Review

CellPress

Exposing synonymous mutations

Ryan C. Hunt*, Vijaya L. Simhadri*, Matthew Iandoli, Zuben E. Sauna, and Chava Kimchi-Sarfaty

Division of Hematology, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA

Synonymous codon changes, which do not alter protein sequence, were previously thought to have no functional consequence. Although this concept has been overturned in recent years, there is no unique mechanism by which these changes exert biological effects. A large repertoire of both experimental and bioinformatic methods has been developed to understand the effects of synonymous variants. Results from this body of work have provided global insights into how biological systems exploit the degeneracy of the genetic code to control gene expression, protein folding efficiency, and the coordinated expression of functionally related gene families. Although it is now clear that synonymous variants are important in a variety of contexts, from human disease to the safety and efficacy of therapeutic proteins, there is no clear consensus on the approaches to identify and validate these changes. Here, we review the diverse methods to understand the effects of synonymous mutations.

An expanding biological footprint of synonymous nucleotide variants

Synonymous nucleotide substitutions in coding regions were historically thought to be of little significance, but they are now the subject of increasing interest to geneticists and pharmacologists [1–4]. Over 50 human diseases have been associated with synonymous mutations [5] and, in a recent survey of 21 429 polymorphisms associated with human disease, nonsynonymous and synonymous variations were determined to have a similar probability of disease association (1.46% versus 1.26%, respectively) in addition to a statistically equivalent effect size, suggesting that the list of disease-causing synonymous mutations will grow [6]. Nonetheless, molecular evolution unmistakably illustrates that most genes tolerate nonsynonymous mutations at a lower rate than their synonymous counterparts [7]. Nature itself may be the best 'experimental' system for engendering fundamental biological principles through comparative genomics, but individual observations continue to highlight the nontrivial nature of this class of genetic variants, including the finding that synonymous and nonsynonymous substitutions introduced within *Salmonella* ribosomal proteins results in remarkably similar distributions of fitness [8]. Hundreds of regions of extreme codon conservation can be found across mammalian genes, areas that are relatively depleted of synonymous but not nonsynonymous substitutions [9]. Thus, purifying selection for synonymous codons can confound classic measurements in evolutionary biology that assume an inflated ratio of nonsynonymous substitutions relative to synonymous changes to be evidence of positive selection [10]. These observations are broadly consistent with the concept that synonymous codons affect the expression and function of the translated protein and, therefore, are under selective pressure [11]. In recent years, there have been significant advances in our understanding of the biochemical, biophysical, and genetic mechanisms by which codon bias is exploited by the translation machinery of cells to control gene expression, the efficiency or speed of protein translation, and the accuracy of protein folding [11–16]. The methodologies that have arisen from these studies are the subject of this review (Figure 1 presents an overview of mechanisms and methods). Given the scope of mechanisms and methodologies applied, we have largely focused on understanding synonymous variants within coding regions of human genes.

Codon usage bias: Nature's cue to the consequence of synonymous codon substitutions

One can study synonymous variants individually or en masse, each approach benefiting from a distinct set of methodologies. Synonymous codons appear globally at nonrandom frequencies, a phenomenon termed 'codon usage bias' (CUB), a subject of intensive investigation. Codon usage can differ across species, within a genome, or even along a single gene. In addition to neutral factors, such as mutational biases and genomic GC content, codon bias has arisen through the optimization of fundamental cellular processes, including the speed and fidelity of translation. The relative contribution of neutral and selective explanations for CUB remains a topic of debate. Evidence for selection is strong in prokaryotes, where isoaccepting tRNA abundance and codon usage frequencies have coevolved for optimal fitness [17]. Although selective forces may account for a smaller share of observed CUB in mammalian genes, a critical mass of evidence has accumulated in recent years arguing for various forms of selective pressure contributing to codon bias in higher eukaryotes. A somewhat unexpected finding is higher constraint at fourfold degenerate sites within hominid genomes relative to murids, despite their discrepant population sizes [18]. Moreover, when careful consideration is

Corresponding authors: Hunt, R.C. (ryan.hunt@fda.hhs.gov); Sauna, Z.E. (zuben.sauna@fda.hhs.gov); Kimchi-Sarfaty, C. (chava.kimchi-sarfaty@fda.hhs.gov).
Keywords: synonymous mutation; synonymous single nucleotide polymorphism; codon usage; mRNA structure; translation speed.
*These authors contributed equally to this work.

0168-9525/
Published by Elsevier Ltd. http://dx.doi.org/10.1016/j.tig.2014.04.006

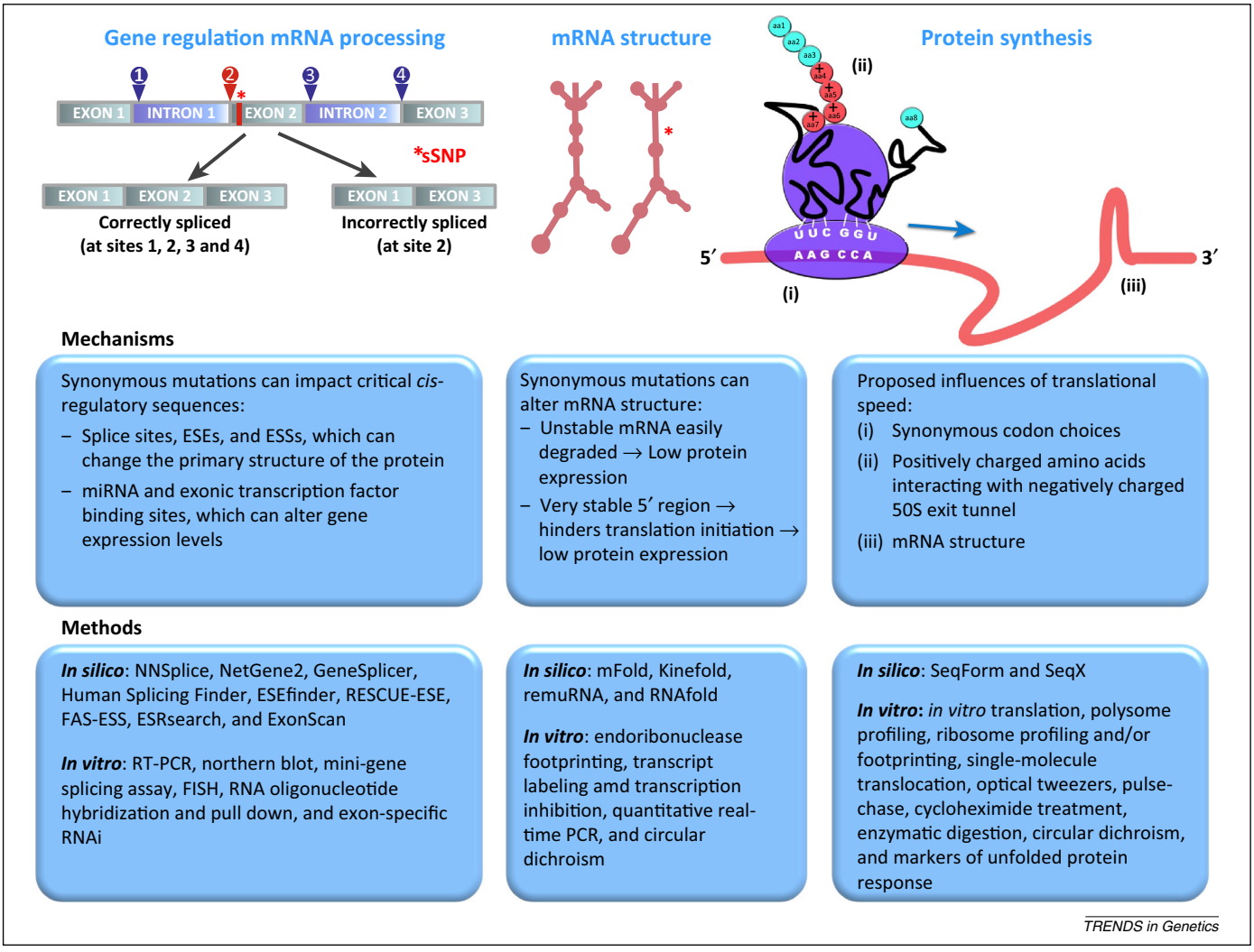

Figure 1. Schematic depiction of the general means by which synonymous codon substitutions may exert a biological effect, and select methods and approaches to delineate their underlying molecular and cellular mechanisms.

given to the nature of human genes analyzed (i.e., using data sets of genes expressed at consistent levels across tissues and physiologic states), selection strengthens for gene characteristics, including CUB, that achieve expression efficiency [19]. A recent report implicated a surprising percentage (approximately 15%) of all human codons in the binding of transcription factors (TF). These conserved exonic TF binding sites impose a selective constraint that partly drives observed codon preferences within mammalian genomes [20]. Of relevance to this discussion, >17% of single nucleotide polymorphisms (SNPs), independent of class of variant (nonsynonymous and synonymous), falling within these regions were determined to alter TF binding. Binding sites for TFs, exonic splice enhances, and other *cis*-regulatory elements can generate local CUB that is distinct from that imposed by translational selection [21], making the notion of universal 'codon optimality' a spurious term. Indeed, codons near intron–exon boundaries in many cases are not thought to be translationally optimal, an evolutionary trade-off to preserve important consensus regulatory sequences [22].

