Reviewed Preprint
v1 • May 19, 2026
Not revised

Reviewed Preprint
v2 • June 4, 2026
Revised by authors

For correspondence:
ichihashi@bio.c.u-tokyo.ac.jp

Competing interests: No competing interests declared

Funding: See page 18

Reviewing editor: Rajan Sankaranarayanan, Centre for Cellular and Molecular Biology, India

Experimental verification of the error minimization theory using non-standard genetic codes constructed in vitro

Ryota Miyachi[a], Norikazu Ichihashi[a,b,c]

[a]Department of Life Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan • [b]Komaba Institute for Science, The University of Tokyo, Tokyo, Japan • [c]Research Center for Complex Systems Biology, Universal Biology Institute, The University of Tokyo, Tokyo, Japan

eLife Assessment

This **valuable** work addresses a longstanding question of how the extant genetic code came to be selected and conserved almost universally across life. Using a mutational approach and a small set of reporters, the authors demonstrate that the mutational impact was similar for non-standard genetic codes. The data provide **solid** support for the claim of having provided experimental verification of the error minimization theory.

https://doi.org/10.7554/eLife.111164.2.sa4

Abstract

All living systems use an almost identical genetic code, the standard genetic code, in which 20 amino acids are assigned to 61 codons non-randomly. According to the error minimization theory, amino acids are arranged to minimize the mutational effect on protein function, while experimental verification remains limited. In this study, we constructed 10 non-standard genetic codes in vitro by reassigning three amino acids (Ala, Ser, and Leu) in vacant codons of the minimal genetic code, which consists of 21 tRNAs. Most of these non-standard genetic codes have a higher cost of amino acid replacement than the standard genetic code, calculated based on three amino acid properties: polar requirement (PR), molecular volume (MV), and hydropathy index (HI). The protein function of three reporter genes expressed using these non-standard genetic codes decreased similarly when random mutations were introduced into the genes, implying that the effect of mutations was similar across all the non-standard genetic codes tested here. This result provides direct experimental evidence that mutational robustness does not significantly change in individual reporter protein activity when the genetic code is altered within the range of mutational cost tested in this study ($Cost_{PR}$: 5.29 – 5.77, $Cost_{MV}$: 1848 – 2348, and $Cost_{HI}$: 3.27 – 5.10), which covers approximately 18.4% (PR), 37.6% (MV), and 50.8% (HI) of possible cost range achievable among one million randomly-generated genetic codes.

Introduction

The genetic code is the set of rules for the translation of nucleotide sequences in mRNA into amino acid sequences in proteins. All known living organisms use a nearly identical genetic code, referred to as the standard genetic code (SGC), representing a universal mechanism shared by almost all extant living systems on Earth (Crick, 1968; Koonin and Novozhilov, 2017; Woese, 1965; Woese et al., 1966). In SGC, 20 amino acids and stop codons are nonrandomly assigned to 64 codons; amino acids with similar physicochemical properties are assigned to neighboring

codons (see Fig. 1A, SGC)(Koonin and Novozhilov, 2017, 2009). The origin and reason for this non-random codon organization remain a major mystery in biology(Crick, 1968; Di Giulio, 2005; Koonin and Novozhilov, 2017, 2009).

One explanation for this nonrandomness in SGC is mutational robustness. According to the error minimization theory, amino acid arrangement in SGC was selected to minimize the functional impact of amino acid substitutions caused by single-nucleotide mutations or translational errors(Freeland et al., 2003; Freeland and Hurst, 1998; Haig and Hurst, 1991). Several theoretical studies have estimated the possible cost of amino acid replacement (referred to as mutational cost in this study) based on the physicochemical properties of amino acids. These analyses consistently demonstrated that SGC exhibits an exceptionally low mutational cost compared with randomly generated genetic codes (Buhrman et al., 2013; Freeland and Hurst, 1998; Gilis et al., 2001; Goodarzi et al., 2004; Haig and Hurst, 1991; Omachi et al., 2023), although some genetic codes still have a lower cost than SGC(Błażej et al., 2018; Novozhilov et al., 2007; Wnętrzak et al., 2018). An alternative explanation for the non-randomness of SGC is a by-product of genetic code expansion(Massey, 2016, 2008). According to this explanation, the arrangement of SGC does not confer a selective advantage over other genetic codes. Despite these many theoretical studies, experimental verification is still limited.

A fundamental obstacle to the experimental verification of the role of genetic code arrangement in mutational robustness is the lack of variation in genetic codes in nature; all extant organisms use a nearly identical genetic code (Massey, 2015; Shulgina and Eddy, 2021). Recently, *E. coli* with designed genetic codes have been developed by removing specific stop codons and rarely used sense codons from the genome (Fredens et al., 2019; Lajoie et al., 2013; Nyerges et al., 2023; Robertson et al., 2025, 2021; Zürcher et al., 2022) and reassigning them to non-canonical or alternative amino acids through the introduction of orthogonal tRNA-aminoacyl-tRNA synthetase (aaRS) pairs (Costello et al., 2024; Shandell et al., 2021). However, these approaches based on living cells require extensive genome engineering, and thus, the number of codons that can be reassigned remains limited.

An alternative approach for genetic code alteration is in vitro reconstitution(Hibi et al., 2020; Miyachi et al., 2025; Shimizu et al., 2001). In this approach, the components required for translation, such as tRNAs and aaRSs, are individually prepared and supplied, allowing simultaneous reassignment of multiple codons. A representative platform for such reconstitution is the PURE system (Shimizu et al., 2001), a reconstituted cell-free translation system composed of purified translation components, including ribosomes, translation factors, aaRSs, amino acids, and energy-regeneration components. In particular, a tRNA-free PURE system (Miyachi et al., 2022), in which endogenous tRNA activity is minimized and defined tRNA sets are supplied externally, enables genetic codes to be reconstructed by controlling the supplied tRNAs. To date, several genetic codes have been reconstituted in vitro, including minimal genetic codes composed of 21 tRNAs that mimic or are derived from SGC, and those incorporating non-canonical amino acids(Forster et al., 2003; Iwane et al., 2016; Jones et al., 2025; McFeely et al., 2023). Translation of peptides and proteins using these reconstituted genetic codes has also been demonstrated (Calles et al., 2019; Fujino et al., 2020; Hibi et al., 2020; Iwane et al., 2016; Miyachi et al., 2025). However, reconstitution of elaborately arranged genetic codes and protein translation by them remains challenging.

In this study, to assess the effect of genetic code arrangement on mutational robustness, we constructed non-standard genetic codes in vitro and compared the effect of mutations on the function of translated proteins. We first established a minimal genetic code, composed of 21 tRNAs with vacant codons, which allows multiple alternative codon assignments to be introduced under otherwise comparable translation conditions. By introducing tRNA variants into these vacant codons, we then constructed 10 non-standard genetic codes with distinct costs of amino acid replacement, along with a near-standard genetic code. Using these genetic codes, we translated reporter gene libraries carrying random mutations and quantified the resulting protein activities.

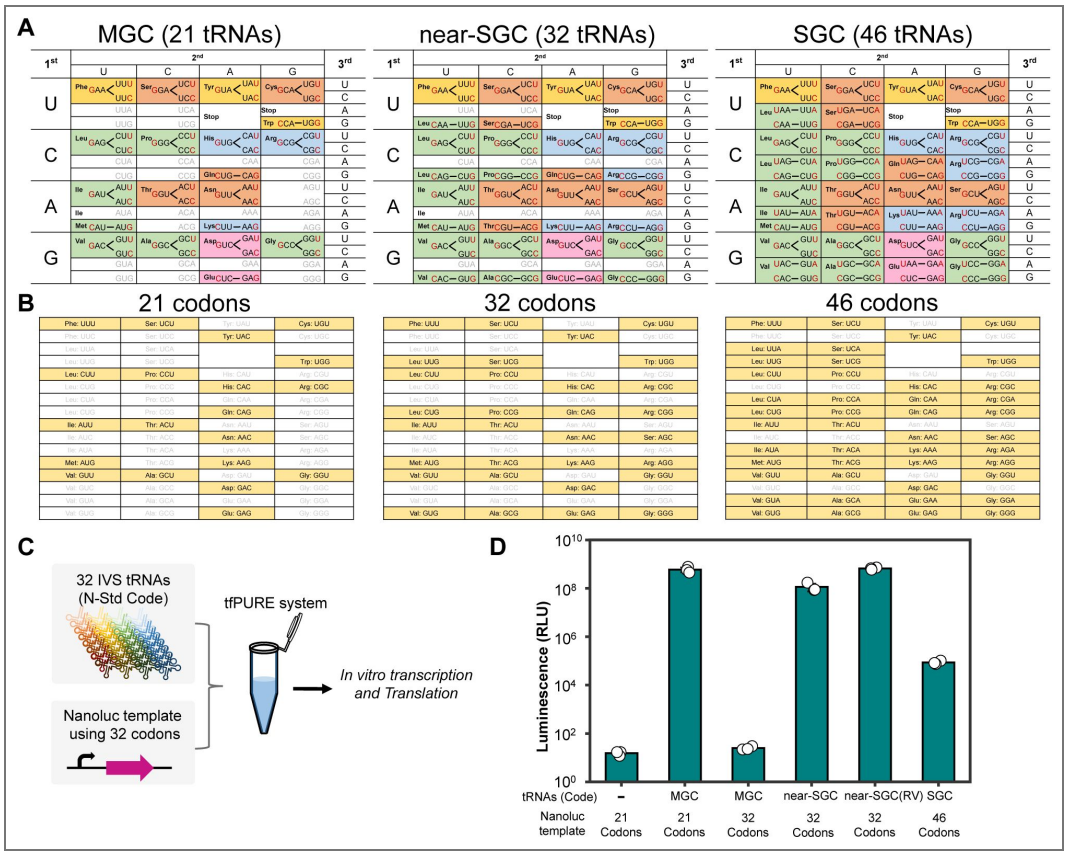

Figure 1. In vitro construction of minimal, near-standard, and standard genetic codes

(A) Anticodons of tRNAs and their corresponding codon assignments in the minimal genetic code (MGC), the near-standard genetic code (near-SGC), and the standard genetic code (SGC). Each codon is colored according to the physicochemical properties of the assigned amino acid: hydrophobic (green), aromatic (yellow), polar uncharged (orange), basic (blue), and acidic (pink). In each box, the anticodons of tRNAs (left) and the corresponding codons (right) are shown. (B) Codon sets used for reporter genes. The 21 codons contain only codons that are usable for MGC. Reporter genes composed of these codons were used in D and the subsequent experiments using non-standard genetic codes. The 32 and 46 codons contain codons that are usable for near-SGC and SGC, respectively. Reporter genes composed of these codons were used in D. (C) Schematic of the translation assay. Reporter genes (NanoLuc, 1 nM) consisting of the 21, 32, or 46-codon were translated in a customized reconstituted translation system lacking endogenous tRNAs (tRNA-free PURE system; tfPURE) supplemented with in vitro-synthesized tRNAs corresponding to MGC, near-SGC, or SGC (IPEN tRNA at 100 ng/μL; all other tRNAs at 12 ng/μL), and T7 RNA polymerase (0.42 U/μL) at 30 °C for 16 h. (D) NanoLuc activity after incubation. In near-SGC (RV), two tRNAs ($tRNA^{Val}$ and $tRNA^{Arg}$) were increased to 100 ng/μL. Each dot represents an independent experiment (n = 3). Bars indicate mean values, and error bars represent standard deviations. Statistical comparisons in (D) were performed using one-way ANOVA followed by Tukey's post hoc test on NanoLuc activity; major comparisons are summarized in Table S8.

Random mutations decreased reporter protein function at similar levels across all genetic codes examined, implying that alterations of the genetic code within the ranges explored in this study have no significant effect on mutational robustness of individual protein activity.

Results

Construction of minimal and standard genetic codes

Before constructing non-standard genetic codes, we first constructed a minimal genetic code (MGC) using 21 tRNAs (20 for 20 amino acids and one for formyl-Met) according to a previously reported method (Fujino et al., 2020; Hibi et al., 2020)(Fig. 1A, left). In this MGC, we used tRNAs with G at the first position of the anticodon for 15 amino acids (Phe, Ser, Tyr, Cys, Leu, Pro, His, Arg, Ile, Thr, Asn, Val, Ala, Asp, and Gly), which can decode codons with C or U at the third position through wobble base pairing (Hibi et al., 2020). We also used tRNAs with C at the first position of the anticodon for 5 amino acids (Trp, Gln, Met, Lys, and Glu), which can decode codons with G at the third position. The expected correspondences between anticodons and the codons they decode are represented as lines in Fig. 1A (left three letters in each box indicate the anticodon of the tRNA, and the right ones indicate the codon).

To ensure fair comparisons, we also constructed genetic codes close to the SGC using the same method (i.e., using unmodified tRNAs). Previously, a genetic code approximating the SGC was constructed in vitro using 32 tRNAs for peptide synthesis (Iwane et al., 2016). Following this method, we introduced 11 additional tRNAs (Leu_CAA, Ser_CGA, Leu_CAG, Pro_CGG, Arg_CCG, Ser_GCU, Thr_CGU, Arg_CCU, Val_CAC, Ala_CGC, and Gly_CCC) into the MGC (Fig. 1A, middle). We refer to this as the near-standard genetic code (near-SGC). Although the near-SGC mimics the SGC in most of the codons, it does not cover codons with A at the third nucleotide position (NNA). We therefore further introduced 14 tRNAs with UNN anticodons that are expected to decode NNA codons to construct SGC consisting of a 46-tRNA set (Fig. 1A, right).

We next examined the translation efficiency of the reconstructed MGC, near-SGC, and SGC. Translation was performed in a custom-made tRNA-free PURE system (tfPURE) (Miyachi et al., 2025), supplemented with 21 tRNAs for MGC, 32 tRNAs for near-SGC, or 46 tRNAs for SGC. As a reporter protein, we used the NanoLuc luciferase, originally derived from deep-sea shrimp (Hall et al., 2012). For MGC, the NanoLuc sequence was designed to use only the 21 corresponding codons (Fig. 1B, 21codons). For near-SGC and SGC, the NanoLuc coding sequences were designed so that the codons available in each genetic code were used with minimal differences in codon counts, while preserving the amino acid sequence (Fig. 1B, 32 codons and 46 codons). After incubation at 30 °C for 16 hours, luminescence was measured as an indicator of translation efficiency. Translation using near-SGC exhibited lower luminescence (1.2×10^8; Fig. 1D, near-SGC) than that using MGC (6.0×10^8; MGC, 21 codons). To improve translation efficiency with near-SGC, we focused on two tRNA concentrations ($tRNA^{Val}_{CAC}$ and $tRNA^{Arg}_{CCU}$), which were suggested to have low activities in a previous study (Iwane et al., 2016). Upon increasing the concentrations of these two tRNAs, the translation increased up to 100 ng/μL (Fig. S1). At 100 ng/μL for each tRNA, the translation efficiency of near-SGC was increased to a level comparable to that of MGC (6.6×10^8; Fig. 1D, near-Std (RV)). The increased concentrations of these two tRNAs were used in all the subsequent experiments. The translation level of SGC, which uses 46 codons with 46 tRNAs, was approximately 100-fold lower (8.6×10^4; Fig. 1D), indicating that the additional tRNAs corresponding to NNA codons may not function as efficiently as other tRNAs. This inefficiency is probably due to the lack of chemical modifications in tRNAs used here, as the anticodon of the native tRNAs that decode NNA are frequently modified in the cell (El Yacoubi et al., 2012; Masuda and Hou, 2024). Such chemical modifications may be required for efficient translation of the NNA codon (Cui et al., 2015; Madore et al., 1999; Tamura et al., 1992).

We also examined whether the lower translation efficiency of the 46-codon NanoLuc template could be explained by sequence-dependent effects, such as codon context or mRNA structure. When the 21-, 32-, and 46-codon NanoLuc templates were translated using native *E. coli* tRNAs in the tfPURE system (Fig. S2), the 46-codon template showed lower activity than the 21- and 32-

codon templates; however, this difference was within approximately two-fold. Accordingly, we decided to use only the 32 codons used in near-SGC (i.e., excluding NNA codons) in the subsequent construction of non-standard genetic codes.

