

# Chapter 2

## Silent (Synonymous) SNPs: Should We Care About Them?

**Ryan Hunt, Zuben E. Sauna, Suresh V. Ambudkar, Michael M. Gottesman, and Chava Kimchi-Sarfaty**

### Abstract

One of the surprising findings of the Human Genome Project was that single nucleotide polymorphisms (SNPs), which, by definition, have a minor allele frequency greater than 1%, occur at higher rates than previously suspected. When occurring in the gene coding regions, SNPs can be synonymous (i.e., not causing a change in the amino acid) or nonsynonymous (when the amino acid is altered). It has long been assumed that synonymous SNPs are inconsequential, as the primary sequence of the protein is retained. A number of studies have questioned this assumption over the last decade, showing that synonymous mutations are also under evolutionary pressure and they can be implicated in disease. More importantly, several of the mechanisms by which synonymous mutations alter the structure, function, and expression level of proteins are now being elucidated. Studies have demonstrated that synonymous polymorphisms can affect messenger RNA splicing, stability, and structure as well as protein folding. These changes can have a significant effect on the function of proteins, change cellular response to therapeutic targets, and often explain the different responses of individual patients to a certain medication.

**Key words:** Single nucleotide polymorphism, messenger RNA splicing, messenger RNA stability, messenger RNA structure, protein folding, synonymous mutations, nonsynonymous mutations, codon frequency, codon usage.

---

### 1. Introduction

A large portion of intraspecies phenotypic variation can be attributed to single nucleotide polymorphisms (SNPs), single base changes occurring at a frequency greater than 1% within a population. In humans, for instance, it is estimated that about 90% of sequence variation can be credited to SNPs (1). It is believed that development and progression of cancer, diabetes, cardiovascular disease, psychiatric issues, and other common diseases are dependent on inputs from multiple loci and environmental cues, which

make them markedly more challenging to study than rare, highly penetrant Mendelian diseases (2) (*see also Chapter 1* in this volume). Association studies have uncovered SNPs at the basis of many disease traits. The development of robust technologies to detect SNPs promises to bolster the library of these functional SNPs. Even if a disease-associated SNP is not discovered to be the causative agent, it often exists in linkage disequilibrium (*see Note 1*) with the disease-causing allele, and can lead investigators to the actual functional risk variant, which may be hundreds of kilobases away (1, 3). Thus, SNPs are major contributors to variability seen in drug response and disease susceptibility, rendering them key players in advancing pharmacogenomics and the development of personalized medicine.

SNPs exist throughout the entire genome, both within coding regions as well as outside coding sequences. SNPs residing outside coding regions can occur in intergenic sequences, 5'- or 3'-untranslated regions, intronic regions, and associated noncoding regions such as promoters and transcription factor binding sites. Those falling within coding regions can be further categorized into two groups: synonymous, or “silent,” (*see Note 2*) and nonsynonymous. A coding-region SNP which alters the transcribed codon such that a different amino acid is incorporated into the polypeptide is referred to as a “nonsynonymous SNP.” Owing to wobble transfer RNA (tRNA) base pairing and redundancy in the genetic code, some (synonymous) SNPs within coding regions do not lead to a change in the amino acid incorporated at the site of their occurrence.

Much of the lack of appreciation for this class of SNPs derives from faulty nomenclature. The ubiquitously popular term “silent SNP” leads one to the view that such a change in deoxynucleotide has no consequence for molecular, cellular, or physiological phenomena. The use of this terminology can easily mislead many audiences, yet many continue to opt for “silent SNP,” and the term remains prevalent in publications to this date (4, 5). With a growing body of evidence invalidating this presupposition, it is prudent to lay this historic term to rest while promoting terminology that reflects current understanding. Here, we advocate for the use of “synonymous” to categorize those SNPs that do not lead to a change in primary polypeptide sequence. Although the concept of similarity is still present, “synonymous SNP” directs one’s attention to the fact that these single base pair changes require *different* tRNAs to decode them; tRNAs that are, however, charged with the same amino acid. Moreover, this title avoids the notion that these SNPs are biologically silent or inconsequential.

Here, we argue that studying synonymous SNPs is a worthwhile endeavor. We also hope to defeat any notion that synonymous SNPs are trivial, as this belief is flawed on many fronts: messenger RNA (mRNA) splicing, mRNA stability, mRNA structure, protein translation and cotranslational protein folding.

## 2. The Impact of Synonymous SNPs

### 2.1. mRNA Splicing

The most well documented way in which synonymous SNPs exert their impact on gene function is via perturbations of mRNA splicing. It has been estimated that 92–94% of all multiexonic human genes are alternatively spliced (6). With the modest number of protein-coding genes revealed by the Human Genome Project, creating sufficient proteome diversity is achieved, in part, through alternative splicing. Ensuring the correct relative abundance of transcript variants is critical for correct proteostasis, as protein isoforms can have null, distinct, antagonizing, or agonistic functions (7, 8). Therefore, synonymous SNPs generating ectopic mRNA splicing can bring about distinct phenotypes, including human disease.