CUB, particularly the existence of selective pressures that help generate codon preferences, underscores the nontrivial nature of synonymous codon choices. There exists a multitude of approaches to quantify CUB. These measurements can be carried out at various levels, from within individual genes or functional domains to whole-genome assessments. Some approaches control for the background genomic nucleotide composition [23,24]. Others seek to understand translational-based impact of codon usage isolated from other influences; such codon bias indices, including the Codon Adaptation Index and Relative Codon Adaptation Index [25], accomplish this latter task. Assessing codon bias may reveal the existence of skewed employment of synonymous codons across a gene or within a given protein domain; these genes or specific loci may warrant particular interest during the investigation of synonymous nucleotide variants.

Identifying candidate synonymous changes for investigation

CUB suggests that synonymous mutations, as a population, have biological consequence, but how are those with particularly high impact identified? A straightforward approach is genotyping of candidate gene(s). Often, however, genetic contributions to disease or phenotypic traits are unknown, a challenge that is being addressed with the advent of increasingly economical high-throughput

genomic technologies. When used within the context of human disease, these technologies can help identify sites of putative pathologic genetic aberrations, including synonymous changes. Genome-wide association studies (GWAS) have gained substantial popularity as a means to study the association of both nonsynonymous and synonymous polymorphic sites with disease or phenotypic categories [17]. Although GWAS can identify novel genetic determinants to disease, the SNPs used are generally thought to be in linkage disequilibrium with the causative mutation as opposed to being, in themselves, pathogenic. To access this information more directly, whole-exome or genome-wide sequencing can be used to search for rare, causal genetic variants. GWAS and next-generation sequencing were critical in revealing synonymous changes impacting miRNA gene regulation (Box 1).

Despite inherent limitations and the risk for false positive and negative results, these approaches have successfully revealed some genetic determinants of disease, with whole-exomic approaches seeing growing clinical utility [19]. Although next-generation sequencing has yielded a wealth of genetic data for researchers, it has also engendered a new challenge: prioritizing disease-causing mutations from millions of identified neutral variants. Potentially pertinent synonymous changes are often filtered out during the initial analysis of large data sets [26]. The tools for scrutinizing nonsynonymous variants can leverage prototypical mutational features concerning the particular amino acid involved, whereas assessing synonymous changes is a less direct task. New tools, such as silent variant analyzer (SilVA) [27], aim to predict the functional impact of synonymous mutations in large-scale genotyping projects. Given a Variant Call Format text file, SilVA combines 26 mutational attributes of each variant, including several genomic features (e.g., splice site enhancers and/or suppressors, CpG and relative position within the transcript, CUB, conservation, and mRNA folding) and computes a 'harmfulness score'. A gene-specific predictor for synonymous and nonsynonymous mutations in coagulation Factor VIII illustrates a general strategy for investigators to score synonymous mutations [28]. Similar to SilVA, by integrating characteristics at the level of DNA and/or RNA alongside more conventional variables, this tool is uniquely equipped to assess this class of genetic variant. Interrogating local codon usage among homologous genes may also reveal a synonymous variant within a region of heightened codon conservation. These sites often mark critical *cis*-regulatory elements, many of which are involved in mRNA processing [9] and, thus, may merit functional investigation.

Ultimately, synonymous mutations can impact host of cellular processes, from gene regulation, mRNA structure and processing (Table 1) to protein synthesis, conformation, and functionality (Table 2); experimental approaches to investigate these possibilities are discussed below.

Methods to investigate the effects of synonymous mutations on the fidelity of splicing

The most obvious and well-characterized manner by which synonymous changes result in clear biological impact is through perturbations to pre-mRNA splicing. Auxiliary exonic splicing elements, namely splicing enhancers (ESEs) and silencers (ESSs), are necessary to fully define exon–intron junctions in higher eukaryotes. Exonic enhancers in particular are thought to be recognized by serine/arginine-rich (SR) proteins that recruit splicing machinery. The high degree of sequence conservation at these various sites has facilitated the development of *in silico* tools to analyze putative sites and regulators of splicing (Table 3). Synonymous nucleotide substitutions can perturb an existing splice site, or less commonly introduce a novel splice site. Either outcome generates aberrant mRNAs that are unstable or encode flawed protein products. Synonymous variants falling within ESEs or ESSs also have the potential to alter the relative abundance of certain mRNA isoforms, which is sufficient to cause several human diseases, including frontotemporal dementia and familial growth hormone deficiency [29,30]. At least 4% of synonymous mutations have an evolutionarily conserved deleterious effect on ESEs [31] and, accordingly, synonymous sites in ESEs evolve more slowly than in other exonic regions [32].

There are several experimental approaches to delineate how synonymous variants modulate spicing. Fresh tissue or cells from affected individuals can be analyzed simply by reverse transcription PCR (RT-PCR). This approach using lymphocytes from a patient with Treacher Collins syndrome implicated a synonymous mutation within *TCOF1* to mediate exon skipping and a pathogenic frameshift within the following exon [33]. However, ready access to clinical samples is often limited, which has spurred the development of other experimental approaches to understand the influence of sequence variation on splicing. One such approach is exemplified by a study of medium-chain acyl-CoA dehydrogenase (MCAD) deficiency [34]. Here, it was shown that a missense mutation within exon 5 of *MCAD* inactivates a critical ESE, resulting in exon skipping and nonsense mediated decay. However, in the presence of a naturally occurring flanking synonymous polymorphic variant, a downstream ESS is inactivated and normal splicing is restored. To define these elements, the target genomic segment in *MCAD* was cloned to create a 'minigene' and transfected into human cell lines. Beyond this, the authors also used a RNA interference and pull-down assay to demonstrate a decrease in heterogeneous nuclear ribonucleoprotein (hnRNP) A1-binding at the ESS in the presence of the synonymous change.

Measuring the impact of synonymous mutations on mRNA structure and stability

Synonymous variations can alter mRNA secondary structure, which in turn can affect the efficiency of translation and rate of mRNA degradation. A general case for the importance of mRNA secondary structure and the role of synonymous variants was made during a study of over 150 individual random synonymous changes introduced into GFP, resulting in 250-fold variation in expression across the library [35]. Over half of the variation in expression levels could be attributed to mRNA folding energy near the start codon, which is in line with the idea that initiation of highly expressed genes is facilitated by less stable secondary structures near the 5′ terminus, whereas weakly

Box 1. An emerging arena of biology susceptible to the influence of synonymous variants: miRNA regulation

miRNAs represent an integral form of metazoan post-transcriptional regulation and are estimated to target 60% of mammalian genes [82]. These generally noncoding RNA molecules base pair with complementary mRNA sequences and typically function to initiate degradation or suppress translation of the target sequence [83,84]. Although highly conserved bases within 5′ 'seed regions' have been shown to be especially important [85], the determinants of miRNA–mRNA interactions are likely more complex than base-pair matching alone. Better understanding of miRNA targeting has spurred the development of various *in silico* tools to help identify consensus domains, motifs, and binding regions within mRNAs (Table I). These algorithms weigh various features, including base pairing, thermodynamics, sequence composition and conservation, number of targets sites per transcript, and miRNA expression level (reviewed in [86–88]). One can ask whether a given synonymous variant lies within a putative miRNA binding site and assess the impact of nucleotide variation therein. However, predicting miRNA targets remains a challenging task. There is no one preeminent target prediction tool, and there is a general tendency towards false positive results relative to experimental observations [89,90]. Integrating independent miRNA target predictions together with miRNA expression into a single score (e.g., ComiR) is a promising strategy, yielding improved target prediction and assessment of SNPs [91]. Further refinement of these tools will benefit from high-throughput mapping and global characterization of miRNA binding sites [92,93]. Any target prediction should be experimentally validated with a reporter construct, mRNA and protein quantitation, ribonuclease protection, inhibitory miRNA-complimentary oligonucleotides, or other methods (described below).

