNPs from GWASs tend to lie near or within known genes; however, most candidate SNPs do not fall into exonic protein-coding regions, suggesting that regulatory functions are at play, such as modulation of transcription factor binding sites (rSNPs). A standard approach to validating rSNPs is the use of reporter gene assays to test promoter activity in heterologous tissues. We suspect, however, that some of the rSNPs implicated as being causative on the basis of observed clinical associations, and subsequently testing positive for the trait on reporter gene assays, may nevertheless be only surrogate markers, with definitive proof of *in vivo* function remaining elusive. *In vitro* reporter gene assays often suggest possible regulatory effects, none of which can be documented *in vivo* in the relevant target tissues.⁴

The variable-number tandem repeat (VNTR) in the serotonin transporter (*SERT*)-linked promoter region is a proposed rSNP. Whereas this important drug target gene has been implicated in numerous CNS disorders and in antidepressant therapy with selective serotonin reuptake inhibitors, often with contradictory results from many studies using the *SERT*-linked promoter region,¹⁶ we have found no evidence for an effect on mRNA expression of *SERT*-linked promoter region in the human brain (autopsies of the pons where the gene is transcribed and mature protein then transported to target brain regions.)¹² If there are regulatory variants in *SERT*, they still await full characterization. Similarly, a 3/4 repeat in the promoter region of *MAOA* (rs71980227) has been shown to associate with mRNA expression, and this has been supported by the results of reporter gene assays. This pVNTR has been extensively studied in more than 100 clinical studies of CNS disorders, showing robust associations. However, an analysis in human brain tissues suggests that the regulatory variant(s) more likely resides in the 3' portion of the gene,¹⁷ probably functioning as an srSNP. In support of this, a recent GWAS of mRNA expression profiles¹⁰ reports strong associations of a series of *MAOA* SNPs with *MAOA* mRNA levels, with the strongest associations occurring in the 3' region of the transcribed *MAOA* region (Figure 3, calculated from ref. 10). This result supports the likelihood that the pVNTR may not be causative, or may be only a minor contributor to overall genetic variability. Even though the entire *MAOA* gene locus consists of a large haplotype block in high linkage disequilibrium (LD), the use of a surrogate SNP within this haplotype block (rather than the functional variant) for clinical association studies captures only part of the genetic impact. Because *MAOA* is an

important target of antidepressant drugs, it will be necessary to reassess the true genetic effect of functional variants (likely srSNPs) on CNS traits and drug response. Given that drug response is a complex trait, failure to use the causative variants may well account for low clinical utility.

Even in the case of srSNPs, for which there is apparently solid evidence of exerting a functional impact, the question of relevance to the specific target tissues often remains open. For example, a SNP in the coding region of *DRD2*, encoding the D2 dopamine receptor that is the main target for antipsychotic drugs, was shown to affect mRNA turnover *in vitro*.¹⁸ However, we were unable to validate this effect in human brain tissues;¹³ instead, we identified srSNPs that affect splicing (see later text). Another critical question is to what extent validated functional variants account for the overall genetic variability in a gene locus, in a given patient population. Often, we find more than one frequent functional SNP (mostly rSNPs and srSNPs), for example, in *DRD2*, which has at least three functional variants.¹³ Many genes that encode drug metabolizing enzymes carry more than one mutation, e.g., *CYP2D6*. Because of this, multimarker and within-gene interaction models are required for such genes, including the phasing of functional SNPs in haplotypes. Detailed molecular genetics studies are needed to resolve these issues—an essential step for optimizing clinical utility. If the interaction between multiple variants in the same gene is complex, further complexity emerges from gene–gene interactions, requiring an understanding of the mechanisms and pathways involved.

THE HIGH PREVALENCE OF ALTERNATIVE OR ABERRANT SPLICING

srSNPs can act through many mechanisms, but heterogeneous nuclear RNA splicing to mature mRNA ranks as one of the most frequently affected events (Figure 1). It is necessary to distinguish between alternative and aberrant splicing. Alternative splicing is a common event affecting a large portion of all polyexonic genes,^{5,8} providing opportunities for generating numerous transcripts or functional protein variants, e.g., more than 10,000 splice variants of the Ca^{2+} channel *CACNA1C*. The splicing apparatus, the spliceosome, consists of a multicomponent protein complex and differs among tissues, with each expressing distinct splice factors that are critically important in shaping tissue physiology.⁸ Hence, alternative splicing can vary more between tissues in the same subject than between the same tissues in different subjects. Genetic variants can alter splicing in many locations along the transcribed region: splice sites, branch points, intronic and exonic splice enhancer and suppressor sites (Figure 5), and anywhere in the transcribed region where a new splice site can be generated (e.g., *CYP3A5*) (ref. 19). We tend to designate all such changes as “aberrant” splicing. However, just as was discussed for other regulatory polymorphisms, genetic factors that have an impact on splicing could have accumulated in such a fashion as to enhance tissue function; therefore, the term “aberrant” should refer only to those that lead to deleterious consequences. Moreover, genetic variants that modulate alternative splicing can affect drug response (e.g., *DRD2* (ref. 13)).

The estimated frequency of prevalence of genetic variants that affect splicing, relative to that of other variants, is still uncertain. On the one hand, high estimates (suggesting that these are involved in >50% of all diseases) could have arisen because of the relative ease with which changes in splicing can be measured by splice-specific quantitative reverse-transcription–PCR methods. On the other hand, the effects of genetic splicing are potentially underestimated because studies are often not carried out in the relevant target tissues (requiring autopsy tissues), in multiple subjects.

We often view events affecting RNA biology as being separate from those affecting DNA, chromatin, and transcription. However, transcription and heterogeneous nuclear RNA

splicing are intricately linked, given that splicing occurs coincidentally with transcription in physical proximity. The rate of transcription alone modulates splicing. Moreover, epigenetic events affecting chromatin structure are now also shown to regulate splicing, via common protein factors assembled in complexes near transcription/splicing sites.²⁰ Therefore, these processes must be viewed together, rather than as isolated events.