Schizophrenia is a complex neurological disorder stemming from frontal and temporal lobe dysfunction. Although complicated by the multifactorial nature of the disease, the genetic underpinnings giving rise to schizophrenia are beginning to be understood. Glutamate neurotransmission, a process tightly linked to schizophrenia (among many other neurological disorders), is largely regulated by metabotropic glutamate receptors (GRMs). Activation of group II GRMs, composed of GRM2 and GRM3, moderates glutamate synaptic levels. A synonymous SNP within *GRM3* was the first SNP discovered showing strong ties to the development of schizophrenia (9). Until recently, the mechanisms underlying this association have remained unclear and negative findings have added doubt to the validity of this SNP's schizophrenic ties (10, 11). A splicing variant of *GRM3* lacking the fourth exon transmembrane domain, *GRM34*, was known to be present in both schizophrenic and control patients. Recent investigation into the aforementioned SNP located within the third exon of *GRM3* revealed that its presence markedly increased (about 20%) levels of the truncated splicing variant *GRM34*. However, this transcript was not detected at statistically significant levels in schizophrenic patients (12). The authors did note that the power of their testing was insufficient, and that this synonymous SNP is still likely to interact with other SNPs to generate schizophrenia. Advancing our understanding of psychophysiological phenomena, such as the development of schizophrenia, requires a holistic analysis of biological inputs. This finding provides evidence that SNPs, notably synonymous SNPs, should fall under that realm.

Synonymous SNPs can also influence the splicing of precursor mRNA (pre-mRNA) without giving rise to a disease phenotype or predisposition. In fact, one study highlights a synonymous mutation's ability to silence the effects of a deleterious mutation.

The medium-chain acyl-CoA dehydrogenase (*MCAD*) gene encodes for MCAD, and deficiency in MCAD can be brought about by a missense mutation in exon 5. An exonic splicing enhancer, which normally antagonizes an exonic splicing silencer, becomes inactivated in the presence of this mutation, thus leading to nonfunctioning MCAD. While it has no effect on *MCAD* splicing by itself, the presence of a single synonymous SNP in *MCAD* exon 5 renders the exonic splicing silencer inactive, thus restoring levels of the functional splice product in the presence of the missense mutation (13).

Synonymous SNPs also contribute to the intricateness of infectious disease. There has been long-standing interest in identifying genetic factors giving rise to individual variability in infectious disease susceptibility and progression. The flavivirus *West Nile virus* is causing a growing global epidemic and has been a point of increasing scientific attention in recent years. In an effort to identify genetic indicators of viral susceptibility, researchers sequenced oligoadenylate synthetase (*OAS*) genes and ribonuclease L (*RNASEL*) genes, two groups of innate viral resistance genes previously identified by mouse models, in 33 individuals hospitalized with *West Nile virus*. Surprisingly, they found no deletions, insertions, or nonsense mutations in any patient. But through analysis of coding-region sequences, they identified a single synonymous polymorphism in the 2',5'-oligoadenylate synthetase like (*OASL*) gene occurring at a significantly high rate in affected individuals. Moreover, because this polymorphism is present within an exonic splice enhancer, it was suggested that the polymorphic transcript undergoes increased splicing, increasing the yield of a truncated *OASL* product, which may impair innate viral immunity (14).

The role of auxiliary elements within exons (exonic splicing enhancers and exonic splicing silencers) in human disease has been the subject of increasing attention in recent years (15). Owing to their capacity to affect mRNA splicing, synonymous SNPs are known to be a part of numerous multifaceted diseases, from multiple sclerosis (16) to autism (17). Synonymous base pair changes in these splicing motif regions can change the splicing patterns of mRNA transcripts directly, or they can alter the penetrance of concurring mutations elsewhere in the gene.

## 2.2. mRNA Stability

The stability of mRNA is intimately linked to gene expression. The ability to monitor and rapidly adjust the half life of mRNA with specificity is critical at both the cellular and the physiological levels. The factors which determine the stability of a given mRNA fall into two categories: characteristics of the transcript itself (*cis* factors) and determinants originating outside of the primary transcript (*trans* factors). The latter group includes elements like growth factors, RNases, RNA binding proteins, ions, dissolved oxygen,

and irradiation (for a review *see* (18)). Here, we direct attention to *cis* factors, particularly synonymous SNPs and their impact on mRNA stability. While most *cis* factors that determine mRNA stability fall within the 3'-untranslated region of a transcript, various influences stemming from coding sequences have been identified.

At least nine susceptibility loci exist for the chronic skin disorder psoriasis (PSORS1–PSORS9). The *CDSN* gene exists within the PSORS1 region and encodes corneodesmosin, an extracellular adhesion protein found in the major component of the epidermis, keratinocytes. Because corneodesmosin is often overexpressed in skin lesions of psoriasis patients, one hypothesis conjectured that risk alleles of *CDSN* predict an mRNA transcript with a heightened half life, thus leading to the overexpression of corneodesmosin. Upon investigation of a risk haplotype, it was revealed that the transcript bearing the haplotype exhibited a twofold increase in mRNA stability. What is more, site-directed mutagenesis revealed that this synonymous SNP was solely responsible for the observed increase in mRNA stability. Further examination revealed that the synonymous SNP altered the transcript's affinity for a cytoplasmic RNA binding protein (19).