Two recent reports highlight the ability of synonymous variants to cause disease through aberrant miRNA regulation. In the study of Crohn's disease, an exonic synonymous SNP (c.313T>C) within the autophagy regulatory gene immunity-related GTPase family M protein (*IRGM*) was found in near-perfect linkage disequilibrium with a susceptibility locus previously identified through GWAS [94]. Using *in silico* methods (SnipMir, RegRNA, and Patrocles), the sequence immediately surrounding c.313T>C was identified as a potential binding site for miR-196. This prediction was verified with tandem affinity purification of biotinylated miR-196 in cells stably expressing Flag-tagged Argonaute [95], a member of the RNA-induced silencing complex. Other complementary approaches were used in this study, including quantification of *IRGM* mRNA in biopsy samples by qRT-PCR and *in situ* hybridization of miR-196. The lack of downregulation of the risk-associated allele, c.313T, resulted in IRGM overexpression and abnormal intracellular processing of Crohn's disease-associated pathogens [96]. To date, one additional study was successful in linking a synonymous mutation with miRNA dysregulation. Through whole-exome sequencing, a subset of melanoma cases harboring a recurrent F17F mutation in BCL2L12 was identified. After exhaustive investigation, it was determined that this synonymous change in *BCL2L12* perturbed hsa-miR-671-5p binding and resulted in heightened BCL2L12 expression, triggering anti-apoptotic signaling [97].

Table I. Examples of software for the analysis of miRNA binding sites

Software	URL	Description	Refs
miRanda	http://www.microrna.org/microrna/home.do	Identifies miRNA targets based on sequence complementarity with local position weighting, thermodynamics of duplexes, and target site conservation	[98]
Diana-microT	http://diana.cslab.ece.ntua.gr/microT/	Searches for targets of a given miRNA or miRNA target sites on a gene. A signal:noise ratio and a precision score are calculated by comparing against outcomes obtained with scrambled sequences carrying the same dinucleotide content as the actual 3′ UTR, which provides insight into the significance of each result	[99]
PITA	http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html	Assesses the potential for miRNA binding by computing the difference in free energy gained from the formation of the miRNA–target duplex against the energetic cost of unpairing the target sequence to make it accessible to the miRNA	[100]
Patrocles	http://www.patrocles.org/	Database of sequence polymorphisms that are predicted to perturb miRNA-mediated gene regulation; Patrocles finder determines whether a synonymous variant might perturb miRNA-mediated gene targeting	[101]
TargetScan and TargetScanS	http://www.targetscan.org/	Predicts targets of miRNAs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA	[85,102]
SnipMir	http://www.microarray.fr:8080/merge/index?action=MISNP	Allows the user to input a sequence in SNP fasta format (i.e., with nucleotide variation) and determine the implications of sequence variation for an individually selected miRNA	–
RegRNA	http://regrna.mbc.nctu.edu.tw/	Assesses a given RNA sequence for over ten distinct functional RNA motifs, including miRNA target sites	[103]
TargetRank	http://hollywood.mit.edu/targetrank/	Predicts gene targets of input miRNA and generates TargetRank scores for each conserved and nonconserved 8mer, 7mer and 6mer matched within the 3′ UTR	[104]
miRror	http://www.proto.cs.huji.ac.il/mirror/	Integrates predictions from 12 miRNA resources that are based on complementary algorithms into a unified statistical framework. Predictions are based on the notion of a combinatorial regulation by groups of miRNAs or genes	[105]
PicTar	http://pictar.mdc-berlin.de/	Predicts targets that are shared among multiple miRNAs across species, filters alignments based on their thermodynamics, and finally computes a hidden Markov model maximum likelihood score for each predicted target	[106]
ComiR	http://www.benoslab.pitt.edu/comir/	Uses user-provided miRNA expression levels and calculates the probability of a given mRNA of being targeted by a set of miRNAs. Integrates four popular scoring schemes (miRanda, PITA, TargetScan, and mirSVR) using a support vector machine trained on *Drosophila* Ago1 IP data	[91,107]

Table 1. Examples of synonymous mutations at the level of mRNA and methods used for their study

Gene symbol	Method(s)	Disease and/or trait	Refs
COMT	*In silico* analysis by RNA mFold, real-time PCR to assess RNA expression, and mRNA degradation analysis	Chronic musculoskeletal pain disorder	[39]
MCAD	qPCR of cells transfected with the minigene, Western blot analysis of pulled-down proteins that are coupled with oligo-RNA to detect nuclear splicing factors, RNA binding assay, RNA affinity chromatography, and gel analysis	MCAD deficiency	[34]
MDR1	qPCR of infected and/or transfected cells to detect mRNA levels	Multidrug resistance of cells	[2]
CHRNE	RT-PCR from leukocytes to detect mRNA fragment length	Congenital myasthenic syndrome	[108]
NSDHL	RT-PCR to detect mRNA fragment length in patient samples	CHILD syndrome	[109]
F9	RT-PCR to detect mRNA fragment length in patient samples	Hemophilia B	[110]
UBE1	*In silico* detection of splicing by ESEfinder; qPCR of lymphoblastoid and fibroblastoid cell lines of patients to detect mRNA levels; MALDI-TOF mass spectrometry analysis	X-Linked infantile spinal muscular atrophy	[111]
LDLR	RT-PCR and qPCR of patients' samples and RT-PCR of minigene transfected into cells to detect mRNA fragment length; nonsense-mediated mRNA decade using cycloheximide	High cholesterol	[112,113]
ALDH1A2	*In silico* detection of splicing by ESEfinder, mRNA in silico prediction by Geneboo program, RT-PCR to evaluate the mRNA levels, and splicing assay using pSPL3 vector	Congenital heart disease	[114]
SCARB1	*In silico* prediction by mFold, mRNA levels by RT-PCR, primer extension, and native gel electrophoresis to probe mRNA structure	Altered lipid levels and cardiovascular risk	[70]
CFHR5	*In silico* prediction by Berkeley Drosophila Genome Project using human DNA sequence	Age-related macular degeneration	[115]
CD44	RT-PCR of minigene transfected into cells with exon-trapping vector followed by sequencing and in silico analysis of splicing sites using ESEfinder	Osteoporosis	[116]
TCOF1	Analysis of size of RT-PCR products using patients' samples and *in silico* analysis of splicing sites using ESEfinder	Treacher Collins syndrome	[33]
PAH	*In silico* analysis of splicing by ESEfinder, *in vitro* splicing assessment by pSXN splicing reporter minigene, and Western blot analysis of pulled-down proteins coupled with oligo-RNA to detect nuclear splicing factors	PKU	[117]
BEST1	*Ex vivo* α-globin-fibronectin-EDB splice assay and *in silico* analysis of splicing sites using RESCUEese	Retinal dystrophy	[118]
THAP1	*In silico* methods to evaluate splicing events by SSF, MaxEnt, NNSPLICE, and Gene Splicer	DYT6 dystonia	[119]
CFTR	*In silico* analysis by mFold, *in vitro* assays to detect mRNA folding and stability	Cystic fibrosis	[40,69]
WT1	RT-PCR of patients' samples to detect mRNA expression level	Pediatric acute myeloid leukemia	[120]
Core	*In silico* analysis by mFold	Susceptibility to Hepatitis C virus infection	[121]
IRGM	In silico assessment of miRNAs binding to different forms of IRGM mRNA using SnipMir, RegRNA, and Patrocles; in vitro miRNA binding to mRNA	Crohn's disease	[96]
ADAMTS13	*In silico* analysis by mFold and Kinefold; *in silico* assessment of codon usage changes (RSCU); and qPCR to detect mRNA expression level in transfected cells with the minigene	Thrombotic thrombocytopenic purpura	[122]
HTRA1	qPCR to detect mRNA expression level in leukocytes	Neovascular age-related macular degeneration	[123]
ATP7A	*In silico* methods to evaluate splicing events by Human Splicing Finder; RT-PCR of patient's sample to detect the mRNA fragment length and mRNA levels	Menkes disease	[124]
MECP2	RT-PCR of patients' samples and *in silico* analysis of splicing sites	Rett syndrome	[125]
ACADM	RT-PCR of cells transfected with minigene to detect mRNA fragment length and for sequencing and *in silico* analysis of splicing sites; Western blot analysis of pulled-down proteins coupled with oligo-RNA to detect nuclear splicing factors	Fatty acid oxidation	[126]
TTP	qPCR, mRNA half-life with transcriptional inhibition by DRB and *in silico* analysis by RNAfold	Breast cancer	[62]

expressed genes tend to have synonymous variations favoring strong or sterically interfering secondary structural elements [36]. A multitude of computational methods exists to predict and evaluate mRNA secondary structure (Box 2), and these have been widely used to predict the structural implications of single nucleotide changes. One should bear in mind that many of these programs generate structures based on potentially faulty assumptions (e.g., one optimal structure in equilibrium) and consider a limited set of parameters (e.g., solely reliant on free energy minimization and nearest neighbor modeling). This makes picking the 'correct' structure difficult, because dissimilar structures can sometimes be energetically indistinguishable. In addition, there is a growing awareness that RNA molecules adopt dynamic complex secondary and tertiary structures, meaning that structure prediction is a moving target [37,38]. Nevertheless, RNA structure prediction continues to advance with the development of algorithms that account for pseudoknots, generate statistical distributions of possible structures, and carry out comparative alignments.