Test of the availability of each vacant codon for Ala, Ser, and Leu

We next aimed to construct multiple non-standard genetic codes (nonSGCs) by introducing tRNAs with altered anticodon sequences into the vacant codons of MGC, excluding NNA codons. Currently, anticodon sequences of tRNAs can be altered only for Ala, Ser, and Leu while maintaining recognition by their respective aminoacyl-tRNA synthetases (aaRSs) because the aaRSs for these amino acids do not recognize the anticodon sequence (Asahara et al., 1993; Francklyn and Schimmel, 1989; Sampson and Saks, 1993). Therefore, we decided to design nonSGCs by reassigning Ala, Ser, and Leu to 10 vacant codons (UUG, UCG, CUG, CCG, CGG, ACG, AGC, GUG, GCG, and GGG) in MGC. Note that one vacant codon (AGG) was assigned to the original amino acid (Arg) because this codon is already distant from the codon used for Arg (CGC) in MGC and does not need to be reassigned.

We tested whether each of the 10 vacant codons is usable for Ala, Leu, and Ser when the corresponding tRNA variant with an altered anticodon is introduced (Fig. 2A). To test this, we constructed a total of 25 genetic codes (nine for Ala, eight for Ser, and eight for Leu) by introducing one tRNA variant with an altered anticodon into MGC (referred to as 21 + 1) and also designed 25 corresponding NanoLuc template DNAs, in which some codons for Ala, Ser, and Leu (two for Ala, three for Ser, and four for Leu) were replaced with the corresponding vacant codons from the 21-codon set shown in Fig. 1B. For example, when assigning Ala to the UUG codon (Fig. 2A, bottom), two GCU codons used for Ala in the template DNA were changed to UUG codons.

Translation reactions were performed in tfPURE supplemented with 21 tRNAs alone (21 code, equivalent to MGC) or 21 tRNAs plus one additional tRNA variant for Ala, Ser, and Leu (21 + 1 code) with the corresponding NanoLuc template (1 nM) at 30 °C for 16 h. NanoLuc activity was low when using only 21 tRNAs (Fig. 2B, gray bars, 21 code), but increased when the corresponding tRNA variant was supplied (pink bars, 21 + 1 code) for most of the tested codons. An exception was observed when Ser was assigned to the GGG codon, in which NanoLuc activity did not increase with 21 + 1 code for reasons that remain unclear. These results demonstrate that the nine tested vacant codons (i.e., excluding GGG) can be reassigned to the three amino acids, allowing the construction of up to 3^9 = 19,683 types of nonSGCs, or approximately 20,000 codes. For GGG, we retained the original glycine assignment in all subsequent experiments. We further optimized the concentration of each additional tRNA variant, particularly for those that exhibited lower translation efficiency (Fig. S3). Based on these results, the concentrations of the following tRNAs were increased to the indicated concentrations in subsequent experiments: $tRNA^{Ala}_{CAA}$ (40 ng/μL), $tRNA^{Ala}_{CGA}$ (60 ng/μL), $tRNA^{Ala}_{CGU}$ (60 ng/μL), $tRNA^{Leu}_{CGU}$ (80 ng/μL), $tRNA^{Leu}_{CGA}$ (80 ng/μL), $tRNA^{Leu}_{GCU}$ (80 ng/μL).

Design of non-standard genetic codes

We next designed nonSGCs. By assigning Ala, Ser, or Leu to the nine vacant codons (gray boxes in Fig. 3A), a total of 19,683 (= 3^9) possible genetic codes can be generated. The objective of this study is to evaluate how the arrangement of the genetic code influences the impact of point mutations on protein function. To guide the selection of genetic codes to construct, we defined three types of mutational cost for each genetic code based on three physicochemical properties of amino acids, polar requirement (PR), molecular volume (MV), and hydropathy index (HI), according to previous studies (Freeland and Hurst, 1998; Haig and Hurst, 1991). PR reflects polarity-related interactions of amino acids and has been used as a classical measure of amino acid similarity in error minimization analyses. MV represents side-chain size and steric volume, which could affect protein packing and structural stability, whereas HI reflects hydrophobicity, which could be closely related to protein folding or hydrophobic core formation. The mutational costs, three values ($Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$) associated with each genetic code, represent the expected average negative effect of a single nucleotide substitution in a target gene on the protein

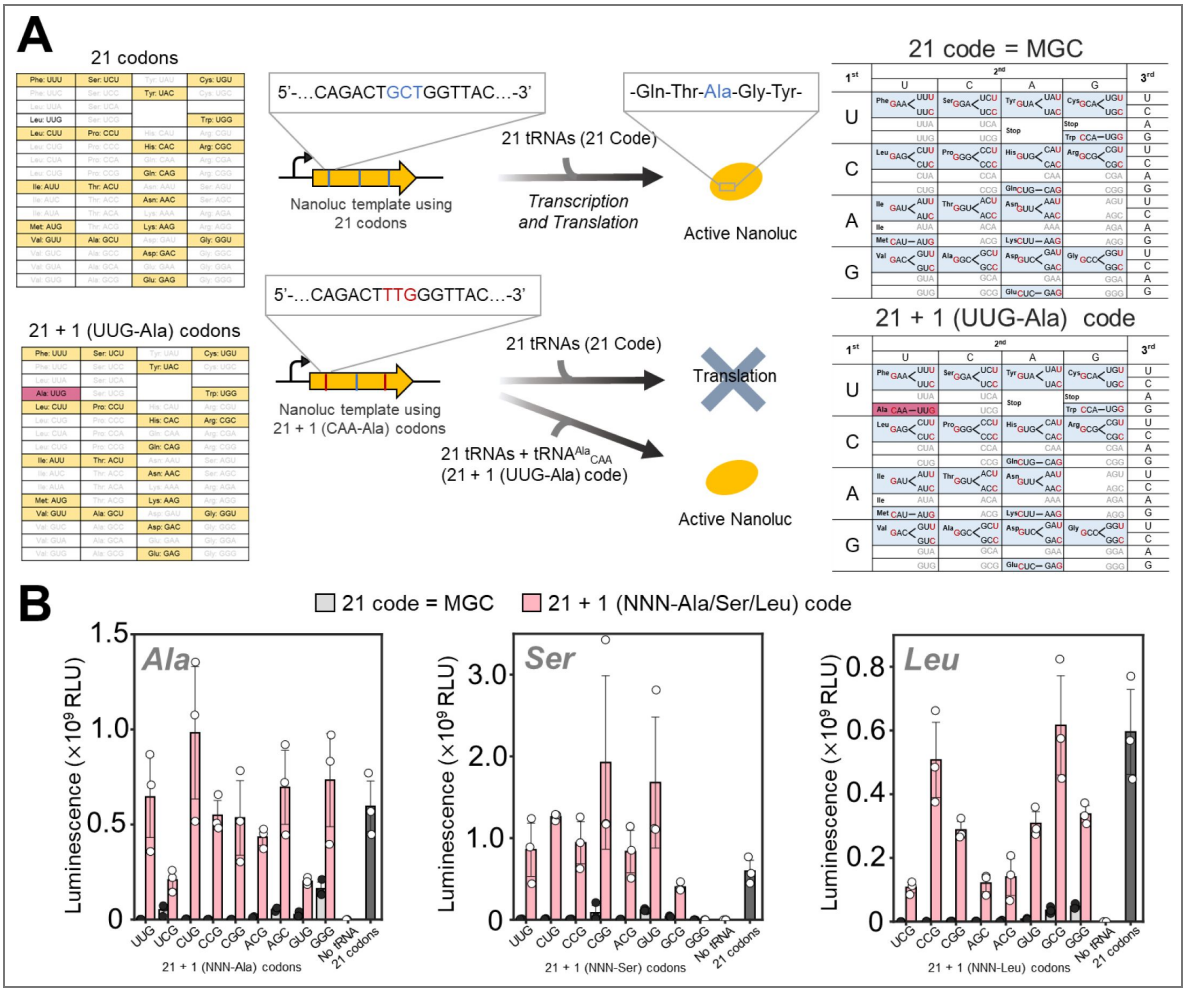

Figure 2. Reassignment experiments to test the availability of 10 vacant codons for Ala, Ser, and Leu

(A) Schematic illustration of reassignment experiments. Translation with the original MGC and NanoLuc template is shown at the top for comparison. An example of Ala reassignment to the UUG codon is shown at the bottom. In this example, three Ala codons in the NanoLuc sequence were replaced with one type of vacant codon (e.g., UUG), generating a 21 + 1 (UUG-Ala) codon set. Similar reassignment experiments were performed for three amino acids (Ala, Ser, and Leu) and nine vacant codons. Specifically, two Ala codons (Ala16 and Ala120), three Ser codons (Ser31, Ser49, and Ser150), or four Leu codons (Leu32, Leu67, Leu144, and Leu170) were replaced. (B) NanoLuc translation results for each codon reassignment experiment. Translation reactions were performed in tfPURE supplemented with a 21-tRNA mixture (600 ng/μL), one tRNA variant (12 ng/μL each), and each NanoLuc template (1 nM) that contains 2 – 4 of a corresponding codon to be tested (21 + 1 NNN-Ala/Ser/Leu codons). Reactions were incubated at 30 °C for 16 h, after which NanoLuc activity was measured. As a control, translation reactions lacking the additional tRNA variant were conducted (21 code, gray bars) and compared to the data with the additional tRNA (21 +1 code, pink bars). Additional controls included translation without any tRNA (no tRNA) and translation using MGC with NanoLuc templates encoded by the original 21 codons (21 codons), both shown for comparison. Each dot represents three technical replicates, and error bars represent standard deviations. For each template, NanoLuc activity in the 21-code and corresponding 21+1-code conditions was compared using Welch's t-test on luminescence. Statistical results are summarized in Table S9.

function, estimated from the change in each physicochemical property of the substituted amino acid. To make the design of non-SGCs more explicit, we show one representative non-SGC together with the near-SGC in Fig. 3B. This comparison illustrates how assignment of Ala, Ser, or Leu to the vacant codon boxes changes the three mutational cost metrics, $Cost_{tPR}$, $Cost_{MV}$, and $Cost_{HI}$.

The three types of mutation costs based on PR, MV, and HI were calculated for each of the 19,683 (= 3^9) possible genetic codes as follows. First, we considered a virtual protein-encoding DNA sequence using the 21 codons shown in Fig. 1B (left) at equal frequencies. We then enumerated the set of codons reachable from each codon via single-nucleotide substitutions and calculated the average change in each physicochemical property of amino acids for each genetic code. In this calculation, mutation biases, including the relative frequencies of transitions and transversions, were incorporated as weighting factors. When mutations produced NNA codons, which are vacant in the genetic codes used here, their substitutions were neglected. The resulting average physicochemical change was defined as the mutational cost for the target genetic code. According to this calculation, a smaller mutational cost indicates that single-nucleotide substitutions lead to smaller changes in the corresponding amino acid properties and thus expectedly smaller changes in protein function, reflecting higher robustness against mutations.

The distributions of mutational costs calculated for each physicochemical property ($Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$) are shown in Fig. 3C. The near-SGC (orange dashed lines) was located in the low-cost region of the distributions for all three properties, consistent with previous theoretical studies (Freeland and Hurst, 1998; Haig and Hurst, 1991). We next selected genetic codes for construction and experimental evaluation. First, we selected the genetic codes exhibiting the maximum and minimum costs for each physicochemical property (Fig. 3B, red dashed lines; $PR_{min/max}$, $MV_{min/max}$, and $HI_{min/max}$). These selected genetic codes are shown in Figs. 3D – 3F, colored according to the physicochemical values of the assigned amino acids. The genetic codes of the minimum costs (PR_{min}, MV_{min}, and HI_{min}) exhibit relatively ordered arrangements of the physicochemical values, whereas those with maximum costs (PR_{max}, MV_{max}, and HI_{max}) exhibit relatively disordered arrangements. Because the PR_{max} and HI_{max} codes were identical, these five extreme genetic codes were selected for construction. Second, we selected five additional genetic codes representing various points across the cost distributions for all three properties (green dashed lines). These 10 nonSGCs (shown in Figs. S4–S6) were constructed by introducing the corresponding additional tRNA sets into MGC and used in subsequent experiments.

Preparation and analysis of random mutation libraries of reporter genes

To compare functional robustness against mutations across different genetic codes, we prepared randomly mutagenized DNA libraries of three reporter genes, β-galactosidase (GAL), firefly luciferase (Luc), and mStayGold (Ando et al., 2023) (mSG), from the original sequences composed of only the 21 codons shown in Fig. 1B (left). DNA libraries were prepared through a three-step PCR procedure (Fig. S7A). In the first PCR, random mutations were introduced by error-prone PCR using Taq DNA polymerase in the presence of Mn^{2+}. The mutation rate was tuned by varying the Mn^{2+} concentration across five levels: 10, 50, 100, 250, and 350 μM. For comparison, we also performed standard PCR using a high-fidelity polymerase (KOD Plus Neo) to prepare a low-mutation library. In the second PCR, a HiBiT tag was attached to the C-terminus using the high-fidelity polymerase for quantitation of translation levels. In the third PCR, Illumina sequencing adapters and unique barcode sequences were added using the high-fidelity polymerase. The products of the second PCR were used directly for translation experiments, whereas the products of the third PCR were subjected to Illumina sequencing to quantify mutation rates and characterize the resulting libraries.

Mutation frequencies at each position were calculated from the sequencing data obtained. The simple average of these values across all positions was defined as the mean per-base mutation rate. When using Taq DNA polymerase, the error rates increased monotonically with increasing Mn^{2+} concentration, ranging from 1.2×10^{-3} to 6.9×10^{-3} (Fig. S7B). In contrast, when using the

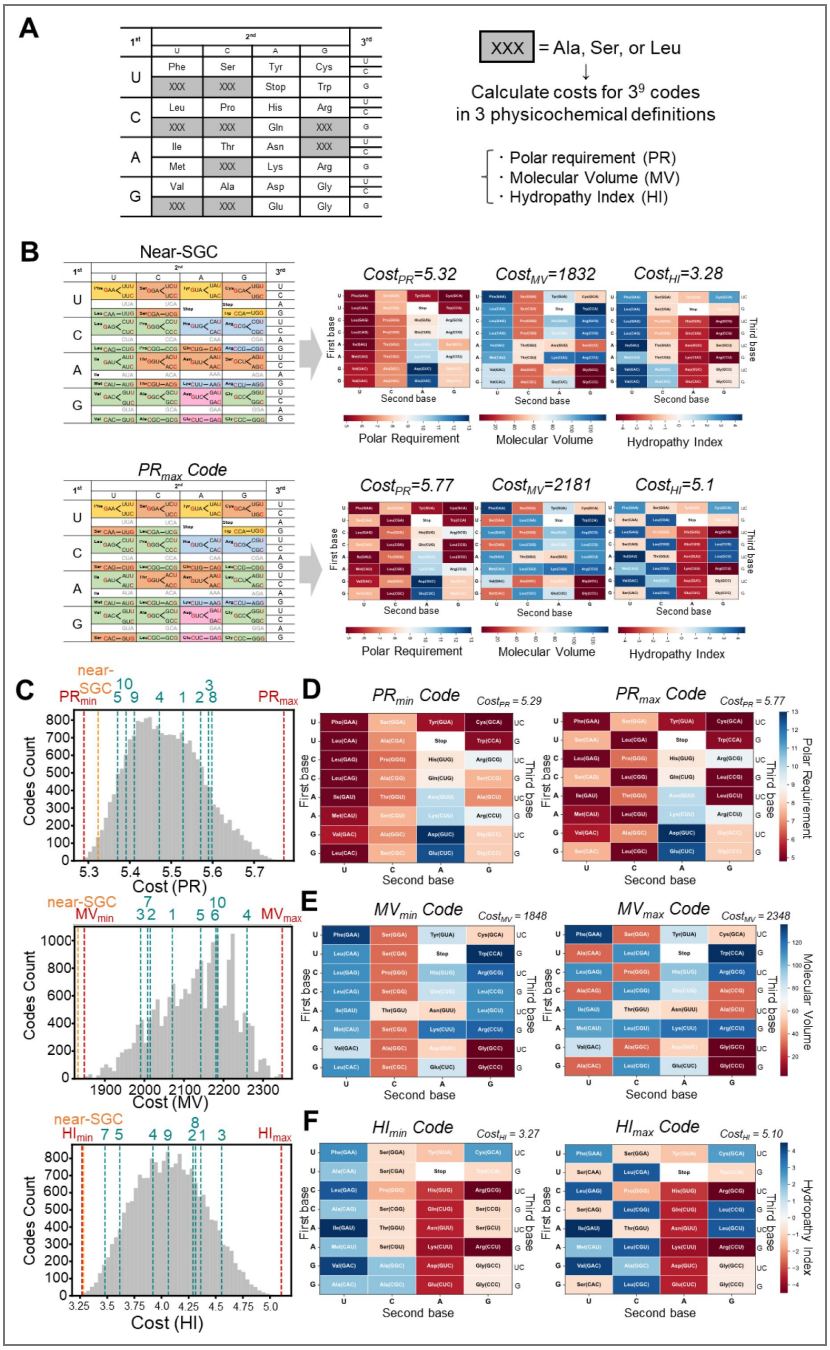

Figure 3. Distribution of mutational costs of reassigned genetic codes

(A) Calculation method of mutational costs for each genetic code based on three physicochemical properties of amino acids. The average change in each of the three physicochemical properties of amino acids upon single-nucleotide substitutions from the 21 codons was calculated (see Methods for details). In the reassigned genetic codes analyzed here, one of three amino acids (Ala, Ser, or Leu) was assigned to each of the nine vacant codons shown in gray, and the costs were calculated for all possible reassignment combinations. (B) Representative comparison of the near-SGC and PR_{max} code. Codon assignment schemes are shown on the left, and heatmap representations of the assigned amino acid values for polar requirement, molecular volume, and hydropathy index are shown on the right. The corresponding $Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$ values are indicated above each heatmap. (C) Distributions of mutational costs for each physicochemical property of amino acids. Dashed lines indicate the cost values of 10 genetic codes selected for experimental construction. Red dashed lines indicate the minimum and maximum cost values for each cost definition, and orange lines indicate the cost values of near-SGC. (D, E, F) Genetic codes exhibiting the minimum and maximum mutational costs based on PR (D), MV (E), and HI (F). The physicochemical values of amino acids assigned to each codon are shown as heatmaps.