ATP-binding-cassette (ABC) transporters comprise a family of integral membrane proteins that function in the ATP-dependent efflux of numerous compounds from cells (20). Studies of *ABCC2* (mutations in this gene have been observed in patients with Dubin–Johnson syndrome) revealed that a synonymous SNP within this peptide may lead to increased mRNA stability and concurrent high levels of *ABCC2* expression. In accord with these findings, individuals heterozygous for this SNP display distinct pharmacokinetics when administered the cholesterol-reducing drug pravastatin, a substrate of *ABCC2* (21).

The six known synonymous SNPs of the human *dopamine receptor D2* (*DRD2*) were investigated for their possible functional consequences. One SNP was determined to decrease mRNA stability and translation. Another, while having no detectable effects on the receptor by itself, voided the influence of the former synonymous SNP in haplotype analysis. This study, while adding to the evidence that synonymous SNPs impact mRNA stability, reminds us that the additive effects of two synonymous SNPs cannot be foreseen by assessment of the function of each SNP individually. Ultimately, these cases demonstrate that gene expression can be significantly impacted by synonymous SNPs via perturbations of mRNA stability. While SNPs within regulatory regions of 5'- and 3'-untranslated regions, on a global view, manipulate mRNA stability to a greater extent, synonymous SNPs should still remain a keen point of interest in mRNA stability studies.

### 2.3. mRNA Structure

In conjunction with determinants stemming from the cellular environment, the structure of any mRNA is largely determined by its primary nucleotide sequence. It is known that mRNA structure is integral to a diverse array of biological processes: splicing characteristics and processing of pre-mRNA, translation control and rhythm, and other regulatory functions dependent on specific mRNA structure. The impact of synonymous SNPs on mRNA structure is well documented. The minor allelic forms of alanyl-tRNA synthetase and replication protein A both contain a single synonymous SNP. In 1999, a study into these two minor alleles revealed marked differences in mRNA secondary structure relative to their major allele counterparts. Distinct nuclease S1 digestion patterns were observed for both polymorphic mRNA transcripts. Additional enzymatic probing showed different protection characteristics achieved using nuclease-resistant phosphorothioate oligonucleotides targeted to regions flanking the SNP site, further highlighting the synonymous-SNP-mediated changes in mRNA structure (22). While cellular or physiological implications were not explored in this instance, more recent work on polymorphisms of the catechol-*O*-methyltransferase (*COMT*) gene has successfully demonstrated that synonymous SNPs can modulate mRNA structure and have downstream effects on protein expression and phenotype.

*COMT* inactivates catecholamine neurotransmitters and its genetic variants show strong associations to individual variation in pain perception (23). Until recently, the molecular mechanisms by which one *COMT* haplotype conferred high sensitivity to experimental pain, while another desensitized pain sensation remained unclear. Prior research had focused on a nonsynonymous SNP that showed reduced enzymatic activity, but the results of this work were often inconsistent. However, when this common SNP was analyzed in concurrence with two other synonymous SNPs, the correlation between haplotype and pain sensitivity became more obvious and a causative molecular mediator emerged: mRNA structure.

Specifically, the haplotype linked to high pain responsiveness (HPS) transcribed the most stable stem loop structure in the region containing the SNPs on the mRNA transcript. Conversely, individuals bearing the low pain responsiveness (LPS) haplotype produced a local stem loop structure with the highest Gibbs free energy. Expression levels were inversely related to the folding potentials, as the HPS haplotype exhibited the lowest protein levels and enzymatic activity (24). Moreover, this work revealed that the three major haplotypes of the *COMT* gene are composed of SNPs that assume an unexpected hierarchy of influence. The nonsynonymous SNP *val*<sup>158</sup>*met* is found in the average pain responsiveness haplotype, but is absent in the HPS and LPS haplotypes. Thus, the synonymous polymorphisms of *COMT* are the

foremost determinants of phenotype. This study demonstrates that haplotypes composed of synonymous SNPs can have profound effects on gene function, and in some cases their effects can be stronger than those of their nonsynonymous counterparts. It was hypothesized that the variation in COMT expression levels among haplotypes was due to differences in protein translation efficiency, which marks the final manner in which synonymous SNPs can exert their influence: protein translation and cotranslational protein folding.

#### **2.4. Protein Folding**

The ways in which mRNA splicing, structure, and stability are perturbed by the presence of synonymous SNPs is relatively well established. Recent work has begun to shed light on how synonymous SNPs have an influence on protein folding and ultimate protein function. Although the hypothesis that “silent” SNPs impact protein folding characteristics has been widespread for some time, hard evidence validating this suggestion has been hard to come by until recently. This idea directly challenges two concepts that have dominated thinking about the role (or lack thereof) of synonymous mutations in biology. The first arises from Anfinsen’s principle (25), which holds that the amino acid sequence of a protein alone determines the three-dimensional structure of that protein. Therefore, it stands to reason that mutations which do not alter amino acid residues would not affect the tertiary structure of the protein or its function. The second is the application of this idea to evolutionary theory. Mutations in the genome are selected for on the basis of the effect on the fitness of the organism. Consequently, synonymous mutations have been thought of as being evolutionarily silent (26). Computational and experimental studies in recent years suggest that neither of these perceptions may be entirely true. There is now mounting evidence for selection against synonymous mutations in several organisms (for reviews see (15, 27)). A very comprehensive recent study (28) used a combination of data mining and computer simulation to determine the selection pressures on synonymous mutations in six model organisms (*Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*) belonging to distinct taxonomic groups. What sets this study apart from earlier studies is the taxonomic range of the organisms studied and the fact that the authors simultaneously assessed a set of ten independent parameters. On the basis of principal component analysis, translational fidelity is the single underlying phenomenon that explains correlations with evolutionary rate. Moreover, the view that the amino acid sequence of a protein *alone* determines the three-dimensional structure has also been challenged over the last two decades (29–31).

