Published in final edited form as:

Pharmacogenomics. 2011 May ; 12(5): 675–680. doi:10.2217/pgs.11.23.

Pharmacogenomic biomarkers: validation needed for both the molecular genetic mechanism and clinical effect

Wolfgang Sadée

Program in Pharmacogenomics, Department of Pharmacology, College of Medicine, The Ohio State University, Columbus, OH 43210, USA

Wolfgang Sadée: Wolfgang.Sadee@osumc.edu

Abstract

Drug therapy can be ineffective or cause adverse reactions in a subset of patients. Pharmacogenomic biomarkers afford the opportunity to optimize an individual's therapy. Yet, few tests are currently part of standard care. To validate biomarkers, clinical replication studies are essential. Equally important, but less appreciated, the genetic mechanisms must also be understood to facilitate translation into clinical use. Representing main contributors to genetic variability, regulatory polymorphisms in particular are still poorly studied (e.g., *5-HTTLPR*). This article focuses on molecular and functional diversity of genetic biomarkers, as a guide to optimal use in personalized medicine.

Keywords

biomarkers; gene expression; pharmacogenomics; serotonin transporter; structural RNA polymorphisms

Insufficient understanding of molecular genetic mechanisms as a barrier to successful development of pharmacogenomic biomarkers

Genetic factors can modulate pharmacokinetic processes (e.g., absorption, distribution, metabolism and excretion [ADME]) and pharmacodynamics (e.g., drug response and toxicity). In particular, genes encoding drug-metabolizing enzymes carry frequent mutations that lead to complete loss of function, reduced activity or in some cases, gain of expressed enzymatic activity. For drugs dependent on a specific polymorphic enzyme for elimination from the body, such as CYP2D6, dramatic effects of mutations on drug levels in the body have been demonstrated. Yet, only a few validated pharmacogenomic tests are routinely used in clinical practice, summarized in a table of validated pharmacogenomic biomarker tests published by the US FDA [101]. Several biomarker tests (now part of the FDA drug label) aim at predicting efficacy or avoiding toxicity, for example, *EGFR* and *KRAS* mutations as indicators of response to cetuximab in targeted cancer chemotherapy. Similarly, a few examples have emerged where polymorphisms in *HLA* genes are highly

© 2011 Future Medicine Ltd

For reprint orders, please contact: reprints@futuremedicine.com

Financial & competing interests disclosure

The author acknowledges support from the National Institute of health General Medical Sciences U01 GM092655. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

predictive of severe adverse drug reactions involving an immune response, for example, *HLA-B*5701* and abacavir toxicity.

We can expect a growing spectrum of pharmacogenomic biomarkers to enter clinical use, but still at a rather slow rate. Translation into clinical practice faces numerous hurdles, in part arising from current approaches to the discovery of genetic variants and their testing in clinical trials. Much effort has been devoted to the optimization of clinical trial design testing the utility of a biomarker, with rigorous statistical criteria and replication requirements being the norm. Conversely, scrutiny of molecular genetics results is considerably more lax, with sparse evidence often sufficient to launch a series of hundreds of clinical association studies that eventually lead to no advance – the genetic variant tested proved to be nonfunctional, and may have been a poor surrogate for any functional variant that could have had an effect on the observed clinical phenotypes. Infatuation with rigorous clinical replication studies while neglecting the underlying biology and genetics has resulted in a large number of initially promising results that fail to replicate because of faulty study design. Often, an initial landmark discovery, such as the link between a traumatic experience early in life, the risk for developing clinical depression and the promoter variant of the serotonin transporter (serotonin transporter linked promoter repeat [*5-HTTLPR*]), engenders a flurry of dozens of replication studies, but the issue is far from settled (see discussion later). Selective reporting, publication bias towards positive results and a trend towards citing the early ‘breakthrough’ paper rather than subsequent contradictory studies all contribute to a field marred by results that often turn out to be spurious.

In this article, some of the hurdles and pitfalls in the development of genetic biomarkers are discussed. The goal is to promote the entering of genetic/genomic biomarkers into clinical trials, preferably after the causative relationship with a biological process or clinical trait is sufficiently well understood.

