# Understanding the contribution of synonymous mutations to human disease

*Zuben E. Sauna and Chava Kimchi-Sarfaty*

Abstract | Synonymous mutations — sometimes called 'silent' mutations — are now widely acknowledged to be able to cause changes in protein expression, conformation and function. The recent increase in knowledge about the association of genetic variants with disease, particularly through genome-wide association studies, has revealed a substantial contribution of synonymous SNPs to human disease risk and other complex traits. Here we review current understanding of the extent to which synonymous mutations influence disease, the various molecular mechanisms that underlie these effects and the implications for future research and biomedical applications.

Splicing
The transcribed precursor RNA consists of exons (which encode amino acids) and introns, which must be edited out: splicing is the process by which this occurs.

Synonymous SNPs
(sSNPs). Single-nucleotide changes that do not result in a change in the amino acid in the translated protein.

Protein therapeutics
Proteins used in the treatment of human diseases that are purified from animal or human sources or, increasingly, manufactured by recombinant DNA technology.

*Laboratory of Hemostasis, Division of Hematology, Center for Biologics Evaluation and Research, Food and Drug Administration, 29 Lincoln Drive, Bethesda, Maryland 20892, USA.*
*e-mails:* zuben.sauna@fda.hhs.gov; chava.kimchi-sarfaty@fda.hhs.gov
doi:10.1038/nrg3051
Published online
31 August 2011

Rapid progress in our understanding of human genetic variation shows that many different types of genetic changes affect complex phenotypes, such as disease risk. Historically, much of the focus has been on mutations that change the amino acid sequence of a protein (non-synonymous mutations), but there is increasing awareness that other types of genetic variations can affect disease risk and treatment outcomes. Owing to the degeneracy of the genetic code, synonymous mutations occurring in the gene-coding regions do not change the amino acid composition of the encoded proteins. In addition, mutations in introns, 3′ and 5′UTRs and other non-coding regions do not alter amino acid sequences. Owing to the dogma that the structure (and therefore function) of proteins is determined by the amino acid sequence[1], synonymous mutations were, until recently, referred to as silent.

A corollary of this perception was that synonymous mutations would have no effect on the fitness of an organism and would be 'neutral' during evolution[2]. Two lines of evidence began to suggest that synonymous mutations could have functional consequences. The first was based on findings that, in many organisms, there is a codon usage bias vis-à-vis synonymous codons, suggesting that even synonymous codons were under evolutionary pressure (see REFS 3,4 for comprehensive Reviews on this subject, which is not discussed here). Second, advances in our understanding of protein synthesis and folding have led to discoveries that provide the mechanistic and conceptual framework to understand this phenomenon. Considerable evidence has accumulated over the past decade to show that synonymous mutations can result in aberrant mRNA splicing, which can lead to human disease[5]. Emerging evidence also suggests that synonymous SNPs (sSNPs) could affect mRNA stability and thus protein expression and enzymatic activity[6]. In addition, although a theoretical case for how codon bias could affect tertiary protein structure was presented several decades ago, it was only demonstrated in 2007 that sSNPs can affect protein conformation and have functional and clinical consequences[7].

Technological innovation in genotyping platforms[8–11] and the development of computational methods to analyse these data have resulted in dramatic advancements in genetic association studies over the past decade; studies that used to be limited to a small number of candidate genes now encompass the entire genome. These genome-wide association studies (GWASs), by taking a hypothesis-free approach, have made it possible to identify loci in genes that were not necessarily previously associated with the disease[12].

In this article, we have generated a compendium of human diseases or clinical conditions associated with synonymous mutations. We also discuss recent studies that are elucidating the mechanisms by which synonymous substitutions bring about changes in the phenotype by affecting splicing accuracy, translation fidelity, mRNA structure and protein folding. Finally, we discuss how an increased understanding of the effect of synonymous mutations could have an impact on future clinical applications, such as pharmacogenetics and protein therapeutics.

## Synonymous mutations and human disease

***Synonymous mutations are implicated in human diseases.*** Despite the perception in many earlier investigations that synonymous mutations are 'silent', which may have led to under-reporting of the effects of synonymous mutations, the list of diseases associated with such mutations is now quite large and continues to grow — a summary is provided in TABLE 1, and Supplementary information S1 (table) gives a detailed list. The synonymous mutations listed in these tables have been identified using hypothesis-driven candidate gene approaches, hypothesis-free GWASs or approaches that use arrays with thousands of genes that are known or predicted to be associated with a disease condition or multi-gene trait (that is, a 'compromise' between the candidate gene and genome-wide approaches)[13]. We have collated this list based only on reports of synonymous mutations in human genes that have a significant association with disease; for example, synonymous mutations in the genomes of pathogens that affect their infectivity are not included. In addition, we have not included reports from the literature that identified both synonymous and non-synonymous mutations, but where the synonymous mutations did not correlate significantly with the disease. Mutations located in the promoter, introns or 3′ and 5′UTRs were also excluded from this list, although they may have similar effects to the synonymous changes discussed in this article. The resulting list (Supplementary information S1 (table)) shows that almost 50 diseases afflicting most organ systems have been associated with synonymous mutations so far.

***The functional importance of synonymous mutations.*** A number of studies illustrate the hazard of the conjecture that non-synonymous mutations have no effect on the phenotype of an organism. For example, three common SNPs in the catechol-*O*-methyltransferase (*COMT*) gene have been associated with regulation of pain perception[14]. As only one of these was a non-synonymous (Val-158-Met) change, it was assumed to be the causal variant[15], despite modest associations with this SNP per se[14]; however, experimental evidence showed that the sSNPs better account for the differences in pain sensitivity[6]. Similarly, two mutations (one synonymous and the other non-synonymous) were found in individuals with Treacher–Collins syndrome. The non-synonymous variation was previously reported to be disease-causing[16]. However, more recent pedigree analysis showed that the non-synonymous mutation was present in five unaffected family members and was thus not pathogenic, whereas the synonymous variant that was present in the patient alone was the causative mutation[17].

Even more fascinating is a recent finding regarding the ΔF508 mutation in cystic fibrosis transmembrane conductance regulator (*CFTR*), which is the mutation that is most frequently associated with cystic fibrosis. For over two decades, the focus of research has been on the protein and the consequences of deletion of phenylalanine at position 508. In the wild-type protein, an isoleucine precedes the phenylalanine, and these amino acids have the codons ATC and TTT, respectively. The ΔF508 mutation is generated by deletion of the last C of the isoleucine codon and the first two Ts of the phenylalanine codon. This forms the codon ATT, which encodes isoleucine, but with a synonymous (C to T) substitution compared with the usual 507 codon. An elegant study by Bartoszewski and co-workers[18] demonstrated that this synonymous change alters the mRNA structure and leads to a misfolded protein. Reverting the ATT codon to the original ATC codon at position 507 in the context of ΔF508 leads to correctly folded mRNA and higher protein levels than observed with the native ΔF508 mutation.