A central precept in protein folding has been that the native conformation of a protein represents the free-energy minimum, i.e., protein folding is under thermodynamic control. In 1987 Alistair Brown and colleagues made a theoretical case for the kinetic control of protein folding (29). They suggested that rare codons could provide translational pauses which may be vital for correct protein folding. In addition, subsequent studies have shown that synonymous mutations may alter RNA secondary structure, which in turn influences the rate of translation (32, 33). Translational pauses at critical sites on the protein permit individual domains of a protein to fold correctly by curtailing disruptive interactions between unfolded regions of the protein. Experimental validation of these ideas came from the laboratory of Anton Komar (30). Consecutive rare codons were substituted with synonymous frequent codons in chloramphenicol acetyltransferase. This resulted in the enzyme having 20% lower specific activity, which was attributed to a change in the kinetics of protein folding. These studies were carried out using an in vitro translation system, raising the question: is this phenomenon physiologically significant? A significant impediment to designing similar studies in vivo is that proteins that are incorrectly folded are eliminated by extremely efficient quality control mechanisms (34).

Evidence for the biological significance of the phenomena described in the preceding sections was provided by an analysis of the SNPs in the *Multidrug Resistance 1* (*MDR1*) gene (31), the product of which (P-glycoprotein, P-gp) is an ABC transporter. P-gp is an ATP-driven efflux pump that extrudes numerous compounds from cells, including chemotherapeutic agents, and has been implicated in the multidrug resistance of many human cancers (for a review see (35)). A large proportion of several populations (e.g., Chinese, Indian, and Malay) carry a haplotype of three SNPs: C1236T-G2677T-C3435T (36, 37). Of these three, G2677T is a nonsynonymous change, while the other two are synonymous. To understand the functional significance of this haplotype that might confer a selective advantage, the three SNPs individually and as a haplotype were studied (31). There was a significant reduction in the extent of reversal of transport by the P-gp modulators verapamil and cyclosporine A in the haplotype, but not in the individual SNPs. Thus, this change was not a consequence of the single nonsynonymous SNP. The altered transport implicated a structural change which was demonstrated using the conformation-sensitive monoclonal antibodies UIC2, and trypsin digestion patterns supported such a conclusion. Changes in the stability of the mRNA as well as splice variants could also have explained these results. However, a full-length mRNA was found and its levels, P-gp levels, and localization were unchanged in the haplotype. This study thus suggested proteins with identical primary sequences could nonetheless exhibit different conformations leading to a change in function.



The hypothesis proposed by Brown (*see* above and (29)) provides a simple framework to understand these observations, and a more detailed hypothesis has recently been advanced (38, 39). A complex protein is composed of several structural motifs, not all of which have the same folding kinetics. Thus, for example,  $\alpha$ -helices are ultrafast folders (39) and fold much faster than, say,  $\beta$ -sheets. Studies have also shown that proteins fold while being translated (39–43) and that “cotranslational folding” leads to intrapeptide contacts during folding different from those that occur during refolding of the covalently intact but unfolded polypeptide (39, 43–45). Furthermore, as has been stated elsewhere, the rate of translation is not uniform and is often controlled by codon bias, mRNA structure, etc. (46). The rate of translation affects protein conformation most significantly when translation occurs between different structural motifs in a multidomain protein. If these domains are composed of fast folders, a change in the rate of translation is unlikely to have a significant effect. If, however, a slow folding domain *follows* a fast folding domain (*see*, e.g., **Fig. 2** in (38)) the slow folder is likely to have competing conformations, one of which may be “stabilized” by the fast folder. However, ribosome stalling could permit the slow folder to be completely folded prior to formation of the subsequent motifs. Thus, the protein could potentially follow alternative branches in a folding pathway, leading to minor conformational changes.

In retrospect, several distinctive properties of P-gp make it an ideal experimental system to experimentally observe alterations in protein conformation as a consequence of synonymous SNPs that alter the rate of translation. Multidrug transporters utilize a single large cavity comprising a “drug-binding pocket” where individual “drug-binding sites” are generated by subtle alterations in the accessibility of different subset(s) of residues for drug-binding (47). Also, multiple drug resistance proteins can alter substrate specificity fairly easily and even subtle changes in the conformation of the drug-binding pocket can result in changes in substrate specificity (*see* (35) for examples). Also, the absence of a single distinctive correctly folded state of the protein could potentially allow the protein to escape the quality control machinery of the cell. Although evidence from P-gp suggests that these SNPs predominantly exert their influence through an altered rate in translation and cotranslational folding, a holistic investigation, assessing these phenomena and those mentioned previously, is necessary to reveal the functional consequences of any synonymous SNP.