Nature of genetic variability

We can distinguish between several classes of functionally distinct polymorphisms, including those directly affecting the protein coding sequence, regulatory polymorphisms that alter gene expression, and sequence variants that alter RNA processing. We have used the acronym SNPs (the most abundant genetic variation) to represent all types of sequence variation, including insertions/deletions, copy number variants and chromosomal abnormalities (where needed, these will be addressed specifically). In addition, not only protein coding genes, but any DNA sequence yielding RNA transcripts are included, allowing for a broader interpretation of what a gene is. Hence, those SNPs capable of altering a gene’s function can be classified into coding region nonsynonymous SNPs (cSNPs) which alter the amino acid sequence directly, regulatory SNPs (rSNPs) which alter transcription and a newly defined class of structural RNA SNPs (srSNPs) which affect RNA processing and functions [1–3], the latter two emerging as critical factors in human diversity (Table 1). A survey of trait/disease-associated SNPs from genome-wide association studies (GWAS) reveals that nonsynonymous SNPs only account for approximately 9%, and SNPs in intergenic regions for approximately 43%, while presumed srSNPs reach approximately 49% (synonymous SNPs 2%, those in 5′- and 3′-UTRs 2% and intronic SNPs 45%) [4]. Evidently, each of these classes of genetic variation affect a different set of biological functions. Moreover, each genetic variant needs to be evaluated in the context of the target tissue where the gene is expressed, as well as environmental conditions that interact with the genetic variants, for example a promoter polymorphism responding to changes in transcription factors. We also need to reflect on a much more complex transcriptome than had been expected, in part revealed through the advent of next-generation sequencing [5]. A

large portion of the genomic sequence is transcribed, while only a small fraction is dedicated to translation into proteins.

Because of the exceedingly diverse functions, the evidence needed to judge a mutation/polymorphism suitable for clinical applications varies greatly between the classes of genetic variants. This article's goal is to provide a guide for evaluating the validity of candidate genetic biomarkers, based on evidence provided on their biological and genetic functions (Table 1).

Coding SNPs

Easily detectable by sequencing, the impact of cSNPs will be manifest in virtually all tissues where the protein is expressed. cSNP mutations tend to lead to physiological defects that are selected against in evolution, resulting in a lower frequency relative to other types of polymorphisms. On the surface, it should be straightforward to identify the functional defect by transfecting cDNAs carrying the cSNP into tissue cultures and measuring the cellular response. However, transfection in heterologous tissues may not accurately represent the effect in native tissue as protein folding and processing is tightly linked to transcription rates and the cellular environment. Moreover, a cSNP can exert its function at the RNA level, as we have demonstrated for the nonsynonymous A118G SNP in the μ opioid receptor [6]. Therefore, one needs to be cautious when predicting the *in vivo* effects in a target tissue from the *in vitro* results.

Regulatory SNPs

Regulatory SNPs alter the transcription of protein coding and noncoding genes, residing mostly in the nontranscribed flanking regions of genes, but also occasionally within the transcribed region. Typically, promoter SNPs are located upstream near the transcription start site, but enhancer and suppressor regions can extend over large distances, making them more difficult to discover. For example, we have identified a set of SNPs 13–20 kb upstream with a strong effect on the expression of the α -5 nicotinic receptor in the human prefrontal cortex [7]. Gene regulation depends on the tissue context, so that any genetic association study involving a rSNP should be founded upon knowledge of what tissue the rSNP is active in. For example, we have determined that a promoter SNP for *VKORC1* – the target of warfarin – is not only responsible for lower expression, but also is sufficient to account for most of the interindividual variability in hepatic *VKORC1* expression [8]. However, this same promoter SNP had no effect on the expression in B lymphocytes and heart tissues; therefore, clinical association studies testing an influence of the *VKORC1* SNP on local vascular effects is not grounded on firm evidence. In addition, rSNPs could exert their effects only under specific environmental conditions, including inflammation, or during early development. A standard approach to validating rSNPs is the use of reporter gene assays testing promoter activity in heterologous tissues. However, *in vitro* reporter gene assays often suggest possible regulatory effects while none can be documented *in vivo* in the relevant target tissues [9]. Rather, analysis of allelic RNA expression applied to human autopsy tissues and cell lines, has proven to be a powerful tool for detecting rSNPs (and also srSNPs), despite the fact that this approach only captures a single time point in an autopsy tissue [1,2,5,9,10]. This illustrates the ambiguity that can remain even in the face of seemingly convincing molecular genetics data.