high-fidelity polymerase, the error rates were much lower (ranging from 3.0×10^{-4} to 4.0×10^{-4}). The mutations observed with the high-fidelity polymerase are thought to have been caused by sequencing errors because the rate is consistent with the reported sequencing error rate (Schirmer et al., 2016) and also the primer-binding regions, which did not undergo PCR amplification, exhibited a similar mutation rate (approximately 5.0×10^{-4}), implying that the actual mutation rate obtained using the high-fidelity polymerase can be lower.

We further analyzed the patterns of nucleotide substitutions observed in the libraries. With Taq polymerase, transitions (A↔G or C↔T) accounted for 61 – 72% of all substitutions, and this bias was consistently observed across all Mn^{2+} concentrations tested (Fig. S7C). This pattern is consistent with previously reported mutation biases associated with this polymerase (Lin-Goerke et al., 1997). Visualization of the positional distribution of mutation rates further confirmed that error-prone PCR introduced mutations throughout the entire gene sequence (Fig. S7D).

Translation of random mutation libraries using near-SGC

To examine the relationship between mutation rates and translated protein function, we performed translation reactions using the near-SGC (RV), containing higher concentrations of two tRNAs ($tRNA^{Val}_{CAC}$ and $tRNA^{Arg}_{CCU}$) as shown in Fig. 1D, and DNA libraries with different mutation rates (Fig. 4A). When using GAL library, we observed a decrease in mean protein activity as the mutation rate increased (Fig. 4B), indicating that a sufficient number of mutations were introduced to evaluate the mutational effect on protein function. Similar trends were observed for Luc and mSG (Figs. 4C and 4D), although the magnitude of the decrease varied among proteins. For example, protein activities decreased to 33%, 2.4%, and 79% for GAL, Luc, and mSG, respectively, at a mutation rate of approximately 0.002 per base, indicating that sensitivity to mutations differs substantially among proteins, consistent with a previous study (Lind et al., 2017).

Translation of random mutation libraries using 10 non-SGCs

We next evaluated the effects of random mutations on protein activity using 10 non-SGCs constructed above. For this experiment, two random mutation libraries were used: a low-mutation library prepared using the high-fidelity polymerase and a high-mutation library prepared using Taq DNA polymerase at a Mn^{2+} concentration that yields mutation rates of 0.002 – 0.005 per base (0.0026 for GAL, 0.0027 for Luc, and 0.0048 for mSG, corresponding to approximately 8.0, 4.5, and 3.3 mutations per gene). We also plotted position-wise non-reference rates along the analyzed regions of each reporter gene, confirming that mutations were broadly distributed across the amplicons (Fig. S8). These libraries were transcribed and translated using 10 non-SGCs selected above, as well as the near-SGC (RV), and the resulting protein activities were measured (Fig. 5A).

We first examined GAL activities. For the low-mutation library, mean protein activity varied by up to 5.4-fold among different genetic codes (Fig. 5B left, green bars). For the high-mutation library, GAL activity decreased overall, while the relative differences in activity among genetic codes observed in the low-mutation library were broadly retained (Fig. 5B right, orange bars). Quantification of protein synthesis levels using the HiBiT assay revealed similar protein synthesis levels both among the different genetic codes and between the low- and high-mutation rates (Fig. S9), indicating that the variation in protein activity among genetic codes and between low- and high-mutation rates was not attributable to differences in translational levels. The variation in protein activity among genetic codes with the low-mutation library was possibly caused by differences in translational error frequencies among genetic codes. Such differences are not surprising because different tRNA sets were used for each genetic code, and each tRNA is expected to have distinct translational accuracies. However, the purpose of this study is to evaluate how the effect of mutations on protein function varies depending on the genetic code; this effect can be extracted even in the presence of different translational error rates by normalizing the protein activity of the mutated library by that of the less-mutated library.

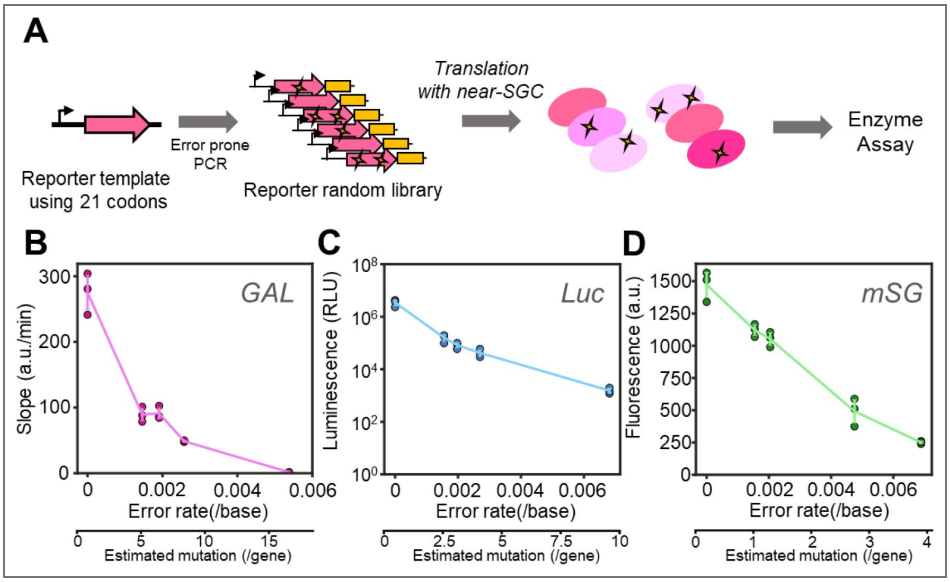

Figure 4. Translation of random mutagenesis libraries with near-SGC

(A) Schematic overview of the protein activity assay using a random mutation library. Reporter genes composed of the 21 codons were subjected to random mutagenesis by error-prone PCR at different Mn^{2+} concentrations to generate DNA libraries, as shown in Fig. S7. These libraries (5 nM) were translated using near-SGC, consisting of a 32-tRNA mixture ($tRNA^{IPEN}$, $tRNA^{Val}_{CAC}$, and $tRNA^{Arg}_{CCU}$ at 100 ng/μL; all other tRNAs at 12 ng/μL) in tfPURE, including T7 RNA polymerase (1.7 U/μL) at 30 °C for 16 h, and each protein activity was measured. (B, C, D) Dependence of β-galactosidase (GAL) (B), firefly luciferase (Luc) (C), and mStayGold (mSG) (D) activity on mutation rate. Note that the vertical axis of panel C (Luc) is on a log scale. Each dot represents the results of three technical replicates, and error bars represent standard deviations. The lower x-axis indicates the estimated number of mutations per gene, calculated by multiplying the mutation rate per base by the coding sequence length of each reporter gene. Spearman's rank correlation coefficients were ρ = −0.90 for GAL, ρ = −1.00 for Luc, and ρ = −1.00 for mSG.

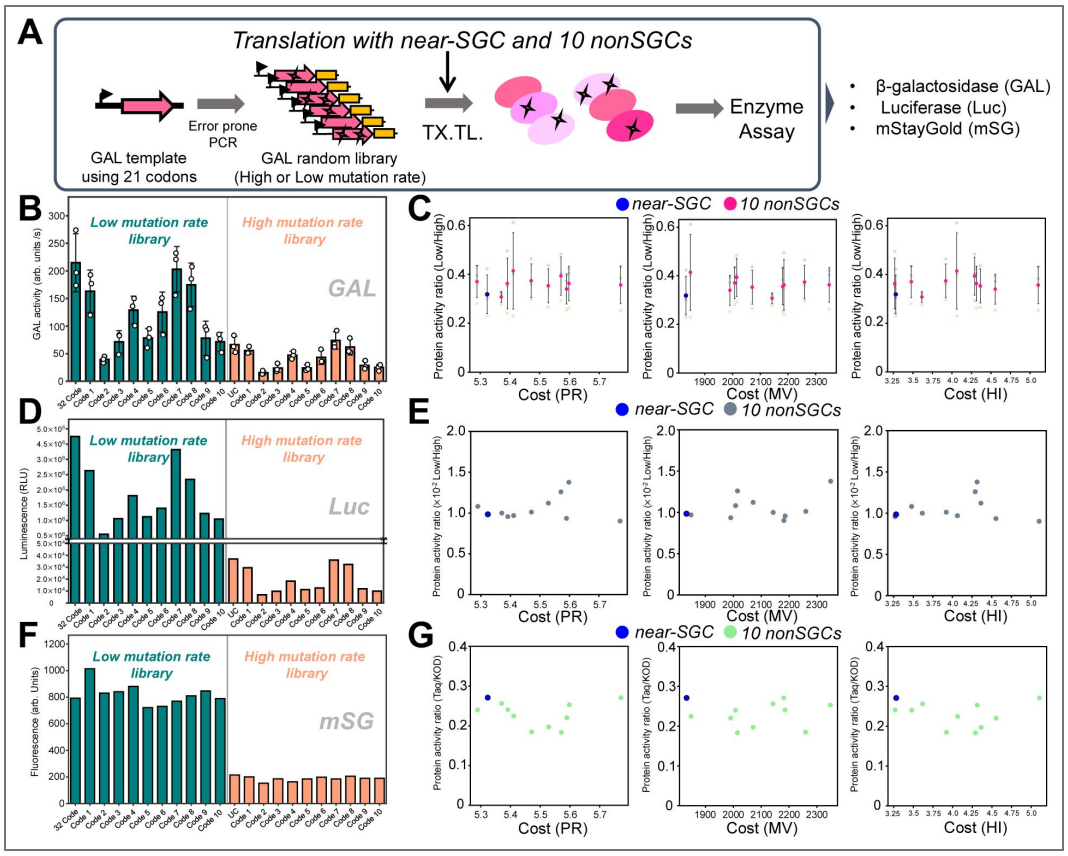

Figure 5. Translation of mutagenized DNA libraries with non-SGCs

(A) Schematic of the experiment for comparing protein activities translated with different genetic codes. Random libraries prepared at low and high mutation rates were translated using either the 10 non-SGCs or the near-SGC (RV). Translation conditions were identical to those described in Fig. 4. (B, D, F) Protein activities of products translated with each genetic code using low- and high-mutation DNA libraries. Activities are shown for β-galactosidase (GAL; B, mutation rate = 2.6×10^{-3} per base), firefly luciferase (Luc; D, mutation rate = 2.7×10^{-3} per base), and mStayGold (mSG; F, mutation rate = 4.8×10^{-3} per base). Quantification of protein synthesis levels for GAL is shown in Fig. S9. (C, E, G) Ratios of protein activity of high-mutation libraries to those of low-mutation libraries, plotted against the corresponding theoretical mutational costs. Data are shown for GAL (C), Luc (E), and mSG (G). Mean values of three technical replicates are shown with standard deviations for GAL. For GAL activity in (B), two-way ANOVA was performed using genetic code and mutation level as factors. Significant main effects of genetic code and mutation level were detected (both $p < 0.0001$), whereas their interaction was not significant. For (C), (E), and (G), Spearman's rank correlation analysis was performed between each mutational cost metric and the high-/low-mutation activity ratio. Statistical details are summarized in Table S10.

To estimate the average activity reduction associated with increased mutational burden under each genetic code, we calculated the ratio of activity obtained from the high-mutation library to that from the corresponding low-mutation library and plotted this ratio against each of the three mutational costs (Fig. 5C). Data before this normalization are shown in Figs. S10–S12. According to the error minimization theory, genetic codes with higher mutational costs tend to reduce the protein activity more upon mutagenesis; thus, a negative correlation between mutational cost and protein activity would be expected. Contrary to this expectation, the effect of mutations on GAL activity remained nearly constant across the ranges of all three mutational costs examined (Fig. 5C).

We performed the same experiment for Luc (Fig. 5D) and mSG (Fig. 5F), and the results were similar to those obtained by GAL. For both proteins, activity with the low-mutation library varied among genetic codes, and decreased with the high-mutation library, although the variation among genetic codes was smaller for mSG. The activity ratios of the high- to low-mutation libraries showed no clear dependence on any of the three mutational costs (Figs. 5E and 5G). Taken together, these results indicate that the mutational robustness of individual reporter protein function did not substantially differ among the genetic codes that exhibit the mutational cost ranging from 5.29 to 5.77 for PR, 1848 to 2348 for MV, and 3.27 to 5.10 for HI.

Discussion

In this study, we constructed 10 non-SGCs by assigning Ala, Ser, and Leu to vacant codons in MGC. These non-SGCs represent mutational costs ranging from 5.29 to 5.77 for $Cost_{PR}$, 1848 to 2348 for $Cost_{MV}$, and 3.27 to 5.10 for $Cost_{HI}$, most of which are higher than those of SGC. The error minimization theory predicts that the deleterious effect of mutations on protein function becomes more severe when using genetic codes with higher mutational cost (Freeland et al., 2003; Freeland and Hurst, 1998; Haig and Hurst, 1991; Massey, 2008). In contrast to this prediction, translation experiments using the non-SGCs and mutagenized libraries revealed that the effects of mutations on protein functions were similar across all genetic codes and reporter genes examined here, suggesting that the mutational robustness of protein activity remained largely unchanged within at least the ranges of mutational cost tested in this study. It should be noted that this conclusion is limited to the activity of individual reporter proteins translated in a reconstituted in vitro system. Therefore, whether similar trends would be observed at the level of cellular fitness or long-term evolution remains an open question. Moreover, our results do not exclude other possible roles of SGC organization. The SGC may have been shaped by multiple factors, including robustness to translational errors, historical constraints during genetic code expansion, biosynthetic or coevolutionary relationships among amino acids, stereochemical interactions, and effects on protein evolvability (Katoh and Suga, 2023; Koonin and Novozhilov, 2017, 2009; Novozhilov et al., 2007; Wong, 2005).

Although protein amounts quantified by the HiBiT tag were comparable among genetic codes, GAL activities differed substantially. This indicates that the activity differences among genetic codes were not primarily attributable to differences in the amount of C-terminally completed translation products. The HiBiT assay does not provide information on the fraction of catalytically active protein, including sequence fidelity or folding state, and therefore cannot distinguish among these possibilities. Detailed characterization of translated products by mass spectrometry would provide further mechanistic insight into how individual non-SGCs affect protein quality. However, the primary objective of the present study was to compare mutation-dependent activity loss across genetic codes. Therefore, we evaluated this effect by normalizing the activity of the high-mutation library to that of the corresponding low-mutation library within each genetic code.