---

### 3. Frequency of SNPs

**Table 2.1** shows that of the 896,454 SNPs documented to date in the human genome (all numbers are from NCBI’s SNP

**Table 2.1**  
**Frequency of various types of single nucleotide polymorphisms (generated using the NCBI database; total number of cases are presented)**

Nucleotide change <sup>a</sup>	Coding synonymous	Coding nonsynonymous	Intron	Splice site	mRNA UTR	Flanking 2000 up/ 500 down	Outside coding regions	Total
A⇌C	3701	8855	438,464	62	13,992	26,104	810,028	1,301,206
A⇌G	29,049	30,783	1,694,999	176	48,035	91,977	2,928,878	4,823,897
A⇌T	1594	4924	360,280	40	9361	18,367	695,762	1,090,328
C⇌G	4416	11,620	447,120	68	14,732	30,048	753,613	1,261,617
C⇌T	31,706	28,366	1,692,412	326	48,426	93,000	2,929,112	4,823,348
G⇌T	3329	9018	440,620	98	13,170	25,745	810,560	1,302,540
A/C/G	75	220	2485	4	130	247	5885	9046
A/C/T	60	114	2166	0	108	191	5570	8209
A/G/T	65	146	2047	3	86	193	5556	8096
C/G/T	78	160	2444	0	135	255	5806	8878
All others (including deletions)	1735	13,542	1,328,009	300	34,160	74,582	1,806,707	3,259,035
Total	75,808	107,748	6,411,225	1077	182,361	360,737	10,757,498	17,896,454

*mRNA UTR* messenger RNA untranslated region

<sup>a</sup>The change in nucleotide, which could occur in either direction

database, reference human genome), about 60% appear on regions outside the coding regions. Of the remaining ones, about 91% of SNPs are in introns, and within the coding regions about 1.5% are nonsynonymous changes, while about 1.1% are synonymous changes. Interestingly, the common changes are cytosine to thymine (or thymine to cytosine) and adenine to guanine (or guanine to adenine), which are transitions and are expected to occur at higher frequencies (48). An interesting feature of this analysis is that when we consider the coding region of the genome alone, the frequencies of synonymous and nonsynonymous SNPs are comparable.

It has often been argued that as the synonymous mutations are “silent” these would show higher rates of mutations; however, this does not appear to be so, at least within the coding regions of the human genome. Moreover the SNPs showing transitional substitutions appear at about twice the rate among nonsynonymous SNPs compared with synonymous SNPs. We endeavored to determine whether this could be explained on the basis of the frequency of each nucleotide in each position within a codon. **Table 2.2** describes the frequency of each nucleotide within the codon. The nucleotides guanine, cytosine, and adenine have about the same frequency, but thymine is significantly less used overall. The frequency is given as a number out of 1,000 codons within the coding region (based on <http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606&aa=1&style=GCG>).

**Table 2.2**  
**Frequency of each nucleotide within the codon**

Nucleotide	1st position	2nd position	3rd position	Total
A	268	310	192	770
C	246	233	298	777
G	311	192	287	790
T	175	265	222	662

Although guanine is the most common base, it is significantly the rarest in the second position and although thymine is significantly less used overall, it is the second most frequent in the second position. Adenine is significantly rarer in the third position among the four bases. However, as synonymous codons most often differ only in their third position, we analyzed the codon frequency from a different perspective with different conclusions. To generate a specific amino acid, there is a need to use one of its codons. Therefore, the frequency of each nucleotide in each codon

position depends primarily on the frequency of the amino acids in the genome. It may well be that the low frequency of thymine in the first position derives directly from the protein sequence. Since synonymous codons most often differ in their third position, it is this position that can be subject to selection on nucleotide preference. However, not all amino acids exhibit a redundancy in codon usage. When we examine only those amino acids that harbor synonymous codons with 3rd position changes, the frequency of adenine is raised from about 19% (as appears in Table 2.2) to about 25% (Table 2.3).

**Table 2.3**  
**Frequency of each nucleotide within the third codon position, when the change does not alter the amino acid (i.e., synonymous substitution)**

Nucleotide	A	C	G	T
Number of appearances within the 3rd position	7,766,153	7,769,044	10,237,953	5,366,408
Percentage	25	25	33	17

Comparing the data in Tables 2.2 and 2.3 demonstrates that guanine appears at a higher frequency in the 3rd position than the other nucleotides and thymine has lower frequency, in situations where more than one codon codes for the same amino acid. These results show that the frequency of nucleotide change in SNPs cannot be attributed to the statistical probability of a particular base pair occurring more frequently at a particular place. It is more likely that natural selection determines the frequency with which the nucleotide changes (depicted in Table 2.1) occur.

---

#### 4. Concluding Remarks

SNPs contribute significantly to genetic variation among humans—on average, one SNP exists per 1.3 kb of human genomic DNA (49). Unlike the onset and progression of rare Mendelian diseases (*see Note 3*), those of common diseases may be influenced substantially by a host of common alleles within the human population (50, 51). Individual variation in drug response is another complex, polygenic phenomenon that arises from common allele determinants. Thus, SNPs are a critical part of overcoming the alarming number of deaths due to fatal drug responses seen in hospitalized patients (52).