OTHER RNA PROCESSES

This review is not intended to delve into the details of RNA biology but rather to raise awareness of its pervasive influence on gene expression. mRNA elongation, turnover, trafficking, and translation are affected by genetic variants, with noncoding RNAs (e.g., microRNAs) playing critical roles (Figure 1). It is likely that most regulatory variants are still extant, awaiting discovery. Nonsynonymous SNPs should be considered as being potential srRNAs; for example, *A118G* in *OPRM1* was initially thought to alter opioid ligand binding with varying results and was subsequently found to affect both mRNA levels and translation *in vitro*.⁶ Analyses of mRNA and proteins are needed for comprehensive studies of translation. Frequently, mRNA levels do not correlate well with protein activity; however, one can narrow the discrepancies by measuring mRNAs associated with ribosomes engaged in translation,¹¹ increasing the correlation between mRNA and protein levels from $r = \sim 0.25$ to ~ 0.65 in the same tissue. This result highlights the importance of RNA localization within the cell,²¹ with profound consequences for biological functions. For the purpose of addressing RNA localization, protein–RNA cross-linking and immunoprecipitation, akin to chromatin immunoprecipitation, offers new avenues to explore RNA biology,²² for example, microRNA-dependent sequestration with argonaute proteins and spliceosomes binding to RNA transcripts. Next-generation sequencing has already proven valuable in studying protein–RNA interactions.

We also predict a critical role for noncoding RNA transcripts, which thus far have been poorly explored in pharmacogenomics. Noncoding and antisense RNAs undergo similar folding dynamics and processing steps, such as splicing. The study of genetic variations in microRNAs has become a vibrant field, with disruption of microRNA biology acknowledged as being linked to diseases.²³

WHAT IS THE IMPORTANCE OF srSNPs IN PHARMACOGENOMICS?

Drug metabolizing enzymes and transporters

The PK/PD characteristics of a drug are subject to variations in genes that encode enzymes and transporters. It is remarkable that several drug metabolism genes with strong relevance to pharmacogenomics have proven highly polymorphic, sometimes with abundant nonsynonymous SNPs causing functional defects. While the high frequency of such mutations in cytochrome P450 (CYP) genes may have a basis in evolution, it has led to expectations that genes that encode membrane transporters or receptors would also carry frequent mutations; however, this has not generally been the case. Several such genes have an extremely low abundance of any type of genetic variant in the protein-coding regions, for example *VMAT2*, *DAT*, and *SERT*. On the other hand, not all CYP enzymes have frequent nonsynonymous SNPs, e.g., *CYP3A4*, even though intersubject variability in enzyme activity is high. As a result, attention has shifted to regulatory polymorphisms.¹ Table 1 summarizes genes that encode drug metabolizing enzymes and transporters shown to carry regulatory SNPs, with focus on srSNPs. We have divided these variants into those with strong validation of mechanism and clinical effect, and others with evidence coming from either mechanistic or clinical studies, but typically not both (Table 1). The distinction may be somewhat arbitrary, and, with the use of less stringent criteria, more examples can be added. Splicing SNPs are overrepresented, reflecting the high prevalence of splicing in the

vast majority of genes. We briefly discuss relevant examples. The alleles representing srSNPs are indicated with their * nomenclature where available. Nonsynonymous SNPs relevant to drug therapy are not discussed here.

CYP2B6*6

A nonsynonymous SNP in exon 4 modulates an exonic enhancer site, generating a nonfunctional splice variant lacking exons 4–6 (ref. 24). The aberrant splice variant displayed a similar turnover such that the total *2B6**6 mRNA levels were unchanged. This example illustrates how mRNA expression arrays could miss important mRNA processes unless the affected splice region is directly measured. In HIV/AIDS patients treated with efavirenz, *CYP2B6**6 was shown to predict clearance and is associated with increased incidence of adverse drug reactions affecting the CNS.²⁵

CYP2C19*2 and *17

Located in exon 5, the *2 allele creates an aberrant splicing site, alters the open reading frame at amino acid 215, and generates a premature stop codon 20 amino acids downstream, resulting in a nonfunctional protein. In therapy with the anticoagulant clopidogrel, *2 carriers experience diminished antiplatelet response and increased risk of major adverse cardiovascular events and stent thrombosis.²⁶ The *17 allele, an rSNP in the promoter region, has been associated with increased CYP2C19 expression and, subsequently, with a significant reduction in stent thrombosis events but also enhanced bleeding.²⁷ The molecular genetics and clinical relevance of *17 need further study.

CYP2D6*4 and *41

The *G>A* polymorphism in *CYP2D6**4 deletes a consensus splicing acceptor site in intron 3, generating a new site one base pair downstream with shifted open reading frame, producing a nonfunctional protein. The minor allele is associated with lower plasma levels of endoxifen and 4-hydroxytamoxifen, active metabolites of tamoxifen, and shorter recurrence-free survival in breast cancer patients treated with tamoxifen.²⁸ *CYP2D6**41, located in intron 6, appears to affect an intronic enhancer element, increasing the level of an alternative splice variant lacking exon 6 with shifted reading frame and producing a nonfunctional protein.²⁹ *CYP2D6**41 shows partial loss of function and is associated with the intermediate metabolizer phenotype. Because of the frequency of nonfunctional *CYP2D6* variants, null alleles with low frequencies are clinically important given the increased likelihood of a subject having “compound heterozygosity,” i.e., carrying two independent null alleles.

CYP3A4

Until recently, none of the sequence variants in *CYP3A4* had been definitively shown to affect expression. Using allelic expression analysis and *in vitro* molecular genetics studies, we have shown that a SNP located in intron 6 (rs35599367, no * designation as yet) decreases hepatic CYP3A4 mRNA levels and enzyme activity, possibly by affecting nascent RNA elongation rate through changes in single-stranded DNA or RNA secondary structure.³⁰ The effect of the SNP is detected in the liver but not in the intestines, thereby illustrating the tissue selectivity of srSNPs. Causing 1.6–6.3-fold lower mRNA expression, and with an allele frequency of 4–8%, this new CYP3A4 allele could prove useful clinically. It is perhaps surprising that this intronic SNP with strong effect size had escaped attention in large-scale studies on *CYP3A4*. However, rs35599367 was initially not selected for inclusion in the GWAS array design because of its relatively low frequency and little or no LD with any other SNP (criteria for GWAS inclusion).