An early achievement that combined computational tools with experimental validation was the functional investigation of polymorphic forms of the enzyme

Table 2. Examples of methods for studying synonymous mutations at the protein level

Gene symbol	Method(s)	Disease and/or trait	Refs
COMT	Expression: Western blot analysis of transfected cells with the minigene	Chronic musculoskeletal pain disorder	[39]
	Function: COMT enzymatic activity		
MDR1	Expression: Western blot and flow cytometry analysis of transfected cells with the minigene	Multi drug resistance	[2,127,128]
	Conformation: Western blot analysis of limited trypsin digested transfected samples and flow cytometry using conformation-sensitive antibodies; immunohistochemistry of intestinal epithelium of inflamed colonic tissue to show differential effectiveness of tacrolimus treatment		
	Function: confocal microscopy and transport using flow cytometry of infected and/or transfected cells; paclitaxel sensitivity in cells using MTT assay		
CHRNE	Expression: AChR expression at the neuromuscular junction by confocal microscopy of muscle biopsy	Congenital myasthenic syndrome	[108]
	Function: clinical neuromuscular transmission by electromyography		
F9	Expression and function: antigen and clotting function of minigene analyzed in transfected cells	Hemophilia B	[110]
UBE1	Expression: Western blot of fibroblastoid cell lines of patients	X-linked infantile spinal muscular atrophy	[111]
SCARB1	Expression: *in vitro* translation and Western blot analysis	Altered lipid levels and cardiovascular risk	[70]
LDLR	Expression: Western blot of transfected cells with the minigene	High cholesterol	[113]
CFTR	Expression: Western blot and *in vitro* translation	Cystic fibrosis	[40,69]
	Folding: Western blot of ALLN-treated samples		
	Protein stability: pulse-chase and cyclohexamide treatment		
	Function: single-channel activity		
Core	Expression: ELISA of HCV-infected patients and Western blot of transfected cells with minigene	Susceptibility to Hepatitis C virus infection	[121]
IRGM	Expression: Western blot analysis of transfected cells with miRNA treated with bafilomycin, *in situ* hybridization, confocal microscopy	Crohn's disease	[96]
	Function: examination of intracellular replication and localization in overexpression of IRGM in bacteria		
ADAMTS13	Expression: Western Blot analysis of intra- and extracellular cells transfected with minigene	Thrombotic thrombocytopenic purpura	[122]
	Folding: Western blot of limited trypsin digested samples		
	Function: protease activity using FRETS-VWF73; *in silico* assessment of amino acid conservation using ConSeq		
BCL2L12	Expression: Western blot	Melanoma	[97]
	Function: Western blot to detect p53 binding, anti-apoptotic activity and assessment of endogenous p53 target gene transcription		
ATP7A	Expression: Western blot	Menkes disease	[124]
HTRA1	Expression: Western Blot analysis of patients' leukocytes and transfected cells, *in vitro* translation Folding: limited trypsin digested samples and conformation-sensitive antibody	Neovascular age-related macular degeneration	[123]
	Function: analysis of phosphorylation		
TTP	Expression: Western blot, polysome-associated RNA analysis, *in vitro* transcription and/or translation	Breast cancer	[62]
	Function: assessment of TTP target half life		

Catechol-*O*-methyltransferase (COMT) and their association with differential pain sensitivity in humans [39]. This work relied heavily on modeling of local mRNA structure that was validated with targeted mutagenesis. This strategy led to the identification of a synonymous SNP as a major source of instability within a native stem-loop structure. The resultant shorter, less stable local stem-loop structure allows for more robust translational efficiency and heightened COMT expression, which decreases pain sensitivity. Additional approaches to infer mRNA structure have been applied to study a synonymous nucleotide change adjacent to the ΔF508 site in cystic fibrosis transmembrane conductance (*CFTR*), a widely disregarded component of the ΔF508 aberration. Using circular dichroism spectroscopy (CD) and an mRNA folding assay based on endoribonuclease digestion/RT-PCR, two large loops in the ΔF508 CFTR mRNA were unveiled, which were conjectured to influence the dynamics of cotranslational folding [40]. Without the synonymous single nucleotide variant, the ΔF508 transcript has wild type-like mRNA structure.

The rates of transcription, processing, and decay collectively determine mRNA abundance within cells. In a recent genome-wide survey, it was shown that mRNA stability may underlie upwards of 40% of disparate gene expression between individuals. Importantly, polymorphic alleles containing guanine and cytosine at the third position in a codon were associated with increased mRNA stability [41]. Therefore, the investigation of mRNA stability is paramount to understanding the impact of synonymous changes. Northern blot analysis and

Table 3. Examples of software for predicting splice sites and splicing regulatory elements

Software	URL	Description	Refs
Analyzer Splice Tool (AST)	http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm	Scores the strength of potential 5′ and 3′ splice sites; is able to quantify the number of H-bonds between U1 and the 5′ splice site and calculate the ΔG of U1/5′ splice site pairing	[129,130]
NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Generates a neural network prediction of 5′ and 3′ splice sites, originally designed to facilitate gene discovery by the gene-finding program Genie	[131]
GeneSplicer	http://ccb.jhu.edu/software/genesplicer/	Detects splice sites in genomic DNA of various eukaryotes	[132]
Human Splicing Finder	http://www.umd.be/HSF/	Predicts effects of mutations on splicing signals and identifies splicing motifs in any human sequence	[133]
Alternative Splice Site Predictor (ASSP)	http://wangcomputing.com/assp/index.html	Predicts alternative isoforms, cryptic splice sites, and constitutive splice sites of coding exons	[134]
NetGene2	http://www.cbs.dtu.dk/services/NetGene2/	Generates neural network predictions of splice sites in humans, *Caenorhabditis elegans*, and *Arabidopsis thaliana* by combining local and global sequence information	[135]
MaxEntScan (MES)	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html	score3ss scores 23 mers using different 3′ splice site models	[136]
		score5ss scores 9 mers using different 5′ splice site models	
SFmap	http://sfmap.technion.ac.il/	Searches within a given sequence for significant binding motifs that are either stored in the database of the web server or defined by the user	[137]
SplicePort	http://spliceport.cbcb.umd.edu/SplicingAnalyser.html	Predicts splice sites for submitted sequences, allows user to see which subsets of features were used in the analysis	[138]
FAS-ESS	http://genes.mit.edu/fas-ess/	Identifies potential exonic spicing silencers and assesses whether a coding SNP abolishes an ESS motif	[139]
ExonScan	http://genes.mit.edu/exonscan/	Scores putative splice cites using an entropy model, combines RESCUE-ESE and FAS (to predict ESEs and ESSs, respectively) with intronic GGGs to predict the likelihood of uncharacterized sequences serving as exons	[136,139,140]
ESRsearch	http://esrsearch.tau.ac.il/	Searches for putative ESRs in a sequence, including both exonic splicing enhancers and silencers	[130,139,140]
RESCUE-ESE	http://genes.mit.edu/burgelab/rescue-ese/	Identifies hexanucleotide candidate ESEs in vertebrate exons. The effect of sequence variation on ESE content can be analyzed by entering both versions of the sequence	[140]
ESEfinder	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home	Searches for ESE motifs that correspond to four key SR proteins based on weighted matrix values	[141]

Box 2. Evaluating synonymous variants using *in silico* tools predictive of mRNA structure

The understanding of rules that govern the folding of RNA secondary and tertiary structures is sufficiently well developed that, for a given RNA sequence, a variety of possible structures can be generated and assigned probability scores. Thus, prediction tools may be useful for assessing the change a synonymous mutation bears on mRNA folding. The most common method of predicting secondary structure is free energy minimization. The Gibbs free energy change quantifies the favorability of a structure at a given temperature. The dominant structure is the one with the lowest Gibbs free energy change in solution at equilibrium. Suboptimal structures may be created by examining all possible pairs or according to their probability in the Boltzmann ensemble. Another parameter that can be taken into consideration is alignment and comparison between similar RNA sequences, which can lead to improvements in the accuracy of the prediction tool. A specific challenge is the identification of cases in which folding involves interactions between stems or stem-loops, forming pseudoknots that can then contribute to a complex tertiary structure (non-nested base pairs). Examples of computational methods are: mFold (http://mfold.rna.albany.edu/?q=mfold), Kinefold (http://kinefold.curie.fr/), remuRNA (http://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/remuRNA/index.htm), and RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). An exemplary product of these tools is depicted in Figure I in which the region surrounding the ΔF508 in CFTR was modeled, highlighting variance both within and between predictive software. These discrepancies originate from variation in their construction and primary consideration(s). mFold models mRNA based on the free energy (ΔG) differences associated with different possible secondary structures. Similarly, RNAfold predicts secondary structures of single-stranded RNA or DNA sequences. Kinefold performs folding simulations of nucleic acids using the folding dynamics algorithm and includes the prediction of pseudoknots. A recently developed tool, remuRNA, aims to measure more sensitively the impact of SNPs on the energy and stability of mRNA structure by aligning the possible structures according to their probability in the Boltzmann ensemble [142]. The program efficiently computes the relative entropy for Boltzmann ensembles of wild type and mutant RNA structures. This measure takes into account any changes in the probability distribution of various mRNA conformational states introduced by a point nucleotide substitution. The method can consider both synonymous and nonsynonymous mutations. Of these, changes by synonymous variations tend to result in greater impact on the predicted secondary structure and the Boltzmann probability of mRNA conformers, demonstrating the importance of evolutionary selection on conserved synonymous sites in the preservation of mRNA secondary structure.

