nature reviews genetics

https://doi.org/10.1038/s41576-025-00850-1

Perspective

Check for updates

Functional synonymous mutations and their evolutionary consequences

Jianzhi Zhang [1] & Wenfeng Qian [2,3]

Abstract

Synonymous mutations are coding mutations that do not alter protein sequences. Commonly thought to have little to no functional consequence, synonymous mutations have been widely used in evolutionary analyses that require neutral markers, including those foundational for the neutral theory. However, recent studies suggest that synonymous mutations can influence nearly every step in the expression of genetic information and may often be strongly non-neutral. We review the extent and mechanisms of these phenotypic and fitness effects and discuss the implications of the functionality and non-neutrality of synonymous mutations for various analyses and conclusions pertinent to genetics, evolution, conservation and disease.

Sections

Introduction

Impact on gene expression

Causes of synonymous codon usage bias

Fitness effects of synonymous mutations

Evolutionary consequences

Synonymous mutations and disease

Conclusions and future studies

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [2]State Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [3]University of Chinese Academy of Sciences, Beijing, China. e-mail: jianzhi@umich.edu; wfqian@genetics.ac.cn

Introduction

In virtually every organism, the 20 amino acids that make up all proteins are encoded by 61 of the 64 possible triplet codons in DNA, with the remaining three codons serving as stop signs in protein synthesis. Two amino acids — methionine (Met) and tryptophan (Trp) — are each encoded by one codon, whereas the other 18 amino acids are each encoded by two to six different codons. Consequently, a single-nucleotide mutation in a coding DNA sequence might not alter the protein amino acid sequence; such mutations are known as synonymous mutations or silent mutations. By contrast, single-nucleotide mutations that alter the protein sequence are called nonsynonymous, missense or replacement mutations. Depending on the relative frequencies of the 61 sense codons and the molecular spectrum of mutation, approximately one-fourth to one-third of all single-nucleotide coding mutations are synonymous.

Synonymous mutations have long been presumed to have minimal phenotypic or fitness effects and regarded as neutral or nearly neutral[1]. Studies of synonymous mutations have contributed substantially to the development of the neutral theory of molecular evolution[2], a landmark evolutionary theory considered the sole conceptual revolution in evolutionary biology since the modern synthesis[3]. The neutral assumption of synonymous mutations has allowed synonymous polymorphisms and substitutions to be used as neutral markers (that is, free from natural selection) in many analyses, including, for example, to infer the rate and molecular spectrum of mutation, estimate effective population size (N_e), test natural selection and date evolutionary events.

The first indication that synonymous mutations may be non-neutral appeared in the mid-1970s, when synonymous codons were found at different frequencies in the few genes sequenced by then[4–7]. It was soon discovered that the relative frequency of a synonymous codon in a gene tends to rise with the relative abundance of the corresponding tRNA in the cell. This phenomenon of synonymous codon usage bias (CUB) is more prominent in highly expressed genes than in lowly expressed ones, prompting the hypothesis that translational efficiency is affected by synonymous mutations and optimized through CUB[8,9]. Synonymous codons over-represented and under-represented in highly expressed genes of a genome are therefore referred to as its preferred codons and unpreferred codons, respectively.

Since the pioneering work described above, synonymous mutations have been found to influence many steps in the expression of genetic information, from transcription to mRNA processing, translation and co-translational protein folding. One could even argue that synonymous mutations alter the protein sequence (stochastically) because they affect translational accuracy. Furthermore, evidence for fitness effects of synonymous mutations is accumulating, thanks in a large part to the advent of genome editing and high-throughput, sequencing-based fitness quantification. These new findings challenge the common assumption that synonymous mutations are phenotypically silent, demanding a broad examination of how the non-neutrality of synonymous mutations impacts various analyses and conclusions pertinent to genetics, evolution, conservation and disease.

In this Perspective, we describe the mechanisms underlying the phenotypic and fitness effects of synonymous mutations and the causes of CUB. We then summarize evidence for the non-neutrality of synonymous mutations and discuss how this non-neutrality can affect various analyses dependent on the neutral assumption of synonymous mutations. Finally, we describe disease-causing synonymous mutations and discuss outstanding questions. Although our focus is on eukaryotes, prokaryotic studies are referred to when relevant.

Impact on gene expression

Synonymous mutations in a gene can affect multiple steps in gene expression and alter the concentration of the functional protein[10,11] (Fig. 1). Below we describe these steps individually.

Transcription

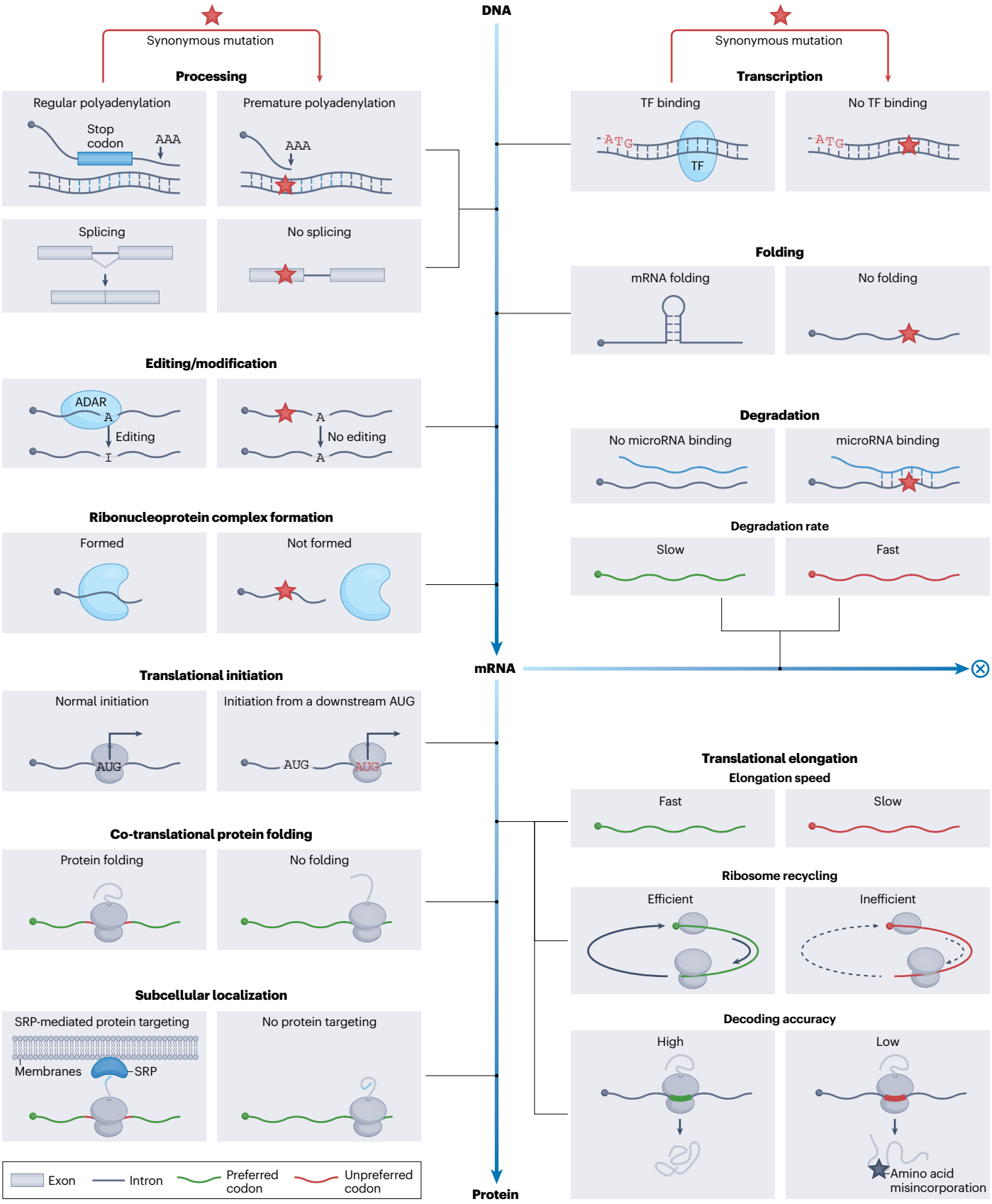
In humans, ~15% of codons are dual-use codons ('duons') that simultaneously specify amino acids and bind to transcription factors[12]. Hence, synonymous mutations in duons can affect transcription factor binding and transcription[12,13]. Note that a biochemical effect may or may not have a fitness effect detectable by natural selection[14], an important distinction in discussing functional synonymous mutations. Two studies suggested that the transcription-factor-binding activities of most duons are not subject to selection[15,16] and that the purported sequence conservation of duons[12] is an artefact of the variation in the percentage of G and C nucleotides (GC content)[16]. However, a subsequent study reported sequence conservation of duons even after controlling for GC content[17]. In yeast, a study of thousands of de novo coding mutations did not find a significant difference between mutations within and outside transcription-factor-binding regions in their effects on the mRNA concentration[18], providing no evidence for the hypothesis of widespread functionality of duons in transcription. An *Escherichia coli* study reported that synonymous mutations in a gene placed on a plasmid could affect the transcription of a neighbouring gene on the same plasmid[19], but the generality of this phenomenon is unknown.

By changing the local GC content, synonymous mutations can affect nucleosome positioning and thereby influence the accessibility of transcription-factor-binding sites and transcription[20,21]. Indeed, in the multicellular fungus *Neurospora crassa*, some mutations that convert unpreferred to preferred codons in reporter genes increase transcription[22]. Across *N. crassa* genes, the nuclear transcript level is positively correlated with the use of preferred codons, prompting the hypothesis of a genome-wide role of CUB on transcription[23]. However, this positive correlation more likely results from natural selection for preferred codons in highly expressed genes than from the boosting of transcription by preferred codons[24] (as discussed below).

mRNA processing, modification and localization

A eukaryotic pre-mRNA must be processed through multiple steps such as polyadenylation and splicing before translation. Efficient polyadenylation signals are under-represented in *N. crassa* coding sequences, especially in the case of highly expressed genes[25], which suggests that

Fig. 1 | Multifaceted effects of synonymous mutations on the expression of genetic information in eukaryotes. The effects of synonymous mutations (indicated by a red star or a change in the colour of mRNA) on gene expression are mapped onto the backbone of the central dogma (blue arrows). Shown are the effects on transcription, RNA processing, RNA folding, RNA editing/modification, ribonucleoprotein complex formation, RNA degradation, translational initiation, translational elongation, co-translational protein folding and protein subcellular localization. These effects can alter the sequence, structure, amount and/or localization of the RNA or protein produced. ADAR (adenosine deaminase acting on RNA) is the enzyme responsible for A-to-I RNA editing in animals. SRP, signal recognition particle; TF, transcription factor.