A further limitation of this study is that the reporter activities were measured at the level of pooled random mutation libraries. Therefore, the high-/low-mutation activity ratio used in this study should be interpreted as the relative reduction in average activity caused by increasing the mutational burden in a heterogeneous mutation pool, rather than as the effect of identical variants before and after additional mutations. This library-averaged approach was chosen because the mutational costs considered here are also defined as average expected

physicochemical effects over many possible single-nucleotide substitutions. In addition, because the non-SGCs constructed in this study were generated by reassigning only Ala, Ser, and Leu, the detectable effects may depend on how frequently mutations involving these amino acids occur in each reporter gene and whether the affected positions are functionally important. If genetic code-dependent effects are restricted to a small subset of deleterious variants, such effects may be masked in pooled activity measurements. Future studies using defined variants or high-throughput genotype–phenotype mapping assays will be required to determine the mutation-specific and position-specific mechanisms underlying genetic code-dependent effects on protein function (Rozhoňová et al., 2024).

It should be noted that the range of mutational costs experimentally tested in this study is limited because we could only reassign three amino acids (Ala, Ser, and Leu) to the vacant codons in MGC. If all 20 amino acids could be reassigned to the nine vacant codons, the possible range of mutational costs would be 5.29 – 8.49 for $Cost_{PR}$, 1826 – 2805 for $Cost_{MV}$, and 3.21 – 6.65 for $Cost_{HI}$ (Fig. S13). The genetic codes constructed in this study cover approximately 15.1%, 51.1%, and 53.3% of these possible cost ranges, for PR, MV, and HI, respectively. Furthermore, if we could assign any amino acids to any codon boxes with a degenerate pattern similar to that of SGC, as assumed in previous theoretical studies (Freeland and Hurst, 1998; Haig and Hurst, 1991), the cost ranges were 2.67 – 4.79 for $Cost_{PR}$, 1952 – 3209 for $Cost_{MV}$, and 4.69 – 12.44 for $Cost_{HI}$ for one million randomly sampled genetic codes. The non-SGCs constructed in this study covered 18.4%, 37.6%, and 50.8% of these possible cost ranges for PR, MV, and HI, respectively (Fig. S14). We consider that the cost ranges examined in this study represent substantial fractions, especially for MV and HI. Although the near-SGC did not necessarily exhibit the lowest cost for each individual physicochemical metric, this does not mean that it is unfavorable in the multidimensional cost space. Because the SGC may reflect a balance among multiple physicochemical constraints rather than optimization of a single property, we also calculated integrated cost indices by combining $Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$ after min–max normalization or z-score normalization. In both integrated indices, the near-SGC showed the lowest overall cost when compared with all 19,683 candidate non-SGCs (Fig. S15), indicating that no candidate non-SGC exhibited a lower combined cost than the near-SGC when the three physicochemical properties were considered jointly. However, we also think that the finding of this study — mutational robustness does not detectably depend on the genetic codes — might change if nonSGCs with broader cost ranges could be constructed.

This limitation in the achievable mutational cost ranges in this study arises from the fact that only three amino acids (Ala, Ser, and Leu) can be reassigned to different codons. The other amino acids cannot be reassigned to different codons because their corresponding tRNAs require the cognate anticodon sequence for recognition by the corresponding aaRS. To construct genetic codes with a broader range of mutational costs, it will be necessary to identify new aaRSs that do not recognize anticodon sequence but still catalyze aminoacylation for the other 17 amino acids. To date, such aaRSs have not been clearly reported, but some aaRSs, such as a mitochondrial aaRS with altered anticodon recognition (Su et al., 2011) and aaRS mutants that shuffle the anticodon recognition domain (Brevet et al., 2003), may be suitable for this purpose. Although numerous engineered orthogonal aaRS–tRNA pairs have been developed to incorporate non-canonical amino acids into proteins, these systems are typically designed to target stop or limited sense codons(Costello et al., 2024; Robertson et al., 2021; Shandell et al., 2021; Zeng et al., 2014). Alternatively, ribozyme-based aminoacylation systems such as flexizymes (Goto and Suga, 2021; Jones et al., 2025; Murakami et al., 2006; Terasaka et al., 2014) offer a more flexible route to assigning diverse amino acids to vacant codons, although protein synthesis using ribozyme-based aminoacylation is still a challenge (Chen et al., 2021).

The results of this study suggest that altering the arrangement of the genetic code does not necessarily lead to a drastic loss of functional robustness against mutations. This finding indicates that genetic codes can be redesigned for specific purposes without severe functional penalties, at least in vitro. This insight paves the way for utilizing engineered genetic codes (Huang et al., 2024; Pines et al., 2017; Zürcher et al., 2022) for broader applications in protein engineering and in

vitro synthetic biology, such as genetic code expansion using non-canonical amino acids (Iwane et al., 2016; Jones et al., 2025; Katoh and Suga, 2022; McFeely et al., 2023; Passioura et al., 2018), and the design of artificial translation systems orthogonal to natural translation systems for biocontainment (Calles et al., 2019; Fujino et al., 2024, 2020; Nyerges et al., 2023; Robertson et al., 2021). Accelerated protein evolution may also be achievable using non-standard genetic codes (Pines et al., 2017; Zürcher et al., 2022) because the structure of the genetic code determines the amino acids accessible through mutation. While some theoretical studies suggest that alternative genetic codes could enhance evolvability (Pines et al., 2017), others have argued that the standard genetic code itself possesses advantageous evolvability features (Firnberg and Ostermeier, 2013; Rozhoňová et al., 2024). The knowledge obtained in this study enhances the usefulness of genetic code engineering in vitro for broader applications.

Method

DNA preparation

All DNA templates encoding reporter proteins were obtained through a fragment synthesis service provided by Twist Bioscience. The NanoLuc reporter genes used for translation under MGC, near-SGC, or SGC were designed to utilize 21, 32, or 46 codons (Fig. 1B). For validation of translation by each anticodon variant tRNA in Fig. 2, NanoLuc reporter genes were designed based on the 21 codon set, in which two alanine, three serine, or four leucine codons were replaced by target codons, producing a total of 25 distinct NanoLuc templates. For assays involving non-SGCs, DNA sequences encoding β-galactosidase (GAL), firefly luciferase (Luc), and mStayGold (mSG) were designed using the 21 codon set. The sequences of all DNA templates used in this study are provided in Supplementary Table S1.

To prepare templates for translation reactions, PCR amplification was performed using primers 1 and 2, which were common to all constructs. For PCR under low-mutation conditions, KOD Plus Neo DNA polymerase (Toyobo) was used. Error-prone PCR conditions are described in a separate section below. To prepare templates with a C-terminal HiBiT tag for protein quantification (Fig. S7, 2^{nd} PCR), a HiBiT tag fragment was first amplified using primers 3 and 4. This fragment was then fused to the coding region of each reporter gene by overlap extension PCR; PCR products encoding GAL, Luc, or mSG were amplified using primers 1 and 5 (GAL), 6 (Luc), or 7 (mSG), respectively, using either KOD Plus Neo for the low-mutation library or Taq DNA polymerase for the higher-mutation libraries, and subsequently fused with the HiBiT fragment using primers 1 and 8. The resulting DNA fragments were purified using the FastGene Gel/PCR Extraction Kit (Nippon Genetics), and DNA concentrations were determined from A260 measurements using a NanoDrop spectrophotometer (Thermo Fisher Scientific). The sequences of all primers used in this study are listed in Supplementary Table S2.

Error-prone PCR

To introduce random mutations into protein-coding DNA, error-prone PCR was performed as described schematically in Fig. S7 (1^{st} PCR). PCR reactions were carried out using Taq DNA polymerase (New England Biolabs) in the manufacturer-recommended Standard Taq Reaction Buffer supplemented with additional $MgCl_2$ and $MnCl_2$. The final concentration of $MgCl_2$ was adjusted to 3.0 mM. To modulate the mutation rates, $MnCl_2$ was added at final concentrations of 10, 50, 100, 250, or 350 μM. PCR amplification was performed for 30 cycles, and the resulting products were used as random DNA libraries for subsequent experiments.

tRNA preparation

Among the 21 tRNAs used in MGC, those corresponding to our previous work (Miyachi et al., 2025, 2022) were used without modification, whereas the remaining 11 tRNAs for near-SGC and 14 for SGC were newly prepared by replacing their anticodon sequences. For these tRNAs, plasmid templates encoding the original sequences of the corresponding 21-tRNAs set were used as templates, and only the anticodon region was replaced by site-directed mutagenesis using

inverse PCR. Anticodon variants of $tRNA^{Ala}$, $tRNA^{Ser}$, and $tRNA^{Leu}$ were prepared in the same manner, with substitutions restricted to the anticodon region. The primer sets used for inverse PCR and the sequences of all tRNAs are listed in Tables S2 and S3.

Plasmids constructed by inverse PCR were used as templates for tRNA transcription. To prepare transcription templates, PCR amplification was performed using a forward primer containing a T7 promoter sequence and a reverse primer designed for run-off transcription carrying a 2'-O-methyl modification (Tables S2 and S3). PCR reactions were carried out using KOD Plus Neo DNA polymerase (Toyobo) according to the manufacturer's instructions. Following amplification, PCR products were treated with DpnI at 37 °C for 2 h to remove template plasmids. The resulting DNA fragments were purified, and DNA concentrations were determined as described above.

Of the 46 tRNAs used in this study, the majority were prepared by in vitro transcription using T7 RNA polymerase. To enable efficient transcription, tRNAs whose native sequences do not begin with guanosine at the 5'-end were redesigned as 5'-G variants unless otherwise noted. Consequently, most tRNAs were synthesized by in vitro transcription regardless of their native 5'-terminal nucleotide. Exceptions were $tRNA^{Asn}_{GUU}$, $tRNA^{Gln}$, $tRNA^{fMet}$, $tRNA^{Ile}_{GAU}$, $tRNA^{Trp}$, and $tRNA^{Pro}_{GGG}$, which were obtained by chemical synthesis (Agilent), as our previous report (Miyachi et al., 2022).

In vitro transcription reactions were performed in a mixture containing 40 mM Tris–HCl (pH 8.0), 10 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 1 U/μL T7 RNA polymerase (Takara Bio), 2 mM each NTP (ATP, GTP, CTP, and UTP), 3 mM GMP, 5 ng/μL template DNA, 2 U/μL inorganic pyrophosphatase (New England Biolabs), and 0.8 U/μL RNasin (Promega). Reactions were incubated at 37 °C for 12 h, and RNA products were purified using the PureLink RNA Mini Kit (Invitrogen). Purified tRNAs were dissolved in water and stored at −80 °C until use. RNA concentrations were determined by absorbance at 260 nm.

Calculation of mutational cost

The mutational costs for each genetic code, *Cost*(*C*), were calculated based on previously proposed formulations of genetic code optimality(Freeland and Hurst, 1998; Haig and Hurst, 1991). In this calculation, the cost reflects the expected change in amino acid property caused by single-nucleotide substitutions from the initial 21-codon sets in this study (Fig. 1B). Details of calculations are as follows. We denote each codon in the 21-codon set as β, and the amino acid encoded by β as $AA(\beta)$. For each β, we identified the set of codons reachable by a single-nucleotide substitution, denoted as $N(\beta)$. Each element of this set is represented as a pair (β', $w_{\beta\to\beta'}$), where β' is a neighboring codon box and $w_{\beta\to\beta'}$ is a weight reflecting the relative likelihood of the corresponding nucleotide substitution. These weights account for differences in substitution probabilities depending on nucleotide position and whether the substitution is a transition or transversion. The substitution weights were adopted from the previous reports(Freeland and Hurst, 1998; Omachi et al., 2023) and are summarized in Table S4.

For each possible substitution from β to β', we calculated the physicochemical distance between the corresponding amino acids, $\Delta(AA(\beta), AA(\beta'))$. This distance was evaluated using three physicochemical properties: polar requirement (PR), molecular volume (MV), and hydropathy index (HI). For each property, the distance was defined as the squared difference between the property values of the two amino acids. For example, for PR, the distance function is defined as:

$$\Delta(\alpha_1, \alpha_2) = (PR(\alpha_1) - PR(\alpha_2))^2 \quad (1)$$

The expected cost contribution of each substitution was calculated by multiplying the amino acid distance by the corresponding substitution weight. The overall mutational cost was then obtained by summing these weighted distances over all single-nucleotide substitutions from all codon boxes in the 21-codon set and dividing by the total number of substitutions considered. Formally, the cost of the genetic code *C* was defined as:

$$\mathrm{Cost}(C) := \frac{\sum_{\beta \in B_{21\ \mathrm{codons}}} \sum_{\left(\beta', w_{\beta \to \beta'}\right) \in N(\beta)} w_{\beta \to \beta'} \cdot \Delta\left(AA(\beta), AA\left(\beta'\right)\right)}{\sum_{\beta \in B_{21\ \mathrm{codons}}} |N(\beta)|} \quad (2)$$

Here, $B_{21\ codons}$ denotes the set of codon boxes fixed in the 21-box code. The denominator represents the total number of possible single-nucleotide substitutions considered, ensuring that the cost corresponds to the average expected physicochemical impact per mutation. Physicochemical property values for each amino acid were taken from the datasets adopted by Haig and Hurst (Haig and Hurst, 1991), and are listed in Table S5.

For integrated cost analysis, $\mathrm{Cost}_{\mathrm{PR}}$, $\mathrm{Cost}_{\mathrm{MV}}$, and $\mathrm{Cost}_{\mathrm{HI}}$ were combined after normalization because the three metrics have different numerical scales. For min–max normalization, each metric $m \in \{PR, MV, HI\}$ cost $Cost_{mnorm}(C)$ was normalized across the set G of candidate non-SGCs as

$$\mathrm{Cost}_{m_{\mathrm{norm}}}(C) = \frac{\mathrm{Cost}_m(C) - \min_{h \in G}(h)}{\max_{h \in G}(h) - \min_{h \in G}(h)} \quad (3)$$

The integrated min–max cost was defined as the equal-weight mean of the three normalized costs. For z-score normalization, each cost was transformed as

$$Z_m(g) = \frac{\mathrm{Cost}_{m_{\mathrm{norm}}}(C) - \mu_{m,G}}{\sigma_{m,G}} \quad (4)$$

where $\mu_{m,G}$ and $\sigma_{m,G}$are the mean and standard deviation across candidate non-SGCs. The integrated z-score cost was defined as the equal-weight mean of the three z-scores. The near-SGC was evaluated using the same normalization parameters calculated from the candidate non-SGC distribution.

Sequence analysis of random libraries

Random mutagenesis libraries (2nd PCR products; Fig. S7) were used as templates for amplicon preparation for sequencing (3rd PCR). Using 40 pM of each library as input, an approximately 500-bp region spanning the coding sequence of each reporter gene was amplified by PCR using KOD Plus Neo DNA polymerase (Toyobo). Primers used for this PCR contained both sample-identifying barcodes and Illumina adapter sequences. For GAL, primers 107–111 were used as forward primers with primer 112 as a common reverse primer. For Luc, primers 113–117 were used as forward primers with primer 118 as a common reverse primer. For mSG, primers 119–123 were used as forward primers with primer 124 as a common reverse primer. PCR products were purified using the FastGene Gel/PCR Extraction Kit (Nippon Genetics). DNA concentrations were determined by absorbance at 260 nm and subjected to Illumina MiSeq sequencing.

Paired-end FASTQ files generated by the Illumina platform were used for all analyses. Libraries were prepared from three reporter genes across five error-prone PCR conditions, with each sample corresponding to an amplicon of approximately 500 bp. Sample-identifying barcode sequences were positioned at the 5' end of read 1 (R1) of the pair-end sequencing, and the R1 reads covering approximately 300 bp from the 5' side of each amplicon were used for non-reference-rate analysis. Demultiplexing was performed using cutadapt, requiring a perfect full-length match to the barcode sequence; only read pairs in which the R1 barcode matched exactly were retained, and all other read pairs were excluded from downstream analyses. Demultiplexed FASTQ files were generated for each sample. Paired-end reads were aligned to the corresponding amplicon reference sequence (generated for each reporter gene and condition) using BWA-MEM. Alignments were processed with SAMtools to generate sorted and indexed BAM files. Position-wise base information was extracted using SAMtools mpileup, applying filters of mapping quality MAPQ ≥ 20 (−q 20) and base quality Phred Q ≥ 30 (−Q 30). Insertions and deletions were excluded from all analyses.