CYP3A5*3

Located in intron 3, *3 creates a cryptic consensus splice site yielding an additional exon 3B and a frame shift with premature stop codons, nonsense-mediated mRNA decay, and nonfunctional protein.¹⁹ *CYP3A5**3 causes low, if any, *CYP3A5* expression in a majority of Caucasians, leading to increased cyclosporine blood concentrations and decreased dose requirement in transplant recipients and increased exposure to several other drugs, such as tacrolimus and paclitaxel.

NAT1 *10 and *11

Located in the 3'-untranslated regions, *10 and *11 appear to increase enzyme activity, through mechanisms that are as yet uncertain, even though several studies have addressed this issue.³¹ Many studies suggest that both *10 and *11 are associated with susceptibility to cancer (see ref. 32), a finding consistent with expectations if one assumes a mechanism whereby increased acetylation of xenobiotics leads to more proximate carcinogenic metabolites. Whereas NAT2 is expressed mainly in the liver, NAT1 expression is widespread and therefore likely to have an impact on drug effects and toxicity in target tissue. However, the nature and magnitude of the effects of *10 and *11 in drug metabolism remain to be resolved; therefore, predictions based on NAT1 genotype are currently not feasible.

ABCB1*13

Among the drug transporters, much work has been focused on multidrug-resistance polypeptide 1, (Pgp, ABCB1). A series of three frequent (~40%) SNPs, including the synonymous *C3435T*, has been implicated in numerous association studies with drug effects as the phenotype, e.g., response to antiretroviral treatment in HIV-infected individuals.³³ However, these results are not uniformly reproduced, and the effect size is often below the range considered useful for clinical applications (with possible exceptions in specific situations, for example, antiviral drug exposure in lymphocytes). Although the potential function of these SNPs remains unclear, we have demonstrated that *C3435T* affects hepatic multidrug-resistance polypeptide 1 mRNA expression, apparently by altering the turnover of the mRNA.³⁴ Others have suggested that the T allele introduces a rare codon, thereby reducing the rate of translation;³⁵ however, a direct proof of this mechanism remains elusive. Nevertheless, the synonymous *C4535T* provides a prominent example of an srSNP that affects mRNA functions.

ABCC2

Of 26 polymorphisms in *ABCC2* screened for association with blood pravastatin concentration, one synonymous SNP located in exon 10 (c.1446 C > G) is associated with 95% higher mRNA levels and a lower pravastatin area under the plasma level–time curve.³⁶ However, it is unclear whether the SNP is causative or merely in high LD with a functional variant, most likely residing in the transcribed portion of the gene.

GENES THAT ENCODE DRUG TARGETS

In the context of pharmacodynamics, drug targets include receptors, enzymes, transporters, and their associated proteins involved in mediating drug effects. Numerous genetic variations have been reported for these gene classes, but their effects on drug response involve intermediary events and parallel pathways. Clinically observed genetic associations can identify risk genes and candidate variants, but causative relationships often remain hidden. Moreover, the understanding of gene–gene interactions is a precondition for identifying valid biomarkers. In cancer, somatic mutations in drug targets, for example, tyrosine kinases, can occur very frequently; but these are not the focus of this review. In the

case of genetic variants that affect regulatory function, tissue selectivity is critical to an understanding of the effects on disease and therapy. At present, few studies have demonstrated a clear link between treatment outcome and regulatory SNPs (Table 2). A main drug target for antipsychotic therapy, the D2 dopamine receptor (encoded by *DRD2*), serves as an example in which a deliberate focus on the discovery of srSNPs has yielded the identification of two splicing SNPs with strong effects on dopaminergic signaling.¹³ Similarly, the statin target HMG-coenzyme A reductase represents a drug receptor of critical importance, for which genetic factors are being explored to account for treatment success or failure.

D2S and D2L splice variants affected by intronic DRD2 SNPs

Coupled largely to inhibitory G proteins, the D2 receptor is alternatively spliced to produce the D2S (short) and D2L (long) isoforms, the latter including exon 6. As a presynaptic receptor, D2S is an autoreceptor that reduces presynaptic dopamine release, whereas D2L resides largely postsynaptically and may even synergize D1 activity (Figure 6). Many genetic association studies have implicated *DRD2* in multiple CNS disorders and in the response to antipsychotics such as clozapine.³⁷ Several variants across the *DRD2* gene locus have emerged as potential causative factors. However, in the absence of knowledge of a definitive molecular mechanism operating *in vivo* in the human brain, there is little confidence for clinical application of these findings. We therefore used allele-specific and splice-isoform-specific mRNA analyses in human autopsy brain tissues, leading to the discovery of a regulatory SNP ~1 kb upstream of the transcription start site (rs12364283) and two intronic SNPs in high LD with each other (rs2283265, rs1076560) that increase inclusion of exon 6—leading to enhanced production of D2L at the expense of D2S.¹³ In collaboration with A. Bertolino at the University of Bari, Italy, we found that the intronic SNPs affect memory processing and brain activity as measured using functional magnetic resonance imaging in the prefrontal–striatal pathway.¹³ The presence of high LD between the two intronic splicing SNPs and a majority of the SNPs selected in previous clinical association studies suggests that much of the evidence of significant clinical association could have come from surrogate markers that converge on alternative splicing as the causative factor. Even a frequently used marker SNP, *Taq1A*, located within an adjacent gene (*ANMK1*), is in substantial LD with the two intronic SNPs, although a role for *ANMK1* cannot be ruled out.

HMGCR (HMG-coenzyme A reductase)

A polymorphism located in intron 13 modulates a splice enhancer/repressor element, reducing the formation of the alternative splice variant lacking exon 13, which has no enzyme activity. Total mRNA levels (including all splice variants) did not change.³⁸ This variant is associated with lower levels of low-density lipoprotein cholesterol;³⁸ however, the influence on the outcome of statin therapy remains to be studied further. Statins are highly effective in reducing cholesterol levels in most patients but prevent only a portion of myocardial infarcts. The discovery of genetic biomarkers that can predict the extent to which an individual can benefit from statin therapy remains an important but as yet unmet goal.