DNA
Synonymous mutation
Processing
Regular polyadenylation
Stop codon
AAA
Premature polyadenylation
AAA
Splicing
No splicing
Editing/modification
ADAR
A
Editing
I
A
No editing
A
Ribonucleoprotein complex formation
Formed
Not formed
Translational initiation
Normal initiation
AUG
Initiation from a downstream AUG
AUG
AUG
Co-translational protein folding
Protein folding
No folding
Subcellular localization
SRP-mediated protein targeting
Membranes
SRP
No protein targeting
Exon
Intron
Preferred codon
Unpreferred codon
Synonymous mutation
Transcription
TF binding
ATG
TF
No TF binding
ATG
Folding
mRNA folding
No folding
Degradation
No microRNA binding
microRNA binding
Degradation rate
Slow
Fast
mRNA
Translational elongation
Elongation speed
Fast
Slow
Ribosome recycling
Efficient
Inefficient
Decoding accuracy
High
Low
Amino acid misincorporation
Protein

synonymous mutations that create polyadenylation signals have been selectively purged. Pre-mRNA splicing depends on splicing donors, acceptors and regulatory sequences. Gains and losses of these elements by synonymous mutations are generally detrimental[26,27]. Many mRNA modifications such as A-to-I editing and m^6A methylation depend on *cis*-regulatory motifs, so they can be affected by synonymous mutations. Indeed, synonymous mutations in tumour suppressor genes were recently reported to promote tumorigenesis by disrupting m^6A-dependent mRNA metabolism[28]. Synonymous mutations can also influence mRNA localization by altering the local GC content[29–31]. More broadly, mRNA processing, modification and localization often depend on the formation of ribonucleoprotein complexes comprising RNA molecules and their specific binding proteins. Synonymous mutations can be non-neutral because they influence RNA–protein interactions[32].

mRNA secondary structure

mRNAs fold to specific secondary structures that function in the head-to-tail intramolecular communication along the mRNA and in translational regulation[33–36]. Synonymous mutations can alter mRNA secondary structures and folding strengths, and thereby be subject to selection[37–39]. In particular, efficient translational initiation relies on a fairly open mRNA structure around the start codon[40,41], and, consistently, synonymous mutations in the first ~40 nucleotides of an open reading frame can influence the amount of protein produced[34]. Additionally, GC-poor synonymous codons are favoured around microRNA target sites in plants, presumably because they reduce mRNA secondary structures and increase the accessibility of microRNA target sites[42].

Translational initiation

Synonymous codon usage can regulate translational initiation, with unpreferred codons reducing initiation by decreasing the binding of initiation factors to transcripts[43]. Synonymous mutations can also generate out-of-frame AUG triplets downstream of the bona fide start codon. In eukaryotes, these triplets compete for translational initiation when they are near the start codon[44]. Indeed, proximal out-of-frame AUG triplets downstream of the start codon are depleted in yeast and human genomes[44]. Similarly, in prokaryotes, the Shine–Dalgarno sequence, which determines the position of translational initiation, can appear in the coding region by synonymous mutations. Such coding-region Shine–Dalgarno sequences are under-represented (especially in highly expressed genes)[45,46] and evolutionarily transient[47], suggesting that they are detrimental.

Translational elongation

Because the waiting time for the cognate tRNA of a codon to arrive at the ribosome A site should be inversely proportional to the relative concentration of the tRNA among all available tRNAs, and because preferred codons have high relative concentrations of cognate tRNAs[8], preferred codons have long been assumed to have elevated translational elongation speed[8,48]. This assumption became verifiable at the genomic scale with the advent of ribosome profiling or Ribo-seq, which uses high-throughput sequencing to determine in vivo ribosome locations on mRNAs with codon resolution[49]. The rationale is that faster translational elongation at a codon should result in fewer captured ribosomes and therefore fewer Ribo-seq reads at the codon. Unexpectedly, elongation was not found to be faster at preferred codons than at unpreferred codons based on early Ribo-seq data[50,51], probably because the cycloheximide treatment does not fully stop the ribosome movement, affecting elongation speed quantification. Later experiments that stopped ribosomes by flash-freezing cells indeed revealed faster elongations of preferred than unpreferred codons[52,53].

The translational 'ramp' hypothesis proposes that translational elongation is slowed by unpreferred codons at the first 30–50 codons of a transcript, serving to minimize downstream ribosomal traffic jams[54]. Although an elevated ribosome density in the corresponding region was initially detected[49], this was subsequently found to be largely an artefact of the cycloheximide treatment[52,53,55]. The use of unpreferred codons in this region may have resulted from natural selection against mRNA folding that impedes efficient translational initiation[56,57]. Alternatively, it may be a by-product of the rapid evolutionary turnover of the 5′ coding region[58]. Further evidence against the 'ramp' hypothesis came from disome sequencing (Disome-seq) experiments (Box 1).

mRNA stability

Using unpreferred codons in a gene promotes its mRNA degradation and thereby reduces the mRNA concentration[59–63]. Specifically, yeast experiments showed that Not5 is recruited to the ribosome E-site when unpreferred codons cause slow translational elongation, which in turn activates the Ccr4-Not complex that degrades the poly-A tail and recruits Dhh1 that promotes de-capping, leading to mRNA degradation[64–66]. This translation-dependent mRNA degradation mechanism probably originated to purge defective mRNA molecules that stall ribosomes, whereas the degradation of mRNAs rich in unpreferred codons may be a by-product. A recent mammalian study showed that synonymous mutations can influence the occupancy of tRNAs at the ribosome P-site, which in turn affects the recruitment of the CCR4-NOT complex subunit CNOT3 (known as Not5 in yeast) to the ribosome E-site, impacting mRNA degradation[67]. Additionally, synonymous mutations can affect mRNA stability by modulating microRNA binding, evidenced by slowed synonymous substitutions at microRNA binding regions[68].

Translational accuracy

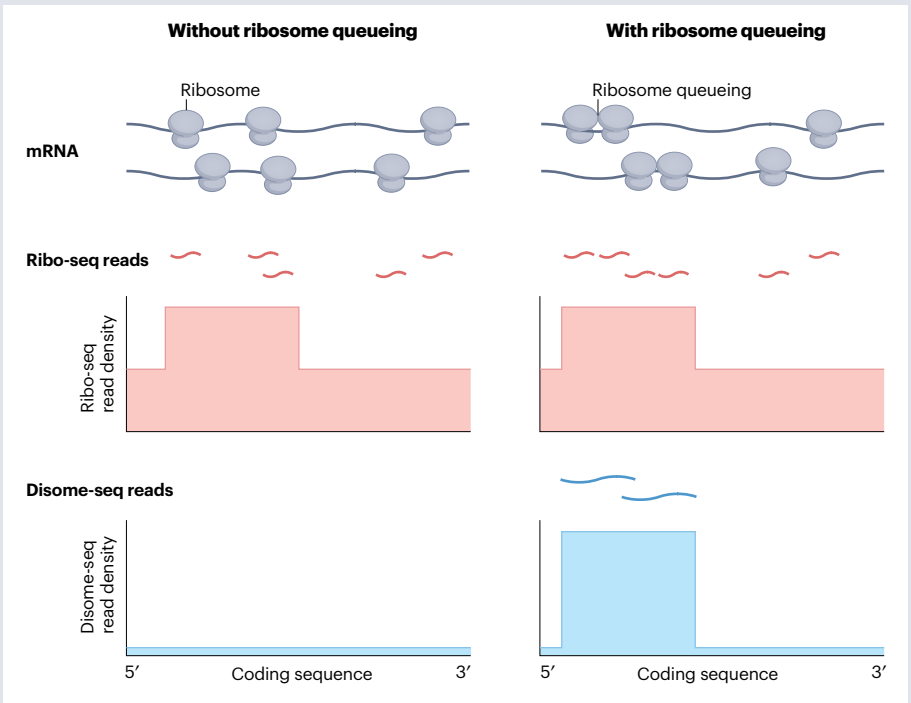
Mistranslation — incorporation into the peptide of amino acids not encoded by the mRNA — is primarily caused by codon–anticodon mispairing in translation[69]. Although ribosomes can generally distinguish cognate from near-cognate tRNAs, the distinction is imperfect[69]. Mistranslations identified from *E. coli* proteomic data[69] show that preferred codons are generally more accurately translated than unpreferred synonymous codons[70], suggesting that synonymous mutations influence translational fidelity. The same is true in *Drosophila melanogaster*[71].

Co-translational protein folding and subcellular localization

Translation kinetics such as pauses at particular regions of a nascent peptide can influence co-translational protein folding[11,72–74]. For instance, synonymous mutations in the crystallin gamma B gene influence translation kinetics and cause protein misfolding and accelerated degradation[75]. Similarly, synonymous mutations affect the relative abundance of translation intermediates when the multidomain protein Suf1 was expressed in *E. coli*[76]. Consistently, translational pauses detected by Disome-seq (Box 1) are enriched in interdomain and loop regions of the protein structure[77,78]. Such pauses presumably provide extra time for co-translational folding of the upstream domain without the interference from downstream residues. Translational pauses downstream of the signal peptide can regulate protein subcellular localization by mediating the recruitment of signal recognition particles that assist in protein translocation across membranes[77,79], and subcellular localization-altering synonymous mutations are known[80].

Box 1 | Disome-seq for probing ribosome queueing

Ribosome sequencing (Ribo-seq) identifies ribosome-protected mRNA fragments and allows estimating relative translational elongation speeds of different regions of an mRNA[49] (see the figure). However, this method does not directly probe ribosome queueing, which is often invoked in translation models (for example, in the 'ramp' hypothesis[54]). Ribosome queueing refers to the phenomenon in which the elongation speed of the 5′-trailing ribosome is reduced because the 5′-trailing ribosome collides with the 3′-leading and paused ribosome. By contrast, disome sequencing (Disome-seq) sequences mRNA fragments that are resistant to RNase due to protection by two stacked ribosomes[77,78,149–151], and mapping these sequencing reads to an mRNA permits probing local ribosome queueing (see the figure).



Disome-seq analyses in yeast, zebrafish and human cells found that ribosome queueing is induced by stop codons at the ribosome A-site, prolines at the ribosome P-site and tracts of positively charged amino acids in the exit tunnel[78,149]. Explicit tests in yeast[78] and human cells[149] showed that unpreferred codons barely induce ribosome queueing, indicating that synonymous codon usage is unlikely a general mechanism for this phenomenon. Interestingly, ribosome queueing is under-represented at the beginning of coding sequences[77,78,149–151], an observation that is inconsistent with the 'ramp' hypothesis. One potential reason why unpreferred codons are decoded slowly, yet do not induce ribosome queueing, is that, in a working cycle of translational elongation, decoding is rapid relative to peptide bond formation and translocation[152]. Consistent with this explanation, when decoding histidine codons becomes a time-consuming step upon the treatment of yeast cells with an inhibitor of histidine biosynthesis, the unpreferred codon of histidine (CAT) induces more ribosome collisions than the preferred codon (CAC)[78].

We stress that, unless otherwise stated, the biochemical effect of a synonymous mutation on gene expression may or may not have a fitness effect detectable by natural selection. Below we discuss the fitness effects of synonymous mutations from two angles: genomic CUB and individual synonymous mutations.

Causes of synonymous codon usage bias

CUB refers to unequal usage of synonymous codons at the genomic scale[81], as opposed to that at specific genomic positions that might have position-specific roles such as altering transcription factor binding or co-translational protein folding[82]. Below we discuss the evolutionary forces responsible for the creation and maintenance of CUB.

Joint forces of mutation, drift and selection

The identities of preferred codons vary among species[83–85]; preferred codons tend to end with G or C in species with high GC content in intergenic regions, suggesting that the interspecific CUB variation is largely driven by mutation bias[83,84] (Fig. 2a). Some deviations in synonymous codon usage from the mutation bias-based expectation may be explained by intrinsic biochemical properties of codon–anticodon pairing[81]. Demands for translational efficiency and accuracy can drive the co-evolution of the tRNA pool and the CUB of various genes in a species[50,70]. Indeed, the generally stronger CUB in more highly expressed genes that cannot be explained by transcription-coupled mutation biases[86,87] suggests that CUB is also shaped by natural selection[8].