There is a consensus that nonsynonymous SNPs and SNPs within regulatory regions have a greater propensity, relative to their synonymous SNP counterparts, to affect gene behavior when analyzed in isolation, i.e., on a wild-type background (48). This understanding is not disputable and is unlikely to be overturned as we continue to document functional SNPs and their cellular implications. Moreover, it is likely that the majority of synonymous SNPs within the human genome do not produce a significant change in gene expression or gene product function, which may account for the widespread reluctance to invest time in studying them. But overlooking synonymous SNPs would be a costly mistake in the quest to see genetics at the forefront of unraveling human disease. Even if only a small percentage of synonymous SNPs generate functional consequences, with the large number of synonymous SNPs across the human genome, these SNPs undoubtedly warrant our attention. While most changes attributed to synonymous SNPs are only subtle, when they are considered in concert with other SNPs and environmental inputs they may become critically important. With that said, instances also abound in which synonymous SNPs act alone to produce phenotypic variation.

As we have argued, a growing body of evidence is revealing that synonymous SNPs can indeed perturb cellular functions and elicit distinct clinical phenotypes. Synonymous SNPs falling within splicing consensus sequences can perturb the normal splicing characteristics of a gene. Synonymous SNPs, often through a change in mRNA structure, can alter the stability of a transcript by changing the ability of RNA binding proteins to recognize the transcript. The secondary structure of mRNA is integral to splicing and processing of pre-mRNA, as well as translation control and rhythm. Therefore, a synonymous SNP altering mRNA structure can influence a host of cellular functions. Finally, these SNPs affect protein translation and cotranslational protein folding, a concept not easily accepted, as it invalidates Anfinsen's principle and demonstrates that a simple thermodynamic model for protein folding is too simplistic. It is debated whether codon usage drives tRNA expression or vice versa, with many recent investigations showing that they coevolve (53). Regardless of the cause, we know organisms show a preference for certain codons/tRNAs over others, with expression profiles mirroring these biases. As a result, the rate at which an amino acid is incorporated into a growing peptide can change significantly from one synonymous codon to the next. During translation, some proteins require programmed translational pauses at specific sites to ensure correct cotranslational folding, which can be accomplished through a stretch of rare codons. Synonymous polymorphisms can alter conventional protein folding kinetics and can lead to a protein product of distinct conformation and activity.

The widespread belief that little or no selective pressure exists on synonymous SNPs has prejudiced our thinking for quite some time. A case in point lies in a common method to assess the evolutionary pressure acting on protein-coding sequences: investigators often calculate the ratio of the rate of nonsynonymous substitutions ( $K_a$ ) to the rate of synonymous substitutions ( $K_s$ ) (54). While  $K_a/K_s$  ratios may be applicable for a majority of coding sequences, the value of this ratio breaks down in areas containing functional synonymous SNPs, as the calculation relies on the assumption that synonymous substitutions are evolutionarily neutral. As we have contended throughout this chapter, the idea that synonymous SNPs are exempt from selective pressure is inconsistent with numerous findings. Some have even noted extraordinary synonymous codon conservation in mammalian coding regions, where synonymous preservation takes precedence over nonsynonymous substitutions (55).

We have presented strong evidence stressing the importance of synonymous SNPs in a span of biological phenomena. With the rapidly growing list of synonymous polymorphisms associated with human diseases (for a review see (56)), these SNPs cannot be ignored by ascribing to them the term “silent.” Rather, synonymous SNPs should be the recipients of increasing attention, especially as we continue to advance novel techniques with which to detect their presence and assay their potential role in human disease and personalized medicine.

---

## 5. Notes

1. Linkage disequilibrium stands for the nonrandom association between two or more alleles (haplotype) in a population, such that combinations of these alleles are more likely to occur together than would be expected by chance. Linkage disequilibrium analysis is a useful method in genetic mapping studies.
2. Owing to the fact that the primary polypeptide sequence appears unchanged, synonymous SNPs have garnered the label “silent” under the belief that their presence is inconsequential to the gene and its corresponding product. Recent evidence shows that this is not really the case and the term “silent” should be generally avoided when referring to synonymous SNPs.
3. Mendelian disease is a popular/common term for a genetic disorder which follows simple Mendelian patterns of inheritance and results from a mutation at a single locus. Most

Mendelian diseases (e.g., cystic fibrosis) are rare, but the frequency may vary depending on the genetic background of a particular population.

## Acknowledgments

This research was supported, in part, by the Intramural Research Program of the National Institutes of Health, National Cancer Institute. Special thanks are expressed to George Leiman, NCI, NIH, and Geetha S., CBER, FDA, for editorial assistance.