***Synonymous mutations can predict disease susceptibility and treatment outcomes.*** Synonymous mutations can occur in a gene that has been directly associated with disease pathogenesis, as has been shown in the case of Treacher–Collins syndrome[17], X-linked infantile spinal muscular atrophy[19], Seckel syndrome[20] and cystic fibrosis[18]. However, association studies also identify synonymous mutations in genes that do not have a known link to the disease mechanism but that are nonetheless clinically important because they can help to predict disease outcomes. For example, an sSNP identified in Wilms' tumour 1 (*WT1*) of patients with childhood acute myeloid leukaemia correlated with improved disease outcomes[21]. Additionally, an sSNP in heat shock 27 kDa protein family member 7 (*HSPB7*) is strongly associated with susceptibility to dilated cardiomyopathy[13].

Finally, the safety and efficacy of small-molecule drugs is affected by mutations in proteins that are involved in drug disposition, such as transporters and metabolizing enzymes[22]. In 2007, it was reported that the synonymous mutations in a haplotype (1,236C to T, 2,677G to T and 3,435C to T) that is commonly found in *ABCB1* affect protein folding and substrate specificity[7]. Since then, a large number of studies have reported that this haplotype is associated with treatment outcomes. Although the *ABCB1* gene product is commonly associated with multidrug resistance during cancer chemotherapy, this molecular pump acts on numerous other drugs as well[23]. For example, the haplotype is associated with prolonged, progression-free survival of patients with metastatic renal cell cancer who were treated with sunitinib[24], as well as improved outcomes for patients with inflammatory bowel disease who were treated with tacrolimus[25]. In the case of the patients with metastatic renal cell cancer, this prolonged survival could be a consequence of the altered affinity of some drugs for ABCB1 as a consequence of the haplotype, which would result in less efficient efflux and increased bioavailability of the tyrosine kinase inhibitor, sunitinib. In the case of the patients with inflammatory bowel disease, the authors suggest that the polymorphisms affect ABCB1 levels or its function in the gastrointestinal tract. Tacrolimus absorption from the gastrointestinal tract is likely to be higher in patients in whom intestinal ABCB1-mediated clearance is reduced.

Interestingly, although it has been demonstrated that the *ABCB1* gene product plays a substantial part in

Table 1 | **Predicted mechanisms for physiological effects**

| Class of human disease | Splicing | RNA structure | Rate of translation | Unknown |
|---|---|---|---|---|
| Vision | ✓ | ✓ | | ✓ |
| Muscle | ✓ | | | ✓ |
| Kidney | ✓ | ✓ | | |
| Skin | | ✓ | | |
| Lung | ✓ | ✓ | ✓ | |
| Heart | | ✓ | | |
| Blood-related | ✓ | ✓ | ✓ | ✓ |
| Bone | ✓ | | | |
| Cancer | ✓ | ✓ | ✓ | ✓ |
| Immune | ✓ | ✓ | | |
| Diabetes | | ✓ | | ✓ |
| Neurological | ✓ | ✓ | | ✓ |
| Longevity | ✓ | ✓ | | ✓ |
| Liver disease | ✓ | ✓ | | ✓ |

The table shows predicted mechanisms by which synonymous mutations associated with different classes of human diseases exert a physiological effect. A detailed list of human diseases associated with synonymous mutations is given in Supplementary information S1 (table). The mechanisms by which synonymous mutations have been shown to affect protein levels and conformation and thus bring about physiological changes are illustrated in FIG. 1. The mechanisms identified here are based on published follow-up computational or experimental studies for the associations listed in Supplementary information S1 (table). Where such follow-up studies were not conducted, we have endeavoured to predict plausible mechanisms based on the position of the mutation, the relative synonymous codon usage (RSCU) change as a consequence of the mutation (BOX 1) or the potential alterations in the mRNA structure.

circumventing chemotherapy and in the absorption, distribution, metabolism and excretion of drugs, the physiological role of this transporter has not been elucidated[26]. The fact that the haplotype (1,236C to T, 2,677G to T and 3,435C to T) of *ABCB1* is associated with susceptibility to a range of diseases could eventually provide some clues about the physiological substrate (or substrates).

***The global importance of synonymous mutations in human disease.*** Although many of the findings described above are individually interesting, a broader question is how important are synonymous mutations genome-wide? A recent study by Chen and co-workers[27] is the only one to date that endeavours to address this issue directly. The authors conducted a survey of 21,429 associations between diseases and SNPs curated from 2,113 reports studying human genetic association, and they concluded that non-synonymous SNPs (nsSNPs) and sSNPs shared similar likelihood and effect size for disease association. Similarly, approximately 1,600 extended regions with highly conserved sSNPs were found in a set of 11,786 genes[28]. This is consistent with the estimate made by Chamary and Hurst[29] that 5–10% of human genes contain at least one region in which silent mutations could be harmful. Thus, global surveys, computational estimates and individual mechanistic studies suggest that sSNPs that have been identified in GWASs should be included in follow-up functional and mechanistic studies.

Non-synonymous SNPs
(nsSNPs). Single-nucleotide changes that result in a change in the amino acid in the translated protein.

Spliceosome
A complex of small nuclear RNAs and protein subunits that removes introns from the transcribed precursor mRNA.

Splicing enhancers
Short nucleotide sequences that flag the boundaries of the exon for the splicing machinery of the cell.

MicroRNAs
(miRNAs). Short RNA molecules that are post-transcriptional regulators. They bind to complementary sequences on mRNAs and result in gene repression.

## Effects on protein levels, structure and function

The genetic studies described above show that synonymous mutations are associated with human disease; however, these studies have not, for the most part, investigated the underlying molecular mechanisms in detail. Experimental and computational studies from several laboratories indicate that there is a range of mechanisms by which synonymous mutations can affect the yields of active, correctly folded protein and thus have an impact on physiological activity. The amount of protein is influenced by transcription, translation and turnover of mRNAs and proteins. Recently, global analysis of the control of gene expression[30] suggested the dominance of regulation at the translation level. Translation rate is a function of initiation, elongation and termination. Initiation is generally thought to be rate-limiting, but defining the relative contribution of each step to protein abundance continues to be a challenging task. A comprehensive characterization of determinants of human protein abundance[31] showed that mRNA abundance explains 25–30% of the variation, whereas another 30–40% of the variation can be accounted for by a combination of various sequence features relating to translation and protein composition.