CLINICAL RELEVANCE OF srSNPs IN DRUG THERAPY

The identification of valid genetic biomarkers for individualized drug therapy is the main objective of pharmacogenomics. Regulatory SNPs may offer paths for evolutionary selection such that these SNPs could accumulate and attain high frequencies. If a fitness advantage emerges from a mutation, a robust phenotypic effect is implicit and is likely to have clinical relevance. Frequently occurring variants through positive selection can

represent a double-edged sword, depending on the conditions. Our diets and environmental exposure have diversified extensively in recent evolutionary time—a factor of importance in the context of CYP enzymes. Also, exposure to drugs can be viewed as a new external challenge. We envisage that these genes that have undergone positive selection could serve as primary drug targets, resulting in strong pharmacogenetic effects primarily derived from regulatory variants. Moreover, for polymorphisms to become useful biomarkers for drug therapy, sufficient allele frequency is an important consideration. Prominent examples include *CYP2D6* and *CYP3A5*. The former carries a frequently occurring SNP (*4) that disrupts *CYP2D6* splicing, preventing the formation of active enzymes (Table 1). Poor metabolizers who are homozygous for two null alleles of *CYP2D6* are at increased risk of adverse drug reactions or fail to activate prodrugs such as codeine and tamoxifen (Table 1). Similarly, an intronic SNP generates a novel splice acceptor site in the gene that encodes *CYP3A5*, affecting the metabolism of several drugs (Table 1). We briefly discuss a few examples that illustrate how the pharmacogenetic biomarker field can advance.

CYP3A5

Among the first reported examples of an srSNP affecting splicing in a context relevant to drug therapy, the intronic *CYP3A5* SNP generates a new splice site, thereby suppressing generation of functional protein.¹⁹ This splice variant has strong effects on drugs that are preferentially metabolized by *CYP3A5*, such as tacrolimus and paclitaxel (Table 1). However, because *CYP3A5* and the more abundantly expressed *CYP3A4* share a high sequence homology and hence have overlapping substrate selectivity, it has been difficult to assign effect size to either of the two enzymes in any given individual. Hepatic expression of both enzymes is highly variable; *CYP3A5* may be the predominant enzyme in some individuals but not in most. Given that the two genes are located adjacent to each other and share a large haplotype block, surrogate SNPs in one could be correlated with the expression of the other. That is, SNP associations in one gene locus could serve as a marker for a functional variant in the other, confounding the assignment of relative contributions to drug metabolism by either *CYP3A4* or *CYP3A5*, or both. Although *CYP3A4* had not been considered polymorphic, a genetic approach to addressing the *CYP3A5*–*CYP3A4* interaction was not feasible.

CYP3A4

CYP3A4 is increasingly being targeted in drug discovery because it is perceived as being nonpolymorphic (although enzyme activity can vary over a 30-fold range). The recently discovered intronic SNP (rs35599367) (ref. 30) has potential clinical implications, with 8–15% of the population being heterozygous and <1% homozygous for the minor allele. Millions of patients take *CYP3A4*-metabolized drugs, putting a substantial number of subjects at potential risk of unexpectedly high drug exposure. A biomarker test using rs35599367 could identify at least a portion of those at risk of strong adverse drug reactions. In a relatively small study, carriers of the minor allele were found to require smaller statin doses to reach cholesterol-reduction goals.³⁰ Therefore, the *CYP3A4* intron 6 SNP is relevant to drug development in clinical trials and in drug therapy with narrow therapeutic margins—for example, anticancer drugs, because heterozygous carriers of the allele could experience increased risk of serious adverse drug reactions with even small increments in the dose administered. Predicting the metabolic capacity of the individual patient could enable the adjustment of doses differentially for rapid and slow metabolizers. Moreover, it will be possible to investigate whether genetic variability in both *CYP3A4* and *CYP3A5* can be used as predictive biomarker in a gene–gene interaction paradigm.

TPH2

Tryptophan hydroxylase 2 catalyzes a rate-limiting step in serotonin biosynthesis in the brain, specifically in the raphe nuclei of the pons. It is therefore a critical factor in all therapies that target serotonergic signaling. We had identified a synonymous SNP in exon 7 (rs7305115; Pro312Pro) that generates an exonic splicing enhancer site, resulting in a twofold increase in expression in the pons, thereby representing a gain of function.³⁹ Independently, this highly prevalent SNP (30–50% allele frequency), or marker SNPs in LD with it, has been implicated in suicidal ideation⁴⁰ and in affecting response to antidepressant drugs.⁴¹ Since the discovery of *TPH2*, numerous clinical association studies have been conducted, typically with different sets of SNPs, making meta-analysis challenging. We propose that the exonic splicing enhancer SNP has potential as a clinically useful biomarker; however, more prospective clinical trials are needed to determine the conditions in which sufficient predictive power can be attained to guide therapy. Another topical question that should be addressed is whether rs7305115 could play a role in suicidal tendencies in juvenile subjects on selective serotonin reuptake inhibitor therapy—a serious issue for adolescents taking antidepressants.

DRD2

Situated at the fulcrum of dopaminergic neurotransmission, *DRD2* has been a candidate gene in numerous clinical association studies, implicating a number of SNPs in diverse disorders and in antipsychotic therapy (see ref. 13). Because of the high LD of several of these *DRD2* SNPs with the two intronic SNPs that affect DS/L splicing, we propose that a majority of these clinical studies are linked to the same causative splicing event. Viewed from this perspective, the literature on the clinical effects of *DRD2* variants is strong. The splicing variants should be genotyped in more clinical trials of antipsychotic agents in order to replicate reported associations with response to clozapine³⁷ and other antipsychotics (Table 2). Also, we have recently completed an analysis of *DRD2* variants in severe cocaine abuse; the two intronic splice variants were found to be associated with heavy cocaine abuse, showing odds ratios of up to 3—a very strong effect size given the complex nature of drug addiction.⁴²

These results prepare the ground for valid clinical studies involving gene–gene interactions, with a number of genes in the dopamine biosynthetic and signaling paths being highly polymorphic. For example, D2S interacts physically with the dopamine transporter (DAT), which also appears to carry more than one regulatory variant; consequently, genetic *DRD2*-*DAT* interaction could have significant clinical impact.