## References

- Collins, F. S., Brooks, L. D. and Chakravarti, A. (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231.
- Glazier, A. M., Nadeau, J. H. and Aitman, T. J. (2002) Finding genes that underlie complex traits. *Science* **298**, 2345–2349.
- Goldstein, D. B. and Weale, M. E. (2001) Population genomics: Linkage disequilibrium holds the key. *Curr. Biol.* **11**, R576–579.
- Gumus-Akay, G., Rustemoglu, A., Karadag, A. and Sunguroglu, A. (2008) Genotype and allele frequencies of MDR1 gene C1236T polymorphism in a Turkish population. *Genet. Mol. Res.* **7**, 1193–1199.
- Sauvage, C., Bierne, N., Lapegue, S. and Boudry, P. (2007) Single nucleotide polymorphisms and their relationship to codon usage bias in the pacific oyster *Crassostrea gigas*. *Gene* **406**, 13–22.
- Wang, E. T., Sandberg, R., Luo, S., Khrebtsukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. and Burge, C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476.
- Hart, M. C. and Cooper, J. A. (1999) Vertebrate isoforms of actin capping protein beta have distinct functions in vivo. *J. Cell. Biol.* **147**, 1287–1298.
- Xing, Y., Xu, Q. and Lee, C. (2003) Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett.* **555**, 572–578.
- Egan, M. F., Straub, R. E., Goldberg, T. E., Yakub, I., Callicott, J. H., Hariri, A. R., Mattay, V. S., Bertolino, A., Hyde, T. M., Shannon-Weickert, C., Akil, M., Crook, J., Vakkalanka, R. K., Balkissoon, R., Gibbs, R. A., Kleinman, J. E. and Weinberger, D. R. (2004) Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12604–12609.
- Marti, S. B., Cichon, S., Propping, P. and Nothen, M. (2002) Metabotropic glutamate receptor 3 (GRM3) gene variation is not associated with schizophrenia or bipolar affective disorder in the German population. *Am. J. Med. Genet.* **114**, 46–50.
- Norton, N., Williams, H. J., Dwyer, S., Ivanov, D., Preece, A. C., Gerrish, A., Williams, N. M., Yerassimou, P., Zammit, S., O'Donovan, M. C. and Owen, M. J. (2005) No evidence for association between polymorphisms in GRM3 and schizophrenia. *BMC Psychiatry* **5**, 23.
- Sartorius, L. J., Weinberger, D. R., Hyde, T. M., Harrison, P. J., Kleinman, J. E., and Lipska, B. K. (2008) Expression of a GRM3 splice variant is increased in the dorsolateral prefrontal cortex of individuals carrying a schizophrenia risk SNP. *Neuropsychopharmacology* **33**, 2626–2634.
- Nielsen, K. B., Sorensen, S., Cartegni, L., Corydon, T. J., Doktor, T. K., Schroeder, L. D., Reinert, L. S., Elpeleg, O., Krainer, A. R., Gregersen, N., Kjems, J. and Andresen, B. S. (2007) Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: A synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am. J. Hum. Genet.* **80**, 416–432.

14. Yakub, I., Lillibridge, K. M., Moran, A., Gonzalez, O. Y., Belmont, J., Gibbs, R. A. and Tweardy, D. J. (2005) Single nucleotide polymorphisms in genes for 2'-5'-oligoadenylate synthetase and RNase L in patients hospitalized with West Nile virus infection. *J. Infect. Dis.* **192**, 1741–1748.
15. Cartegni, L., Chew, S. L. and Krainer, A. R. (2002) Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285–298.
16. Fedetz, M., Matesanz, F., Caro-Maldonado, A., Fernandez, O., Tamayo, J. A., Guerrero, M., Delgado, C., Lopez-Guerrero, J. A. and Alcina, A. (2006) OAS1 gene haplotype confers susceptibility to multiple sclerosis. *Tissue Antigens* **68**, 446–449.
17. Solis-Anez, E., Delgado-Luengo, W., Borjas-Fuentes, L., Zabala, W., Arraiz, N., Pineda, L., Portillo, M. G., Gonzalez-Ferrer, S., Chacin, J. A., Pena, J., Montiel, C., Morales, A., Rojas de Atencio, A., Canizales, J., Gonzalez, R., Miranda, L. E., Abreu, N., and Delgado, J. (2007) [Molecular analysis of the GABRB3 gene in autistic patients: An exploratory study]. *Invest Clin.* **48**, 225–242.
18. Ross, J. (1995) mRNA stability in mammalian cells. *Microbiol. Rev.* **59**, 423–450.
19. Capon, F., Allen, M. H., Ameen, M., Burden, A. D., Tillman, D., Barker, J. N. and Trembath, R. C. (2004) A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum. Mol. Genet.* **13**, 2361–2368.
20. Jones, P. M. and George, A. M. (2004) The abc transporter structure and mechanism: Perspectives on recent research. *Cell. Mol. Life Sci.* **61**, 682–699.
21. Niemi, M., Arnold, K. A., Backman, J. T., Pasanen, M. K., Godtel-Armbrust, U., Wojnowski, L., Zanger, U. M., Neuvonen, P. J., Eichelbaum, M., Kivisto, K. T. and Lang, T. (2006) Association of genetic polymorphism in ABC2 with hepatic multidrug resistance-associated protein 2 expression and pravastatin pharmacokinetics. *Pharmacogenet. Genomics* **16**, 801–808.
22. Shen, L. X., Basilion, J. P. and Stanton, V. P., Jr. (1999) Single-nucleotide polymorphisms can cause different structural folds of mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7871–7876.
23. Diatchenko, L., Slade, G. D., Nackley, A. G., Bhalang, K., Sigurdsson, A., Belfer, I., Goldman, D., Xu, K., Shabalina, S. A., Shagin, D., Max, M. B., Makarov, S. S., and Maixner, W. (2005) Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum. Mol. Genet.* **14**, 135–143.
24. Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., Maixner, W. and Diatchenko, L. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930–1933.
25. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223–230.
26. Kimura, M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276.
27. Chamary, J. V., Parmley, J. L. and Hurst, L. D. (2006) Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98–108.
28. Drummond, D. A. and Wilke, C. O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352.
29. Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R. and Brown, A. J. (1987) The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J. Mol. Biol.* **193**, 413–417.
30. Komar, A. A., Lesnik, T. and Reiss, C. (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**, 387–391.
31. Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V. and Gottesman, M. M. (2007) A “Silent” Polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525–528.
32. Ivanov, I. G., Saraffova, A. A. and Abouhaidar, M. G. (1997) Unusual effect of clusters of rare arginine (AGG) codons on the expression of human interferon alpha 1 gene in *Escherichia coli*. *Int. J. Biochem. Cell. Biol.* **29**, 659–666.
33. Parmley, J. L. and Hurst, L. D. (2007) How do synonymous mutations affect fitness? *Bioessays* **29**, 515–519.
34. Bukau, B., Weissman, J. and Horwich, A. (2006) Molecular chaperones and protein quality control. *Cell* **125**, 443–451.
35. Ambudkar, S. V., Dey, S., Hrycyna, C. A., Ramachandra, M., Pastan, I. and Gottesman, M. M. (1999) Biochemical, cellular, and pharmacological aspects of the multidrug