Structural RNA SNPs

This type of genetic variant may be at least as prevalent as rSNPs [1–3]. srSNPs reside in genomic regions transcribed into RNAs, and therefore present in each nascent RNA transcript, both protein coding and noncoding. Countless biological pathways contribute to RNA processing, and the RNA transcripts support numerous functions, including structural,

regulatory, sensory and catalytic [1,11]. Both intronic and exonic SNPs can alter RNA processing, including splicing, while only exonic SNPs can affect subsequent processes such as turnover of mature mRNAs, translation and biological functions. Among the critical RNA processes that can be altered by srSNPs, alternative or aberrant splicing alone have been implicated in numerous genetic diseases [12].

As discussed with rSNPs, srSNPs can be highly tissue- and context-dependent, because each tissue expresses a different complement of RNA-interacting factors, such as the multimeric spliceosome [11,12]. The L-type channel CACNA1C hnRNA is processed by at least 12 alternative splice sites, generating thousands of possible mRNA variants with two pharmacological types: one expressed in heart muscle with low sensitivity to dihydropyridines, and the other expressed in smooth muscle with high sensitivity. By measuring all splice variants in human heart tissues, we have found high expression of the dihydropyridine-sensitive forms in three of 65 subjects [13], demonstrating large interindividual differences in splicing; a possible indicator of unexpected verapamil responses. Similarly, alternative splicing of the dopamine D2 receptor mRNA to yield functionally distinct short (dopamine D2 receptor, short splice isoform) and long (dopamine D2 receptor, long splice isoform) isoforms is a normal process that differs between brain tissues, such as the prefrontal cortex and putamen [14]. We have further demonstrated that two frequent intronic SNPs that shift the splicing process in favor of D2L also affect memory processing [14]. More recently, we have determined that the same intronic SNPs are highly over-represented in heavy cocaine abusers who had taken a fatal overdose (odds ratio ~3) [15]. These results demonstrated how alternative splicing can affect phenotype, and specifically drug response. Therefore, it is imperative to study RNA processing in relevant target tissues, testing whether and where srSNPs exert their effect, and under what conditions.

Taken together, applying strict criteria to the validation of cSNPs, rSNPs and srSNPs is equally critical to rigorous clinical trial design, as is the need for the replication of clinical association results. In view of the growing number of candidate SNPs gleaned from GWAS, with credible associations to traits and diseases, but with mostly low effect size, a better understanding of the causative relationships represents the next logical step to take advantage of the plethora of new discoveries. We understand that most of the common diseases and therapy outcomes are affected by multiple genes. If we wish to take advantage of gene–gene interactions in the design of valid biomarkers, the underlying biology must be understood for optimal results.

We can no longer afford countless replications of a single polymorphism in clinical association studies without firm evidence that this variant is indeed functional and critical to the trait tested. For example, several thousand association studies have been conducted with an intronic insertion/deletion polymorphism of *ACE* (encoding angiotensin-converting enzyme), with largely conflicting results and sparse, if any, evidence that the SNP has biological function. Searching for regulatory SNPs in the *ACE* locus, we have been unable to detect any effect of the insertion/deletion polymorphism on mRNA expression, and our meta-analysis also failed to support clinical effects [16]. Rather, we have identified a set of SNPs approximately 3-kb upstream of the transcription start site with a strong effect on expression and high allele frequency (but only in African–Americans), conveying a high risk of coronary artery disease in hypertensive patients [16]. Despite the growing evidence disputing a true function, or a role as a useful surrogate marker, association studies continue to rely on the *ACE* insertion/deletion polymorphism. One might wish for an obligatory indepth study of the extant literature before further clinical trials are implemented. Given the high complexity of the genetic basis of any clinical effect, definitive answers are often not attainable, but the most highly educated guess is the least one can do. By further exploring

fundamental genetics, we have a better chance of shedding light on the ‘missing heritability’ [17] in health, disease and therapy.