OPRM1

The μ -opioid receptor carries a nonsynonymous SNP (*A118G*, see Figure 2) that is thought to affect ligand binding, mRNA expression, and translation.⁶ Numerous studies have addressed the clinical relevance of this variant, with mixed results. A metaanalysis of clinical association studies that tested *A118G* as a factor in susceptibility to opioid-induced analgesia has failed to confirm previous positive results.⁴³ On the other hand, significant associations were found between *A118G* and response to naltrexone in the treatment of alcoholism.⁴⁴ More recently, Shabalina *et al.*⁷ have identified additional variants and haplotypes in *OPRM1*, including rs563649. This SNP, located in a new exon upstream of the reference sequence, alters a structurally conserved internal ribosome entry site that often affects mRNA levels and translation efficiency. Secondary folding is critical for internal ribosome entry site functions, and rs563649 provides a conspicuous example of an srSNP. This study further implicates rs563649 in pain perception,⁷ indicating that the genetic variability of *OPRM1* has yet to be fully resolved. At present, the published genetic

association results fall short of yielding a valid biomarker for clinical use—a caveat for many reported polymorphisms in drug target genes.

RETROVIEW AND PROJECTION

Recent large-scale studies have shifted focus onto regulatory polymorphisms, which we subdivide here on a functional basis into rSNPs (transcription) and srSNPs (affecting RNA functions). This distinction is important because different approaches need to be taken to discover the causative factors and evaluate the validity of the evidence for clinical applications. We propose that srSNPs, including those affecting the ubiquitous splicing of the transcriptome, are at least as prevalent as rSNPs, the latter, in turn, being thought to be more prevalent than coding SNPs (affecting protein structure/function directly). We have focused here on srSNPs because this type of genetic variant has not previously been considered a functional class of its own, with numerous mechanisms all occurring at the RNA level. Splice variations have received the most attention because they are now acknowledged to be highly prevalent in disease, but other regulatory mechanisms may also be abundant, including sequestration and subcellular transport and the structural and mechanistic regulation of translation into functional protein.

As a result of historical bias, synonymous or “silent” SNPs in coding exons and intronic SNPs have often been neglected in studies. However, with the advent of powerful genome-wide methodologies for revealing genetic factors in expression regulation, we expect this knowledge gap to close rapidly. Whereas the current emphasis is still on finding candidate genes and marker polymorphisms, we must strive to identify all functional variants in a given gene that occur with sufficient frequency in at least one ethnic population, so that the genetic penetrance with respect to a clinical trait can be fully tested. We should avoid repeating clinical association studies numerous times with surrogate markers for prominent candidate genes, such as *ACE*, *CETP*, and *5HT2A*; rather, we should focus on molecular genetics and physiological pathways in order to advance the frontiers of clinical genetics and pharmacogenomics.

Such advances can play out well in pharmacogenomics because drugs typically target a relatively well-understood biochemical process or network, representing a subset of the underlying disease etiology. As a result, the effect size of a genetic variant with respect to drug response can reach sufficient levels for clinical utility (odds ratios ~3 or higher). However, regulatory polymorphisms have yet to be systematically explored, even in the most intensely studied pharmacogenes, such as *CYP2D6*, *2C9*, and *2C19*. Their clinical relevance has become apparent in recent clinical trials, but the question lingers as to whether the available tests represent most of the relevant genetic variants and sufficiently predict interindividual variability to drug response. The current literature indicates that this may not be the case. Failure to measure all frequent functional polymorphisms, or the use of imperfect surrogate marker SNPs, can introduce additional noise, thereby reducing the predictive power of a clinical test. Regulatory polymorphisms, both rSNPs and srSNPs, will likely prove essential to fill gaps in the “missing heritability,”⁹ with the promise of discovering pharmacogenomic biomarkers of increased predictive power and therefore of enhanced value for personalized health care.

Acknowledgments

This study was supported by grants from the National Institute on Drug Abuse (DA022199) and the National Institute of General Medical Sciences (GM092655).

References

1. Johnson AD, Wang D, Sadec W. Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacol Ther.* 2005; 106:19–38. [PubMed: 15781120]
2. Sadec W. Measuring cis-acting regulatory variants genome-wide: new insights into expression genetics and disease susceptibility. *Genome Med.* 2009; 1:116. [PubMed: 20090894]
3. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* 2009; 106:9362–9367. [PubMed: 19474294]
4. Johnson AD, et al. Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. *Pharmacogenet Genomics.* 2008; 18:781–791. [PubMed: 18698231]
5. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell.* 2009; 136:777–793. [PubMed: 19239895]
6. Zhang Y, Wang D, Johnson AD, Papp AC, Sadec W. Allelic expression imbalance of human mu opioid receptor (OPRM1) caused by variant A118G. *J Biol Chem.* 2005; 280:32618–32624. [PubMed: 16046395]
7. Shabalina SA, et al. Expansion of the human mu-opioid receptor gene architecture: novel functional variants. *Hum Mol Genet.* 2009; 18:1037–1051. [PubMed: 19103668]
8. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet.* 2007; 8:749–761. [PubMed: 17726481]
9. Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature.* 2009; 461:747–753. [PubMed: 19812666]
10. Zeller T, et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE.* 2010; 5:e10693. [PubMed: 20502693]
11. Ingolia NT, Ghaemmhami S, Newman JR, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science.* 2009; 324:218–223. [PubMed: 19213877]
12. Lim JE, Papp A, Pinsonneault J, Sadec W, Saffen D. Allelic expression of serotonin transporter (SERT) mRNA in human pons: lack of correlation with the polymorphism SERTLPR. *Mol Psychiatry.* 2006; 11:649–662. [PubMed: 16432527]
13. Zhang Y, et al. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. *Proc Natl Acad Sci USA.* 2007; 104:20552–20557. [PubMed: 18077373]
14. Ge B, et al. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet.* 2009; 41:1216–1222. [PubMed: 19838192]
15. Heap GA, et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet.* 2010; 19:122–134. [PubMed: 19825846]
16. Risch N, et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA.* 2009; 301:2462–2471. [PubMed: 19531786]
17. Pinsonneault JK, Papp AC, Sadec W. Allelic mRNA expression of X-linked monoamine oxidase a (MAOA) in human brain: dissection of epigenetic and genetic factors. *Hum Mol Genet.* 2006; 15:2636–2649. [PubMed: 16893905]
18. Duan J, et al. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet.* 2003; 12:205–216. [PubMed: 12554675]
19. Kuehl P, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet.* 2001; 27:383–391. [PubMed: 11279519]
20. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science.* 2010; 327:996–1000. [PubMed: 20133523]
21. Holt CE, Bullock SL. Subcellular mRNA localization in animal cells and why it matters. *Science.* 2009; 326:1212–1216. [PubMed: 19965463]
22. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell.* 2010; 141:129–141. [PubMed: 20371350]