***RNA processing and post-transcriptional regulation.*** In the primary transcription products of human genes (namely, precursor mRNAs (pre-mRNAs)), exons are separated by non-coding introns, which could represent over 90% of a transcription unit. The spliceosome constitutes the cellular machinery that regulates and executes the removal of introns (splicing) with a high degree of precision[32]. This is accomplished by the binding of splicing regulatory proteins to short nucleotide sequences in the exons (exonic splicing enhancers), which, in turn, direct the spliceosome to the correct exon–intron boundary (FIG. 1a, left panel). Efficient splicing has limited tolerance of mutations in the exonic splicing enhancers, even if they have no effect on protein coding[33]. Thus, a disruption of the spliceosome was the first molecular mechanism that explained the association of synonymous mutations with human disease (FIG. 1a, right panel). This topic has been comprehensively reviewed[3,5,34] and will not be discussed in detail here.

MicroRNAs (miRNAs) are important post-transcriptional regulators; it is estimated that these molecules target about 60% of mammalian genes[35]. In humans, many miRNA target sites are in the 3′UTRs of the mRNA and so are beyond the scope of this article. However, recent work suggests that a synonymous mutation in the coding region of immune-related GTPase family M (*IRGM*) alters an miRNA binding site[36]. The 313 C to T polymorphism in *IRGM* resulted in reduced binding of miR-196. Previous studies have shown that polymorphisms in *IRGM*, including the 313T allele, are associated with Crohn's disease[37]. The recent findings suggest that loss of miRNA-mediated regulation of the 313T *IRGM* allele results in sustained expression of the IRGM protein, which could be the underlying mechanism for the disease risk.