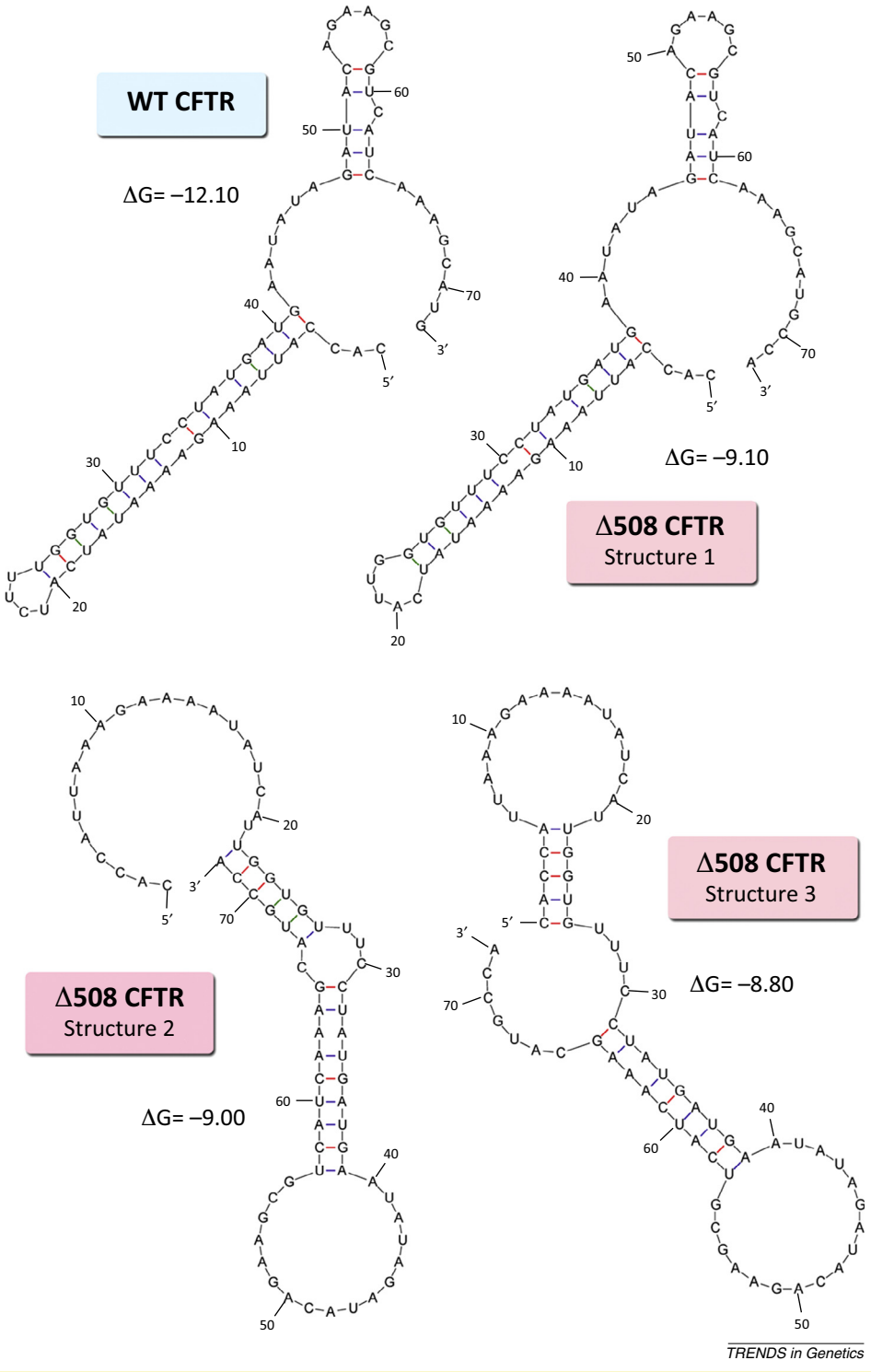

Figure I. *In silico* modeling of RNA folding. Shown here are local structures generated by mFold of wild type cystic fibrosis transmembrane conductance (CFTR) and I507-ATT ΔF508 CFTR, the most common cystic fibrosis-associated allele, which carries a synonymous variant in the neighboring isoleucine. Note that divergent local structures can be generated in spite of similar thermodynamic stability. The sequence used for modeling is: (wild type) CACCAUUAAAGAAAAUAUCAUCUUUGGUGUUUCCUAUGAUGAAUAUAGAUACAGAAGCGUCAUCAAAGCAUG; (I507-ATT ΔF508 CFTR): CACCAUUAAAGAAAAUAUCAUUGGUGUUUCCUAUGAUGAAUAUAGAUACAGAAGCGUCAUCAAAGCAUGCCA.

qRT-PCR can be used to quantitate mRNA transcript levels. The assumption that changes in overall abundance of mRNA are reflective solely of mRNA degradation is potentially faulty, which can be circumvented by techniques that more directly measure decay. These generally use a combination of transcriptional inhibitors [e.g., 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB)] and various transcript-labeling techniques. Using a highly sensitive ribonuclease protection assay, a single synonymous mutation, C957T, in the human dopamine receptor D2 (*DRD2*) gene was found to significantly accelerate mRNA decay [42]. RNA structure prediction using

mFold supported the likelihood that this synonymous mutation favors an mRNA secondary structure lacking a critical stability element. Remarkably, the addition of a second synonymous change, one that bears no functional impact alone, has the capacity to rescue the C957T variant and restore mRNA stability. This comprehensive study highlights the utility of a multifaceted approach, as a spectrum of methods including *in vitro* translation, RNase protection assays, gene GC3 content analysis as well as *in silico* tools collectively proved fruitful in deciphering the impact of this synonymous mutation.

Methods for measuring the effects of synonymous mutations on the rate of translation and protein expression, conformation, and function

Rate of translation

Translation rate is an important determinant of translation fidelity and protein folding. This is exemplified by expressing aggregation-prone eukaryotic proteins in bacteria, experimentally slowing translation through ribosomal mutagenesis results in greater folding efficiency and mitigates aggregation [43]. However, translation rate varies even within a single transcript, as chain elongation experiences stops and pauses. The proposed mechanisms contributing to this phenomenon are diverse and debated [13] but include: (i) mRNA structure: ribosomal processivity is strongly modulated by local mRNA stability and structural motifs [36,44], features that are, in turn, subject to the influence of synonymous nucleotide variation; (ii) codon usage: the cellular pools of isoaccepting tRNAs vary substantially among synonymous codons in a species and tissue-specific manner [45,46], giving rise to the notion that translational rate can be modified by codon choice [47]. Infrequent codons (for which there generally exists a less abundant supply cognate tRNAs [17]) can be used to achieve local control of translation and coordinate co-translational folding [48]. The clustering of rare codons within the N termini of bacterial genes has been observed for some time [49]. The enrichment of these codons has been proposed to slow translation and allow for the proper spacing of adjacent ribosomes under a 'ramp' model of translation [50], thus preventing ribosomal 'traffic jams', translational stalling, and truncated or misfolded proteins [12,51–53]. Others have instead interpreted the same rare 5′ codons as representing selection for RNA structure permissive of translation initiation [44,54]; and (iii) charge of the nascent polypeptide: positively charged residues incorporated into the emerging polypeptide can interact with the negatively charged exit tunnel of the ribosome, an electrostatic interaction conjectured to impact ribosomal speed [55,56]. Although some have argued that positively charged amino acids at N termini help establish a translational ramp, this pattern may merely serve to orient the topology of membrane proteins [57].

It should be noted that whether synonymous codon choices in themselves serve to alter ribosomal speed is an orthodoxy that, although widely advertised and supported in the literature [12,15,13,16], has recently been challenged by transcriptome-wide ribosome profiling: CUB towards frequent codons in highly expressed genes, instead of modulating translation speed, may derive from cellular fitness associated with maintaining a balanced supply of tRNAs [56,58,59]. Nonetheless, given recent examples implicating 'nonoptimal' codon usage in nature being critical for determining proper translational rates, cotranslational folding, and protein functionality [60,61], translational speed remains a topic worthy of attention.