23. Meola N, Gennarino VA, Banfi S. microRNAs and genetic diseases. *Pathogenetics*. 2009; 2:7. [PubMed: 19889204]
24. Hofmann MH, et al. Aberrant splicing caused by single nucleotide polymorphism c.516G>T [Q172H], a marker of CYP2B6*6, is responsible for decreased expression and activity of CYP2B6 in liver. *J Pharmacol Exp Ther*. 2008; 325:284–292. [PubMed: 18171905]
25. Ribaud HJ, et al. Effect of CYP2B6, ABCB1, and CYP3A5 polymorphisms on efavirenz pharmacokinetics and treatment response: an AIDS Clinical Trials Group study. *J Infect Dis*. 2010; 202:717–722. [PubMed: 20662624]
26. Shuldiner AR, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA*. 2009; 302:849–857. [PubMed: 19706858]
27. Tiroch KA, et al. Protective effect of the CYP2C19 *17 polymorphism with increased activation of clopidogrel on cardiovascular events. *Am Heart J*. 2010; 160:506–512. [PubMed: 20826260]
28. Kiyotani K, et al. Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. *J Clin Oncol*. 2010; 28:1287–1293. [PubMed: 20124171]
29. Toscano C, et al. Impaired expression of CYP2D6 in intermediate metabolizers carrying the *41 allele caused by the intronic SNP 2988G>A: evidence for modulation of splicing events. *Pharmacogenet Genomics*. 2006; 16:755–766. [PubMed: 17001295]
30. Wang D, Guo Y, Wrighton SA, Cooke GE, Sadec W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J*. 2010 e-pub ahead of print 13 April 2010.
31. Walker K, Ginsberg G, Hattis D, Johns DO, Guyton KZ, Sonawane B. Genetic polymorphism in N-Acetyltransferase (NAT): Population distribution of NAT1 and NAT2 activity. *J Toxicol Environ Health B Crit Rev*. 2009; 12:440–472. [PubMed: 20183529]
32. Jiao L, et al. Haplotype of N-acetyltransferase 1 and 2 and risk of pancreatic cancer. *Cancer Epidemiol Biomarkers Prev*. 2007; 16:2379–2386. [PubMed: 18006927]
33. Fellay J, et al. Swiss HIV Cohort Study. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet*. 2002; 359:30–36. [PubMed: 11809184]
34. Wang D, Johnson AD, Papp AC, Kroetz DL, Sadec W. Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. *Pharmacogenet Genomics*. 2005; 15:693–704. [PubMed: 16141795]
35. Kimchi-Sarfaty C, et al. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science*. 2007; 315:525–528. [PubMed: 17185560]
36. Niemi M, et al. Association of genetic polymorphism in ABCC2 with hepatic multidrug resistance-associated protein 2 expression and pravastatin pharmacokinetics. *Pharmacogenet Genomics*. 2006; 16:801–808. [PubMed: 17047488]
37. Hwang R, et al. Dopamine D2 receptor gene variants and quantitative measures of positive and negative symptom response following clozapine treatment. *Eur Neuropsychopharmacol*. 2006; 16:248–259. [PubMed: 16278074]
38. Burkhardt R, et al. Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler Thromb Vasc Biol*. 2008; 28:2078–2084. [PubMed: 18802019]
39. Lim JE, Pinsonneault J, Sadec W, Saffen D. Tryptophan hydroxylase 2 (TPH2) haplotypes predict levels of TPH2 mRNA expression in human pons. *Mol Psychiatry*. 2007; 12:491–501. [PubMed: 17453063]
40. Ke L, Qi ZY, Ping Y, Ren CY. Effect of SNP at position 40237 in exon 7 of the TPH2 gene on susceptibility to suicide. *Brain Res*. 2006; 1122:24–26. [PubMed: 17011525]
41. Tzvetkov MV, Brockmüller J, Roots I, Kirchheiner J. Common genetic variations in human brain-specific tryptophan hydroxylase-2 and response to antidepressant treatment. *Pharmacogenet Genomics*. 2008; 18:495–506. [PubMed: 18496129]
42. Moyer RA, et al. Intronic polymorphisms affecting alternative splicing of human dopamine D2 receptor are associated with cocaine abuse. *Neuropsychopharmacology*. 2011; 19:76–83.