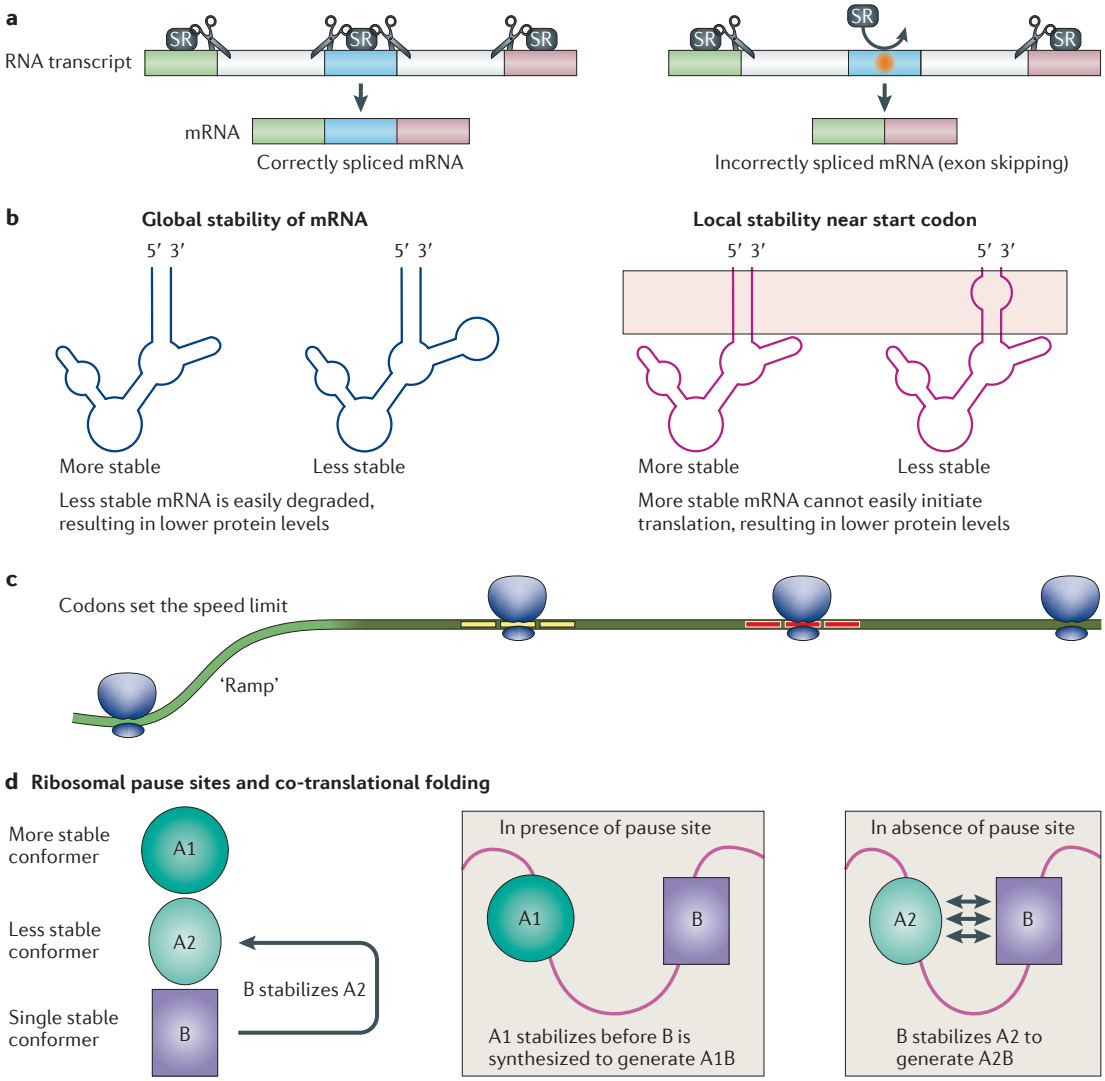


Figure 1 | **Molecular mechanisms by which synonymous mutations can change in protein levels or conformation.** **a** | During normal RNA processing, the spliceosome edits the transcription product to remove the introns (grey bars). The splicing regulatory proteins (black boxes labelled 'SR') have a central role in flagging the ends of the exons (green, blue and pink bars) for excision. The splicing regulatory proteins recognize short sequences in the exon. Synonymous mutations (right panel, orange shading) can result in a failure of the splicing regulatory proteins to interact with some exons, so their boundaries are not recognized by the spliceosome machinery. Exon skipping results in truncated mRNAs, which can result in non-functional proteins. **b** | Both the global stability (left panel) and local stability (right panel) of mRNA are important and can determine protein levels. When synonymous mutations result in a less stable mRNA (see main text for examples), the result can be reduced levels of the protein, which can sometimes have clinical consequences. The local stability of mRNA structures near the start codon, however, controls protein levels by initiation of translation. Very stable structures in this region impede translation, resulting in lower protein levels. **c** | Genome-wide trends in codon usage have been used to determine some 'rules' that show how control of the kinetics of translation optimizes both the speed and accuracy of translation. The first 50–60 codons use rare tRNAs to generate a 'ramp' that prevents subsequent ribosomal traffic jams. Following the ramp, the protein is translated at full speed. This can be accomplished either by using more frequent codons or by using autocorrelated codons — that is, when amino acids recur in the sequence, they use the same codon, allowing the tRNA to be recycled (yellow and red dashes). Synonymous mutations that affect the kinetics of translation can either slow down the rate of protein synthesis or lead to protein misfolding, which, in turn, can result in proteotoxicity. **d** | A model has recently been developed that explains how synonymous mutations that generate so-called translation 'pause sites' can result in alternative conformers during co-translational folding. This example shows two regions, A and B, of a hypothetical protein. B only has one stable conformation, but A has two conformers, A1 and A2, which are more and less stable, respectively. However, B can stabilize A2. If the protein undergoes co-translational folding and if A and B are permitted to fold independently, the A1B form is generated; independent folding can be facilitated by a ribosome pause site between the A and B. However, if B can fold fast — before A has folded — A2B can form. The presence or absence of pause sites between A and B can influence which of these two forms (A1B or A2B) of the protein will be synthesized. Synonymous mutations that alter the pause site could therefore result in a different conformation of the protein. Part **d** is based on a hypothesis presented by Tsai *et al.*[68].

Exon skipping
RNA splicing that results in one or more exons being eliminated from the final mRNA.

Another way in which a synonymous mutation might influence protein levels is by altering mRNA degradation. For example, a synonymous mutation in dopamine receptor D2 (*DRD2*) results in a less stable mRNA secondary structure, which correlates with increased mRNA degradation and a lower protein level[38]. However, mRNA secondary structures can also have many other effects.

***Translation initiation.*** Nackley *et al.*[6] were among the first to provide strong experimental evidence that mRNA secondary structure modulates protein expression, which, in turn, has a physiological consequence. For *COMT* (also see previous section), these investigators demonstrated that the synonymous changes resulted in altered mRNA structures and the most stable structure was associated with the lowest protein levels, reduced enzyme activity and greater sensitivity to pain. A plausible explanation for decreased protein levels accompanying more stable mRNA structures is provided by studies on mRNA structural stability and translation initiation[39,40]. In one of these studies[39], a library of 154 copies of *GFP* with random changes at synonymous sites was evaluated for protein expression. There was a 250-fold variation in protein levels across the library, and the stability of mRNA secondary structure near the start codon explained more than half of this variation.

The model is that more stable local structures in this region of the mRNA hinder translation initiation, resulting in reduced protein expression (FIG. 1b); this is supported by additional studies[41,42]. Importantly, these findings indicate that, in addition to the nature of the synonymous mutation, the location of the mutation in the gene is important.

***Early translation elongation.*** To understand the possible effects of synonymous mutations on translation elongation, we first need to consider the concept of codon bias (reviewed in REF. 4); although the genetic code is degenerate, synonymous codons are not used in equal frequencies. There is evidence in some organisms that codons that are used preferentially will correspond to more abundant tRNAs[43] and that codon usage and tRNA content may have co-evolved[44]. It has, however, not been determined whether codon usage has driven tRNA evolution or the other way around. Regardless, differences in tRNA levels can affect the rate of translation elongation[45], and recent studies have tried to understand how this can affect protein levels and protein folding. BOX 1 discusses relative synonymous codon usage associated with synonymous mutations that cause disease

It has been suggested that in bacteria, archaea and eukaryotes, codon usage and tRNA abundance have co-evolved to optimize translation efficiency[46–48]. Although there is some truth to this perspective, it is too simplistic, and studies in higher organisms suggest that using both rare and frequent codons may be necessary to balance the amount of protein that is produced with correct protein folding (discussed further below).

Genome-wide studies of codon usage have provided notable advances and should be distinguished from earlier work in which the kinetic control of protein folding was often investigated in individual genes or proteins, often using cell-free systems[49–53]. Now, data mining and computational modelling enable global patterns to be identified, which are then tested in a proof-of-principle experimental system. For example, such an approach was used to determine whether there are conserved patterns in the distribution of rare and frequent codons across individual mRNAs and in the transcriptome[54]. A consistent feature of mRNA transcripts was a cluster of rare codons at the beginning of the sequence. Thus, it would be expected that translation of the first 30–50 codons would be slow. These so-called 'ramps' that slow translation elongation immediately after initiation are found in many genes but are more frequently associated with highly expressed genes. This is consistent with the proposed function of the ramp, which is to space the ribosomes on the mRNA[54]. A suitable distance between adjacent ribosomes would prevent ribosome congestion, which could cause ribosome stalling and lead to misfolded or truncated proteins[55,56]. This view is consistent with a genome-wide analysis of translation based on deep sequencing of ribosome-protected mRNA fragments[57], which demonstrated a change in the rate of translation from early to late translation elongation. Together with the studies discussed above, which show that a more stable local mRNA structure at the beginning of genes impedes translation initiation[41,42], it thus appears likely that the first 50 codons of a gene are important for the efficiency and fidelity of translation, but multiple mechanisms may be operational (FIG. 1c).

***Later translation elongation.*** The slow ramp is often followed by rapid translation elongation (FIG. 1c), and recent studies have provided tantalizing clues about how this may be accomplished. One suggestion is that, when the same amino acid recurs in a sequence, the same codon is likely to be used[58]. The underlying biophysical principle for this phenomenon is suggested to be the advantage conferred by the reuse of a tRNA (FIG. 1c); it has been shown that aminoacyl tRNA synthetases form complexes associated with ribosomes[59–61]. The idea that choice of codon is not only influenced by overall codon frequency is further illustrated in BOX 1, which shows that in individual instances, either a rare-to-frequent or frequent-to-rare alteration in the codon can be associated with a disease condition.

***Co-translational folding.*** A significant body of literature suggests that the concept of the primary amino acid sequence of a protein containing all of the information for the three-dimensional structure of a protein is an oversimplification[62–64]. The secondary structural elements of a protein exhibit cooperativity during folding into the tertiary structure[62], and many proteins assume secondary and tertiary structure concomitant with synthesis[65–67]. The different domains of a protein stabilize

Box 1 | **Synonymous codon usage and human disease**

Codon usage is one way in which a kinetic control of translation elongation is imposed. As discussed in the main text, the speed of translation is often important for accurate folding of the protein and to reduce misfolded or aggregated molecules, which have adverse consequences. Synonymous mutations that affect codon usage can disrupt this process. Sharp and colleagues[91] introduced the measurement of relative synonymous codon usage (RSCU) and suggested that a change in RSCU might be associated with a change in local translation elongation rates. This concept was later supported by experimental work[51,92,93] that demonstrated that the effect of a single codon change can be quite pronounced.