Both experimental and statistical methods have been used to estimate the effect of translation speed. Early methods to measure the speed of translation relied on *in vitro* cell free translation systems. One can alternatively isolate mRNA species according to their ribosome abundance through sucrose gradient fractionation and quantitate polysome-associated RNA (i.e., transcript species with high ribosome density, a proxy for translational speed and/or efficiency). This strategy showed that a single synonymous mutation in a novel tumor suppressor gene (*TTP*) is sufficient to impair translational efficiency and promote breast tumor progression [62]. Polysome profiling is limited by lack of codon-level resolution and can be unreliable due to ribosomal binding at the 5′ untranslated region of some genes [59], confounding this type of analysis. Increasingly sophisticated methods are being used to study the biophysical mechanisms by which the information content of synonymous mutations affects protein synthesis. Moreover, investigating a single synonymous mutation does not indicate how prevalent specific mechanisms are, their relative importance, or whether there are any global patterns. To this end, deep sequencing of ribosome-protected mRNA fragments ('footprints') have been used to generate a genome-wide picture of translation at nucleotide resolution, studies that have revealed elevated 5′ ribosomal density independent of codon sequence length [63]. Conventional ribosome profiling does not directly convey the kinetics of translational elongation, but this can be addressed by protocols that stop translation initiation while allowing for translation run-off [59].

Finally, the dynamics of translation with a single ribosome can be studied by at least two independent methods. The first uses ribosomes immobilized on a slide with biotin-streptavidin linkage and protein synthesis is followed as a fluorescence trace using labeled tRNAs [64]. The second method follows individual ribosomes as they translate single messenger RNA hairpins tethered by optical tweezers [65]. These novel approaches have not yet been applied for the study of synonymous mutations.

Protein levels

Techniques for accurately determining protein levels are central to the effective evaluation of synonymous mutations. Traditional methods to measure the amount of a specific protein include spectroscopy or immunoblotting, whereas methods for global quantification of protein levels include absolute protein expression profiling (APEX) [66] and parallel metabolic pulse labeling [67]. The APEX method, pioneered in bacterial and yeast proteomes, uses mass spectrometry in combination with observed protein distributions to produce a peptide count for cells, and has been shown to agree with results from western blotting, flow cytometry, and 2D gel electrophoresis. Parallel metabolic pulse labeling was used to study mRNA and protein

abundance as well as turnover in mammalian cells, using fluorescence to measure mRNA and protein levels. Interestingly, in bacteria and yeast, differences in protein half-lives were explained 73% and 47%, respectively, by mRNA half-lives [66], whereas mammalian cells showed no correlation between mRNA and protein half-lives despite a correlation observed between mRNA and protein levels. This underscores the possibility of synonymous variations altering mRNA or protein stability without necessarily changing the abundance of either mRNA or protein.

Protein stability

Subtle changes resulting from the introduction of synonymous codon changes have the potential to affect protein stability and half-life. This can be inferred crudely by subjecting proteins to repetitive freeze–thaw cycles [60] and assessing their relative structural preservation. A more precise assessment can be performed by treating with cycloheximide followed by protein quantitation through time [68]. Pulse-chase experiments, typically using radioactive or fluorescent labeling approaches, are a classic yet reliable technique to assess protein stability. Radiolabeling two synonymous variants of ΔF508 CFTR provided evidence for substantial reduction in transmembrane protein stability, whereas a construct of CFTR in which a synonymous change was experimentally reverted, showed stability similar to wild type CFTR [69]. Another pulse-chase experiment using ^{125}I, ^{3}H cholesteryl (CE) radio-labeled HDL was used to show that the stability of wild type scavenger receptor class B type I (SCARB1) was significantly greater than that of a synonymous variant [70]. An alternative approach to assessing protein half-life in real time that does not require translation inhibitors, antibodies, or radioactive labeling is the so-called 'bleach-chase method' [71]. Here, protein turnover in a cell is assessed using a pulse of light to render a portion of fluorescently tagged protein in a cell nonfluorescent, an inversion of the classic pulse-chase experiment. The determination of protein half-life is then accomplished through time-lapse microscopy as 'bleached' protein is removed from cells via degradation and dilution.

Protein aggregation and unfolded protein accumulation

Protein misfolding due to synonymous mutations can be recognized by cellular quality control machinery and degraded, but such errors can also result in aggregation. The accumulation of toxic misfolded proteins and compromised proteostasis underlies many human diseases [51,72]. A small pool of proteins prone to misfolding can promote a cascading effect, resulting in proteotoxicity [73]. Given this, it is not surprising that mistranslation-induced protein misfolding has been proposed to exert a selective constraint that patterns coding-sequence evolution [74]. Across taxa, residues buried within the protein molecule, sites that are thought to be influential for determining structure, are encoded by translationally optimal codons [75]. In *Escherichia coli*, there is a greater degree of codon adaptation within genes whose protein products fold mostly independent of chaperonin aide, whereas those dependent on its function have not developed such bias [76], supporting the idea that synonymous mutations can influence protein folding.

Experimentally, aggregation and misfolding can be assessed through the use of protease inhibitors, which can potentiate the accumulation of such proteins [69]. Quantification of endoplasmic reticulum (ER)-stress markers is another strategy to infer protein misfolding resulting from synonymous mutations. Common stress markers indicative of the unfolded protein response include three types of ER-resident sensor proteins, inositol-requiring protein-1 (IRE1), protein kinase RNA-like ER kinase (PERK), and activating transcription factor 6 (ATF6), activated by immunoglobin binding protein (BiP), which initiate a series of signaling cascades [77,78]. Lastly, promising avenues to more directly study cotranslational folding have emerged, and it is now possible to generate a homogenous population of translationally arrested ribosomal nascent chain complexes at specific residues of interest [79]

Protein conformation

Anfinsen's principle holds that the amino acid sequence itself is sufficient to direct native, thermodynamically driven protein folding. However, synonymous mutations have indeed been shown to impair or alter cotranslational protein folding, which can lead to conformational differences [3,5] and change the functionality of the protein product [2]. There are established methods to infer changes to protein conformation that have broader applicability and are less onerous than atomic level techniques, such as NMR or X-ray crystallography (which have not been used in the study of synonymous mutations). Enzymatic degradation has long been used to assess differences in protein conformation, with the idea that dissimilar access to cleavage sites is reflective of discrepant protein structures. One can directly hydrolyze peptide bonds with limited trypsin or chymotrypsin digestion, which has been successfully used to compare synonymous variants of multidrug resistance protein 1 (MDR1) and the circadian clock protein FREQUENCY (FRQ) in *Neurospora* [2,60] or enzymatically cleave glycoside bonds (e.g., *N*-linked glycans with PNGase). Other techniques, with proven application in the study of nonsynonymous mutations, should be extended for synonymous variants as well. In addition to assessing mRNA structure as previously mentioned, the chirality of alpha helices and beta sheets allows circular dichroism to determine the degree of foldedness or changes in secondary structures between synonymous variants [40]. Differential scanning calorimetry can assess thermal stability and heat-induced macromolecule transitions, generating thermodynamic parameters that are revealing of the structural state and intermolecular associations [80,81]. Subtle differences in protein conformation can also be inferred through differential immunodetection, made possible by the isolation of antibodies specific for an epitope type (e.g., linear versus conformational) or specific conformational states (e.g., active versus inactive).

Concluding remarks

A little over a decade ago, the term 'silent' was used to describe synonymous mutations. It is now evident that such mutations can and do have a role in human health

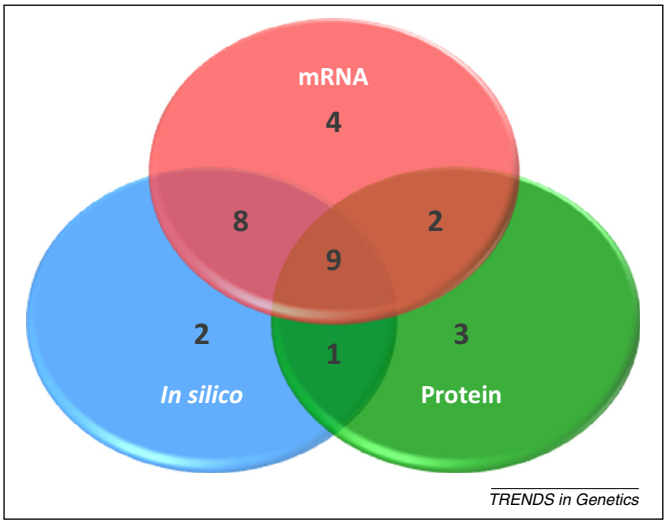

Figure 2. Diagram depicting recent scientific literature (since 2006) focusing on synonymous polymorphisms and/or mutations in human disease and the general category of methods used. 'Protein' refers to any work measuring translation, expression, function, or conformational analysis of protein; 'RNA' refers to any work investigating mRNA splicing, structure, stability, or decay; '*In silico*' refers to any work using predictive computational software. For a full list of publications comprising these categories, see Table S1 in the supplementary material online.

and disease. This revised understanding of a key concept in biology has been possible, in part, due to technological advances and the application of classic methods in novel ways. In a short time, evidence from diverse fields has converged to revise the dogma that synonymous mutations are biologically insignificant. Indeed, synonymous changes can influence multiple levels of cellular biology, from DNA and RNA to protein-based features. Elucidating the exact mechanistic underpinnings of these changes requires the use of an extensive range of methods (Figure 2). As a scientific community, we have made great headway in exposing the involvement of synonymous mutations in a variety of biological and clinical contexts. Given our growing understanding of the contribution of synonymous mutations to biological phenomena, attempts to assess the impact of these mutations should be broad and comprehensive, as outlined by the methods discussed here.