For each nucleotide position i, the mutation count n_i was defined as the total number of bases in the pileup that differed from the reference base, and the total coverage at that position was denoted as d_i. The position-wise error rate was calculated as:

$$\text{error rate}_i = \frac{n_i}{d_i} \tag{5}$$

The representative error rate for a given sample (i.e., a specific PCR condition) was defined as the simple average across all positions in the analyzed region (length L):

$$\text{error rate}_{\text{sample } A} = \frac{1}{L}\sum_{i=1}^{L} \text{error rate}_i \tag{6}$$

To characterize the mutation spectrum, substitution types were counted among identified mutations by tallying the frequency of each reference-to-alternate base change (e.g., A>C). In addition, for each reporter gene, primer-binding regions (19 bp) were defined based on the known primer sequences. Mutation rates and substitution spectra were calculated separately for primer-binding regions and non-primer regions. Primary comparisons were performed using profiles from non-primer regions, whereas primer-region profiles were used to assess the background mutation rate under low-mutation conditions (using KOD Plus Neo).

tRNA-free PURE system (tfPURE) preparation

All components of the laboratory-made PURE system were prepared following previously established procedures. Briefly, individual factors were expressed with histidine tags and purified by affinity chromatography, followed by gel-filtration chromatography, as described previously (Miyachi et al., 2025). Additionally, EF-Tu was subjected to two successive rounds of affinity purification in a stringent buffer to eliminate contaminating tRNAs, as described previously (Miyachi et al., 2025, 2022). Ribosomes were purified by a butyl column and sucrose cushion method and further processed by ultrafiltration to remove residual tRNAs, using the same strategy as in our previous work (Miyachi et al., 2025, 2022). The complete composition and concentrations of tfPURE used in this study are provided in Table S6.

Cell-free translation with reconstituted tRNAs

Cell-free translation reactions were performed using tfPURE supplemented with defined mixtures of in vitro-synthesized tRNAs and reporter gene DNA templates. The concentrations of DNA templates are specified in the corresponding figure legends, and the concentrations of tRNAs used for each genetic code are summarized in Table S7. Reaction mixtures were incubated at 30°C for 16 h.

Following incubation (or during incubation for mSG), reporter protein production was quantified by three different methods. For firefly luciferase (Luc) or NanoLuc, a 1 µL aliquot of the translation reaction was mixed with 30 µL of Luciferase Assay Reagent (Promega) or 50 µL of NanoLuc Assay Reagent (Promega), respectively, and luminescence was measured using a GloMax luminometer (Promega). For mSG, fluorescence signals were monitored continuously for up to 16 h using an Mx3005P real-time PCR system (Agilent Technologies) during the translation reaction. The fluorescence intensity at 0 h was subtracted from that at 16 h, and the resulting difference was used as a measure of mSG activity. For GAL (Matsuura et al., 2011), a 1 µL aliquot of the translation reaction was added to 9 µL of GAL reaction buffer (final concentrations: 50 mM HEPES–KOH, pH 7.4; 1 mM $MgCl_2$; 5 mM DTT; 5 µM TokyoGreen-βGal (Sekisui Medical, Japan)). Fluorescence was recorded at 1-min intervals for 60 min using an Mx3005P instrument with FAM detection settings. GAL activity was determined from the slope of the linear region of the fluorescence time course.

Data availability

All data are presented in the figures.

Acknowledgements

We thank Ms. Kayo Aoyama and Ayu Saito for their technical support.

Additional information

Funding

This work was supported by JST, CREST Grant Number JPMJCR20S1, Japan, and Kakenhi Grant Numbers 22H05402, 24H01111, and 23KJ0815.

Author contribution

RM and NI planned the experiments and wrote the manuscript. RM performed the experiments.

Funding

Funder	Grant reference number	Author
MEXT | Japan Science and Technology Agency (JST)	JPMJCR20S1	Ryota Miyachi

Author ORCID iDs

Norikazu Ichihashi: https://orcid.org/0000-0001-7087-2718

Additional files

Supporting information. Includes supplemental figures (Figures S1 – S12) and tables (Tables S1 – S7).

Note

This reviewed preprint has been updated to correct the corresponding author's email address.

References

Ando R, Shimozono S, Ago H, Takagi M, Sugiyama M, Kurokawa H, Hirano M, Niino Y, Ueno G, Ishidate F, *et al.* (2023) StayGold variants for molecular fusion and membrane-targeting applications. *Nature Methods* **21**:648-656 https://doi.org/10.1038/s41592-023-02085-6 | PubMed

Asahara H, Himeno H, Tamura K, Hasegawa T, Watanabe K, Shimizu M (1993) Recognition Nucleotides of Escherichia coli tRNALeu and Its Elements Facilitating Discrimination from tRNASer and tRNATyr. *Journal of Molecular Biology* **231**:219-229 https://doi.org/10.1006/jmbi.1993.1277 | PubMed

Błażej P, Wnętrzak M, Mackiewicz D, Mackiewicz P (2018) Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm. *PLOS One* **13**:e0201715 https://doi.org/10.1371/journal.pone.0201715 | PubMed

Brevet A, Chen J, Commans S, Lazennec C, Blanquet S, Plateau P (2003) Anticodon Recognition in Evolution. *Journal of Biological Chemistry* **278**:30927-30935 https://doi.org/10.1074/jbc.m302618200

Buhrman H, Van Der Gulik PTS, Klau GW, Schaffner C, Speijer D, Stougie L. (2013) A Realistic Model Under Which the Genetic Code is Optimal. *Journal of Molecular Evolution* **77**:170-184 https://doi.org/10.1007/s00239-013-9571-2 | PubMed

Calles J, Justice I, Brinkley D, Garcia A, Endy D (2019) Fail-safe genetic codes designed to intrinsically contain engineered organisms. *Nucleic acids research* **47** https://doi.org/10.1093/nar/gkz745 | PubMed

Chen J, Chen M, Zhu TF (2021) Translating protein enzymes without aminoacyl-tRNA synthetases. *Chem* **7** https://doi.org/10.1016/j.chempr.2021.01.017

Costello A, Peterson AA, Chen PH, Bagirzadeh R, Lanster DL, Badran AH (2024) Genetic Code Expansion History and Modern Innovations. *Chemical Reviews* **124**:11962-12005 https://doi.org/10.1021/ACS.CHEMREV.4C00275 | PubMed

Crick FHC (1968) The origin of the genetic code. *Journal of Molecular Biology* **38**:367-379 https://doi.org/10.1016/0022-2836(68)90392-6 | PubMed

Cui Z, Stein V, Tnimov Z, Mureev S, Alexandrov K (2015) Semisynthetic tRNA complement mediates in vitro protein synthesis. *Journal of the American Chemical Society* **137**:4404-4413 https://doi.org/10.1021/ja5131963 | PubMed

Di Giulio M. (2005) The origin of the genetic code: theories and their relationships, a review. *Biosystems* **80**:175-184 https://doi.org/10.1016/j.biosystems.2004.11.005 | PubMed

El Yacoubi B, Bailly M, De Crécy-Lagard V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Review of Genetics* **46**:69-95 https://doi.org/10.1146/annurev-genet-110711-155641 | PubMed

Firnberg E, Ostermeier M (2013) The genetic code constrains yet facilitates Darwinian evolution. *Nucleic Acids Research* **41**:7420-7428 https://doi.org/10.1093/NAR/GKT536 | PubMed

Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, Blacklow SC (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proceedings of the National Academy of Sciences of the United States of America* **100**:6353-6357 https://doi.org/10.1073/pnas.1132122100 | PubMed

Francklyn C, Schimmel P (1989) Aminoacylation of RNA minihelices with alanine. *Nature* **337**:478-481 https://doi.org/10.1038/337478a0 | PubMed

Fredens J, Wang K, de la Torre D, Funke LFH, Robertson WE, Christova Y, Chia T, Schmied WH, Dunkelmann DL, Beránek V, *et al.* (2019) Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**:514-518 https://doi.org/10.1038/s41586-019-1192-5 | PubMed

Freeland SJ, Hurst LD (1998) The genetic code is one in a million. *Journal of Molecular Evolution* **47**:238-248 https://doi.org/10.1007/pl00006381 | PubMed

Freeland SJ, Wu T, Keulmann N (2003) The case for an error-minimizing standard genetic code. *Origins of Life and Evolution of the Biosphere* **33**:457-477 https://doi.org/10.1023/A:1025771327614 | PubMed

Fujino T, Sonoda R, Higashinagata T, Mishiro-Sato E, Kano K, Murakami H (2024) Ser/Leu-swapped cell-free translation system constructed with natural/in vitro transcribed-hybrid tRNA set. *Nature Communications* **15**:1-10 https://doi.org/10.1038/s41467-024-48056-z | PubMed

Fujino T, Tozaki M, Murakami H (2020) An Amino Acid-Swapped Genetic Code. *ACS synthetic biology* **9**:2703-2713 https://doi.org/10.1021/acssynbio.0c00196 | PubMed

Gilis D, Massar S, Cerf NJ, Rooman M (2001) Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biology* **2**:RESEARCH0049 https://doi.org/10.1186/GB-2001-2-11-RESEARCH0049 | PubMed

Goodarzi H, Nejad HA, Torabi N (2004) On the optimality of the genetic code, with the consideration of termination codons. *Biosystems* **77**:163-173 https://doi.org/10.1016/j.biosystems.2004.05.031 | PubMed

Goto Y, Suga H (2021) The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. *Accounts of Chemical Research* **54**:3604-3617 https://doi.org/10.1021/acs.accounts.1c00391 | PubMed

Haig D, Hurst LD (1991) A quantitative measure of error minimization in the genetic code. *Journal of Molecular Evolution* **33**:412-417 https://doi.org/10.1007/BF02103132 | PubMed

Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, Otto P, Zimmerman K, Vidugiris G, MacHleidt T, *et al.* (2012) Engineered luciferase reporter from a deep-sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chemical Biology* **7**:1848-1857 https://doi.org/10.1021/cb3002478 | PubMed

Hibi K, Amikura K, Sugiura N, Masuda K, Ohno S, Yokogawa T, Ueda T, Shimizu Y (2020) Reconstituted cell-free protein synthesis using in vitro transcribed tRNAs. *Communications Biology* **3** https://doi.org/10.1038/s42003-020-1074-2 | PubMed

Iwane Y, Hitomi A, Murakami H, Katoh T, Goto Y, Suga H (2016) Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nature chemistry* **8**:317-325 https://doi.org/10.1038/nchem.2446 | PubMed

Jones CA, Makovsky CA, Haney AK, Dutra AC, McFeely CAL, Cropp TA, Hartman MCT (2025) Removing redundancy of the NCN codons in vitro for maximal sense codon reassignment. *Chemical Science* **16**:8932 https://doi.org/10.1039/D4SC06740A | PubMed

Katoh T, Suga H (2023) A comprehensive analysis of translational misdecoding pattern and its implication on genetic code evolution. *Nucleic Acids Research* **51**:10642-10652 https://doi.org/10.1093/NAR/GKAD707 | PubMed

Katoh T, Suga H (2022) In Vitro Genetic Code Reprogramming for the Expansion of Usable Noncanonical Amino Acids. *Annual Review of Biochemistry* **91**:221-243 https://doi.org/10.1146/annurev-biochem-040320-103817 | PubMed

Koonin E V., Novozhilov AS (2017) Origin and Evolution of the Universal Genetic Code. *Annual Review of Genetics* **51**:45-62 https://doi.org/10.1146/annurev-genet-120116-024713 | PubMed

Koonin E V., Novozhilov AS (2009) Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* **61**:99-111 https://doi.org/10.1002/IUB.146 | PubMed

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, *et al.* (2013) Genomically recoded organisms expand biological functions. *Science* **342**:357-360 https://doi.org/10.1126/SCIENCE.1241459 | PubMed

Lind PA, Arvidsson L, Berg OG, Andersson DI (2017) Variation in Mutational Robustness between Different Proteins and the Predictability of Fitness Effects. *Molecular Biology and Evolution* **34**:408-418 https://doi.org/10.1093/molbev/msw239 | PubMed

Lin-Goerke JL, Robbins DJ, Burczak JD (1997) PCR-based random mutagenesis using manganese and reduced DNTP concentration. *BioTechniques* **23**:409-412 https://doi.org/10.2144/97233BM12 | PubMed

Madore E, Florentz C, Giegé R, Sekine SI, Yokoyama S, Lapointe J (1999) Effect of modified nucleotides on Escherichia coli tRNA(Glu) structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *European Journal of Biochemistry* **266**:1128-1135 https://doi.org/10.1046/j.1432-1327.1999.00965.x | PubMed

Massey SE (2016) The neutral emergence of error minimized genetic codes superior to the standard genetic code. *Journal of Theoretical Biology* **408**:237-242 https://doi.org/10.1016/j.jtbi.2016.08.022 | PubMed

Massey SE (2015) Genetic Code Evolution Reveals the Neutral Emergence of Mutational Robustness, and Information as an Evolutionary Constraint. *Life* **5**:1301 https://doi.org/10.3390/life5021301 | PubMed

Massey SE (2008) A Neutral Origin for Error Minimization in the Genetic Code. *Journal of Molecular Evolution* **67**:510-516 https://doi.org/10.1007/s00239-008-9167-4 | PubMed

Masuda I, Hou YM (2024) A tRNA modification pattern that facilitates interpretation of the genetic code. *Frontiers in Microbiology* **15**:1415100 https://doi.org/10.3389/fmicb.2024.1415100 | PubMed

Matsuura T, Hosoda K, Ichihashi N, Kazuta Y, Yomo T (2011) Kinetic Analysis of β-Galactosidase and β-Glucuronidase Tetramerization Coupled with Protein Translation. *The Journal of Biological Chemistry* **286**:22028 https://doi.org/10.1074/JBC.M111.240168 | PubMed

McFeely CAL, Shakya B, Makovsky CA, Haney AK, Ashton Cropp T, Hartman MCT (2023) Extensive breaking of genetic code degeneracy with non-canonical amino acids. *Nature Communications* **14**:5008 https://doi.org/10.1038/s41467-023-40529-x | PubMed

Miyachi R, Masuda K, Shimizu Y, Ichihashi N (2025) Simultaneous in vitro expression of minimal 21 transfer RNAs by tRNA array method. *Nature Communications* **16**:7418 https://doi.org/10.1038/s41467-025-62588-y | PubMed

Miyachi R, Shimizu Y, Ichihashi N (2022) Transfer RNA Synthesis-Coupled Translation and DNA Replication in a Reconstituted Transcription/Translation System. *ACS synthetic biology* **11**:2791-2799 https://doi.org/10.1021/acssynbio.2c00163 | PubMed

Murakami H, Ohta A, Ashigai H, Suga H (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3**:357-359 https://doi.org/10.1038/nmeth877 | PubMed

Novozhilov AS, Wolf YI, Koonin E V (2007) Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biology Direct* **2**:24 https://doi.org/10.1186/1745-6150-2-24 | PubMed

Nyerges A, Vinke S, Flynn R, Owen S V., Rand EA, Budnik B, Keen E, Narasimhan K, Marchand JA, Baas-Thomas M, *et al.* (2023) A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**:720 https://doi.org/10.1038/S41586-023-05824-Z | PubMed

Omachi Y, Saito N, Furusawa C (2023) Rare-event sampling analysis uncovers the fitness landscape of the genetic code. *PLOS Computational Biology* **19**:e1011034 https://doi.org/10.1371/JOURNAL.PCBI.1011034 | PubMed

Passioura T, Liu W, Dunkelmann D, Higuchi T, Suga H (2018) Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code. *Journal of the American Chemical Society* **140**:11551-11555 https://doi.org/10.1021/jacs.8b03367 | PubMed

Pines G, Winkler JD, Pines A, Gill RT (2017) Refactoring the Genetic Code for Increased Evolvability. *mBio* **8** https://doi.org/10.1128/MBIO.01654-17 | PubMed