The role of selection in shaping CUB is well established for microorganisms, but it has been controversial for species with relatively small N_e such as vertebrates[81,86]. For example, distinct codon usage in genes across different functional categories, previously thought to optimize translational efficiency, is now believed to have arisen from the non-adaptive process of GC-biased gene conversion[88]. However, the preferential use of more accurately translated codons at evolutionarily more-conserved sites supports selection-induced CUB even in vertebrates[70].

It is often stated that genetic drift must have contributed to CUB because unpreferred codons are not completely excluded from genomes[81,89]. However, the cause of the above phenomenon may not be genetic drift. Experiments showed that exclusively using preferred codons in a highly expressed gene reduces the cellular translational

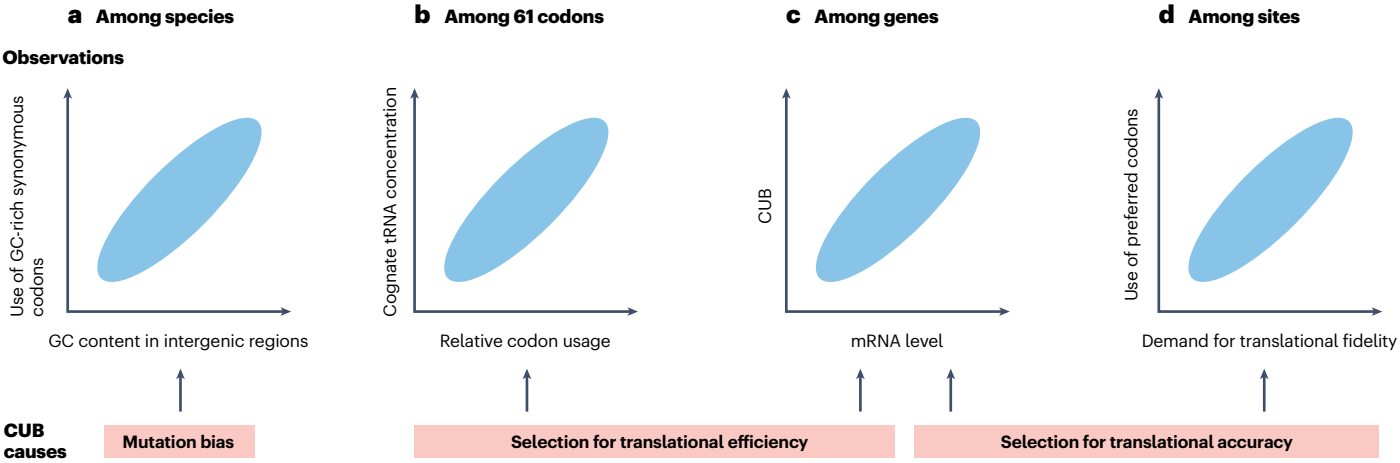


Fig. 2 | Various patterns of codon usage bias and their causes. a, Mutation bias creates a positive correlation between the genomic GC content in intergenic regions and use of GC-rich synonymous codons among species. **b**, Selection for translational efficiency creates a positive correlation between the relative use of a synonymous codon and the relative concentration of its cognate tRNA among the 61 sense codons. **c**, Selections for translational efficiency and accuracy create a positive correlation between the expression level of a gene and its extent of codon usage bias (CUB). **d**, Selection for translational accuracy creates a positive correlation between the demand for translational fidelity at a codon site and the use of a preferred codon at the site among sites of the same amino acid. Blue ellipses indicate positive correlations.

efficiency; instead, optimizing codon usage according to the relative concentrations of cognate tRNAs leads to improved cellular translational efficiency[50,90] (Fig. 2b). Furthermore, unpreferred codons may be used for various reasons, such as maintaining or avoiding certain *cis*-regulatory elements functioning in transcription or pre-RNA processing (Fig. 1). Nevertheless, genetic drift is still likely a player in shaping CUB because some synonymous mutations are presumably effectively neutral. Hence, CUB is jointly affected by mutation, drift and selection[81,89].

Selective agents of CUB

Evidence supports the existence of at least two selective agents of CUB. First, preferred codons are selectively favoured because they speed up translational elongation and improve the overall efficiency of ribosomes[50,52,53,71,89–91]. Second, preferred codons are favoured because they lower the probability of mistranslation and thereby reduce the production of nonfunctional or toxic proteins[70,71,92,93]. In both cases, the selection intensifies with the gene expression level, leading to more preferred codons in more highly expressed genes (Fig. 2c). Additionally, selection for translational accuracy explains the observation in the vast majority of >1,000 taxa surveyed from all domains of life that, compared with unpreferred codons, preferred codons more frequently occupy positions demanding high translational accuracy[70,93–95] (Fig. 2d).

That using unpreferred codons causes mRNA degradation and thereby lowers the mRNA levels[59–63] does not fully explain the positive correlation between the percentage of preferred codons and expression level across genes[62]. Specifically, between yeast genes with <39% preferred codons and those with >70% preferred codons, the mRNA level differs by ~100 times, but the mRNA degradation rate differs only fourfold[61], implying that variation in the mRNA degradation rate explains only a small fraction of the mRNA level variation among genes. Therefore, the stronger CUB of more highly expressed genes is primarily a result of natural selection.

Trade-offs among multiple effects of a synonymous mutation

Selection for translational efficiency and accuracy results in related but nonidentical outcomes of CUB, because the elongation speed of a codon is determined by the relative abundance of its cognate tRNA among all tRNAs in a cell[52,53], whereas the translational accuracy of a codon is determined by the relative abundance of its cognate tRNA among its cognate and near-cognate tRNAs[70]. Furthermore, for a given codon, a trade-off exists between the elongation speed and translational accuracy[33,69,96], probably because increasing the elongation speed reduces the time for kinetic proofreading, and vice versa. Nevertheless, this rate-accuracy conflict can be partially resolved by modulating the elongation speed of individual codons via downstream mRNA secondary structures[33]. More broadly, the pleiotropic effects of synonymous mutations on various steps of gene expression mean that simultaneous optimizations of multiple features may be difficult and that the evolutionary fate of a synonymous mutation depends on the sum of its multifaceted effects.

Fitness effects of synonymous mutations

The fact that synonymous mutations can affect nearly every step in gene expression suggests the possibility that they have fitness effects. Below we review population genetic inferences as well as experimental measures of the fitness effects of synonymous mutations.

Population genetic inferences

By comparing synonymous polymorphism and divergence, Akashi estimated in *Drosophila simulans* that synonymous mutations from preferred to unpreferred codons have an average selection coefficient (s) between $-3.6/N_e$ and $-1.3/N_e$, suggesting that synonymous mutations are under weak selection[97]. The distribution of fitness effects (DFE) of synonymous mutations can be inferred by comparing the site frequency spectrum between synonymous mutations and control neutral mutations. In *D. melanogaster*, s was estimated to be about $-70/N_e$ for 23% of synonymous mutations at fourfold degenerate sites[98]. In humans[99],

~30% of synonymous mutations were estimated to be under weak-to-strong negative selection, with $s<-1/N_e$. Additionally, compared with intronic single-nucleotide polymorphisms (SNPs), human synonymous SNPs predicted to increase and decrease the relevant mRNA's half-life show site frequency spectrum-based signatures of positive and negative selection, respectively[100]. Nevertheless, the estimations discussed here typically relied on simplifying assumptions, including those about the demography and/or neutrality of control sites that are often violated or difficult to verify. Furthermore, population genetic estimates may not be directly compared with experimental estimates of fitness effects, described below, because population genetic estimates represent an average effect of many synonymous mutations over a long evolutionary time in a changing natural environment, whereas experimental estimates are typically obtained for individual synonymous mutations in a specific, laboratory environment. Although a population genetic estimate of *s* can in principle be obtained from any species, experimental measurement of *s* has thus far been obtained almost exclusively in microorganisms, because creating many mutants and assaying their fitness remains daunting for multicellular organisms.

Experimental estimates from case studies

Natural selection can sense a fitness effect with a magnitude exceeding $1/N_e$. However, even the most powerful experiment today cannot reliably detect $|s|<0.001$, and many studies have sensitivities on the order of 0.01. Because N_e is typically $>10^7$ for the microbes used in experimental fitness quantification, experiments are orders of magnitude less sensitive than nature in detecting fitness differentials. In other words, laboratory measures are expected to underestimate the fraction of mutations that are non-neutral to varying degrees, depending on the experiment. For this reason, fractions of non-neutral mutations reported by different studies are usually not directly comparable. For the same reason, it is often useful to compare the DFEs of synonymous and nonsynonymous mutations in the same study, because the experimental sensitivity is the same for the two types of mutations and because most nonsynonymous mutations are known to be strongly deleterious[101,102].

A number of case studies, each focusing on a set of non-randomly-chosen mutations, reported substantial fitness effects of synonymous mutations[103–108]. For example, Agashe et al. made dozens of synonymous mutations in the *fae* gene, which encodes the highly expressed formaldehyde-activating enzyme FAE, of the bacterium *Methylobacterium extorquens*[103]. Interestingly, using only unpreferred synonymous codons in the gene lowered the fitness by 40%, whereas using only preferred codons reduced the fitness by 90%. Both the mRNA level and protein level per mRNA molecule were lower in these mutants than in the wild type, and increasing *fae* expression partially recovered the fitness loss of the mutants. Note that this study estimated the combined effects of multiple synonymous mutations rather than their individual effects.

Experimental estimates from systematic studies

Systematic studies, often comparing the DFEs of synonymous and nonsynonymous mutations[18,109–114], have addressed the generalizability of observations made in case studies[103–108] that some synonymous mutations are strongly non-neutral. Many systematic studies focused on viral genomes, and we estimated from five such studies that, on average, 43% of synonymous mutations are significantly non-neutral[115–119]. One of the first systematic studies in cellular organisms constructed 38 synonymous and 88 nonsynonymous mutants of *Salmonella typhimurium* by engineering individual random point mutations in two ribosomal protein genes[109]. By having each mutant compete with the wild type in a poor medium, the authors measured *s* with a high precision of 0.003. They reported that *s* ranged from 0 to −0.0279 for synonymous mutations and from 0 to −0.0763 for nonsynonymous mutations (Fig. 3a). About 95% of synonymous and 94% of nonsynonymous mutations were significantly non-neutral. The DFEs of synonymous and nonsynonymous mutations differ significantly, with a longer left tail for nonsynonymous than synonymous mutations (Fig. 3a). The authors did not find that mutating preferred codons synonymously to be significantly more deleterious than mutating unpreferred codons synonymously, but they observed a significant negative correlation between *s* and the absolute value of the synonymous mutational effect on mRNA folding strength, suggesting that altering mRNA folding strength is deleterious[109]. However, a subsequent study detected no significant fitness effects of synonymous mutations but substantial effects of nonsynonymous mutations in three genes of *S. typhimurium* required for rapid growth on L-arabinose[113].