We calculated the change in RSCU values as a consequence of synonymous mutations that cause human diseases. The RSCU values were calculated as $RSCU = SN_c / N_a$, where $N_c$ is the frequency of a particular codon c within the human genome, and $N_a$ is the frequency of the amino acid a represented by the codon c. The amino acid a represented by the codon c may also have alternative representations by other codons; $S$ is the number of synonymous codons for a. The frequencies were obtained from the table available at the Codon Usage Database. The results are in given in Supplementary information S2 (table) with a selection of the results for the five genes shown below by way of a demonstration.

The ΔRSCU represents a change in the RSCU values at a gene of interest as a consequence of mutations in the gene. A negative ΔRSCU value — that is, the mutation introduces a rarer codon — might be associated with a slower local rate of translation elongation compared with the wild type (with an opposite effect predicted for a positive ΔRSCU). The changes in RSCU associated with disease conditions exhibit both positive values (in 21 cases) and negative values (in 25 cases). This suggests that synonymous mutations might either increase or decrease local translation rates, which is consistent with the concept that synonymous codons regulate protein levels and conformation by multiple mechanisms. In addition, as discussed in the main text, the location of the mutation can also be crucial, and this is not captured by the ΔRSCU. Also, it should be noted that the association between RSCU and translation elongation rates is based on *in vitro* experiments[51,92,93], and it is not yet clear what the physiological consequences of RSCU changes would be.

| Disease | Gene | Reference SNP number | Location (sequence range of exon) | Codon change | | RSCU change | | ΔRSCU | Amino acid |
|---|---|---|---|---|---|---|---|---|---|
| | | | | From | To | From | To | | |
| Pulmonary sarcoidosis | Caspase recruitment domain 15 (*CARD15*) | rs1861759 | mRNA position 1866, exon 4 (752–2567) | CGT | CGG | 0.48 | 1.21 | 0.73 | R |
| Haemophilia B | *F9* | Not known | Exon 5 | GTG | GTA | 1.85 | 0.47 | –1.38 | V |
| Non-small-cell lung carcinoma | Epidermal growth factor receptor (*EGFR*) | rs2293347 | mRNA position 3228, exon 27 (3193–3360) | GAC | GAT | 1.07 | 0.93 | –0.14 | D |
| Cervical and vulvar cancer | Interleukin-2 (*IL2*) | rs2069763 | mRNA position 169, exon 1 (1–202) | CTG | CTT | 2.37 | 0.79 | –1.58 | L |
| Adult and child attention deficit/ hyperactivity disorder (ADHD) | Neurotrophin 3 (*NTF3*) | rs6332 | mRNA position 502, exon 2b (230–1335) or position 368, exon 2a (1–1168) (different splice variants) | CCG | CCA | 0.45 | 1.11 | 0.65 | P |

each other and influence the folding process. For example, a less stable domain could be stabilized by being in proximity with another domain.

Co-translational folding appears to be regulated by the choice and position of codons. Fast folding elements of the protein structure, such as α-helices, have fewer rare codons than slow folding β-sheets, and genes that code for larger, more complex proteins tend to have a higher proportion of rare codons, and these occur at strategic locations, such as between domains[68]. This evidence supports the idea that choice of codons can either limit or extend the time for the secondary structural elements of a protein to interact, which, in turn, affects the final conformation of the protein[68]. For example, a ribosomal pause site that slows down translation between two domains would allow both domains to fold independently; the removal of the pause site by a synonymous mutation would favour the cooperative folding of the two domains, resulting in an altered structure[68] (FIG. 1d). Synonymous mutations that disrupt the highly synchronized process of protein folding can lead to misfolding, aggregation and lower levels of protein synthesis[69].

***Proteostasis.*** The studies described above show that the use of a particular codon can be important for the correct folding of a protein. Misfolding that occurs as a result of a synonymous mutation may have consequences beyond the functionality of the particular protein in which the mutation occurred[70]. A vivid experimental demonstration of this is provided by the work of Morimoto and colleagues[71]. A single protein that is prone to misfolding can lead to a cascade effect of misfolding in other proteins and, eventually,

Relative synonymous codon usage
(RSCU). This is a simple measure of non-uniform usage of synonymous codons in a coding sequence. RSCU values show the number of times a particular codon is observed relative to the number of times that the codon would be observed if all the codons for a given amino acid had the same probability.

to proteotoxicity. The results of genome-wide analyses of the fitness consequences of synonymous mutations demonstrate that codon usage enforces accurate translation to form a correctly folded protein, which, in turn, may be crucial for maintaining proteostasis. Consequently, there has been a subtle but significant shift away from the conventional perception that evolution of protein-coding genes is primarily determined by functional constraints. Thus, some investigators have argued that selection can act on the 'biophysical processes', even when there is little or no change in protein function per se[72,73]. More broadly, Wilke and Drummond[73,74] posit that evolution of genes is strongly shaped by processes such as protein synthesis, protein folding, and so forth, as well as the end result (that is, the activity of the protein). This view is supported by a principal component analysis designed to identify the biological basis of selective pressures on coding sequence evolution. In this study, a set of ten pairwise correlations was evaluated, and translation infidelity (that is, selection against protein misfolding) was identified as the single underlying cause[74]. Global studies such as these suggest that post-transcriptional processes that are governed by both synonymous and non-synonymous mutations are under selection pressure, because mutations at these sites can perturb the highly conserved network of biological pathways that maintain the proteome for normal cellular function[70,75].

## Clinical and technological implications

Although there may still be some disagreement about the overall importance of synonymous mutations, it is now indisputable that such mutations can and do have a role in human health and disease (Supplementary information S1 (table)). This has considerable implications in the practice of medicine and in biotechnology.

***Pharmacogenetics.*** In addition to affecting disease risk, it is becoming increasingly evident that SNPs have a major role in individual differences observed in how patients respond to medication, whether the medication will cause adverse effects, how their disease will progress, and so forth[69]. These advances have prompted the call for personalized medicine — tailoring clinical practice to the individual rather than basing it on population responses. Although a detailed discussion of personalized medicine is beyond the scope of this article, there are some issues that are particularly relevant to synonymous mutations. Most of the research and discussion in the literature on the subject of personal pharmacogenetics focuses on variations in the enzymes and transporters that determine the disposition of small molecule drugs. We argue that sSNPs in these proteins should be given due consideration. As a case in point, synonymous mutations in the transporter ABCB1 are implicated in drug resistance to chemotherapeutic agents[7]. In addition, there is evidence emerging that sSNPs and nsSNPs in proteins that are drug targets also affect the efficacy and safety of small-molecule drugs[69].