Serotonin transporter polymorphism 5-HTTLPR, stressful life events & risk of depression: lack of association or lack of function?

The serotonin transporter (also known as 5-HTT, SERT and SLC6A4) represents a main target for antidepressant drugs (e.g., selective serotonin reuptake inhibitors) and has been implicated in numerous CNS disorders. A frequent regulatory polymorphism, the variable number tandem repeat in 5-HTTLPR has served as a biomarker assessing risk for various CNS disorders and outcome with antidepressant therapy. A clinical 5-HTTLPR biomarker test is already in use as a guide to therapy in psychiatric practice. However, in spite of hundreds of association studies yielding often contradictory results, the validity of 5-HTTLPR as a biomarker remains uncertain. This example may well be a case of runaway replication studies with insufficient molecular genetic support.

In 2003, Caspi *et al.* published a landmark paper that stood out as a paradigm for the interaction between a genetic variant (5-HTTLPR), environmental factors (stress in early life) and depression [18]. The study purports to demonstrate that the 5-HTTLPR variant predisposes an individual to a greater risk of depression in early adult life, in individuals who had experienced traumatic or stressful events early in life. This study had built on earlier studies by Lesch *et al.* on 5-HTT mRNA expression in lymphocytes, and on reporter genes assays testing the 5-HTTLPR promoter region [19]. However, in a meta-analysis of accumulating studies on the paradigm 5-HTTLPR–stress–depression, while supporting the influence of early stressful events on later risk of depression, Risch *et al.* failed to find a significant genetic association, raising questions about the validity of 5-HTTLPR as a key factor in psychiatric disorders [20].

Positive clinical association results are seemingly buttressed by the common assumption that 5-HTTLPR affects expression in serotonergic neurons in the raphe nuclei, where the 5-HTT gene is expressed and the protein is then disseminated throughout the brain. To test this in the raphe nuclei from human autopsies, we measured allelic 5-HTT mRNA expression; a specific and sensitive means for determining the influence of the 5-HTTLPR allele [21]. In contrast to our expectations, we failed to detect any effect on mRNA expression by this promoter polymorphism. Furthermore, we were unable to repeat the finding of an association between 5-HTTLPR and mRNA levels in lymphocytes as published by Lesch [19]. Other studies similarly fail to confirm that the 5-HTTLPR is active, or results are contradictory (see references in [21]). While one cannot exclude the possibility that 5-HTTLPR is functional under certain conditions (for example, under stress or during early development), the available evidence is weak at best or argues against an active role.

However, the statistical assessment of countless clinical associations with 5-HTTLPR relies on the assumption that its promoter function is proven – representing the prior knowledge. This is clearly not the case; therefore, much more stringent criteria need to be applied before the statistical significance of this genetic association can be accepted as strong evidence for a causative relationship. Multiple hypotheses testing and publication bias for positive results are eminent problems, in view of the absence of a validated molecular genetic mechanism underlying a clinical phenotype. Failure to attain statistical significance for 5-HTTLPR in the Risch *et al.* meta-analysis [20] serves as a reminder that we must strive to understand the underlying biology before launching hundreds of clinical studies. As this example further demonstrates, none of the *in vitro* molecular studies are by themselves definitive, as regulatory polymorphisms are context dependent, and human brain autopsies only capture a single condition at the time of death. We have also analyzed neonatal pontine tissues, again

without finding any effect of *5-HTTLPR* on mRNA expression [Lim JE & Saffen D, Unpublished Data], but we cannot exclude the possibility that *5-HTTLPR* is active under certain conditions not yet tested. Certainly, some of the evidence from clinical association studies appears to be strong (excluding the environmental component considered by Risch *et al.* in their meta-analysis). Taken together, the use of *5-HTTLPR* as a clinical biomarker guiding therapeutic decisions appears to be premature at best until we understand the causative link between genotype and clinical trait [20].