43. Walter C, Lötsch J. Meta-analysis of the relevance of the OPRM1 118A>G genetic variant for pain treatment. *Pain*. 2009; 146:270–275. [PubMed: 19683391]
44. Oslin DW, et al. A functional polymorphism of the mu-opioid receptor gene is associated with naltrexone response in alcohol-dependent patients. *Neuropsychopharmacology*. 2003; 28:1546–1552. [PubMed: 12813472]
45. Iyer L, et al. Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. *Clin Pharmacol Ther*. 1999; 65:576–582. [PubMed: 10340924]
46. Djordjevic N, Ghotbi R, Jankovic S, Aklillu E. Induction of CYP1A2 by heavy coffee consumption is associated with the CYP1A2 -163C>A polymorphism. *Eur J Clin Pharmacol*. 2010; 66:697–703. [PubMed: 20390257]
47. Zhu H, et al. A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. *Hum Mol Genet*. 2007; 16:1765–1772. [PubMed: 17517690]
48. Shen YC, et al. Effects of DRD2/ANKK1 gene variations and clinical factors on aripiprazole efficacy in schizophrenic patients. *J Psychiatr Res*. 2009; 43:600–606. [PubMed: 18926547]
49. Nackley AG, et al. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*. 2006; 314:1930–1933. [PubMed: 17185601]
50. Grohmann M, et al. Alternative splicing and extensive RNA editing of human TPH2 transcripts. *PLoS ONE*. 2010; 5:e8956. [PubMed: 20126463]

Figure 1.

Functional classification of polymorphisms. **(a)** The three main types of polymorphisms (single-nucleotide polymorphisms (SNPs)) categorized by function: transcription (regulatory SNPs (rSNPs)), RNA processes (structural RNA SNPs (srSNPs)), and protein activity (nonsynonymous coding SNPs (cSNPs)). The term “SNP” is used vicariously for all types of genetic variation. Transcription rates are affected by rSNPs that reside typically in 5'-upstream regions but can also occur within the transcribed regions and 3'-downstream. srSNP must reside in the transcribed region if it is to affect RNA processing. **(b)** Canonical structure of pre-mRNA, divided into functional domains. 3'-Untranslated regions (UTR) and 5' UTR: 3' and 5' UTR, containing regulatory signals and protein/RNA docking sites (such as microRNA targets in the 3' UTR and internal ribosomal entry sites in the 5'UTR). Note that, typically, many transcripts are generated from the same gene locus, with multiple transcription initiation and polyadenylation sites, splicing, and antisense transcripts. Introns may give rise to functional noncoding RNAs, such as microRNAs.

Figure 2.

Perturbations of the computed folding profiles caused by single-nucleotide polymorphisms (SNPs) inserted *in silico* at each position of the mRNA coding region of the μ -opioid receptor (OPRM1). At each base position, a nucleotide substitution was inserted (transitions only (Ura $\leftrightarrow$ Thy or Ade $\leftrightarrow$ Gua); transversions yield similar results), and the resultant change in the folding pattern was calculated using Mfold. The x axis represents the base position in OPRM1 cDNA (reference sequence AY521028), and the y axis represents the relative changes in predicted folding patterns of the mRNA structure (expressed in terms of minimum energy conformations in relative units as compared with the wild type, set at 0) (for calculations, see ref. 5). The location of the nonsynonymous SNP A118G is indicated. The extent of the induced deviation does not necessarily correspond to any functional change; only a few substitutions occur in nature, and these would mostly remain silent. However, each substitution has the potential to affect function.

Figure 3.

Single-nucleotide polymorphism (SNP) association values with *MAOA* mRNA expression in monocytes.¹⁰ *MAOA* mRNA levels were measured using microarrays; all the SNPs reside in the *MAOA* gene locus, representing *cis*-expression quantitative trait loci. The SNPs are aligned along the *MAOA* gene locus (not shown to scale). Each point shows the association *P* value, which reaches extremely high significance ($P = 10^{-126}$). This is an indication of the pervasive *cis*-effect on *MAOA* expression. The most significant associations lie in the 3'-region of *MAOA*. The broad distribution of low *P* values reflects the large haplotype block with high linkage disequilibrium, consistent with positive evolutionary selection of *MAOA*.

Figure 4.

Example of the application of RNA-seq technology to identify allelic expression: myelin basic protein (MBP). RNA was extracted from two human autopsy samples of the prefrontal cortex of the brain. RNA transcript readings (~50 bp each, horizontal bars) were obtained by deep-sequencing (SOLiD System Life Technologies, Foster City, CA) and aligned along the basic myelin protein (*MBP*) gene locus, using the Integrative Genomics Viewer (IGV, v. 1.4). In the upper tissue sample, an uneven number of allele counts identifies a case of possible allelic expression imbalance (AEI). This must be validated with additional *MBP* marker single-nucleotide polymorphisms in the same sample and by allelic reverse-transcription PCR methods.

Figure 5.

Schematics of alternative splicing. Polymorphisms that affect splicing can reside in many locations along the transcribed region. *Trans*-acting proteins bind to *cis*-regulatory sequences in pre-mRNA to direct constitutive and alternative RNA splicing. Constitutive splicing typically involves the small nuclear ribonucleoproteins U1 and U2 and the U4/U5/U6 complex. Alternative splicing can include additional regulatory proteins, such as NOVA1, that recognize *cis*-regulatory elements to direct the inclusion or exclusion of alternative exons. Genetic variants occurring in any of these regulatory elements can influence the splicing pathways (intronic and exonic splice enhancer/suppressor sites).

Figure 6.

Structure of the *DRD2* gene locus showing exon 6, which undergoes alternative splicing, and two intronic single-nucleotide polymorphisms that promote exon 6 exclusion. D2S and D2L are the prevalent splice variants in the central nervous system, but several more splice variants exist and may be more abundant in other tissues.