Sane et al. measured the fitness effects of 17 synonymous and 39 nonsynonymous random mutations in *E. coli* in 16 different environments[114]. They observed that the fraction of mutations with $|s|>0.1$ in an environment was on average 36% for synonymous and 45% for nonsynonymous mutations, although the mean *s* of a mutation across environments was not significantly different between synonymous and nonsynonymous mutations. Similar patterns were observed when the analysis was expanded to 138 synonymous and 382 nonsynonymous random mutations[120].

Using CRISPR editing, Shen et al. created 8,341 synonymous, nonsynonymous or nonsense mutants of 21 yeast genes with diverse functions and expression levels[18]. Because their experimental procedure prohibited the study of essential genes, the authors focused on nonessential genes with relatively large fitness defects upon deletion. They reported a mean *s* of −0.012 for synonymous mutations and of −0.015 for nonsynonymous mutations. About 77.2% of synonymous mutations and 77.4% of nonsynonymous mutations were significantly non-neutral. Although the DFEs look overall similar between synonymous and nonsynonymous mutations (Fig. 3b), they are statistically significantly different; mutations with $s<-0.05$ are overwhelmingly nonsynonymous. In both Fig. 3a and Fig. 3b only one mode is apparent in each DFE, and the nonsynonymous DFEs exhibit a longer left tail relative to the synonymous DFEs. In the yeast study, Shen et al. discovered that the mRNA level of the gene mutated in a mutant frequently deviates from the wild-type level, with 25% of mutants deviating by at least 20%. When the mutant mRNA level is below the wild-type level, a positive correlation was observed between the mutant mRNA level and mutant fitness, suggesting that the fitness effects of both synonymous and nonsynonymous mutations in this study were attributable at least in part to a reduction in the mRNA level. Furthermore, the mutant mRNA level correlated positively with the abundance of preferred codons of the mutated gene, suggesting a role for translation-dependent mRNA degradation in determining the fitness effects of the coding mutations studied. In estimating mutant fitness, Shen et al. used a wild-type control that was constructed separately from the mutants, which raised doubts about their results[121]. However, new experiments in which the control and mutants were constructed simultaneously in the same test tube confirmed the original finding[122].

Widespread non-neutrality is also evident for 'immunizing' synonymous mutations placed in or near the protospacer adjacent motif to avoid re-cleavage of the genome after CRISPR editing. Specifically, Yang et al. monitored the growth of about 50,000 CRISPR-edited

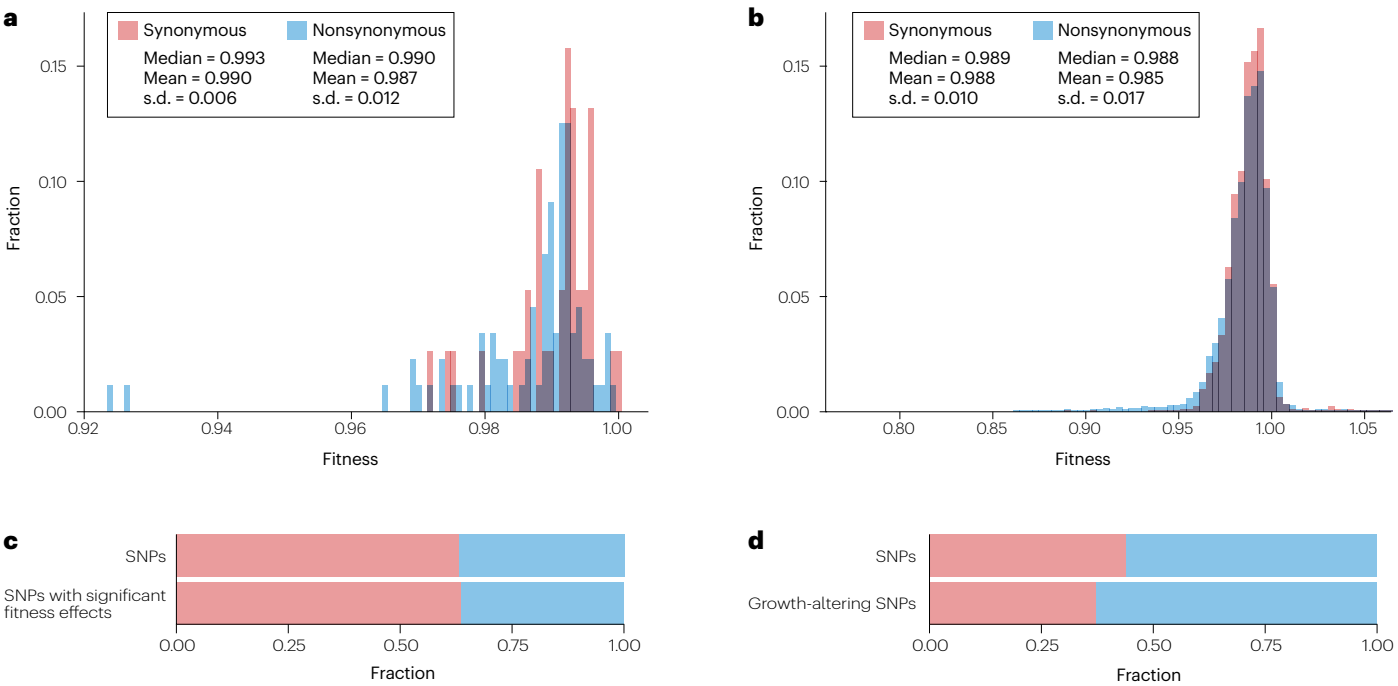


Fig. 3 | Fitness effects of synonymous and nonsynonymous mutations or polymorphisms in four systematic studies. a, Overlaid *Salmonella typhimurium* fitness distributions of 38 synonymous and 88 nonsynonymous mutants of two ribosomal protein genes. **b**, Overlaid *Saccharomyces cerevisiae* fitness distributions of 1,866 synonymous and 6,306 nonsynonymous mutants of 21 genes. **c**, Relative fractions of synonymous and nonsynonymous single-nucleotide polymorphisms (SNPs) considered between *S. cerevisiae* strains BY and RM (top bar), and relative fractions of synonymous and nonsynonymous SNPs with a significantly fitter RM allele than BY allele (bottom bar). **d**, Relative fractions of synonymous and nonsynonymous SNPs considered between *S. cerevisiae* strains YJM975 and RM (top bar), and relative fractions of synonymous and nonsynonymous SNPs with significant effects on the growth rate differences between the two strains (bottom bar). In parts **a** and **b**, the fitness distributions of synonymous and nonsynonymous mutants are significantly different ($P = 0.03$ and 1.3×10^{-6}, respectively) according to the Kolmogorov–Smirnov test. The relative fractions of synonymous and nonsynonymous SNPs are not significantly different between the top and bottom bars ($P = 0.85$) in part **c**, but they are significantly different ($P = 0.03$) in part **d**, according to a χ-squared test. Part **a** adapted with permission from ref. 109, AAAS. Part **b** adapted from ref. 18, Springer Nature.

E. coli strains, each with a target synonymous mutation in one of 30 genes along with one to five immunizing synonymous mutations[123]. They reported that 9% of mutants, each with more than one synonymous mutation, were significantly different from the control in growth rate, and this value ranged from 1% to 26% across the 30 genes. The authors noted that the above percentages are likely conservative estimates, because the control was not the wild type but mutants of a lowly expressed gene similarly created by CRISPR editing. By using more replicates, which led to reduced fitness measurement errors, the authors additionally estimated that 22% of nearly 1,700 sets of immunizing edits have significant growth effects. These findings require consideration of the potential effects of immunizing synonymous mutations when interpreting results from CRISPR editing with such mutations.

Large fitness effects of synonymous mutations are also evident in synthetic biology. Nyerges et al. attempted to synthesize an *E. coli* genome that uses only 57 of the 64 codons[124]. They encountered numerous instances of strongly deleterious synonymous codon swaps and discovered that replacements of (multiple) synonymous codons cause frequent changes in gene expression and the generation of new promoters and antisense RNAs.

The above studies measured the fitness effects of de novo mutations, which likely differ in DFE from polymorphisms, because polymorphisms are affected by selection in addition to mutation. For example, compared with neutral mutations, deleterious mutations are selectively purged so have reduced probabilities to appear as polymorphisms. To our knowledge, two studies have systematically analysed the fitness effects of natural polymorphisms in yeast. Sharon et al. constructed *Saccharomyces cerevisiae* mutants by individually replacing the allele in a laboratory strain (BY) with that in a vineyard strain (RM) at 16,006 sites[125]. They reported that 572 of these polymorphisms had significant positive fitness effects in a glucose medium and that the probability to show up in this subset is similar between synonymous and nonsynonymous polymorphisms (Fig. 3c). Synonymous polymorphisms with significant positive fitness effects were enriched in genes with strong CUB. Because the ancestral relationship between BY and RM alleles is unknown, whether the polymorphism-generating mutation is beneficial or deleterious is undetermined here. Complementary to the reverse genetic approach used above, She and Jarosz used a forward genetic approach — single-nucleotide resolution quantitative trait locus mapping — to identify 370 causal nucleotide variants responsible for growth rate (as a proxy for fitness) differences between a vineyard strain and a clinical strain of *S. cerevisiae* in 26 media[126]. The probability for a synonymous SNP to affect growth was 76% of that for a nonsynonymous SNP (Fig. 3d). Furthermore, the median effect size (measured by

the fraction of growth rate variance explained) was only slightly lower for synonymous than nonsynonymous causal SNPs.

Systematic studies are not necessarily unbiased, because they might have focused on certain genes that are particularly important or highly expressed, such as ribosomal protein genes, but some systematic studies[114,120,124–126] are agnostic to such biases because they examine a random subset of mutations or polymorphisms in the genome. Note that many deep mutational scanning studies, even when they serve their original purposes well, do not aim for nor allow reliable estimation of the DFE of synonymous mutations, because they express focal genes in non-native systems (for example, using non-native promoters)[127–129], measure protein expression or activity but not fitness[128], lack appropriate controls (for example, assuming a zero median fitness effect of synonymous mutations)[129], or have relatively large measurement errors[129]. Therefore, comparing synonymous and nonsynonymous mutations from such studies[130] can be misleading[131].

Beneficial synonymous mutations

The estimated DFEs have shown that, although the vast majority of non-neutral synonymous mutations are deleterious, a small fraction are advantageous[18,110,114,120,123]. Furthermore, synonymous beneficial mutations have been observed in experimental evolution[106,108,110,132] and other studies[133,134]. For instance, when subjecting *M. extorquens* strains containing synonymous mutations in *fae* to experimental evolution, Agashe et al. observed several highly beneficial synonymous changes in the evolved populations that each increased fitness by >10%, probably by raising mRNA and protein levels[108]. In another example, Bailey et al. evolved a *Pseudomonas fluorescens* population in a glucose minimal medium and observed the accumulation of two synonymous mutations in the *gtsB* gene that increased fitness by 7% and 9%, respectively, likely by boosting the *gtsB* mRNA level[135]. A subsequent study found that many random synonymous mutations in *gtsB* have detectable benefits, likely by promoting the transcription of *gtsB* and downstream genes in the same operon[111]. Nevertheless, beneficial synonymous substitutions are much rarer than beneficial nonsynonymous substitutions in experimental evolution[110,136]. Apart from a potential reporting bias due to the common belief that synonymous mutations are neutral, clonal interference in asexual experimental evolution is likely a major reason. Specifically, due to clonal interference, only the fittest mutant among simultaneously present advantageous mutants can reach fixation. Because more nonsynonymous than synonymous mutations are expected by chance and because the fitness effects may be greater for beneficial nonsynonymous than beneficial synonymous mutations, clonal interference exacerbates the difference in fixation probability between synonymous and nonsynonymous mutations[110]. Given the suspected difference in effect size between beneficial synonymous and beneficial nonsynonymous mutations, we predict that the contribution of beneficial synonymous mutations to adaptation is larger as the population approaches the fitness optimum.