***Protein therapeutics.*** Innovative protein therapeutics is a rapidly growing segment of the pharmaceutical market[76]. However, little attention has been paid to the consequences of introducing synonymous mutations into protein therapeutics[76]. This is of particular concern because genetic modifications to improve the expression of recombinant proteins — such as codon optimization — are now routinely performed. Although examples of codon optimization resulting in spectacular increases in protein levels have been published[77], it is not clear whether this increase is generally true for most genes. A recent multi-gene study expressed 97 codon-optimized human genes in *Escherichia coli*: 20% of these genes showed a decrease in expression following optimization, and the increase in expression in the rest was not particularly high, averaging between two- and threefold in the different classes of genes studied[78]. Moreover, many of the codon optimization strategies developed for *E. coli* may not be as effective when host cells from other organisms are used. Bacteria have a set of very strongly preferred codons, whereas mammals do not.

It should be borne in mind that some mutations that result in the alteration to a more frequent codon (that is, a natural experiment in optimization) can result in a disease condition (BOX 1). In addition to generating misfolded proteins that are less active, synonymous substitutions can induce aggregation[74] or present new epitopes on the protein, leading to immunogenicity[79]. Immunogenicity is currently a substantial impediment to the successful development of therapeutic proteins, as the development of inhibitory antibodies can significantly reduce the efficacy of these drugs[80]. The evidence presented here that synonymous mutations are not necessarily silent should provide some caution in the use of codon optimization strategies, because a maximal rate of translation does not always result in optimal protein production. However, increased understanding of the rules that govern codon usage and the underlying biophysics could lead to improvements in these strategies.

## Future directions

Recent genetic, biophysical and computational studies have revealed that the process by which genes make proteins is vastly more complex than could have been envisioned even 5 years ago. Thus, multiple emerging mechanisms have been identified that provide an indication of the means by which synonymous mutations can affect physiological changes (FIG. 1 presents a by no means comprehensive list). However, it has not been possible to experimentally demonstrate the molecular mechanism (or mechanisms) by which even one synonymous mutation causes a human disease. This is despite the existence of an extensive list of synonymous mutations linked to human disease (Supplementary information S1 (table)) and is true even in the case of two studies that have endeavoured to delineate the molecular mechanisms by which synonymous mutations in specific genes affect physiological outcomes[6,7]. Plausible cases have been made that synonymous mutations in the *COMT* gene (described above and in REF. 6) affect pain sensitivity by affecting mRNA stability and

Proteostasis
(Protein homeostasis). This process is necessary for normal cellular function and is maintained by a highly conserved network of biological pathways.

Codon optimization
The use of the host organism's codon bias when proteins are often expressed in a foreign host.

Optical tweezers
Highly focused laser beams that can be used to hold and move microscopic objects, sometimes at the level of a single molecule.

that the altered substrate specificity of the multidrug-resistance protein ABCB1 is mediated by the kinetics of translation[7]. However, in neither of these studies was it possible to obtain direct experimental evidence for the mechanisms suggested, principally because of technical limitations. Some important biophysical advances suggest that this may soon change. For example, optical tweezers have been used to follow the step-by-step translation of a single mRNA molecule by a single ribosome, and the development of such approaches has the potential to provide precise information about the kinetics of translation[81]. Similarly, a study published earlier this year exploits the use of fluorescently tagged proteins and advances in mass spectrometry to quantify protein levels in intact cells[82]. This is important because the strategy does not limit the study of gene regulation to transcription processes (such as high-throughput mRNA quantification) and so permits the study of variation that is governed by post-transcriptional events.

In addition, the rapid convergence of progress in many different areas has resulted in interesting new findings that foreshadow emerging areas of study. Host–parasite interactions engender complex physiological as well as evolutionary relationships, and studies on synonymous mutations may help to improve understanding of the co-evolution of hosts and parasites. The practical applications of these studies include the identification of novel drug targets, the determination of risk factors in susceptible populations and the development of more effective immunization strategies. This is demonstrated by findings that synonymous mutations in either the host or the pathogen can influence the susceptibility to disease: for example, synonymous mutations have been implicated in the development of drug-resistant strains of *Mycobacterium tuberculosis*[83], and synonymous mutations in the TIR-domain-containing adaptor-like protein (TIRAP) in the host can increase susceptibility to tuberculosis in some populations[84].

In this Review, we have discussed the relative abundance of tRNAs that code for the same amino acid. In addition, new evidence suggests important roles for the regulation of tRNAs mediated by isodecoders (tRNAs that share the same anticodon but have different sequences). More than 270 isodecoder genes have been identified in the human genome[85], and understanding how (and whether) their sequence diversity is reflected in tRNA function will be an important area of research. Also, tRNA post-transcriptional modifications are likely to have an important role in gene expression, as considerably more genetic information is allocated to tRNA modifications than to tRNA genes[86].

Finally, GWASs are finding that, besides the synonymous mutations in the coding regions of genes, mutations in non-coding and intergenic regions are associated with diseases[87,88]. Even mutations in so-called gene deserts have been associated with disease risk[8,89,90], and determining the functional basis of these associations opens up a challenging new area of study that will require improvements in our understanding of how nucleotide sequences convey far more information than the primary sequence of a protein. Given the level of interest and the pace of progress, it will not be long before our understanding of the mechanisms by which synonymous mutations affect human disease find applications in drug development as well as the practice of clinical medicine.