This article demonstrates how rigorous validation of both clinical genetic association results and molecular genetics need to be considered in the most effective clinical trial design. Use of a poor surrogate marker of a functional genetic variant greatly reduces the power of any clinical trial, and should be considered a faulty study design. However, for a vast majority of candidate SNPs, including those emerging from genome-scale surveys, the underlying mechanisms remain hidden. A significant effort must be mounted to resolve the functions and mechanisms for each polymorphism with sufficient effect size to gather clinical interest.

Another critical question needs to be addressed: to what extent do 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 polymorphism within the same gene locus, for example in *DRD2*, which has at least three functional variants [14]. Many genes encoding drug-metabolizing enzymes carry more than one mutation, for example *CYP2D6*. Thus, multimarker and within-gene interaction models are required for such genes, including the phasing of functional SNPs in haplotypes. Detailed molecular genetic 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.

Problems with the use of surrogate markers are rapidly compounded when considering the interactions of two or more genes, a logical next step that cannot rely on GWAS results alone, because the number of possible combinations is simply too large. Currently, there is a tendency to sum up – or multiply – the effects of a set of candidate SNPs in plausible disease risk genes, with the expectation to generate a biomarker panel with greater predictive power. Yet, even a thorough use of systems biology to sort out relevant gene–gene interactions can still fall short of yielding robust predictors valid in different populations. Only by bringing together basic genetic and clinical elements can we advance individualized therapy and disease prevention.

Conclusion & future perspective

Pharmacogenomic biomarkers will assume a prevalent role in individualized therapy, and specifically in early treatment or prevention, as predictors of disease risk and treatment outcomes. An increasing number of therapies will link a diagnostic test with a drug ('theranostics'). The predictive value of pharmacogenomic biomarkers will gradually increase with the use of multigene panels, built from insights obtained from both molecular genetic studies and validated with clinical trials, increasingly relying on whole-genome sequencing.

Bibliography

Papers of special note have been highlighted as:

- of interest

▪ ▪ of considerable interest

- 1▪. Johnson AD, Wang D, Sadec W. Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacol Ther.* 2005; 106:19–38. Provides detailed information on genetic factors involved in the regulation of gene expression. [PubMed: 15781120]
2. Sadec W. Measuring *cis*-acting regulatory variants genome-wide: new insight into expression genetics and disease susceptibility. *Genome Med.* 2009; 1:116. [PubMed: 20090894]
3. Sadec W, Wang D, Papp AC, et al. Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. *Clin Pharmacol Ther.* 2011; 89(3):355–365. [PubMed: 21289622]
- 4▪ ▪. Hindorff LA, Sethupathy P, Junkins HA, 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. Discusses the relative prevalence of genetic variations affecting human traits, as gleaned from genome-wide association study results. [PubMed: 19474294]
5. Heap GA, Yang JH, Downes K, 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]
6. Zhang Y, Wang DX, Johnson AD, Papp AC, Sadec W. Allelic expression imbalance of human μ opioid receptor (*OPRM1*) caused by variant A118G. *J Biol Chem.* 2005; 280:32618–3224. [PubMed: 16046395]
7. Smith RM, Alachkar H, Papp AC, et al. Nicotinic $\alpha 5$ receptor subunit mRNA expression is associated with distant 5' upstream polymorphisms. *Eur J Hum Gen.* 2011; 19(1):76–83.
8. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadec W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (*VKORC1*) affects gene expression and warfarin dose requirement. *Blood.* 2008; 112:1013–1021. [PubMed: 18523153]
9. Johnson AD, Zhang Y, Papp AC, 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]
10. Ge B, Pokholok DK, Kwan T, 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]
- 11▪. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell.* 2009; 136:777–793. Highlights the importance of RNA in cell function and disease. [PubMed: 19239895]
- 12▪ ▪. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet.* 2007; 8:749–761. Documents the importance of splicing and provides insight into the sequence codes directing tissue-selective splicing. [PubMed: 17726481]
13. Wang D, Papp AC, Binkley PF, Johnson JA, Sadec W. Highly variable mRNA expression and splicing of L-type voltage-dependent calcium channel α subunit 1C (*CACNA1C*) in human heart tissues. *Pharmacogenet Genomics.* 2006; 16:735–745.
- 14▪. Zhang Y, Bertolino A, Fazio L, 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. Reports on the effect of splice variants in the dopamine D2 receptor, a gene linked to psychiatric disorders and response to antipsychotic therapies. [PubMed: 18077373]
- 15▪. Moyer RA, Wang D, Papp AC, et al. Intronic polymorphisms affecting alternative splicing of human dopamine D2 receptor are associated with cocaine abuse. *Neuropsychopharmacology.* 2011; 36(4):753–762. Reports on the effect of splice variants in the dopamine D2 receptor, a gene linked to psychiatric disorders and response to antipsychotic therapies. [PubMed: 21150907]
16. Johnson AD, Gong Y, Wang D, et al. Promoter polymorphisms in *ACE* (angiotensin-I converting enzyme) associated with clinical outcomes in hypertension. *Clin Pharmacol Ther.* 2009; 85:36–44. [PubMed: 18946466]
- 17▪ ▪. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009; 461:747–753. Analyzes some of the reasons why we have as yet failed to understand the genetic factors believed to underlie the main multigenic diseases. [PubMed: 19812666]