- transporter. *Annu. Rev. Pharmacol. Toxicol.* **39**, 361–398.
36. Pauli-Magnus, C. and Kroetz, D. L. (2004) Functional implications of genetic polymorphisms in the multidrug resistance gene MDR1 (ABCB1). *Pharm. Res.* **21**, 904–913.
  37. Kimchi-Sarfaty, C., Marple, A. H., Shinar, S., Kimchi, A. M., Scavo, D., Roma, M. I., Kim, I. W., Jones, A., Arora, M., Gribar, J., Gurwitz, D., and Gottesman, M. M. (2007) Ethnicity-related polymorphisms and haplotypes in the human ABCB1 gene. *Pharmacogenomics* **8**, 29–39.
  38. Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M. and Nussinov, R. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.* **383**, 281–291.
  39. Komar, A. A. (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **34**, 16–24.
  40. Clarke, D. T., Doig, A. J., Stapley, B. J. and Jones, G. R. (1999) The alpha-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7232–7237.
  41. Kiho, Y. and Rich, A. (1964) Induced enzyme formed on bacterial polyribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 111–118.
  42. Fedorov, A. N. and Baldwin, T. O. (1995) Contribution of cotranslational folding to the rate of formation of native protein structure. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1227–1231.
  43. Fedorov, A. N. and Baldwin, T. O. (1997) Cotranslational protein folding. *J. Biol. Chem.* **272**, 32715–32718.
  44. Batey, S., Scott, K. A. and Clarke, J. (2006) Complex folding kinetics of a multidomain protein. *Biophys. J.* **90**, 2120–2130.
  45. Kowarik, M., Kung, S., Martoglio, B. and Helenius, A. (2002) Protein folding during cotranslational translocation in the endoplasmic reticulum. *Mol. Cell.* **10**, 769–778.
  46. Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V. and Gottesman, M. M. (2007) The sounds of silence: Synonymous mutations affect function. *Pharmacogenomics* **8**, 527–532.
  47. Schumacher, M. A. and Brennan, R. G. (2003) Deciphering the molecular basis of multidrug recognition: Crystal structures of the staphylococcus aureus multidrug binding transcription regulator QacR. *Res. Microbiol.* **154**, 69–77.
  48. Keller, I., Bensasson, D. and Nichols, R. A. (2007) Transition-transversion bias is not universal: A counter example from grasshopper pseudogenes. *PLoS Genet.* **3**, e22.
  49. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q. and Lander, E. S. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**, 231–238.
  50. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.
  51. Lander, E. S. (1996) The new genomics: Global views of biology. *Science* **274**, 536–539.
  52. Lazarou, J., Pomeranz, B. H. and Corey, P. N. (1998) Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* **279**, 1200–1205.
  53. Higgs, P. G. and Ran, W. (2008) Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Mol. Biol. Evol.* **25**, 2279–2291.
  54. Hurst, L. D. (2002) The ka/ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486.
  55. Schattner, P. and Diekhans, M. (2006) Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res.* **34**, 1700–1710.
  56. Charnary, J. V. and Hurst, L. D. (2009) How Trivial DNA Changes Can Hurt Health. *Sci Am.* **30**, 46–53.

Single Nucleotide Polymorphisms

Methods and Protocols

Komar, A.A. (Ed.)

2009, XIV, 464 p. 100 illus., 18 illus. in color., Hardcover

ISBN: 978-1-60327-410-4

A product of Humana Press