1. Anfinsen, C. B. Principles that govern folding of protein chains. *Science* **181**, 223–230 (1973).
2. Kimura, M. Preponderance of synonymous changes as evidence for neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
3. Chamary, J. V., Parmley, J. L. & Hurst, L. D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Rev. Genet.* **7**, 98–108 (2006).
**This is an excellent early Review from a molecular evolution perspective that brought the importance of synonymous mutations on human health and disease to a larger audience.**
4. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Rev. Genet.* **12**, 32–42 (2011).
**This is a recent Review that explains global patterns in codon bias and how they influence protein folding.**
5. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
6. Nackley, A. G. *et al.* Human catechol-*O*-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930–1933 (2006).
**The first study to elucidate a detailed molecular mechanism based on mRNA structure for how synonymous mutations can have physiological consequences.**
7. Kimchi-Sarfaty, C. *et al.* A "silent" polymorphism in the *MDR1* gene changes substrate specificity. *Science* **315**, 525–528 (2007).
**The first study to provide evidence that synonymous changes that do not affect mRNA levels have clinical consequences.**
8. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genet.* **37**, 549–554 (2005).
9. Kennedy, G. *et al.* Large-scale genotyping of complex DNA. *Am. J. Hum. Genet.* **71**, 204 (2002).
10. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* **1**, 109–111 (2004).
11. Steemers, F. J. *et al.* Whole-genome genotyping with the single-base extension assay. *Nature Methods* **3**, 31–33 (2006).
12. Manolio, T. A., Brooks, L. D. & Collins, F. S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118**, 1590–1605 (2008).
13. Stark, K. *et al.* Genetic association study identifies *HSPB7* as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genet.* **6**, e1001167 (2010).
14. Mannisto, P. T. & Kaakkola, S. Catechol-*O*-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol. Rev.* **51**, 593–628 (1999).
15. Winterer, G. & Goldman, D. Genetics of human prefrontal function. *Brain Res. Rev.* **43**, 134–163 (2003).
16. Ellis, P. E., Dawson, M. & Dixon, M. J. Mutation testing in Treacher Collins syndrome. *J. Orthodont.* **29**, 293–297 (2011).
17. Macaya, D. *et al.* A synonymous mutation in *TCOF1* causes Treacher Collins syndrome due to mis-splicing of a constitutive exon. *Am. J. Med. Genet. A* **149**, 1624–1627 (2009).
18. Bartoszewski, R. A. *et al.* A synonymous single nucleotide polymorphism in ΔF508 *CFTR* alters the secondary structure of the mRNA and the expression of the mutant protein. *J. Biol. Chem.* **285**, 28741–28748 (2010).
**This paper reveals a very interesting finding. A common amino acid deletion has long been thought to be responsible for cystic fibrosis; this study shows that a coincident synonymous mutation results in a change in mRNA structure and protein level and may be responsible for the polymorphism.**
19. Ramser, J. *et al.* Rare missense and synonymous variants in UBE1 are associated with X-linked infantile spinal muscular atrophy. *Am. J. Hum. Genet.* **82**, 188–193 (2008).
20. O'Driscoll, M., Ruiz-Perez, V. L., Woods, C. G., Jeggo, P. A. & Goodship, J. A. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nature Genet.* **33**, 497–501 (2003).
21. Ho, P. A. *et al. WT1* synonymous single nucleotide polymorphism rs16754 correlates with higher mRNA expression and predicts significantly improved outcome in favorable-risk pediatric acute myeloid leukemia: a report from the children's oncology group. *J. Clin. Oncol.* **29**, 704–711 (2011).
22. Hassan, H. E., Myers, A. L., Coop, A. & Eddington, N. D. Differential involvement of P-glycoprotein (ABCB1) in permeability, tissue distribution, and antinociceptive activity of methadone, buprenorphine, and diprenorphine: *in vitro* and *in vivo* evaluation. *J. Pharm. Sci.* **98**, 4928–4940 (2009).
23. Ambudkar, S. V., Kim, I. W. & Sauna, Z. E. The power of the pump: mechanisms of action of P-glycoprotein (ABCB1). *Eur. J. Pharm. Sci.* **27**, 392–400 (2006).
24. van der Veldt, A. A. M. *et al.* Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin. Cancer Res.* **17**, 620–629 (2011).
25. Herrlinger, K. R. *et al.* ABCB1 single-nucleotide polymorphisms determine tacrolimus response in patients with ulcerative colitis. *Clin. Pharmacol. Ther.* **89**, 422–428 (2011).
26. Sauna, Z. E., Kim, I. W. & Ambudkar, S. V. Genomics and the mechanism of P-glycoprotein (ABCB1). *J. Bioenerg. Biomembr.* **39**, 481–487 (2007).
27. Chen, R., Davydov, E. V., Sirota, M. & Butte, A. J. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* **5**, e13574 (2010).