Robertson WE, Funke LFH, de la Torre D, Fredens J, Elliott TS, Spinck M, Christova Y, Cervettini D, Böge FL, Liu KC, *et al.* (2021) Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**:1057-1062 https://doi.org/10.1126/SCIENCE.ABG3029 | PubMed

Robertson WE, Rehm FBH, Spinck M, Schumann RL, Tian R, Liu W, Gu Y, Kleefeldt AA, Day CF, Liu KC, *et al.* (2025) Escherichia coli with a 57-codon genetic code. *Science* https://doi.org/10.1126/SCIENCE.ADY4368 | PubMed

Rozhoňová H, Martí-Gómez C, McCandlish DM, Payne JL (2024) Robust genetic codes enhance protein evolvability. *PLOS Biology* **22**:e3002594 https://doi.org/10.1371/JOURNAL.PBIO.3002594 | PubMed

Sampson JR, Saks ME (1993) Contributions of discrete tRNASer domains to aminoacylation by E.coli seryl-tRNA synthetase: a kinetic analysis using model RNA substrates. *Nucleic Acids Research* **21**:4467-4475 https://doi.org/10.1093/nar/21.19.4467 | PubMed

Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17**:125 https://doi.org/10.1186/s12859-016-0976-y | PubMed

Shandell MA, Tan Z, Cornish VW (2021) Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* **60**:3455-3469 https://doi.org/10.1021/ACS.BIOCHEM.1C00286 | PubMed

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19** https://doi.org/10.1038/90802 | PubMed

Shulgina Y, Eddy SR (2021) A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* **10** https://doi.org/10.7554/eLife.71402 | PubMed

Su D, Lieberman A, Lang BF, Simonović M, Söll D, Ling J (2011) An unusual tRNAThr derived from tRNAHis reassigns in yeast mitochondria the CUN codons to threonine. *Nucleic Acids Research* **39**:4866-4874 https://doi.org/10.1093/nar/gkr073 | PubMed

Tamura K, Himeno H, Asahara H, Hasegawa T, Shimizu M (1992) In vitro study of E.coli tRNAArg and tRNALys identity elements. *Nucleic Acids Research* **20**:2335-2339 https://doi.org/10.1093/nar/20.9.2335 | PubMed

Terasaka N, Hayashi G, Katoh T, Suga H (2014) An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nature Chemical Biology* **10**:555-557 https://doi.org/10.1038/nchembio.1549 | PubMed

Wnętrzak M, Błazej P, Mackiewicz D, Mackiewicz P. (2018) The optimality of the standard genetic code assessed by an eight-objective evolutionary algorithm. *BMC Evolutionary Biology* **18**:192 https://doi.org/10.1186/s12862-018-1304-0 | PubMed

Woese CR (1965) On the evolution of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **54**:1546 https://doi.org/10.1073/PNAS.54.6.1546 | PubMed

Woese CR, Dugre DH, Saxinger WC, Dugre SA (1966) The molecular basis for the genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **55**:966-974 https://doi.org/10.1073/PNAS.55.4.966 | PubMed

Wong JTF (2005) Coevolutlon theory of genetic code at age thirty. *BioEssays* **27**:416-425 https://doi.org/10.1002/BIES.20208 | PubMed

Zeng Y, Wang W, Liu WR (2014) Toward Reassigning the Rare AGG Codon in Escherichia coli. *Chembiochem : a European journal of chemical biology* **15**:1750 https://doi.org/10.1002/cbic.201400075 | PubMed

Zürcher JF, Robertson WE, Kappes T, Petris G, Elliott TS, Salmond GPC, Chin JW (2022) Refactored genetic codes enable bidirectional genetic isolation. *Science* **378**:516-523 https://doi.org/10.1126/SCIENCE.ADD8943 | PubMed

Peer reviews

Reviewer #1 (Public review):

[Editors' note: This version has been assessed by the Reviewing Editor without further input from the original reviewers. The authors have addressed the comments raised in the previous round of review satisfactorily and toned down the comments as advised.]

In this manuscript, the authors investigate the relationship between genetic codes and their robustness to single-point mutations. They construct ten alternative genetic codes by reassigning nine codons to Leu, Ser, or Ala, and assess mutational robustness using three

reporter proteins subjected to error-prone PCR. This represents an interesting experimental approach to addressing the hypothesis that the standard genetic code is optimized for mutational robustness.

https://doi.org/10.7554/eLife.111164.2.sa3

Reviewer #2 (Public review):

The study addresses the long-standing question in molecular biology and genetics: why has nature selected the current genetic code (SGC, or standard genetic code)? The authors have tested 'error minimization theory', one of the prevailing hypotheses to explain this. Their approach is to create a minimum genetic code (MGC) and its variants (3^9 theoretical possible codes). Using three parameters to quantify the effect of mutations (Polarity, volume, and hydropathy), they computationally test the cost of these genetic codes (3^9) by simulations. Finally, they test this cost experimentally using an in vitro translation system with 10 select genetic code variants with a range of costs (low to high). They use three randomly mutated reporter genes for this purpose - beta-galactosidase, luciferase, and mSG. They find no correlation between the cost of the genetic code and the reporters' output. Based on these observations, they suggest that error-minimization theory may not explain the current egocentric code.

The question they are asking is very exciting, and their approach is solid. The authors are very careful in their analyses and conclusions.

https://doi.org/10.7554/eLife.111164.2.sa2

Reviewer #3 (Public review):

Summary:

In this manuscript, Miyachi and Ichihashi investigate whether the arrangement of the genetic code affects mutational robustness. Using an in vitro minimal genetic code with vacant codons, they constructed 10 non-standard genetic codes by reassigning Ala, Ser, and Leu, generating codes with replacement costs that were generally higher than those of the standard genetic code across several amino acid property measures. They then tested how random mutations affected the activity of reporter proteins translated under these altered codes. Although error minimization theory predicts that higher-cost codes should make mutations more harmful, the authors report that protein function declined to a similar extent across all codes examined, suggesting that mutational robustness remains largely unchanged within the range of genetic code alterations tested here.

Strengths:

This is an interesting study that investigates one of the most fundamental and intriguing questions in molecular evolution: the emergence of the genetic code, which is nearly universal across nature. The in vitro approach is a powerful aspect of the work and provides an opportunity to examine this phenomenon experimentally at a depth that has previously been inaccessible.

https://doi.org/10.7554/eLife.111164.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***

> ***Reviewer #1 (Public review):***
>
> *In this manuscript, the authors investigate the relationship between genetic codes and their robustness to single-point mutations. They construct ten alternative genetic codes by reassigning nine codons to Leu, Ser, or Ala, and assess mutational robustness using three reporter proteins subjected to error-prone PCR. This represents an interesting experimental approach to addressing the hypothesis that the standard genetic code is optimized for mutational robustness.*

We sincerely thank the reviewer for the positive evaluation of our experimental approach. We are encouraged that the reviewer recognizes the value of constructing multiple non-standard genetic codes *in vitro* and using them to experimentally examine the relationship between genetic code arrangement and mutational robustness. In the revised manuscript, we have further clarified the scope of our experimental system and the interpretation of the results, particularly emphasizing that our conclusions concern the mutational robustness of individual reporter protein activity measured in an *in vitro* translation system.

> *Major comment:*
>
> *While I find the experimental design valuable, I am not fully convinced by the authors' conclusion that "alterations of the genetic code within the ranges explored in this study have no significant effect on mutational robustness". The current analysis is based on the functional output of three individual reporter proteins. Given that cellular systems involve far more complex interactions, it would be more appropriate to limit this conclusion to mutational robustness at the level of individual protein activity, rather than making broader generalizations.*

We thank the reviewer for this important comment. We agree that our original wording was broader than what can be directly supported by the present experiments. Because our analysis is based on the functional outputs of three individual reporter proteins translated in a reconstituted *in vitro* system, the results do not directly address mutational robustness at the level of the cellular system, protein interaction networks, or organismal fitness.

Accordingly, we have revised the manuscript to limit our conclusion to the mutational robustness of individual reporter protein activity. In the revised Abstract, Results, and Discussion, we now state that within the experimentally tested range of non-standard genetic codes, we did not detect a dependence of the mutation-induced decrease in reporter protein activity on mutational cost. We have also added a statement in the Discussion noting that cellular systems involve many additional layers, including protein–protein interactions, metabolic networks, quality-control systems, and growth selection, and that whether genetic code arrangement affects robustness at these higher biological levels remains an important question for future work.

Specifically, we have added this explanation and the new experiment to the revised manuscript as follows.

Abstract

“This result provides direct experimental evidence that mutational robustness does not significantly change in individual reporter protein activity when the genetic code is altered within the range of mutational cost tested in this study…”

Introduction

"Random mutations decreased reporter protein function at similar levels across all genetic codes examined, implying that alterations of the genetic code within the ranges explored in this study have no significant effect on mutational robustness of individual protein activity."

Result

"Taken together, these results indicate that mutational robustness of individual reporter protein function did not substantially differ among the genetic codes…"

Discussion

"…suggesting that mutational robustness of protein activity remained largely unchanged within at least the ranges of mutational cost tested in this study. It should be noted that this conclusion is limited to the activity of individual reporter proteins translated in a reconstituted *in vitro* system. Therefore, whether similar trends would be observed at the level of cellular fitness or long-term evolution remains an open question."

> *Specific comments*
>
> *(1) tRNA modification and expression efficiency (Page 5, line 131)*

The authors attribute the observed inefficiency to the lack of chemical modifications in the tRNAs used. However, gene expression efficiency can also be strongly influenced by DNA sequence design. To better support this claim, it would be helpful to compare luciferase activity when expressed using native E. coli tRNAs. This comparison could clarify whether the observed effects are due to tRNA modification status or other sequence-dependent factors.

We thank the reviewer for this important suggestion. We agree that the translation efficiency of NanoLuc templates with 21-, 32-, and 46-codons may be affected not only by the chemical modification of tRNAs but also by sequence-dependent factors, such as codon context and mRNA structure.

To examine this possibility, we performed an additional comparison using native *E. coli* tRNAs in the tfPURE system. When the NanoLuc templates encoded with 21, 32, or 46 codons were translated using native *E. coli* tRNAs, the observed luminescence values were 1.2×10^{10}, 0.78×10^{10}, and 0.60×10^{10}, respectively. Thus, the 46-codon NanoLuc template showed lower activity than the 21- and 32-codon templates even with native tRNAs, indicating that sequence-dependent effects indeed contribute to translation efficiency.

However, the difference among these templates with native *E. coli* tRNAs was within approximately two-fold. This effect was much smaller than the marked decrease observed when the 46-codon template was translated using the *in vitro* prepared 46 tRNAs SGC system. Therefore, while sequence-dependent effects cannot be excluded, the inefficient translation in the reconstructed 46 tRNAs SGC is likely to be mainly attributable to the limited functionality of unmodified tRNAs decoding NNA codons.

We have revised the manuscript to clarify this interpretation and have added the new comparison using native *E. coli* tRNAs.

"We also examined whether the lower translation efficiency of the 46-codon NanoLuc template could be explained by sequence-dependent effects, such as codon context or mRNA structure. When the 21-, 32-, and 46-codon NanoLuc templates were translated using native *E. coli* tRNAs in the tfPURE system (Figure 1–figure supplement 2), the 46-codon template showed lower activity than the 21- and 32-codon templates; however, this difference was within approximately two-fold. Accordingly, we decided to use only the 32 codons used in near-SGC (i.e., excluding NNA codons) in the subsequent construction of non-standard genetic codes."

> *(2) Discrepancy between expression level and activity (Figure S7 vs Figure S8).*
>
> *Although GAL expression levels appear similar across different genetic codes (Figure S7), their activities differ substantially (Figure S8), even in the low-mutation library. This discrepancy warrants further investigation. Possible explanations include differences in protein folding efficiency or translational error rates, as mentioned by the authors in the main text.*
>
> *To address this, the authors could analyze the protein products using mass spectrometry. If this is not feasible due to low expression levels, alternative approaches such as SDS-PAGE (e.g., with radiolabeling or Western blotting) would still provide valuable information. Additionally, comparing activity after in vitro refolding could help distinguish between folding defects and sequence-level errors. While I understand that the primary aim of this study is to compare mutational robustness across genetic codes, discussing these observations would significantly enhance the mechanistic insight of the work.*

We agree that the discrepancy between similar GAL expression levels and different GAL activities across genetic codes is important for interpreting the results.

In our experiment, GAL protein amounts were quantified using a C-terminal HiBiT tag. Because the HiBiT tag was fused to the C-terminus of GAL, this assay indicates that the amount of C-terminally completed GAL products did not differ substantially among genetic codes. However, we agree that this assay does not evaluate the sequence fidelity, amino acid misincorporation patterns, or folding state of the translated products. Therefore, the observed differences in GAL activity despite similar HiBiT signals may reflect genetic code-dependent differences in translational error rates, amino acid misincorporation, protein folding efficiency, or other effects on the fraction of catalytically active protein.

We have revised the Discussion to explicitly describe this interpretation and to clarify that detailed mechanistic dissection of these baseline activity differences, for example by mass spectrometry, SDS-PAGE/Western blotting, or refolding analysis, is an important future direction but beyond the scope of the present study. We also clarified that the main analysis in this study uses the ratio of activity from the high-mutation library to that from the corresponding low-mutation library within each genetic code.

We have added this explanation to the revised manuscript as follows.

“Although protein amounts quantified by the HiBiT tag were comparable among genetic codes, GAL activities differed substantially. This indicates that the activity differences among genetic codes were not primarily attributable to differences in the amount of C-terminally completed translation products. The HiBiT assay does not provide information on the fraction of catalytically active protein, including sequence fidelity or folding state, and therefore cannot distinguish among these possibilities. Detailed characterization of translated products by mass spectrometry would provide further mechanistic insight into how individual non-SGCs affect protein quality. However, the primary objective of the present study was to compare mutation-dependent activity loss across genetic codes. Therefore, we evaluated this effect by normalizing the activity of the high-mutation library to that of the corresponding low-mutation library within each genetic code.”

> *(3) Protein expression analysis for additional reporters.*
>
> *Since protein expression levels are critical for interpreting reporter activity, similar analyses should also be performed for luciferase (Luc) and mSG in both high- and low-mutation libraries. This would ensure that differences in activity are not confounded by variations in protein abundance.*

We agree that protein abundance is an important factor for interpreting reporter activity. In this study, we performed HiBiT-based protein quantification for GAL because GAL showed the largest variation in absolute activity among genetic codes, even in the low-mutation library. This analysis showed that the amount of C-terminally completed GAL products was broadly comparable among genetic codes and between low- and high-mutation libraries, indicating that the observed GAL activity differences were not primarily attributable to differences in total protein abundance.

For all three reporters, our main analysis was based on the ratio of activity from the high-mutation library to that from the corresponding low-mutation library within each genetic code. This normalization was intended to evaluate mutation-dependent activity loss while reducing the influence of code-specific baseline differences in expression level or protein quality. We believe that the data are sufficient to evaluate the effect of mutations on protein activities. Nevertheless, we agree that protein quantification for Luc and mSG would provide useful information regarding variation in the baseline levels of reporter activity, and this is an important direction for future work.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *The study addresses the long-standing question in molecular biology and genetics: why has nature selected the current genetic code (SGC, or standard genetic code)? The authors have tested 'error minimization theory', one of the prevailing hypotheses to explain this. Their approach is to create a minimum genetic code (MGC) and its variants (3^9 theoretical possible codes). Using three parameters to quantify the effect of mutations (Polarity, volume, and hydropathy), they computationally test the cost of these genetic codes (3^9) by simulations. Finally, they test this cost experimentally using an in vitro translation system with 10 select genetic code variants with a range of costs (low to high). They use three randomly mutated reporter genes for this purpose - beta-galactosidase, luciferase, and mSG. They find no correlation between the cost of the genetic code and the reporters' output. Based on these observations, they suggest that error-minimization theory may not explain the current egocentric code.*
>
> *The question they are asking is very exciting, and their approach is solid. The authors are very careful in their analyses and conclusions.*

We sincerely thank the reviewer for the positive assessment of our study and for the helpful suggestions. We are encouraged that the reviewer found the question exciting and the approach solid. In the revised manuscript, we have clarified the rationale for using the MGC/near-SGC framework, added further analyses and explanations of the mutational cost calculations, and revised the wording of our conclusions to more explicitly define the scope and limitations of the present experimental system.