Disclaimer

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.

Acknowledgments

We thank Nobuko Katagiri, CBER/FDA for expert assistance with mRNA folding tools and Avi Kimchi, NCBI, NIH for editorial assistance and insightful discussions. This work was supported by funds from the Laboratory of Hemostasis, Center for Biologics Evaluation and Research (C.K-S. and Z.E.S.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2014.04.006.

References

1 Chamary, J.V. and Hurst, L.D. (2009) The price of silent mutations. *Sci. Am.* 300, 46–53
2 Kimchi-Sarfaty, C. *et al.* (2007) A 'silent' polymorphism in the MDR1 gene changes substrate specificity. *Science* 315, 525–528
3 Sauna, Z.E. and Kimchi-Sarfaty, C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* 12, 683–691
4 Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42
5 Sauna, Z.E. and Kimchi-Sarfaty, C. (2001) *Synonymous Mutations as a Cause of Human Genetic Disease*, John Wiley & Sons
6 Chen, R. *et al.* (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* 5, e13574
7 Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18, 486–487
8 Lind, P.A. *et al.* (2010) Mutational robustness of ribosomal protein genes. *Science* 330, 825–827
9 Schattner, P. and Diekhans, M. (2006) Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res.* 34, 1700–1710
10 Parmley, J.L. and Hurst, L.D. (2007) How common are intragene windows with KA > KS owing to purifying selection on synonymous mutations? *J. Mol. Evol.* 64, 646–655
11 Spencer, P.S. *et al.* (2012) Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* 422, 328–335
12 Tuller, T. *et al.* (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354
13 Dana, A. and Tuller, T. (2012) Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.* 8, e1002755
14 Czech, A. *et al.* (2010) Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations. *Mol. Biosyst.* 6, 1767–1770
15 Cannarozzi, G. *et al.* (2010) A role for codon order in translation dynamics. *Cell* 141, 355–367
16 Novoa, E.M. and Ribas de, P.L. (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28, 574–581
17 Dong, H. *et al.* (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260, 649–663
18 Eory, L. *et al.* (2010) Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol. Biol. Evol.* 27, 177–192
19 Line, S.R. *et al.* (2013) Translational signatures and mRNA levels are highly correlated in human stably expressed genes. *BMC Genomics* 14, 268
20 Stergachis, A.B. *et al.* (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342, 1367–1372
21 Parmley, J.L. and Hurst, L.D. (2007) Exonic splicing regulatory elements skew synonymous codon usage near intron-exon boundaries in mammals. *Mol. Biol. Evol.* 24, 1600–1603
22 Warnecke, T. and Hurst, L.D. (2007) Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. *Mol. Biol. Evol.* 24, 2755–2762
23 Urrutia, A.O. and Hurst, L.D. (2001) Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics* 159, 1191–1199
24 Novembre, J.A. (2002) Accounting for background nucleotide composition when measuring codon usage bias. *Mol. Biol. Evol.* 19, 1390–1394
25 Fox, J.M. and Erill, I. (2010) Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Res.* 17, 185–196
26 Richards, C.S. *et al.* (2008) ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* 10, 294–300
27 Buske, O.J. *et al.* (2013) Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 29, 1843–1850
28 Hamasaki-Katagiri, N. *et al.* (2013) A gene-specific method for predicting hemophilia-causing point mutations. *J. Mol. Biol.* 425, 4023–4033
29 Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437

30 Solis, A.S. *et al.* (2008) Splicing fidelity, enhancers, and disease. *Front. Biosci.* 13, 1926–1942
31 Caceres, E.F. and Hurst, L.D. (2013) The evolution, impact and properties of exonic splice enhancers. *Genome Biol.* 14, R143
32 Parmley, J.L. *et al.* (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* 23, 301–309
33 Macaya, D. *et al.* (2009) A synonymous mutation in TCOF1 causes Treacher Collins syndrome due to mis-splicing of a constitutive exon. *Am. J. Med. Genet. A* 149A, 1624–1627
34 Nielsen, K.B. *et al.* (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
35 Kudla, G. *et al.* (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258
36 Griswold, K.E. *et al.* (2003) Effects of codon usage versus putative 5′-mRNA structure on the expression of Fusarium solani cutinase in the *Escherichia coli* cytoplasm. *Protein Expr. Purif.* 27, 134–142
37 Ding, Y. *et al.* (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700
38 Rouskin, S. *et al.* (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705
39 Nackley, A.G. *et al.* (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314, 1930–1933
40 Bartoszewski, R.A. *et al.* (2010) A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J. Biol. Chem.* 285, 28741–28748
41 Duan, J. *et al.* (2013) Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci. Rep.* 3, 1318
42 Duan, J. *et al.* (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* 12, 205–216
43 Siller, E. *et al.* (2010) Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* 396, 1310–1318
44 Bentele, K. *et al.* (2013) Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675
45 Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2, 13–34
46 Dittmar, K.A. *et al.* (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* 2, e221
47 Lavner, Y. and Kotlar, D. (2005) Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 345, 127–138
48 Zhang, G. *et al.* (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 16, 274–280
49 Chen, G.F. and Inouye, M. (1990) Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.* 18, 1465–1473
50 Mitarai, N. *et al.* (2008) Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J. Mol. Biol.* 382, 236–245
51 Balch, W.E. *et al.* (2008) Adapting proteostasis for disease intervention. *Science* 319, 916–919
52 Ehrenberg, M. *et al.* (2010) Maximum rrn promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase. *Biochimie* 92, 12–20
53 Powers, E.T. and Balch, W.E. (2008) Costly mistakes: translational infidelity and protein homeostasis. *Cell* 134, 204–206
54 Goodman, D.B. *et al.* (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479
55 Lu, J. and Deutsch, C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* 384, 73–86
56 Charneski, C.A. and Hurst, L.D. (2013) Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11, e1001508
57 Charneski, C.A. and Hurst, L.D. (2014) Positive charge loading at protein termini is due to membrane protein topology. not a translational ramp. *Mol. Biol. Evol.* 31, 70–84
58 Qian, W. *et al.* (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 8, e1002603
59 Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
60 Zhou, M. *et al.* (2013) Non-optimal codon usage affects expression. structure and function of clock protein FRQ. *Nature* 495, 111–115
61 Xu, Y. *et al.* (2013) Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* 495, 116–120
62 Griseri, P. *et al.* (2011) A synonymous polymorphism of the Tristetraprolin (TTP) gene. an AU-rich mRNA-binding protein. affects translation efficiency and response to Herceptin treatment in breast cancer patients. *Hum. Mol. Genet.* 20, 4556–4568
63 Ingolia, N.T. *et al.* (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
64 Chen, C. *et al.* (2013) Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat. Struct. Mol. Biol.* 20, 582–588
65 Wen, J.D. *et al.* (2008) Following translation by single ribosomes one codon at a time. *Nature* 452, 598–603
66 Lu, P. *et al.* (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25, 117–124
67 Schwanhausser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
68 Turgeon, D. *et al.* (2001) Relative enzymatic activity, protein stability, and tissue distribution of human steroid-metabolizing UGT2B subfamily members. *Endocrinology* 142, 778–787
69 Lazrak, A. *et al.* (2013) The silent codon change I507-ATC->ATT contributes to the severity of the DeltaF508 CFTR channel dysfunction. *FASEB J.* 27, 4630–4645
70 Constantineau, J. *et al.* (2010) A synonymous variant in scavenger receptor. class B. type I gene is associated with lower SR-BI protein expression and function. *Atherosclerosis* 210, 177–182
71 Eden, E. *et al.* (2011) Proteome half-life dynamics in living human cells. *Science* 331, 764–768
72 Gusella, J.F. and MacDonald, M.E. (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* 1, 109–115
73 Gidalevitz, T. *et al.* (2006) Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* 311, 1471–1474
74 Drummond, D.A. and Wilke, C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352
75 Zhou, T. *et al.* (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol. Biol. Evol.* 26, 1571–1580
76 Warnecke, T. and Hurst, L.D. (2010) GroEL dependency affects codon usage: support for a critical role of misfolding in gene evolution. *Mol. Syst. Biol.* 6, 340
77 Zhang, K. and Kaufman, R.J. (2006) Protein folding in the endoplasmic reticulum and the unfolded protein response. *Handb. Exp. Pharmacol.* 172, 69–91
78 Zhang, K. and Kaufman, R.J. (2008) From endoplasmic-reticulum stress to the inflammatory response. *Nature* 454, 455–462
79 Evans, M.S. *et al.* (2005) Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat. Methods* 2, 757–762
80 Chiu, M.H. and Prenner, E.J. (2011) Differential scanning calorimetry: an invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. *J. Pharm. Bioallied Sci.* 3, 39–59
81 Johnson, C.M. (2013) Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* 531, 100–109
82 Friedman, R.C. *et al.* (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105
83 Fang, Z. *et al.* (2013) The sequence structures of human microRNA molecules and their implications. *PLoS ONE* 8, e54215
84 Bartel, D.P. (2004) MicroRNAs: genomic, biogenesis, mechanism, and function. *Cell* 116, 281–297
85 Lewis, B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell* 115, 787–798