Evolutionary consequences

Because many evolutionary analyses and conclusions depend on the assumption that synonymous mutations are neutral, findings of non-neutral synonymous mutations demand a re-examination of these analyses and conclusions. Below we discuss analyses and conclusions that are potentially affected if synonymous mutations are broadly non-neutral and those that are unlikely to be affected.

Impact on the neutral theory of molecular evolution

The neutral theory holds that most interspecific nucleotide differences and intraspecific nucleotide polymorphisms are selectively neutral rather than adaptive[2]. Note that the theory does not require most mutations to be neutral. In fact, the neutral theory acknowledges the prevalence of deleterious mutations, which are selectively purged so contribute little to nucleotide divergence and polymorphism. Hence, the discovery of many detrimental synonymous mutations, analogous to the discovery of many detrimental nonsynonymous mutations, does not by itself shake the standing of the neutral theory.

Although King and Jukes used the presumable neutrality of synonymous mutations to support the neutral theory in their foundational paper[137], this is not a necessary condition for the neutral theory, as long as some synonymous mutations are neutral. That the nonsynonymous substitution rate (d_N) of a gene is typically much lower than the corresponding synonymous substitution rate (d_S) is commonly interpreted as evidence for the neutral theory, because, under the neutral assumption of synonymous mutations, $d_N<d_S$ can be explained without invoking positive selection[2]. If synonymous mutations are mostly deleterious, $d_N<d_S$ can be explained by stronger purifying selection against nonsynonymous than synonymous mutations; again, there is no need to invoke positive selection. That is, the common interpretation of the observation of $d_N<d_S$ remains valid.

Impact on parameter estimation in population genetics and molecular evolution

The estimation of many parameters in population genetics and molecular evolution relies on nucleotide changes that are selectively neutral. For example, the rate and molecular spectrum of mutation are often estimated from synonymous polymorphisms or substitutions. This strategy inevitably leads to an underestimation of the mutation rate if many synonymous mutations are detrimental. In fact, it was noticed more than 40 years ago that the mutation rate estimated from synonymous substitutions in a mammalian α-globin gene is approximately one-half that estimated from substitutions in an α-globin pseudogene[138]. Similarly, the molecular spectrum of mutation inferred from synonymous substitutions may be biased. Fortunately, with the reduced cost of DNA sequencing, mutation rate and spectrum can now be estimated by genome sequencing of parents and offspring, or by mutation accumulation followed by genome sequencing, in which repeated bottlenecks render the impact of natural selection minimal[139].

N_e is a fundamental population genetic parameter pertinent to many evolutionary processes and theories. It is also key to conservation biology because it summarizes the demographic history of a wild population, predicts the risk of inbreeding and assesses the effectiveness of human-aided genetic management in the future. The N_e of a diploid population is typically estimated by dividing its neutral genetic diversity (π) by 4μ, in which μ is the neutral mutation rate per site per generation. If μ is reliably estimated, for example from mutation accumulation or genome sequencing of parents and offspring, whereas π is estimated from synonymous polymorphisms, as is often the case, N_e will be underestimated when many synonymous mutations are detrimental. N_e estimated using the coalescent theory may share the same problem if the estimation relies on the neutral assumption of synonymous mutations. The extent of N_e underestimation depends on the DFE of synonymous mutations and is expected to be larger when the actual N_e is larger, because the selection on non-neutral synonymous mutations intensifies with N_e.

Evolutionary events such as speciation and gene or genome duplication are sometimes dated by dividing the number of synonymous substitutions per site accumulated in evolution by the synonymous mutation rate, under the assumption that synonymous mutations are neutral. When the mutation rate is reliably estimated, evolutionary time would be underestimated if synonymous mutations are on average deleterious.

Impact on selection tests

A variety of statistical tests are commonly used to detect positive or negative selection from DNA sequences. There are generally three classes of such tests, depending on whether the data used are intraspecific polymorphisms, interspecific divergences or both. Tests using intraspecific polymorphisms include those based on the site frequency spectrum, haplotype structure or population differentiation. Because these tests do not typically distinguish between synonymous and nonsynonymous changes, they are not affected by the non-neutrality of synonymous mutations. This said, significant results from these tests do not always implicate selection because they could also be caused by demographic changes. Tests based on interspecific divergences typically compare d_S and d_N, so they can be influenced if synonymous mutations are non-neutral (Box 2).

Two commonly used selection tests are based on both intraspecific polymorphisms and interspecific divergences: the Hudson–Kreitman–Aguadé (HKA) test[140] and the McDonald–Kreitman test[141]. The HKA test does not generally assume the neutrality of synonymous mutations, so its results are not affected if the said mutations are non-neutral. The McDonald–Kreitman test compares the number of polymorphisms with that of divergences at two types of sites — synonymous and nonsynonymous — and detects deviation from independence in a 2×2 contingency table[141]. Positive selection promoting the fixation of nonsynonymous mutations is inferred when the ratio of the number of nonsynonymous changes to that of synonymous changes is significantly greater between species than within species, under the neutral assumption of synonymous mutations. However, if synonymous mutations are non-neutral, the interpretation of the McDonald–Kreitman test can be complicated. For instance, selection against detrimental synonymous mutations may hinder synonymous divergence more than polymorphism[97], which could lead to a spurious inference of positive selection.

Box 2 | Testing natural selection by d_N/d_S

The most common selection test in molecular evolution consists of comparing d_N/d_S, in which d_N is the nonsynonymous substitution rate and d_S is the synonymous substitution rate, with 1, under the assumption that synonymous mutations are neutral. Positive selection acting on nonsynonymous mutations is inferred when d_N/d_S significantly exceeds 1, whereas negative selection is inferred when d_N/d_S is significantly smaller than 1. Are the above inferences still valid if many synonymous mutations are detrimental? One could argue that, because in addition to altering the protein sequence, nonsynonymous mutations can also affect the expression of genetic information just as synonymous mutations do, the d_N/d_S-based selection test remains valid with the qualification that the potential selection detected is related only to the protein sequence. Supporting the above view, synonymous and nonsynonymous mutations in yeast have similar effects on the mRNA level of the gene mutated[18]. However, synonymous and nonsynonymous mutations may not have equal effects on all aspects of gene expression. For example, they affect mRNA folding strength by different amounts, because the GC content, and thereby the mutation spectrum, differ between synonymous and nonsynonymous sites[38,39]. More studies are needed to investigate under what conditions the d_N/d_S-based selection test is valid and whether it is possible to exclude non-neutral synonymous sites in computing d_N/d_S (ref. 153).

Additionally, d_N/d_S is often compared across genes in a genome[154]. In this case, because codon usage usually varies with the gene expression level, the synonymous mutation spectrum, and thereby the distribution of fitness effects, are expected to vary among genes of different expression levels. Future studies are needed to evaluate to what extent a comparison in d_N/d_S informs the variation in the intensity of selection on protein sequences across genes. Furthermore, d_N/d_S may be computed for a set of genes or all genes in a genome and then compared among conspecifics or different species[155]. This comparison should be valid, provided that the species being compared are sufficiently closely related such that their synonymous mutation spectra and distributions of fitness effects are similar.

Synonymous mutations and disease

That synonymous mutations have functional or fitness effects implies that they can cause disease, although we caution that the majority of such effects have been discovered in non-human studies. Furthermore, causing disease and lowering fitness are not the same thing, because disease is not defined according to fitness. For example, a human mutation that reduces fertility by 1% would be strongly deleterious and selectively purged, but it would have no clinical relevance. Conversely, cancer-causing mutations are clearly pathogenic, but they may have little fitness effects if their pathogenicity occurs only after the reproductive age.

The number of nonsynonymous mutations known to be associated with disease far exceeds that of synonymous mutations, likely because an average nonsynonymous mutation is more likely than an average synonymous mutation to cause disease, nonsynonymous mutations outnumber synonymous mutations and nonsynonymous mutations are considered in disease studies more often than synonymous mutations. When last reviewed in 2011, synonymous mutations had been associated with nearly 50 human diseases that affect most organ systems[142] (see Box 3 for examples and mechanisms of pathogenesis). Genome-wide association studies conducted in the past decade have drastically improved the discovery of genetic variants associated with disease. How different are synonymous and nonsynonymous SNPs in their probabilities of association with disease? Chen et al. surveyed across 21,429 disease–SNP associations curated from 2,113 publications, finding that the percentage of disease-associated synonymous SNPs is only slightly lower than that of nonsynonymous SNPs[143]; for comparison, that of nonsense SNPs is twice as high, whereas that of SNPs in 3′ untranslated regions is about one-half as high. It is almost certain that some of the disease-associated SNPs are not causal but are in linkage disequilibrium with causal variants, but this problem is unlikely to be more severe for synonymous than nonsynonymous SNPs, because significant associations with nonsynonymous SNPs are probably preferentially reported compared with those with synonymous SNPs. Furthermore, the average effect size is virtually identical

between synonymous and nonsynonymous disease-associated SNPs[143]. Karczewski et al. used exome data to perform genome-wide association studies for 4,529 diseases and traits in nearly 400,000 participants of the UK Biobank[144]. They found the fraction of SNPs exhibiting at least one significant association to be the highest for loss-of-function SNPs, intermediate for nonsynonymous SNPs and lowest for synonymous SNPs. Of note, the difference in the fraction is small between synonymous and nonsynonymous SNPs and is clearly smaller than that between nonsynonymous and loss-of-function SNPs.

An exome study reported that de novo synonymous mutations disrupting exonic splicing regulatory sequences and those within DNase I hypersensitivity sites (potentially influencing transcription factor binding) in frontal cortex tissues are significantly enriched in patients with autism spectrum disorders and schizophrenia, respectively[145]. The estimated contribution of these synonymous mutations to disease liability is comparable with that of de novo protein-truncating mutations and much greater than that of de novo nonsynonymous mutations[145].

Additionally, somatic synonymous mutations can drive tumorigenesis. For example, from cancer genomic data, Supek et al. discovered an excess of synonymous mutations in oncogenes[146]. They estimated that about 20% of synonymous mutations in all known oncogenes in cancer genome projects have been positively selected (for example, cancer-causing drivers) and approximately half of synonymous drivers alter mRNA splicing. A more recent pan-cancer analysis of synonymous mutations found patterns of synonymous mutations indicative of their roles in tumorigenesis such as their enrichment in known cancer genes and repeatability in cancer genomes that resemble nonsynonymous mutations[147]. Synonymous mutations in cancer are catalogued in SynMICdb[147].