> *(1) The rationale for using MGC instead of SGC: It is unclear why the authors rely on the MGC for this analysis when the central question concerns the SGC. If the goal is to evaluate whether the SGC minimizes mutational cost, a more direct approach would be to generate alternative variants of the SGC itself and compare their mutational cost distributions. At present, it is difficult to assess whether conclusions drawn from this comparison are fully relevant to the stated biological question.*

We thank the reviewer for this important comment. We agree that directly constructing alternative variants of the SGC by changing amino acid assignment from SGC would be the most straightforward approach to testing whether the SGC minimizes mutational cost. However, this approach is currently not feasible in our reconstituted translation system for two reasons.

First, our attempt to construct a 46-tRNA SGC-like system revealed that translation using the 46-codon NanoLuc template was approximately 100-fold less efficient than translation using the MGC or near-SGC (Fig. 1). This low activity likely reflects inefficient decoding of NNA codons by *in vitro*-prepared tRNAs, which lack native post-transcriptional modifications. Because this system did not provide sufficient translational activity for systematic reporter assays, we restricted subsequent experiments to the 32-codon near-SGC framework, excluding NNA codons. We now describe this technical limitation more explicitly in the revised manuscript.

Second, the MGC framework provides vacant codons that can be reassigned by adding anticodon-variant tRNAs. This feature is essential for constructing multiple genetic code variants in parallel under controlled *in vitro* conditions. We, therefore, constructed the near-SGC-based non-SGC by adding each tRNA variant to the MGC as an experimentally tractable model system to verify whether differences in genetic code arrangement affect mutation-induced decreases in reporter protein activity.

We have added this explanation to the revised manuscript as follows.

"We first established a minimal genetic code, composed of 21 tRNAs with vacant codons, which allows multiple alternative codon assignments to be introduced under otherwise comparable translation conditions."

Despite this technical limitation, we believe that the central conclusion of this study—that mutational robustness in individual reporter protein activity does not change significantly when the genetic code is altered within the range of mutational costs tested here—remains well-supported by the present results.

> *(2) The mutational cost analysis appears biologically oversimplified because all amino acid substitutions are treated equivalently. The analysis assumes that all mutations contribute equally to fitness consequences, which does not reflect biological reality. In natural proteins, the impact of an amino acid substitution depends strongly on its structural and functional context. For example, substitutions affecting catalytic residues, ligand-binding interfaces, phosphorylation sites, or other regulatory motifs can severely impair protein function even when associated changes in polarity, hydropathy, or volume are minimal. Conversely, substitutions in structurally permissive or functionally dispensable regions may have little or no measurable effect despite larger physicochemical differences. Therefore, changes in polarity, hydropathy, and volume alone do not necessarily predict functional consequences.*

We agree that the mutational cost used in this study is a simplified measure and does not capture the full biological complexity of amino acid substitutions. As the reviewer pointed out, the functional consequence of a substitution depends strongly on its structural and functional context, including whether the affected residue is involved in catalysis, ligand binding, protein–protein interactions, regulatory motifs, folding, or structurally permissive regions.

In this study, we used physicochemical-property-based mutational costs because this type of definition has been widely used in classical formulations of the error minimization theory. Our aim was therefore not to construct a comprehensive predictor of protein fitness effects, but to experimentally test whether the conventional theoretical cost metrics used to discuss genetic code optimality are reflected in the average mutation-induced decrease in reporter protein activity. We have now clarified this rationale in the revised manuscript.

"It should be noted that this conclusion is limited to the activity of individual reporter proteins translated in a reconstituted *in vitro* system. Therefore, whether similar trends

would be observed at the level of cellular fitness or long-term evolution remains an open question."

> *(3) It is not clear why they increased the concentration of the two tRNAs in near-SGC. Have they maintained the same tRNA concentrations in experiments explained in Fig 5 for all 10 genetic codes tested?*

We apologize that the rationale for increasing the concentrations of $tRNA^{Val}_{CAC}$ and $tRNA^{Arg}_{CCU}$ was not sufficiently clear in the original manuscript. As we wrote in the previous manuscript, "To improve translation efficiency with near-SGC, we focused on two tRNA concentrations ($tRNA^{Val}_{CAC}$ and $tRNA^{Arg}_{CCU}$), which were suggested to have low activities in a previous study (Iwane et al., 2016)," we tested whether increasing their concentrations would improve translation efficiency. As shown in Figure 1–figure supplement 1, NanoLuc activity increased as the concentrations of these two tRNAs were raised and used at 100 ng/μL for $tRNA^{Val}_{CAC}$ and $tRNA^{Arg}_{CCU}$ in the optimized near-SGC, referred to as near-SGC (RV), and in all subsequent experiments. Additional anticodon-variant tRNAs required for each non-SGC were used at optimized concentrations determined from Figure 2–figure supplement 1. For each genetic code, the same tRNA composition and concentrations were used for the low- and high-mutation libraries (See Supplementary Table S7). To clarify this point, we added the sentence, "The increased concentrations of these two tRNAs were used in all the subsequent experiments," in the corresponding part.

> ***Reviewer #3 (Public review):***
>
> *In this manuscript, Miyachi and Ichihashi investigate whether the arrangement of the genetic code affects mutational robustness. Using an in vitro minimal genetic code with vacant codons, they constructed 10 non-standard genetic codes by reassigning Ala, Ser, and Leu, generating codes with replacement costs that were generally higher than those of the standard genetic code across several amino acid property measures. They then tested how random mutations affected the activity of reporter proteins translated under these altered codes. Although error minimization theory predicts that higher-cost codes should make mutations more harmful, the authors report that protein function declined to a similar extent across all codes examined, suggesting that mutational robustness remains largely unchanged within the range of genetic code alterations tested here.*
>
> *Strengths:*
>
> *This is an interesting study that investigates one of the most fundamental and intriguing questions in molecular evolution: the emergence of the genetic code, which is nearly universal across nature. The in vitro approach is a powerful aspect of the work and provides an opportunity to examine this phenomenon experimentally at a depth that has previously been inaccessible.*
>
> *Weaknesses:*
>
> *However, the authors' use of random mutation libraries has certain limitations that prevent the study from realizing its full potential to uncover the mechanisms governing the molecular evolution of the genetic code.*

We sincerely thank the reviewer for the positive evaluation of our study and for recognizing the strength of the *in vitro* approach. We are encouraged that the reviewer considers this system a powerful way to experimentally address the emergence of the genetic code.

We also appreciate the reviewer's constructive comments regarding the limitations of random mutation libraries. We agree that pooled random libraries do not allow us to assign functional effects to individual mutations or to fully uncover the molecular mechanisms underlying mutational robustness. In the revised manuscript, we therefore clarify that our

conclusions concern the library-averaged effects of random mutations on individual reporter protein activity, rather than the effects of specific mutations or cellular-level fitness. To address this limitation, we have added explanations of the scope and limitations of the present approach.

> *(1) Statistical analyses are missing for several of the manuscript's main claims. This issue applies throughout the paper, including, but not limited to, Figures 1D, 2B, 4B-D, and 5B.*

We thank the reviewer for this important comment. We agree that statistical analyses are necessary to support the major claims of the manuscript. We have therefore added statistical analyses appropriate for the purpose and experimental design of each figure.

For Fig. 1D, we performed one-way ANOVA followed by Tukey's post hoc test on NanoLuc activity to compare translation efficiencies among the MGC, near-SGC, near-SGC (RV), and SGC conditions. This analysis showed a significant overall difference among conditions (one-way ANOVA, $p < 0.0001$). Tukey's post hoc test showed that near-SGC was significantly lower than MGC, that near-SGC (RV) significantly improved near-SGC translation, and that near-SGC (RV) was not significantly different from MGC. In contrast, the 46-tRNA SGC remained significantly less efficient than near-SGC (RV). We have summarized the major comparisons in Supplementary Table S8.

For Fig. 2B, we compared NanoLuc activity between the 21-code control and the corresponding 21+1-code condition for each codon reassignment using Welch's t-test on luminescence. This analysis was added to statistically support whether each anticodon-variant tRNA increased NanoLuc translation from the corresponding reassigned template. The statistical results are summarized in Supplementary Table S9.

For Fig. 4B–D, we converted mutation rates per base to estimated numbers of mutations per gene and performed Spearman's rank correlation analysis to evaluate whether reporter activity decreased monotonically with increasing mutational load. This analysis showed strong negative monotonic trends between mutation rate (estimated mutation number) and reporter activity for all three reporters ($\rho = -0.90$ to -1.00), supporting that the random mutation libraries reduced protein activity in a mutation-load-dependent manner.

For Fig. 5B, replicate-level data were available for GAL, and we therefore performed two-way ANOVA using genetic code and mutation level as factors. This analysis detected significant main effects of genetic code and mutation level, indicating that GAL activity differed among genetic codes and decreased in the high-mutation library. However, no significant interaction between genetic code and mutation level was detected, indicating that the magnitude of mutation-induced activity reduction was not strongly code-dependent under the conditions examined.

Finally, because the central claim of Fig. 5C, 5E, and 5G is that mutational cost does not systematically predict mutation-induced activity loss, we performed Spearman's rank correlation analysis between each mutational cost metric and the high-/low-mutation activity ratio. No significant correlations were detected for any reporter or cost metric (Spearman's $\rho = -0.23$ to 0.25), supporting the conclusion that mutational cost did not show a detectable monotonic relationship with mutation-induced activity loss within the tested range.

We have added these statistical analyses to the revised manuscript. The following sentences were added to the figure legends:

Fig. 1

"Statistical comparisons in (D) were performed using one-way ANOVA followed by Tukey's post hoc test on NanoLuc activity; major comparisons are summarized in Table S8."

Fig. 2

"For each template, NanoLuc activity in the 21-code and corresponding 21+1-code conditions was compared using Welch's t-test on luminescence. Statistical results are summarized in Table S9."

Fig. 4

"Spearman's rank correlation coefficients were ρ = −0.90 for GAL, ρ = −1.00 for Luc, and ρ = −1.00 for mSG"

Fig. 5

"For GAL activity in (B), two-way ANOVA was performed using genetic code and mutation level as factors. Significant main effects of genetic code and mutation level were detected (both $p < 0.0001$), whereas their interaction was not significant. For (C), (E), and (G), Spearman's rank correlation analysis was performed between each mutational cost metric and the high-/low-mutation activity ratio. Statistical details are summarized in Table S10."

> *(2) In Figure 2A, the authors modify the NanoLuc gene by reassigning Ala, Leu, or Ser to new codons and elegantly show that the in vitro availability of the corresponding tRNAs is important for protein function. However, the functional importance of the specific modified positions within NanoLuc is not clear. As a result, it is difficult to determine what the expected consequences of these codon changes should be, which in turn limits the interpretation of the observed changes in protein activity. To improve the interpretability of this experiment, the authors should report exactly how many codons were modified in each variant and, ideally, examine the effect of progressively increasing the number of reassigned codons.*

We agree that the exact positions and numbers of codon replacements should be clearly reported. In the revised manuscript, we have added a list of the modified amino acid positions. In brief, two Ala codons, three Ser codons, or four Leu codons were replaced with the target vacant codon; the modified positions were Ala16 and Ala120, Ser31, Ser49, and Ser150, and Leu32, Leu67, Leu144, and Leu170, respectively.

We also agree that progressively increasing the number of reassigned codons would provide additional mechanistic insight. However, the purpose of Fig. 2 was to test whether each vacant codon could be decoded by the corresponding anticodon-variant tRNA to produce functional NanoLuc, rather than to analyze the positional contribution of each replacement. We previously performed such progressive codon replacement analysis for one reassigned codon, ACG, in a related study (Miyachi et al., 2025), and the results supported the same qualitative interpretation. Although we did not repeat this progressive analysis for all codons in the present study, we expect that the qualitative interpretation of Fig. 2 would not be substantially changed.

We have revised the figure text to clarify the scope of the experiment and added the detailed codon replacement information.

"(A) Schematic illustration of reassignment experiments. Translation with the original MGC and NanoLuc template is shown at the top for comparison. An example of Ala reassignment to the UUG codon is shown at the bottom. In this example, three Ala codons in the NanoLuc sequence were replaced with one type of vacant codon (e.g., UUG), generating a 21 + 1 (UUG-Ala) codon set. Similar reassignment experiments were performed for three amino acids (Ala, Ser, and Leu) and nine vacant codons. Specifically, two Ala codons (Ala16 and Ala120), three Ser codons (Ser31, Ser49, and Ser150), or four Leu codons (Leu32, Leu67, Leu144, and Leu170) were replaced."

> *(3) The calculations presented in Figure 3 raise an interesting conceptual question: why does the near-standard genetic code not exhibit the lowest cost? One possible explanation is that the standard genetic code evolved under multiple competing constraints and is therefore not expected to be optimal for any single cost metric, while still achieving strong overall performance. In this context, it would be informative if the authors combined the three cost measures into a single integrated index and examined whether the near-SGC performs more favorably when all three dimensions are considered together. Such an analysis could add important depth to the study.*

We agree that the near-SGC is not necessarily expected to minimize each individual cost metric, because the standard genetic code may reflect multiple competing physicochemical, translational, biosynthetic, and evolutionary constraints rather than optimization of a single property.

To address this point, we added an integrated cost analysis combining the three physicochemical cost metrics, $\mathrm{Cost_{PR}}$, $\mathrm{Cost_{MV}}$, and $\mathrm{Cost_{HI}}$. Because these three metrics have different numerical scales, we normalized each metric before integration. We used two types of integrated indices.

First, for each metric $m \in \{PR, MV, HI\}$, we calculated a min–max normalized cost,

$$Cost_{m_{norm}}(C) = \frac{Cost_m(C) - min_{h\in G}(h)}{max_{h\in G}(h) - min_{h\in G}(h)}$$

Where G denotes the set of 19,683 candidate non-SGCs generated by assigning Ala, Ser, or Leu to the nine vacant codon boxes. We then defined the integrated min–max cost as

$$Cost_{\mathrm{int,minmax}}(C) = \frac{1}{3}\left(Cost_{PR_{norm}}(C) + Cost_{MV_{norm}}(C) + Cost_{HI_{norm}}(C)\right)$$

Second, we calculated a z-score-normalized cost for each metric,

$$Z_m(C) = \frac{Cost_{m_{norm}}(C) - \mu_{m,G}}{\sigma_{m,G}}$$

Where $\mu_{m,G}$ and $\sigma_{m,G}$ are the mean and standard deviation of $Cost_{m_{norm}}$ across the candidate non-SGCs. The integrated z-score cost was then defined as

$$C_{\mathrm{intz}}(C) = \frac{1}{3}(Z_{PR}(C) + Z_{MV}(C) + Z_{HI}(C)).$$

Using both integrated indices, the near-SGC ranked first when compared with all 19,683 candidate non-SGCs; in other words, no candidate non-SGC showed a lower integrated cost than the near-SGC. The integrated min–max cost of the near-SGC was 0.01525, whereas the lowest value among candidate non-SGCs was 0.12301. Similarly, the integrated z-score cost of the near-SGC was −2.47947, whereas the lowest candidate value was −1.90838.

We have added this integrated cost analysis as Supplementary Figure 5–figure supplement 7. We have also revised the Discussion to note that the near-SGC does not necessarily minimize every individual physicochemical cost, but performs most favorably when PR, MV, and HI are considered comprehensively. This result is consistent with the idea that the standard genetic code may represent a compromise among multiple constraints rather than optimization of a single physicochemical property.

“We consider that the cost ranges examined in this study represent substantial fractions, especially for MV and HI. Although the near-SGC did not necessarily exhibit the lowest cost for each individual physicochemical metric, this does not mean that it is unfavorable in the multidimensional cost space. Because the SGC may reflect a balance among multiple physicochemical constraints rather than optimization of a single property, we also calculated integrated cost indices by combining Cost_PR, Cost_MV, and Cost_HI after min–max

normalization or z-score normalization. In both integrated indices, the near-SGC showed the lowest overall cost when compared with all 19,683 candidate non-SGCs (Figure 5–figure supplement 7), indicating that no candidate non-SGC exhibited a lower combined cost than the near-SGC when the three physicochemical properties were considered comprehensively."