Table 1
Validated and proposed srSNPs and rSNPs in genes that encode drug metabolizing enzymes

Gene	SNP	Location, MAF	Function and mechanism	Association studies	Refs.
<i>Validated srSNPs and rSNPs in genes that encode drug metabolizing enzymes^a</i>					
<i>CYP2C19</i>	*2, srSNP, rs4244285 G > A	Exon 5, 6–45%	Aberrant splice site, altered open reading frame, premature stop codon, nonfunctional protein	Reduces clopidogrel activation and therapy outcome	26
<i>CYP2D6</i>	*4, srSNP, rs3892097 G > A	Intron 3, 2–24%	Deletes consensus splicing acceptor site, shifted open reading frame, nonfunctional protein	Reduces tamoxifen activation and therapy outcome	28
<i>CYP2D6</i>	*41, srSNP, rs28371725 G > A	Intron 6, 2–15%	Intronic splicing enhancer element, increased skipping of exon 6, nonfunctional splice variant	Intermediate-metabolizer phenotype	29
<i>CYP3A5</i>	*3, srSNP, rs776746 A > G	Intron 3, 3–64%	Cryptic consensus splicing site, exon 3B with shifted open reading frame, premature stop codons, nonsense-mediated mRNA decay	Low CYP3A5 expression, affects cyclosporine, tacrolimus and paclitaxel pharmacokinetics and therapy outcome	See refs. in ref. 25
<i>UGT1A1</i>	*28, rSNP, 6, 7, or 8 TA repeats	Promoter, 4–33% (7-repeat)	Variable TA repeats at TATA box site of promoter region	7 or 8 repeat: lower transcription than 6 repeats, causing Gilbert's syndrome and modulating drug response and toxicity	45
<i>Proposed srSNPs and rSNPs in genes that encode drug metabolizing enzymes^b</i>					
<i>CYP2B6</i>	*6, srSNP rs3745274 G > T	Exon 4, 15–47%	Exonic splicing enhancer element, skipped exon 4–6, nonfunctional protein	Efavirenz pharmacokinetics and increase in CNS adverse events	24,25
<i>CYP3A4</i>	rs35599367 C > T, srSNP	Intron 6, 6–8%	Decreased CYP3A4 mRNA level, possibly affecting nascent RNA elongation rate	CYP3A4-metabolized statin dose requirement for optimal cholesterol control	30
<i>NAT1</i>	*10, srSNP rs1057126 T > A	3'UTR, 11–53%	SNP increases protein expression	Increased enzyme activity, associated with cancer risk	31,32
<i>NAT1</i>	*11, srSNPs rs7254666, rs28359534, rs8190862, rs8190863	3'UTR, 1–3%	Haplotype increases protein expression	Increased enzyme activity, associated with cancer susceptibility	31,32
<i>CYP1A2</i>	*1F, rSNP rs762551 C > A	Intron 1	Higher inducibility of SNP allele	The presence of the A allele results in higher CYP1A2 activity in smokers and heavy coffee drinkers	46
<i>CYP2C19</i>	*17, rSNP	Promoter	Increased transcription	Drug pharmacokinetics, drug response and clinical outcomes	27
<i>ABCB1 (MDR1)</i>	*13, srSNP, rs1045642 C > T	Exon 26, 12–60%	Changed single-stranded RNA structure, affects mRNA stability	Antiretroviral treatment outcome, cyclosporine bioavailability, methadone dose requirement	34,35
<i>ABCC2</i>	c.1446 C>G, synonymous Thr-482, srSNP	Exon 10	Increased mRNA level, unknown mechanism	Pravastatin pharmacokinetics	37

ABC, ATP-binding cassette transporter; CNS, central nervous system; *CYP*, cytochrome P450; MAF, minor allele frequency (the range provides approximations in different ethnic groups); NAT1, N-acetyltransferase 1; rSNP, regulatory single-nucleotide polymorphism; SNP, single-nucleotide polymorphism; srSNP, structural RNA single-nucleotide polymorphism; UGT, UDP-glucuronosyltransferase; UTR, untranslated region.

^a Inclusion in this list does not imply that the variant is currently a valid biomarker for drug therapy with accepted clinical utility. The list of references is not intended to cover all relevant studies. An rSNP is included for purposes of comparison with srSNPs.

^b Proposed srSNPs and rSNPs in genes that encode drug metabolizing enzymes and transporters, where either the underlying mechanism, or the clinical effect size, or both, require further validation. The list of references is not intended to cover all relevant studies. Some rSNPs are added for purposes of comparison.

Table 2

Validated rsSNPs in genes that encode drug targets (receptors, enzymes, and transporters)

Gene	SNP	Location, MAF	Function and mechanism	Association studies	Reference(s)
<i>LDLR</i>	rs688 C > T, synonymous Asn-591	Exon 12, 0–45%	Decreased LDLR exon 12 splicing	Increased LDL in premenopausal women	47
<i>DRD2</i>	rs6277 (C957T) synonymous Pro-319	Exon 7, 3–53%	Decreased mRNA stability/translation <i>in vitro</i>	Poor response to antipsychotics, methadone	18
<i>DRD2</i>	rs2283265 G > T rs1076560 G > T (in high LD)	Intron 5/6, 9–49%	Decreased DRD2 exon 6 splicing efficiency	Cognition and attention, risk for substance abuse	13,48
<i>OPRM1</i>	rs1799971 (A118G) Asn40Asp rs563649 (5'UTR)	Exon 1, 0–49%	Decreased mRNA expression and translation	Response to naltrexone in alcoholism, pain sensitivity	5,6,43,44
<i>COMT</i>	Haplotype rs6269, rs4633, rs4818, rs4680		Decreased translation and enzyme activity	Predictive of pain sensitivity	49
<i>TPH2</i>	rs4290270 A > T synonymous Ala-375	Exon 9, 30–59%	Altered TPH2 mRNA splicing and editing	NA	50
<i>TPH2</i>	rs7305115A > G synonymous Pro-312	Exon 7, 29–48%	Linked to allelic expression, TPH2 splicing	Risk for metabolic syndrome in patients with major depression, suicidal tendency	39,40,41
<i>HMGCR</i>	rs3846662 A > G	Intron 13	Intronic splicing enhancer/repressor element, enhances inclusion of exon 13	Associated with lower level of LDL cholesterol	38

This list provides select examples only. None of these assays is currently being used in routine pharmacogenomics biomarker testing. The list of references is not intended to cover all relevant studies.

COMT, catechol O-methyl transferase; *DRD2*, dopamine D2 receptor; *HMGCR*, 3-hydroxy-3-methylglutaryl-CoA reductase; *LD*, linkage disequilibrium; *LDL*, low-density lipoprotein; *LDLR*, LDL receptor; MAF, minor allele frequency; NA, not available; *OPRM1*, μ -opioid receptor; *TPH2*, tryptophan hydroxylase.