Conclusions and future studies

Synonymous mutations are not all silent; case studies and systematic analyses have shown that not only are many synonymous mutations strongly non-neutral ($|s| \gg 1/N_e$), but they also have experimentally measurable s ($|s| > 0.001$). If synonymous and nonsynonymous mutations are similar in their impact on various steps of gene expression (Box 2), and if the only difference between synonymous and nonsynonymous mutations is that the latter alter the protein sequence, the finding from many systematic analyses that nonsynonymous mutations are overall only moderately more deleterious than synonymous mutations suggests that altering the protein sequence may not be the main contributor to many nonsynonymous mutations' fitness effects. However, the species so far used in the experimental estimation of DFE are limited largely to microorganisms. Because natural selection intensifies with N_e, wild types should be better optimized in species with larger N_e. Consequently, the fraction of synonymous mutations that are deleterious is expected to rise with N_e. Nonetheless, for deleterious mutations with $|s|$ greater than a threshold far exceeding the inverse of N_e of any species under comparison, their fraction should be approximately constant across these species if N_e is the only relevant variable. However, N_e is rarely the only relevant variable in interspecific comparisons. For example, the importance of differential gene expression among tissues in multicellular organisms may mean different selections on synonymous mutations between unicellular and multicellular organisms beyond those predicted solely by their difference in N_e.

The moderate difference in fitness effects between synonymous and nonsynonymous mutations experimentally discovered seems inconsistent with the much lower nucleotide diversity and substitution rate at nonsynonymous than synonymous sites (for example, d_N/d_S is on the order of 0.1 in many species). However, we must emphasize that the experimentally estimated DFEs of synonymous and nonsynonymous mutations (Fig. 3a,b) cannot reliably predict d_N/d_S, because, under the neutral theory, the critical part of DFEs for predicting d_N/d_S is when $|s|$ is on the order of $1/N_e$ or smaller, which is unfortunately invisible to any experimental fitness measurement[18]. It remains possible, for example, that the fraction of neutral and nearly neutral mutations among nonsynonymous mutations is ~10% of that among synonymous mutations, which will result in a d_N/d_S of ~0.1. Nonetheless, under the above scenario, the DFEs of synonymous and nonsynonymous mutations would differ greatly in neutral and nearly neutral regions, but only moderately otherwise, a possibility that has not been previously considered. Furthermore, a comparison between human and chimpanzee genome sequences revealed that the substitution rate at synonymous sites divided by that in (presumably neutral) intergenic regions is 0.635 for autosomes and 0.547 for the X chromosome[148], suggesting that the fraction of synonymous mutations that are deleterious is most likely below 50% in great apes, contrasting the estimate of at least 75% in yeast from lab-based DFE quantification[18], although these fractions may not be directly comparable, as cautioned earlier. Another possibility, which has received some empirical support, is that the fitness effects of synonymous mutations

Box 3 | Examples of pathogenic synonymous mutations

The most common cause of cystic fibrosis is a three-nucleotide (CTT) deletion in the *CFTR* gene, which encodes the cystic fibrosis transmembrane conductance regulator. The deletion consists of the third position of an Ile codon (ATC) and the first two positions of the adjacent Phe codon (TTT). The result of this deletion is the loss of the Phe residue and a synonymous conversion of ATC (Ile) to ATT (Ile), which change the mRNA secondary structure and translation dynamics[156], leading to co-translational misfolding and degradation of CFTR. It had been assumed that the loss of Phe is the cause of the disease until the discovery that the true culprit is the synonymous change. Specifically, converting ATT back to ATC, without reinstating the missing Phe, largely restores the CFTR level and function[156,157].

Congenital myasthenic syndromes are a group of disorders characterized by muscle weakness, and they are caused by defects in the transmission of signals from nerve cells to muscles. Congenital myasthenic syndromes often result from mutations in genes encoding the nicotinic acetylcholine receptor subunits, such as *CHRNE*, which encodes the ε-subunit. A patient was found to carry a homozygous synonymous mutation at the third position of a Gly codon in *CHRNE*. This mutation creates a new splice donor site located four nucleotides upstream of the normal site, leading to a deletion and generating a frameshift in exon 9 followed by a premature termination codon[158].

Crohn's disease is a type of inflammatory bowel disease. A CTG>TTG (Leu) synonymous mutation in the autophagy gene *IRGM* is associated with the disease[159]. This mutation abolishes the downregulation of *IRGM* by a family of microRNAs[160], which subsequently compromises the autophagy-mediated control of intracellular replication of Crohn's disease-associated adherent-invasive *Escherichia coli*[160].

Glossary

A-to-I editing
Enzymatic alteration of RNA molecules consisting of the conversion of adenosines (A) to inosines (I) at specific positions.

Clonal interference
Competition among genotypes with different beneficial mutations in an asexual population.

Coalescent theory
A mathematical theory of population genetics that traces all alleles of a gene sampled from a population to a single ancestral copy.

Codon usage bias
(CUB). The phenomenon in which synonymous codons of an amino acid are unequally used in a genome.

Cycloheximide
A fungicide often used to block eukaryotic translational elongation in experiments.

Deep mutational scanning
An experimental approach for measuring the effects of individual nucleotides in a DNA segment by creating many mutants of the DNA followed by high-throughput functional/fitness assays of the mutants.

Effective population size
(N_e). Number of individuals in an ideal (that is, Wright–Fisher) population that results in the same amount of genetic drift as in the actual population considered.

Fitness
A quantitative representation of the ability of an individual to pass its genome to the next generation.

Genome-wide association studies
Investigations of genome-wide sets of genetic variants in groups of individuals to find variants associated with traits of interest.

Kinetic proofreading
Mechanism that allows enzymes, particularly those involved in DNA replication, RNA transcription and protein synthesis, to enhance their fidelity by discriminating between correct and incorrect substrates. The accuracy of this process is higher than expected solely based on the difference in activation energy between forming correct products and incorrect products.

Linkage disequilibrium
Non-random association of alleles of different loci in a population.

m^6A methylation
Methylation at the nitrogen-6 position of adenosines at specific positions in an RNA molecule.

MicroRNA
Single-stranded, non-coding RNA molecules of 21 to 23 nucleotides that bind to mRNAs to cause mRNA degradation or suppress mRNA translation.

Modern synthesis
Prevailing evolutionary theory developed in the 1930s to the 1940s by combining Darwin's theory of evolution by natural selection with a population-oriented view of Mendelian genetics.

mRNA folding strength
Reduction in free energy of a folded mRNA molecule relative to its unfolded form.

Mutation accumulation
A genetic experiment in which a population of organisms is propagated through repeated population bottlenecks to minimize the impact of natural selection.

Mutation bias
A phenomenon in which some types of mutation occur more frequently than other types.

Near-cognate tRNAs
tRNAs carrying an amino acid that differs from that encoded by the codon of concern and whose anticodon does not pair with the codon at exactly one nucleotide position.

Neutral mutations
Genetic changes having selection coefficients that are smaller than the inverse of the effective population size.

Neutral theory
An evolutionary theory positing that most interspecific differences and intraspecific polymorphisms at the DNA or protein sequence level are selectively neutral rather than adaptive.

Nonsynonymous mutations
Point mutations in a coding sequence that alter the encoded protein sequence.

Nonsynonymous substitution rate
(d_N). Number of nonsynonymous nucleotide substitutions per nonsynonymous site between two homologous coding sequences.

Nucleosome
The basic structural unit of DNA packaging in eukaryotes consisting of a segment of DNA wound around eight histone proteins.

Polyadenylation
Part of the process of mRNA maturation in which a nascent RNA transcript is cleaved at a particular site and subsequently becomes the object of the addition of a poly-A tail.

Preferred codons
Codons that are used more frequently than the average of all codons of the same amino acid in highly expressed genes of a genome.

Protospacer adjacent motif
A short DNA sequence of usually 2–6 nucleotides required for a Cas nuclease in the CRISPR system to cut; it is generally found 3–4 nucleotides downstream from the cut site.

Pseudogene
A nonfunctional segment of DNA that is derived from a previously functional gene.

Quantitative trait locus mapping
A method to determine the chromosomal regions or genetic variants affecting the variation of a quantitative trait among individuals of a species.

Ribosome E-site
5′-most of the three tRNA binding sites in a ribosome that allows a deacylated tRNA to exit. The 3′-most of the three tRNA binding sites in a ribosome that selects charged tRNA molecules during protein synthesis is known as A-site.

Ribosome P-site
Middle of the three tRNA binding sites in a ribosome that holds the tRNA linked to the nascent polypeptide chain.

Selection coefficient
(s). The difference in relative fitness between a mutant and the wild type.

Shine–Dalgarno sequence
Sequence motif in mRNA that recruits ribosomes through interaction with the anti-Shine–Dalgarno sequence in 16S ribosomal RNA of prokaryotes.

Site frequency spectrum
Distribution of the allele frequencies of a set of loci such as single-nucleotide polymorphisms in a population or sample of individuals.

Splicing
A step in pre-mRNA processing that removes introns and joins exons to form mature mRNAs.

Synonymous mutations
Point mutations in a coding sequence that do not alter the encoded protein sequence.

Synonymous substitution rate
(d_S). Number of synonymous nucleotide substitutions per synonymous site between two homologous coding sequences.

Unpreferred codons
Codons that are used less frequently than the average of all codons of the same amino acid in highly expressed genes of a genome.

Box 4 | Computational prediction of pathogenetic synonymous mutations

Compared with the abundance of software for predicting nonsynonymous mutations that are pathogenic[161], the field for computational prediction of pathogenic synonymous mutations is relatively young[162,163]. Two groups of methods — one specifically for synonymous mutations and the other for any type of point mutations — exist. The first group comprises DDIG-SN[164], regSNPs-splicing[165], seDSM[166], SilVA[167], Syntool[168] and TraP[169], whereas the second group includes, for example, CADD[170], DANN[171], FATHMM-MKL[172], PhD-SNP^{g} (ref. 173) and PredictSNP2 (ref. 174). Almost all these predictions use the information of sequence conservation, although different methods also consider different additional features such as pre-mRNA splicing, mRNA folding and codon-usage bias. Comparison of several methods identified three top performers (TraP, SilVA and FATHMM-MKL) and led to the development of an ensemble method named PrDSM that combines these methods[162]. Although PrDSM performs better than its three constituents individually, its overall performance is still not very high (area under curve=0.786)[162]. One bottleneck in the development of better computational predictions is the set of synonymous mutations with accurate annotations of effects on fitness or disease (that is, ground truth). With increasing awareness of larger-than-expected influences of synonymous mutations on fitness and pathogenesis, we expect that more efforts will be dedicated to generating the ground truth and developing better computational predictions.

are less environment-dependent than nonsynonymous mutations; as a result, a higher fraction of synonymous than nonsynonymous mutations are neutral across environments and thereby can reach fixation[18]. Because the natural environment frequently changes, DFE differences between synonymous and nonsynonymous mutations shown in a constant laboratory environment may not recapitulate their difference in nature. In the same vein, to what extent s inferred under simple population genetic models reflects the truth requires further studies.

Given the many mechanisms by which synonymous mutations can influence gene expression, which mechanisms are the main causes of their fitness effects? If one measures the fitness effects and various phenotypic effects of many synonymous mutations, the fitness effect can be modelled as a mathematical function of various phenotypic effects, allowing estimating the relative contributions of different phenotypic effects to the fitness effect, subject to the caveat that measures of different phenotypic effects have different precisions. Note, however, that the potential findings from a set of synonymous mutations may have little relevance to individual mutations due to the large variation among mutations in their mechanisms of fitness effects.