86 Min, H. and Yoon, S. (2010) Got target? Computational methods for microRNA target prediction and their extension. *Exp. Mol. Med.* 42, 233–244
87 Alexiou, P. *et al.* (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* 25, 3049–3055
88 Rajewsky, N. (2006) microRNA target predictions in animals. *Nat. Genet.* 38 (Suppl.), S8–S13
89 Lim, L.P. *et al.* (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773
90 Bentwich, I. (2005) Prediction and validation of microRNAs and their targets. *FEBS Lett.* 579, 5904–5910
91 Coronnello, C. *et al.* (2012) Novel modeling of combinatorial miRNA targeting identifies SNP with potential role in bone density. *PLoS Comput. Biol.* 8, e1002830
92 Chi, S.W. *et al.* (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486
93 Hong, X. *et al.* (2009) Immunopurification of Ago1 miRNPs selects for a distinct class of microRNA targets. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15085–15090
94 Parkes, M. *et al.* (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* 39, 830–832
95 Nonne, N. *et al.* (2010) Tandem affinity purification of miRNA target mRNAs (TAP–Tar). *Nucleic Acids Res.* 38, e20
96 Brest, P. *et al.* (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.* 43, 242–245
97 Gartner, J.J. *et al.* (2013) Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13481–13486
98 Enright, A.J. *et al.* (2003) MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1
99 Maragkakis, M. *et al.* (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.* 37, W273–W276
100 Kertesz, M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284
101 Hiard, S. *et al.* (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res.* 38, D640–D651
102 Lewis, B.P. *et al.* (2005) Conserved seed pairing. often flanked by adenosines. indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20
103 Chang, T.H. *et al.* (2013) An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* 14 (Suppl 2), S4
104 Nielsen, C.B. *et al.* (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 13, 1894–1910
105 Friedman, Y. *et al.* (2010) MiRror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics* 26, 1920–1921
106 Krek, A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500
107 Coronnello, C. and Benos, P.V. (2013) ComiR: combinatorial microRNA target prediction tool. *Nucleic Acids Res.* 41, W159–W164
108 Richard, P. *et al.* (2007) A synonymous CHRNE mutation responsible for an aberrant splicing leading to congenital myasthenic syndrome. *Neuromuscul. Disord.* 17, 409–414
109 Saito, M. and Ishiko, A. (2008) A novel silent mutation in the NSDHL gene causing CHILD syndrome as a result of aberrant splicing. *Br. J. Dermatol.* 159, 1204–1206
110 Knobe, K.E. *et al.* (2008) Why does the mutation G17736A/Val107Val (silent) in the F9 gene cause mild haemophilia B in five Swedish families? *Haemophilia* 14, 723–728
111 Ramser, J. *et al.* (2008) Rare missense and synonymous variants in UBE1 are associated with X-linked infantile spinal muscular atrophy. *Am. J. Hum. Genet.* 82, 188–193
112 Tveten, K. *et al.* (2009) Functional analysis of the synonymous R385R mutation in the low-density lipoprotein receptor gene. *Genet. Test. Mol. Biomarkers* 13, 243–248
113 Kulseth, M.A. *et al.* (2010) Analysis of LDLR mRNA in patients with familial hypercholesterolemia revealed a novel mutation in intron 14. which activates a cryptic splice site. *J. Hum. Genet.* 55, 676–680
114 Pavan, M. *et al.* (2009) ALDH1A2 (RALDH2) genetic variation in human congenital heart disease. *BMC Med. Genet.* 10, 113
115 Narendra, U. *et al.* (2009) Genetic analysis of complement factor H related 5. CFHR5. in patients with age-related macular degeneration. *Mol. Vis.* 15, 731–736
116 Vidal, C. *et al.* (2009) Effects of a synonymous variant in exon 9 of the CD44 gene on pre-mRNA splicing in a family with osteoporosis. *Bone* 45, 736–742
117 Dobrowolski, S.F. *et al.* (2010) The phenylalanine hydroxylase c.30C>G synonymous variation (p.G10G) creates a common exonic splicing silencer. *Mol. Genet. Metab.* 100, 316–323
118 Davidson, A.E. *et al.* (2010) A synonymous codon variant in two patients with autosomal recessive bestrophinopathy alters in vitro splicing of BEST1. *Mol. Vis.* 16, 2916–2922
119 Groen, J.L. *et al.* (2010) DYT6 dystonia: mutation screening, phenotype, and response to deep brain stimulation. *Mov. Disord.* 25, 2420–2427
120 Ho, P.A. *et al.* (2011) 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
121 Budkowska, A. *et al.* (2011) Synonymous mutations in the core gene are linked to unusual serological profile in hepatitis C virus infection. *PLoS ONE* 6, e15871
122 Edwards, N.C. *et al.* (2012) Characterization of coding synonymous and non–synonymous variants in ADAMTS13 using ex vivo and in silico approaches. *PLoS ONE* 7, e38864
123 Jacobo, S.M. *et al.* (2013) Age-related macular degeneration-associated silent polymorphisms in HtrA1 impair its ability to antagonize insulin-like growth factor 1. *Mol. Cell. Biol.* 33, 1976–1990
124 Moller, L.B. *et al.* (2013) A silent nucleotide substitution in the ATP7A gene in a child with Menkes disease. *Mol. Genet. Metab.* 110, 490–492
125 Sheikh, T.I. *et al.* (2013) A synonymous change. p.Gly16Gly in MECP2 Exon 1 causes a cryptic splice event in a Rett syndrome patient. *Orphanet J. Rare Dis.* 8, 108
126 Bruun, G.H. *et al.* (2013) A synonymous polymorphic variation in ACADM exon 11 affects splicing efficiency and may affect fatty acid oxidation. *Mol. Genet. Metab.* 110, 122–128
127 Kwon, W.S. *et al.* (2009) G-T haplotype (2677G>T/A and 3435C>T) of ABCB1 gene polymorphisms is associated with ethnic differences to paclitaxel sensitivity in cancer cells with different gene expression pattern. *Cancer Lett.* 277, 155–163
128 Herrlinger, K.R. *et al.* (2011) ABCB1 single-nucleotide polymorphisms determine tacrolimus response in patients with ulcerative colitis. *Clin. Pharmacol. Ther.* 89, 422–428
129 Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15, 7155–7174
130 Carmel, I. *et al.* (2004) Comparative analysis detects dependencies among the 5′ splice-site positions. *RNA* 10, 828–840
131 Reese, M.G. *et al.* (1997) Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323
132 Pertea, M. *et al.* (2001) GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* 29, 1185–1190
133 Desmet, F.O. *et al.* (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37, e67
134 Wang, M. and Marin, A. (2006) Characterization and prediction of alternative splice sites. *Gene* 366, 219–227
135 Hebsgaard, S.M. *et al.* (1996) Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res.* 24, 3439–3452
136 Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394

137 Paz, I. *et al.* (2010) SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res.* 38, W281–W285
138 Dogan, R.I. *et al.* (2007) SplicePort: an interactive splice-site analysis tool. *Nucleic Acids Res.* 35, W285–W291
139 Wang, Z. *et al.* (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* 119, 831–845
140 Fairbrother, W.G. *et al.* (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013
141 Cartegni, L. *et al.* (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31, 3568–3571
142 Salari, R. *et al.* (2013) Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic Acids Res.* 41, 44–53