18. Caspi A, Sugden K, Moffitt TE, et al. Influence of life stress on depression: moderation by a polymorphism in the *5-HTT* gene. *Science*. 2003; 301:386–389. [PubMed: 12869766]
19. Lesch KP, Bengel D, Heils A, et al. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*. 1996; 274:1527–1531. [PubMed: 8929413]
20. Risch N, Herrell R, Lehner T, 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. Disputes previous findings regarding the effect of *5-HTTLPR* on depression, but is in turn put in question by several other studies. A definitive answer remains elusive. [PubMed: 19531786]
21. 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]

Website

101. US FDA. Table of pharmacogenomic biomarkers in drug labels.
www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm

Executive summary

Insufficient understanding of the molecular genetic mechanisms as a barrier to successful development of pharmacogenomic biomarkers

- Pharmacogenomic biomarkers will increasingly serve as predictors of disease risk and treatment outcomes in clinical practice.
- Molecular genetics and biological pathways must be understood in order to take full advantage of biomarker tests.

Nature of genetic variability

- Genetic variants include SNPs, insertions/deletions, copy number variants and chromosomal rearrangements (all referred to here as SNPs). Variants altering a gene's function can be grouped into regulatory SNPs, those affecting RNA biology and structure, and nonsynonymous SNPs that alter the protein sequence.
- Each group requires a distinct strategy for discovery and analysis of phenotypic effects.

Serotonin transporter polymorphism (5-HTTLPR) stressful life events & risk of depression: lack of association or lack of function?

- *5-HTTLPR* is an example in which the underlying molecular mechanisms are not understood, and the debate over its utility as a biomarker remains unresolved.

Conclusion & future perspective

- Surrogate genetic markers often yield spurious results and should be replaced by validated polymorphisms.
- Gene–gene interactions must be understood in order to resolve the ‘missing heritability’ of disease. Development of multigene panels requires the use of validated variants and known interactions to become optimally useful.

Table 1

Types of genetic variation grouped by distinct functions.

	Regulatory variants; regulatory SNPs	Structural RNA variants; structural RNA SNPs	Coding region variants; coding region nonsynonymous SNPs
Location	Mostly 5'-upstream of transcription sites, but can be distributed over large genomic distances	Reside in the transcribed portion of a gene; can be in introns or exons	Reside in or alter protein coding regions
Molecular mechanism	Alter promoter, enhancer and suppressor regions	Alter RNA sequence, folding, elongation, processing, editing, turnover, trafficking and translation	Alter protein sequence, folding, stability, trafficking and post-translational modifications
Biological effect	Alter gene expression	RNA function: regulatory, sensory, structural and protein expression for coding mRNAs	Alter protein functions and activity

For simplicity, the term SNP is used to include all sequence variation, from SNPs to insertion/deletions, variable number tandem repeats and copy number variations. Chromosomal rearrangements (e.g., aneuploidy, inversion, transversion, translocation) also contribute to genetic variability and similarly affect gene expression, RNA processing and protein sequence.