As mentioned, selection tests that rely on comparisons between numbers of synonymous and nonsynonymous polymorphisms and/or divergences may be invalid or inaccurate if a substantial fraction of synonymous mutations are non-neutral. There is an urgent need to understand the conditions under which such tests are invalid and the type of erroneous inferences that may result. More broadly, reliable methods are needed for identifying neutrally evolving sites for selection tests and other analyses such as N_e estimation.

Finally, synonymous mutations should be considered in studying disease mechanisms, and more efforts are needed for computational predictions of pathogenic synonymous mutations (Box 4). Given the rapid growth in the type, scale and precision of genomic experiments and data, we can expect that the next decade will see more studies that clarify the functional, fitness and pathogenic effects of synonymous mutations.

Published online: 20 May 2025

References

1. Agashe, D. in *Single Nucleotide Polymorphisms* (eds Sauna, Z. E. & Kimchi-Sarfaty, C.) 15–36 (Springer, 2022).
2. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
3. Zhang, J. in *Evolution Since Darwin: The First 150 Years* (eds Bell, M. A. et al.) 87–118 (Sinauer, 2010).
4. Fiers, W. et al. A-protein gene of bacteriophage MS2. *Nature* **256**, 273–278 (1975).
5. Air, G. M. et al. Gene *F* of bacteriophage X174. Correlation of nucleotide sequences from the DNA and amino acid sequences from the gene product. *J. Mol. Biol.* **107**, 445–458 (1976).
6. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. The primary structure of rabbit β-globin mRNA as determined from cloned DNA. *Cell* **10**, 571–585 (1977).
7. Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* **8**, r49–r62 (1980).
8. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
9. Gouy, M. & Gautier, C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* **10**, 7055–7074 (1982).
10. Hanson, G. & Coller, J. Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **19**, 20–30 (2018).
11. Liu, Y., Yang, Q. & Zhao, F. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu. Rev. Biochem.* **90**, 375–401 (2021).
12. Stergachis, A. B. et al. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367–1372 (2013).
13. Birnbaum, R. Y. et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068 (2012).
14. Doolittle, W. F., Brunet, T. D., Linquist, S. & Gregory, T. R. Distinguishing between "function" and "effect" in genome biology. *Genome Biol. Evol.* **6**, 1234–1237 (2014).
15. Agoglia, R. M. & Fraser, H. B. Disentangling sources of selection on exonic transcriptional enhancers. *Mol. Biol. Evol.* **33**, 585–590 (2016).
16. Xing, K. & He, X. Reassessing the "duon" hypothesis of protein evolution. *Mol. Biol. Evol.* **32**, 1056–1062 (2015).
17. Chen, J. et al. Prevalent uses and evolution of exonic regulatory sequences in the human genome. *Nat. Sci.* **3**, e20220058 (2023).
18. Shen, X., Song, S., Li, C. & Zhang, J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731 (2022).
19. Rodriguez, A. et al. Synonymous codon substitutions modulate transcription and translation of a divergent upstream gene by modulating antisense RNA production. *Proc. Natl Acad. Sci. USA* **121**, e2405510121 (2024).
20. Sekinger, E. A., Moqtaderi, Z. & Struhl, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**, 735–748 (2005).
21. Tillo, D. & Hughes, T. R. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinform.* **10**, 442 (2009).
22. Zhou, Z. et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl Acad. Sci. USA* **113**, E6117–E6125 (2016).
23. Zhao, F. et al. Genome-wide role of codon usage on transcription and identification of potential regulators. *Proc. Natl Acad. Sci. USA* **118**, e2022590118 (2021).
24. Qian, W. & Zhang, J. Codon usage bias and nuclear mRNA concentration: correlation vs. causation. *Proc. Natl Acad. Sci. USA* **118**, e2104714118 (2021).
25. Zhou, Z., Dang, Y., Zhou, M., Yuan, H. & Liu, Y. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *eLife* **7**, e33569 (2018).
26. Savisaar, R. & Hurst, L. D. Exonic splice regulation imposes strong selection at synonymous sites. *Genome Res.* **28**, 1442–1454 (2018).
27. Chamary, J. V., Parmley, J. L. & Hurst, L. D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98–108 (2006).
28. Lan, Y. et al. Synonymous mutations promote tumorigenesis by disrupting m^{6}A-dependent mRNA metabolism. *Cell* **188**, 1828–1841.e15 (2025).
29. Courel, M. et al. GC content shapes mRNA storage and decay in human cells. *eLife* **8**, e49708 (2019).
30. Mordstein, C. et al. Codon usage and splicing jointly influence mRNA localization. *Cell Syst.* **10**, 351–362.e8 (2020).

31. Zuckerman, B., Ron, M., Mikl, M., Segal, E. & Ulitsky, I. Gene architecture and sequence composition underpin selective dependency of nuclear export of long RNAs on NXF1 and the TREX complex. *Mol. Cell* **79**, 251–267.e6 (2020).
32. Savisaar, R. & Hurst, L. D. Both maintenance and avoidance of RNA-binding protein interactions constrain coding sequence evolution. *Mol. Biol. Evol.* **34**, 1110–1126 (2017).
33. Yang, J. R., Chen, X. & Zhang, J. Codon-by-codon modulation of translational speed and accuracy via mRNA folding. *PLoS Biol.* **12**, e1001910 (2014).
34. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
35. Pop, C. et al. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* **10**, 770 (2014).
36. Lai, W. C. et al. mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat. Commun.* **9**, 4328 (2018).
37. Gaither, J. B. S. et al. Synonymous variants that disrupt messenger RNA structure are significantly constrained in the human population. *GigaScience* **10**, giab023 (2021).
38. Park, C., Chen, X., Yang, J. R. & Zhang, J. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc. Natl Acad. Sci. USA* **110**, E678–E686 (2013).
39. Shabalina, S. A., Ogurtsov, A. Y. & Spiridonov, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34**, 2428–2437 (2006).
40. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl Acad. Sci. USA* **107**, 3645–3650 (2010).
41. Gu, W., 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).
42. Gu, W., Wang, X., Zhai, C., Xie, X. & Zhou, T. Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. *Mol. Biol. Evol.* **29**, 3037–3044 (2012).
43. Barrington, C. L. et al. Synonymous codon usage regulates translation initiation. *Cell Rep.* **42**, 113413 (2023).
44. Li, K., Kong, J., Zhang, S., Zhao, T. & Qian, W. Distance-dependent inhibition of translation initiation by downstream out-of-frame AUGs is consistent with a Brownian ratchet process of ribosome scanning. *Genome Biol.* **23**, 254 (2022).
45. Li, G. W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
46. Yang, C., Hockenberry, A. J., Jewett, M. C. & Amaral, L. A. N. Depletion of Shine-Dalgarno sequences within bacterial coding regions is expression dependent. *G3* **6**, 3467–3474 (2016).
47. Hockenberry, A. J., Jewett, M. C., Amaral, L. A. N. & Wilke, C. O. Within-gene Shine–Dalgarno sequences are not selected for function. *Mol. Biol. Evol.* **35**, 2487–2498 (2018).
48. Varenne, S., Buc, J., Lloubes, R. & Lazdunski, C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**, 549–576 (1984).
49. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
50. Qian, W., Yang, J. R., Pearson, N. M., Maclean, C. & Zhang, J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, e1002603 (2012).
51. Charneski, C. A. & Hurst, L. D. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* **11**, e1001508 (2013).
52. Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S. & Press, W. H. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.* **11**, e1005732 (2015).
53. Weinberg, D. E. et al. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* **14**, 1787–1799 (2016).
54. Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
55. Pelechano, V., Wei, W. & Steinmetz, L. M. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* **161**, 1400–1412 (2015).
56. Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Bluthgen, N. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9**, 675 (2013).
57. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479 (2013).
58. Sejour, R., Leatherwood, J., Yurovsky, A. & Futcher, B. Enrichment of rare codons at 5′ ends of genes is a spandrel caused by evolutionary sequence turnover and does not improve translation. *eLife* **12**, RP89656 (2024).
59. Bazzini, A. A. et al. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).
60. Mishima, Y. & Tomari, Y. Codon usage and 3′ UTR length determine maternal mRNA stability in zebrafish. *Mol. Cell* **61**, 874–885 (2016).
61. Presnyak, V. et al. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
62. Chen, S. et al. Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level. *Mol. Biol. Evol.* **34**, 2944–2958 (2017).
63. Boel, G. et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **529**, 358–363 (2016).
64. Radhakrishnan, A. et al. The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. *Cell* **167**, 122–132.e9 (2016).
65. Buschauer, R. et al. The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science* **368**, eaay6912 (2020).
66. Veltri, A. J. et al. Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. *eLife* **11**, e76038 (2022).
67. Zhu, X., Cruz, V. E., Zhang, H., Erzberger, J. P. & Mendell, J. T. Specific tRNAs promote mRNA decay by recruiting the CCR4-NOT complex to translating ribosomes. *Science* **386**, eadq8587 (2024).
68. Guo, X. et al. Selection and mutation on microRNA target sequences during rice evolution. *BMC Genomics* **9**, 454 (2008).
69. Mordret, E. et al. Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. *Mol. Cell* **75**, 427–441 (2019).
70. Sun, M. & Zhang, J. Preferred synonymous codons are translated more accurately: proteomic evidence, among-species variation, and mechanistic basis. *Sci. Adv.* **8**, eabl9812 (2022).
71. Wu, X., Xu, M., Yang, J. R. & Lu, J. Genome-wide impact of codon usage bias on translation optimization in *Drosophila melanogaster*. *Nat. Commun.* **15**, 8329 (2024).
72. Sharma, A. K. & O'Brien, E. P. Non-equilibrium coupling of protein structure and function to translation-elongation kinetics. *Curr. Opin. Struct. Biol.* **49**, 94–103 (2018).
73. Stein, K. C. & Frydman, J. The stop-and-go traffic regulating protein biogenesis: how translation kinetics controls proteostasis. *J. Biol. Chem.* **294**, 2076–2084 (2019).
74. Moss, M. J., Chamness, L. M. & Clark, P. L. The effects of codon usage on protein structure and folding. *Annu. Rev. Biophys.* **53**, 87–108 (2024).
75. Buhr, F. et al. Synonymous codons direct cotranslational folding toward different protein conformations. *Mol. Cell* **61**, 341–351 (2016).
76. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
77. Arpat, A. B. et al. Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *Genome Res.* **30**, 985–999 (2020).
78. Zhao, T. et al. Disome-seq reveals widespread ribosome collisions that promote cotranslational protein folding. *Genome Biol.* **22**, 16 (2021).
79. Pechmann, S., Chartron, J. W. & Frydman, J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat. Struct. Mol. Biol.* **21**, 1100–1105 (2014).
80. Fernández-Calero, T. et al. The transcriptional activities and cellular localization of the human estrogen receptor alpha are affected by the synonymous Ala87 mutation. *J. Steroid Biochem. Mol. Biol.* **143**, 99–104 (2014).
81. Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
82. Callens, M., Pradier, L., Finnegan, M., Rose, C. & Bedhomme, S. Read between the lines: diversity of nontranslational selection pressures on local codon usage. *Genome Biol. Evol.* **13**, evab097 (2021).
83. Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L. & McAdams, H. H. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl Acad. Sci. USA* **101**, 3480–3485 (2004).
84. Hershberg, R. & Petrov, D. A. General rules for optimal codon choice. *PLoS Genet.* **5**, e1000556 (2009).
85. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).
86. Duret, L. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12**, 640–649 (2002).
87. Duret, L. & Mouchiroud, D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96**, 4482–4487 (1999).
88. Pouyet, F., Mouchiroud, D., Duret, L. & Semon, M. Recombination, meiotic expression and human codon usage. *eLife* **6**, e27344 (2017).
89. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
90. Chen, F. et al. Dissimilation of synonymous codon usage bias in virus–host coevolution due to translational selection. *Nat. Ecol. Evol.* **4**, 589–600 (2020).
91. Andersson, S. G. & Kurland, C. G. Codon preferences in free-living microorganisms. *Microbiol. Rev.* **54**, 198–210 (1990).
92. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
93. Akashi, H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**, 927–935 (1994).
94. Stoletzki, N. & Eyre-Walker, A. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* **24**, 374–381 (2007).
95. Zhou, T., Weems, M. & Wilke, C. O. Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol. Biol. Evol.* **26**, 1571–1580 (2009).
96. Johansson, M., Zhang, J. & Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl Acad. Sci. USA* **109**, 131–136 (2012).
97. Akashi, H. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076 (1995).
98. Lawrie, D. S., Messer, P. W., Hershberg, R. & Petrov, D. A. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* **9**, e1003527 (2013).