28. Schattner, P. & Diekhans, M. Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res.* **34**, 1700–1710 (2006).
29. Chamary, J. V. & Hurst, L. D. The price of silent mutations. *Sci. Am.* **300**, 46–53 (2009).
30. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
31. Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* **6**, 400 (2010).
32. Czech, A., Fedyunin, I., Zhang, G. & Ignatova, Z. Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations. *Mol. Biosyst.* **6**, 1767–1772 (2010).
33. Pagani, F., Raponi, M. & Baralle, F. E. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA* **102**, 6368–6372 (2005).
34. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* **8**, 749–761 (2007).
35. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
36. Brest, P. *et al.* A synonymous variant in *IRGM* alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nature Genet.* **43**, 242–245 (2011).
**This study demonstrates the presence of an miRNA binding site in an exon, a synonymous change in which alters suceptibility to Crohn's disease.**
37. Parkes, M. *et al.* Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genet.* **39**, 830–832 (2007).
38. Duan, J. B. *et al.* Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* **12**, 205–216 (2003).
39. Kudla, G., Lipinski, L., Caffin, F., Helwak, A. & Zylicz, M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* **4**, 933–942 (2006).
40. Gu, W. J., Zhou, T. & Wilke, C. O. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.* **6**, e1000664 (2010).
41. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
**This important study, which was carried out using synthetic genes, shows the importance mRNA structure near the start codon and the consequenes of synonymous mutations in this region.**
42. Studer, S. M. & Joseph, S. Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol. Cell* **22**, 105–115 (2006).
43. Dong, H. J., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).
44. Duret, L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* **16**, 287–289 (2000).
45. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Struct. Mol. Biol.* **16**, 274–280 (2009).
46. Bonekamp, F., Dalboge, H., Christensen, T. & Jensen, K. F. Translation rates of individual codons are not correlated with transfer-RNA abundances or with frequencies of utilization in *Escherichia coli*. *J. Bacteriol.* **171**, 5812–5816 (1989).
47. Higgs, P. G. & Ran, W. Q. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Mol. Biol. Evol.* **25**, 2279–2291 (2008).
48. Ikemura, T. Codon usage and transfer-RNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).
49. Komar, A. A. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **34**, 16–24 (2009).
50. Purvis, I. J. *et al.* The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo* — a hypothesis. *J. Mol. Biol.* **193**, 413–417 (1987).
**This study presents the first challenge to the concept that only the sequence of amino acids determines the final conformation of a protein.**
51. Komar, A. A., Lesnik, T. & Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett.* **462**, 387–391 (1999).
52. Fedorov, A. N. & Baldwin, T. O. Contribution of cotranslational folding to the rate of formation of native protein-structure. *Proc. Natl Acad. Sci. USA* **92**, 1227–1231 (1995).
53. Clarke, D. T., Doig, A. J., Stapley, B. J. & Jones, G. R. The α-helix folds on the millisecond time scale. *Proc. Natl Acad. Sci. USA* **96**, 7232–7237 (1999).
54. Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
55. Ehrenberg, M., Dennis, P. P. & Bremer, H. Maximum *rrn* promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase. *Biochimie* **92**, 12–20 (2010).
56. Powers, E. T. & Balch, W. E. Costly mistakes: translational infidelity and protein homeostasis. *Cell* **134**, 204–206 (2008).
57. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
58. Cannarrozzi, G. *et al.* A role for codon order in translation dynamics. *Cell* **141**, 355–367 (2010).
59. Deutscher, M. P. The eukaryotic aminoacyl-transfer RNA-synthetase complex — suggestions for its structure and function. *J. Cell Biol.* **99**, 373–377 (1984).
60. Kaminska, M. *et al.* Dynamic organization of aminoacyl-tRNA synthetase complexes in the cytoplasm of human cells. *J. Biol. Chem.* **284**, 13746–13754 (2009).
61. Mirande, M., Lecorre, D. & Waller, J. P. A complex from cultured Chinese-hamster ovary cells containing 9 aminoacyl-transfer RNA-synthetases — thermolabile leucyl-transfer RNA-synthetase from the Tsh1 mutant-cell line is an integral component of this complex. *Eur. J. Biochem.* **147**, 281–289 (1985).
62. Sosnick, T. R. Kinetic barriers and the role of topology in protein and RNA folding. *Protein Sci.* **17**, 1308–1318 (2008).
63. Jha, S. & Komar, A. A. Birth, life and death of nascent polypeptide chains. *Biotechnol. J.* **6**, 623–640 (2011).
64. Fedyukina, D. V. & Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **40**, 337–359 (2011).
65. Netzer, W. J. & Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343–349 (1997).
66. Nicola, A. V., Chen, W. & Helenius, A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nature Cell Biol.* **1**, 341–345 (1999).
67. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F. U. Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly luciferase. *Nature Struct. Biol.* **6**, 697–705 (1999).
68. Tsai, C. J. *et al.* Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.* **383**, 281–291 (2008).
69. Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V. & Gottesman, M. M. Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer. *Cancer Res.* **67**, 9609–9612 (2007).
70. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
71. Gidalevitz, T., Ben-Zvi, A., Ho, K. H., Brignull, H. R. & Morimoto, R. I. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311**, 1471–1474 (2006).
72. Pal, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nature Rev. Genet.* **7**, 337–348 (2006).
73. Wilke, C. O. & Drummond, D. A. Signatures of protein biophysics in coding sequence evolution. *Curr. Opin. Struct. Biol.* **20**, 385–389 (2010).
74. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
**This is an interesting global computational study conducted in a range of taxa using evolutionary simulation. It shows that selective pressures are exerted as a consequence of protein misfolding resulting from synonymous changes.**
75. Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age-onset proteotoxicity. *Science* **313**, 1604–1610 (2006).
76. Krejsa, C., Rogge, M. & Sadee, W. Protein therapeutics: new applications for pharmacogenetics. *Nature Rev. Drug Discov.* **5**, 507–521 (2006).
77. Ward, N. J. *et al.* Codon optimization of human factor VIII cDNAs leads to high-level expression. *Blood* **117**, 798–807 (2011).
78. Maertens, B. *et al.* Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. *Protein Sci.* **19**, 1312–1326 (2010).
79. De Groot, A. S. & Scott, D. W. Immunogenicity of protein therapeutics. *Trends Immunol.* **28**, 482–490 (2007).
80. Wang, J. H. *et al.* Neutralizing antibodies to therapeutic enzymes: considerations for testing, prevention and treatment. *Nature Biotech.* **26**, 901–908 (2008).
81. Wen, J. D. *et al.* Following translation by single ribosomes one codon at a time. *Nature* **452**, 598–603 (2008).
82. Eden, E. *et al.* Proteome half-life dynamics in living human cells. *Science* **331**, 764–768 (2011).
83. Plinke, C. *et al.* embCAB sequence variation among ethambutol-resistant *Mycobacterium tuberculosis* isolates without embB306 mutation. *J. Antimicrob. Chemother.* **65**, 1359–1367 (2010).
84. Dissanayeke, S. R. *et al.* Polymorphic variation in TIRAP is not associated with susceptibility to childhood TB but may determine susceptibility to TBM in some ethnic groups. *PLoS ONE* **4**, e6698 (2009).
85. Geslain, R. & Pan, T. Functional analysis of human tRNA isodecoders. *J. Mol. Biol.* **396**, 821–831 (2010).
86. Gustilo, E. M., Franck, A. P. F. & Agris, P. F. tRNA's modifications bring order to gene expression. *Curr. Opin. Microbiol.* **11**, 134–140 (2008).
87. Glinskii, A. B. *et al.* Identification of intergenic *trans*-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. *Cell Cycle* **8**, 3925–3942 (2009).
88. Treutlein, J. *et al.* Genome-wide association study of alcohol dependence. *Arch. Gen. Psychiat.* **66**, 773–784 (2009).
89. Haiman, C. A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature Genet.* **39**, 638–644 (2007).
90. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39**, 645–649 (2007).
91. Sharp, P. M., Tuohy, T. M. & Mosurski, K. R. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **14**, 5125–5143 (1986).
92. Bonekamp, F. & Jensen, K. F. The AGG codon is translated slowly in *E. coli* even at very low expression levels. *Nucleic Acids Res.* **16**, 3013–3024 (1988).
93. Folley, L. S. & Yarus, M. Codon contexts from weakly expressed genes reduce expression *in vivo*. *J. Mol. Biol.* **209**, 359–378 (1989).



### Acknowledgements
We thank B. Golding for helpful discussions and G. S. Pandey for comments and help in the manuscript preparation. The assistance of E. Ayalp and I. Kimchi in generating the table in box 1 and J. Sauna for the illustrations in figure 1 is gratefully acknowledged. This work was supported by funds from the Laboratory of Hemostasis (C.K.-S.) and the Center for Biologics Evaluation and Research, US Food and Drug Administration's Modernization of Science programme (Z.E.S.). The findings and conclusions in this article have not been formally disseminated by the US Food and Drug Administration and should not be construed to represent any Agency determination or policy.


### Competing interests statement
The authors declare no competing financial interests.

## FURTHER INFORMATION
**Zuben E. Sauna's homepage:** http://www.fda.gov/BiologicsBloodVaccines/ScienceResearch/BiologicsResearchAreas/ucm246804.htm
**Chava Kimchi-Sarfaty's homepage:** http://www.fda.gov/BiologicsBloodVaccines/ScienceResearch/BiologicsResearchAreas/ucm127059.htm
**Codon usage table:** http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606

## SUPPLEMENTARY INFORMATION
**See online article:** S1 (table) | S2 (table)

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**