99. Huang, Y. F. & Siepel, A. Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease. *Genome Res.* **29**, 1310–1321 (2019).
100. Dhindsa, R. S., Copeland, B. R., Mustoe, A. M. & Goldstein, D. B. Natural selection shapes codon usage in the human genome. *Am. J. Hum. Genet.* **107**, 83–95 (2020).
101. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
102. Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
103. Agashe, D., Martinez-Gomez, N. C., Drummond, D. A. & Marx, C. J. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol. Biol. Evol.* **30**, 549–560 (2013).
104. Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A. & Clark, P. L. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl Acad. Sci. USA* **117**, 3528–3534 (2020).
105. Frumkin, I. et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc. Natl Acad. Sci. USA* **115**, E4940–E4949 (2018).
106. Kristofich, J. et al. Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genet.* **14**, e1007615 (2018).
107. Carlini, D. B. & Stephan, W. In vivo introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. *Genetics* **163**, 239–243 (2003).
108. Agashe, D. et al. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. *Mol. Biol. Evol.* **33**, 1542–1553 (2016).
109. Lind, P. A., Berg, O. G. & Andersson, D. I. Mutational robustness of ribosomal protein genes. *Science* **330**, 825–827 (2010).
110. Bailey, S. F., Alonso Morales, L. A. & Kassen, R. Effects of synonymous mutations beyond codon bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments. *Genome Biol. Evol.* **13**, evab141 (2021).
111. Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *eLife* **8**, e45952 (2019).
112. Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* **31**, 1581–1592 (2014).
113. Lind, P. A., Arvidsson, L., Berg, O. G. & Andersson, D. I. Variation in mutational robustness between different proteins and the predictability of fitness effects. *Mol. Biol. Evol.* **34**, 408–418 (2017).
114. Sane, M., Diwan, G. D., Bhat, B. A., Wahl, L. M. & Agashe, D. Shifts in mutation spectra enhance access to beneficial mutations. *Proc. Natl Acad. Sci. USA* **120**, e2207355120 (2023).
115. Carrasco, P., de la Iglesia, F. & Elena, S. F. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol.* **81**, 12979–12984 (2007).
116. Domingo-Calap, P., Cuevas, J. M. & Sanjuan, R. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* **5**, e1000742 (2009).
117. Peris, J. B., Davis, P., Cuevas, J. M., Nebot, M. R. & Sanjuan, R. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* **185**, 603–609 (2010).
118. Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401 (2004).
119. Cuevas, J. M., Domingo-Calap, P. & Sanjuan, R. The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol.* **29**, 17–20 (2012).
120. Sane, M., Parveen, S. & Agashe, D. Mutation bias alters the distribution of fitness effects of mutations. Preprint at *bioRxiv* https://doi.org/10.1101/2024.03.24.586369 (2025).
121. Kruglyak, L. et al. Insufficient evidence for non-neutrality of synonymous mutations. *Nature* **616**, E8–E9 (2023).
122. Shen, X., Song, S., Li, C. & Zhang, J. Further evidence for strong nonneutrality of yeast synonymous mutations. *Mol. Biol. Evol.* **41**, msae224 (2024).
123. Yang, D. D., Rusch, L. M., Widney, K. A., Morgenthaler, A. B. & Copley, S. D. Synonymous edits in the *Escherichia coli* genome have substantial and condition-dependent effects on fitness. *Proc. Natl Acad. Sci. USA* **121**, e2316834121 (2024).
124. Nyerges, A. et al. Synthetic genomes unveil the effects of synonymous recoding. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.16.599206 (2024).
125. Sharon, E. et al. Functional genetic variants revealed by massively parallel precise genome editing. *Cell* **175**, 544–557.e16 (2018).
126. She, R. & Jarosz, D. F. Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. *Cell* **172**, 478–490.e15 (2018).
127. Hietpas, R. T., Jensen, J. D. & Bolon, D. N. Experimental illumination of a fitness landscape. *Proc. Natl Acad. Sci. USA* **108**, 7896–7901 (2011).
128. Amorosi, C. J. et al. Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *Am. J. Hum. Genet.* **108**, 1735–1751 (2021).
129. Weile, J. et al. A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* **13**, 957 (2017).
130. Dhindsa, R. S. et al. A minimal role for synonymous variation in human disease. *Am. J. Hum. Genet.* **109**, 2105–2109 (2022).
131. Shen, X., Song, S., Li, C. & Zhang, J. On the fitness effects and disease relevance of synonymous mutations. Preprint at *bioRxiv* https://doi.org/10.1101/2022.08.22.504687 (2022).
132. McGrath, K. M. et al. Fitness benefits of a synonymous substitution in an ancient EF-Tu gene depend on the genetic background. *J. Bacteriol.* **206**, e0032923 (2024).
133. Zwart, M. P. et al. Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1-β-lactamase. *Heredity* **121**, 406–421 (2018).
134. Faheem, M., Zhang, C. J., Morris, M. N., Pleiss, J. & Oelschlaeger, P. Role of synonymous mutations in the evolution of TEM β-lactamase genes. *Antimicrob. Agents Chemother.* **65**, e00018–e00021 (2021).
135. Bailey, S. F., Hinz, A. & Kassen, R. Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nat. Commun.* **5**, 4076 (2014).
136. Chen, P. & Zhang, J. The loci of environmental adaptation in a model eukaryote. *Nat. Commun.* **15**, 5672 (2024).
137. King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
138. Miyata, T. & Yasunaga, T. Rapidly evolving mouse alpha-globin-related pseudo gene and its evolutionary history. *Proc. Natl Acad. Sci. USA* **78**, 450–453 (1981).
139. Katju, V. & Bergthorsson, U. Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. *Genome Biol. Evol.* **11**, 136–165 (2019).
140. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
141. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
142. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683–691 (2011).
143. 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).
144. Karczewski, K. J. et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK biobank exomes. *Cell Genom.* **2**, 100168 (2022).
145. Takata, A., Ionita-Laza, I., Gogos, J. A., Xu, B. & Karayiorgou, M. De novo synonymous mutations in regulatory elements contribute to the genetic etiology of autism and schizophrenia. *Neuron* **89**, 940–947 (2016).
146. Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324–1335 (2014).
147. Sharma, Y. et al. A pan-cancer analysis of synonymous mutations. *Nat. Commun.* **10**, 2569 (2019).
148. Lu, J. & Wu, C. I. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc. Natl Acad. Sci. USA* **102**, 4063–4067 (2005).
149. Han, P. et al. Genome-wide survey of ribosome collision. *Cell Rep.* **31**, 107610 (2020).
150. Meydan, S. & Guydosh, N. R. Disome and trisome profiling reveal genome-wide targets of ribosome quality control. *Mol. Cell* **79**, 588–602.e6 (2020).
151. Guydosh, N. R. & Green, R. Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* **156**, 950–962 (2014).
152. Aitken, C. E. & Puglisi, J. D. Following the intersubunit conformation of the ribosome during translation in real time. *Nat. Struct. Mol. Biol.* **17**, 793–800 (2010).
153. Rahman, S., Kosakovsky Pond, S. L., Webb, A. & Hey, J. Weak selection on synonymous codons substantially inflates *dN*/*dS* estimates in bacteria. *Proc. Natl Acad. Sci. USA* **118**, e2023575118 (2021).
154. Zhang, J. & Yang, J. R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015).
155. Chen, P. & Zhang, J. Antagonistic pleiotropy conceals molecular adaptations in changing environments. *Nat. Ecol. Evol.* **4**, 461–469 (2020).
156. 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).
157. Lazrak, A. et al. The silent codon change I507-ATC→ATT contributes to the severity of the ΔF508 CFTR channel dysfunction. *FASEB J.* **27**, 4630–4645 (2013).
158. Richard, P. et al. A synonymous *CHRNE* mutation responsible for an aberrant splicing leading to congenital myasthenic syndrome. *Neuromuscul. Disord.* **17**, 409–414 (2007).
159. Parkes, M. et al. Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
160. 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. *Nat. Genet.* **43**, 242–245 (2011).
161. Li, J. et al. Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res.* **46**, 7793–7804 (2018).
162. Cheng, N. et al. Comparison and integration of computational methods for deleterious synonymous mutation prediction. *Brief. Bioinform.* **21**, 970–981 (2020).
163. Lin, B. C., Katneni, U., Jankowska, K. I., Meyer, D. & Kimchi-Sarfaty, C. In silico methods for predicting functional synonymous variants. *Genome Biol.* **24**, 126 (2023).
164. Livingstone, M. et al. Investigating DNA-, RNA-, and protein-based features as a means to discriminate pathogenic synonymous variants. *Hum. Mutat.* **38**, 1336–1347 (2017).
165. Zhang, X. et al. regSNPs-splicing: a tool for prioritizing synonymous single-nucleotide substitution. *Hum. Genet.* **136**, 1279–1289 (2017).
166. Wang, L. et al. Deleterious synonymous mutation identification based on selective ensemble strategy. *Brief. Bioinform.* **24**, bbac598 (2023).
167. Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N. & Brudno, M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* **29**, 1843–1850 (2013).
168. Zhang, T. et al. Syntool: a novel region-based intolerance score to single nucleotide substitution for synonymous mutations predictions based on 123,136 individuals. *BioMed Res. Int.* **2017**, 5096208 (2017).

169. Gelfman, S. et al. Annotating pathogenic non-coding variants in genic regions. *Nat. Commun.* **8**, 236 (2017).
170. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
171. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
172. Shihab, H. A. et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543 (2015).
173. Capriotti, E. & Fariselli, P. PhD-SNP^g: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* **45**, W247–W252 (2017).
174. Bendl, J. et al. PredictSNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput. Biol.* **12**, e1004962 (2016).

Acknowledgements
The authors thank S. Song, S. Zhang and J. Kong for technical assistance, and members of our laboratories for valuable comments. J.Z. was supported by the U.S. National Institutes of Health (R35GM139484) and W.Q. was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA28030402), the Biological Breeding-National Science and Technology Major Project (2023ZD04073) and Guangdong Provincial Key Laboratory of Synthetic Genomics (2023B1212060054).

Author contributions
The authors contributed equally to all aspects of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Genetics* thanks Laurent Duret, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Related links
SynMICdb: http://SynMICdb.dkfz.de