> *(4) It is difficult to assess the consequences of the random mutations presented in Figure 4 on reporter gene function based solely on the reported "error rate/base" parameter. In particular, the x-axis in Figure 4B should be converted into the estimated number of mutations per gene. This would make the results more intuitive and would allow the reader to better evaluate the expected degree of disruption to protein function.*

We agree that the mutation rate per base alone does not provide an intuitive sense of the expected mutational burden for each reporter gene. We therefore added a second x-axis to Fig. 4B–D showing the estimated number of mutations per gene. This value was calculated by multiplying the mutation rate per base by the coding sequence length of each reporter gene.

We retained the original mutation rate per base axis to preserve the direct link to the sequencing-based mutation rate measurement, while adding the estimated mutations per gene axis to improve interpretability. We have revised the figure and figure 4 legend accordingly.

"The lower x-axis indicates the estimated number of mutations per gene, calculated by multiplying the mutation rate per base by the coding sequence length of each reporter gene."

> *(5) A central limitation of the random mutagenesis libraries used in Figure 5, which also underlie one of the manuscript's main claims, is that the exact mutations and their distribution across the reporter genes are not reported. In addition, protein activity is measured only at the level of the entire library, without directly linking individual mutations to their functional consequences. This substantially limits mechanistic interpretation. In my view, this issue can only be addressed convincingly if the authors test a set of defined variants carrying specific mutations and directly evaluate their functional effects.*
>
> *(6) Related to the previous point, in Figures 5C, 5E, and 5G, the authors present the ratio between low-mutation-rate and high-mutation-rate libraries. However, because each library contains a different collection of mutations, it is unclear what can be inferred from these comparisons. To overcome this limitation, the authors should assess the effects of altered genetic codes on specific, defined mutations rather than on heterogeneous mutation pools alone.*
>
> *(7) Along the same lines, in Figures 5C, 5E, and 5G, it is unclear why the effects of random mutations would be expected to correlate with the three calculated cost metrics, given that the positions, identities, and functional relevance of the mutations within the genes are not known. Without this information, the biological meaning of these correlations remains difficult to evaluate.*

We agree that using pooled random mutation libraries does not allow us to directly link individual mutations to their functional consequences. We also agree that testing defined variants carrying specific mutations would provide a more direct and mechanistic understanding of how each genetic code affects the functional impact of particular amino acid substitutions. However, the purpose of the present study was different from such a defined-variant analysis. Our aim was to experimentally test whether the conventional mutational cost metrics used in error minimization theory predict the average effect of random mutational loads on protein activity. Because these theoretical costs are themselves defined as average expected physicochemical effects over many possible single-nucleotide substitutions, we reasoned that pooled random mutation libraries provide an appropriate

first experimental framework to evaluate whether such average-cost metrics are reflected in the average functional output of translated proteins.

We agree that low- and high-mutation libraries do not contain identical sets of mutations. Therefore, the high-/low-mutation activity ratio should not be interpreted as the effect of the same individual variants before and after additional mutations. Rather, it represents the relative reduction in average activity caused by increasing the mutational burden in a heterogeneous mutation pool under each genetic code. We have revised the text to clarify this interpretation.

We also agree that the positions, identities, and functional relevance of individual mutations are not resolved in this pooled assay. This limitation prevents us from assigning mechanistic effects to specific substitutions. At the same time, using a small set of defined variants would introduce its own selection bias, because the conclusions could strongly depend on which mutations and which protein positions were chosen. Therefore, we consider the random-library approach to be a useful first step for testing library-averaged effects, whereas systematically defined variant analysis or genotype-resolved activity assays will be necessary to reveal mutation-specific mechanisms in future studies.

In response to the reviewer's concern, we have revised the Discussion to explicitly limit our conclusion to library-averaged effects on individual reporter protein activity. We now state that this approach does not identify the functional effects of individual mutations and that future studies using defined variants or high-throughput genotype–phenotype mapping will be required to determine how specific substitutions contribute to genetic code-dependent mutational robustness.

Result

"To estimate the average activity reduction associated with increased mutational burden under each genetic code, we calculated the ratio of activity obtained from the high-mutation library to that from the corresponding low-mutation library and plotted this ratio against each of the three mutational costs (Fig. 5C)."

Discussion

"A further limitation of this study is that the reporter activities were measured at the level of pooled random mutation libraries. Therefore, the high-/low-mutation activity ratio used in this study should be interpreted as the relative reduction in average activity caused by increasing the mutational burden in a heterogeneous mutation pool, rather than as the effect of identical variants before and after additional mutations. This library-averaged approach was chosen because the mutational costs considered here are also defined as average expected physicochemical effects over many possible single-nucleotide substitutions. In addition, because the non-SGCs constructed in this study were generated by reassigning only Ala, Ser, and Leu, the detectable effects may depend on how frequently mutations involving these amino acids occur in each reporter gene and whether the affected positions are functionally important. If genetic code dependent effects are restricted to a small subset of deleterious variants, such effects may be masked in pooled activity measurements. Future studies using defined variants or high-throughput genotype–phenotype mapping assays will be required to determine the mutation-specific and position-specific mechanisms underlying genetic code dependent effects on protein function (Rozhoňová et al., 2024)."

> *(8) For each mutagenesis library, the number of variants, the average number of mutations per variant, and the distribution of mutation positions should be reported clearly and transparently. These details are important for evaluating the strength of the conclusions.*

We agree that a more transparent characterization of the random mutagenesis libraries is necessary for evaluating the strength and limitations of our conclusions.

In the revised manuscript, we have added the estimated number of mutations per gene to the Results section. This value was calculated by multiplying the mutation rate per base by the coding sequence length of each reporter gene. For the high-mutation libraries used in Fig. 5, the estimated numbers of mutations per gene were approximately 8.0 for GAL, 4.5 for Luc, and 3.3 for mSG. We also added position-wise mutation profiles along each reporter gene (Figure 4–figure supplement 2), in addition to the heatmap shown in the original manuscript. These analyses clarify the mutational burden of each library and show that mutations were broadly distributed across the analyzed regions (approximately 300 nt in the middle of each gene) of the reporter genes.

Regarding the number of variants, the translation reactions were performed using 5 nM DNA template in a 5 μL reaction, corresponding to approximately 1.5×10^{10} DNA molecules. However, this value represents the total number of DNA molecules introduced into the reaction and does not directly indicate the number of unique full-length sequence variants, because multiple molecules can share the same genotype, and our sequencing analysis was designed to quantify mutation frequencies and positional distributions rather than to reconstruct full-length genotypes of individual library members. Therefore, we do not infer the exact number of unique variants in each library. Instead, we report the average mutation burden and position-wise non-reference rate distributions.

We have revised the Results and added Supplementary Figure 4–figure supplement 2 accordingly.

“For this experiment, two random mutation libraries were used: a low-mutation library prepared using the high-fidelity polymerase and a high-mutation library prepared using Taq DNA polymerase at a Mn^{2+} concentration that yields mutation rates of 0.002 – 0.005 per base (0.0026 for GAL, 0.0027 for Luc, and 0.0048 for mSG, corresponding to approximately 8.0, 4.5, and 3.3 mutations per gene). We also plotted position-wise non-reference rates along the analyzed regions of each reporter gene, confirming that mutations were broadly distributed across the amplicons (Figure 4–figure supplement 2).”

> *(9) Because only three amino acids were manipulated in the non-standard genetic codes, it remains unclear whether these particular amino acids occupy positions in the reporter proteins that are especially important for function and therefore likely to generate strong phenotypic effects. More broadly, it is not clear whether the assay is sufficiently sensitive to detect the effects of only a subset of deleterious variants within a pooled library. This point should be addressed more explicitly.*

We agree that this is an important limitation of the present study. Because our non-SGCs were constructed by reassigning only Ala, Ser, and Leu, the mutation-dependent effects that can differ among genetic codes are limited to mutations involving these reassigned codons or amino acid substitutions affected by these assignments. Therefore, the sensitivity of the assay depends on how frequently such substitutions occur in the reporter genes and whether the affected Ala, Ser, and Leu-related positions are functionally important.

We have revised the Discussion to address this point more explicitly. In the revised manuscript, we now state that the absence of a detectable cost-dependent effect may reflect not only the limited cost range examined, but also the limited set of reassigned amino acids, the position-dependent importance of Ala/Ser/Leu residues in the reporter proteins, and the sensitivity limit of pooled activity measurements. We further note that future studies using genotype-resolved activity assays (defined variants) will be required to determine whether specific amino acid substitutions or specific protein positions exhibit stronger genetic code-dependent effects.

"A further limitation of this study is that the reporter activities were measured at the level of pooled random mutation libraries. Therefore, the high-/low-mutation activity ratio used in this study should be interpreted as the relative reduction in average activity caused by increasing the mutational burden in a heterogeneous mutation pool, rather than as the effect of identical variants before and after additional mutations. This library-averaged approach was chosen because the mutational costs considered here are also defined as average expected physicochemical effects over many possible single-nucleotide substitutions. In addition, because the non-SGCs constructed in this study were generated by reassigning only Ala, Ser, and Leu, the detectable effects may depend on how frequently mutations involving these amino acids occur in each reporter gene and whether the affected positions are functionally important. If genetic code-dependent effects are restricted to a small subset of deleterious variants, such effects may be masked in pooled activity measurements. Future studies using defined variants or high-throughput genotype–phenotype mapping assays will be required to determine the mutation-specific and position-specific mechanisms underlying genetic code-dependent effects on protein function (Rozhoňová et al., 2024)."

> ***Recommendations for the authors:***
>
> ***Reviewing Editor Comments:***
>
> *While we suggest that you address all the technical points raised by the reviewers, you may specifically want to limit the conclusion of the study to mutational robustness at the level of individual protein activity, rather than making broader generalizations. Also, the statistical analysis needs to be strengthened, as indicated in the reviews.*

We thank the Reviewing Editor for these important suggestions. We agree that the conclusion of the original manuscript was broader than what can be directly supported by the present experiments. In the revised manuscript, we have therefore limited our conclusion to mutational robustness at the level of individual reporter protein activity measured in a reconstituted *in vitro* translation system. We now explicitly state that our results do not directly address robustness at the level of cellular fitness, protein interaction networks, or long-term evolution.

We have also strengthened the statistical analyses throughout the manuscript. Specifically, we added one-way ANOVA followed by Tukey's post hoc test for Fig. 1D, Welch's t-tests for Fig. 2B, Spearman's rank correlation analyses for Fig. 4B–D and Fig. 5C/E/G, and two-way ANOVA for GAL activity in Fig. 5B. These analyses have been incorporated into the revised Results, figure legends, and supplementary information.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) Discuss other alternative hypotheses if the error minimization theory is unlikely.*

We thank the reviewer for this helpful suggestion. We think that the absence of a detectable relationship between mutational cost and reporter protein activity in our assay should not be interpreted as excluding all possible roles of error minimization in the evolution of the genetic code. Our results specifically address one aspect of the error minimization theory:

whether physicochemical-property-based mutational cost predicts the average effect of random point mutations on individual reporter protein activity within the experimentally accessible range of non-SGCs tested here.

In the revised Discussion, we have clarified that the organization of the SGC may have been shaped by multiple factors, including robustness to translational errors, historical constraints associated with genetic code expansion, biosynthetic or coevolutionary processes, stereochemical interactions, and the evolvability of proteins. Our results suggest that the contribution of mutational robustness at the level of individual protein activity may be limited within the range examined here, but they do not exclude the possibility that the SGC provides advantages under other forms of error, at the level of translation fidelity, cellular fitness, or long-term evolution.

We have added a short discussion to clarify this point without expanding the scope of the manuscript beyond the present experimental results.

“It should be noted that this conclusion is limited to the activity of individual reporter proteins translated in a reconstituted *in vitro* system. Therefore, whether similar trends would be observed at the level of cellular fitness or long-term evolution remains an open question. Moreover, our results do not exclude other possible roles of SGC organization. The SGC may have been shaped by multiple factors, including robustness to translational errors, historical constraints during genetic code expansion, biosynthetic or coevolutionary relationships among amino acids, stereochemical interactions, and effects on protein evolvability (Katoh and Suga, 2023; Koonin and Novozhilov, 2017, 2009; Novozhilov et al., 2007; Wong, 2005).”

> *(2) A brief description of the PURE translation system can be provided for people from outside the field.*

We have added a brief description of the PURE system in the Introduction to make the experimental platform more accessible to readers outside the field. Specifically, we now explain that the PURE system is a reconstituted cell-free translation system composed of purified translation factors, ribosomes, aminoacyl-tRNA synthetases, tRNAs, amino acids, and energy-regeneration components. We also clarify that, in this study, we used a tRNA-free version of the PURE system, in which defined synthetic tRNA sets were supplied externally to reconstruct each genetic code.

Introduction

“A representative platform for such reconstitution is the PURE system (Shimizu et al., 2001), a reconstituted cell-free translation system composed of purified translation components, including ribosomes, translation factors, aaRSs, amino acids, and energy-regeneration components. In particular, a tRNA-free PURE system (Miyachi et al., 2022), in which endogenous tRNA activity is minimized and defined tRNA sets are supplied externally, enables genetic codes to be reconstructed by controlling the supplied tRNAs.”

> *(3) Figure 5D and F - Technical replicates are provided only for GAL. A similar approach should be taken for LUC and mSG.*

We agree that replicate-level measurements for Luc and mSG would further improve reliability. However, repeating the full translation experiments for these reporters was not feasible in the current revision, as each experiment requires large amounts of freshly prepared tRNA-free PURE system and multiple defined tRNA mixtures for every genetic code variant tested. Given these material and technical constraints, we were unable to perform additional biological replicates within the scope of this revision. We would like to emphasize, however, that the GAL replicates shown in Fig. 5D and F are fully consistent across independent experiments, providing direct evidence for the reproducibility of the assay itself.

Furthermore, the key metric in our analysis, the activity ratio between high- and low-mutation groups within each genetic code, is an internally normalized measure that is inherently less sensitive to between-experiment variability than absolute activity values. The correlation analyses further showed no significant relationship between mutational cost and this ratio across all three reporters, and this conclusion is consistent regardless of which reporter is examined. Together, we believe these results provide a robust basis for the conclusions drawn, even in the absence of full replication for Luc and mSG.

> *(4) Provide statistical analysis wherever it is relevant (e.g, to support a lack of correlation).*

We have strengthened the statistical analyses throughout the revised manuscript. In particular, to support the lack of detectable correlation between mutational cost and mutation-induced activity loss, we performed Spearman's rank correlation analyses between each mutational cost metric and the high-/low-mutation activity ratio for all three reporters. No significant correlations were detected for any reporter or cost metric. In addition, we added statistical analyses for other relevant figures, including one-way ANOVA followed by Tukey's post hoc test for Fig. 1D, Welch's t-tests for Fig. 2B, Spearman's rank correlation analyses for Fig. 4B–D, and two-way ANOVA for GAL activity in Fig. 5B.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *(1) In line 122, the phrase "as evenly as possible" is ambiguous and should be explained more precisely.*

We thank the reviewer for pointing this out. We have revised the phrase "as evenly as possible" to describe the codon design more precisely. Specifically, we now state that the NanoLuc coding sequences were designed so that the codons available in each genetic code were used with minimal differences in codon counts, while preserving the amino acid sequence of NanoLuc.

"For near-SGC and SGC, the NanoLuc coding sequences were designed so that the codons available in each genetic code were used with minimal differences in codon counts, while preserving the amino acid sequence (Fig. 1B, 32 codons and 46 codons)."

> *(2) For Figure 1D, a Western blot or another protein gel-based assay would be helpful to exclude the possibility that the observed differences arise from variation in translation efficiency rather than differences in protein activity.*

We agree that a protein gel-based assay such as Western blotting would in principle allow us to distinguish differences in translated protein amount from differences in specific activity, and we understand why such data would be informative. However, we would like to clarify that the primary purpose of Fig. 1D was to evaluate the overall functional translation output of each reconstructed genetic code, rather than to determine the mechanistic basis of any observed differences. In this context, NanoLuc luminescence serves as an integrated readout of the entire translation process, encompassing both translational efficiency and protein folding/activity. Crucially, regardless of whether the observed differences in NanoLuc luminescence reflect lower protein yield, reduced specific activity, or a combination of both, the conclusion of Fig. 1D remains the same. Although we did not perform Western blotting in this study, we believe that such an analysis would not change this interpretation and that the current data are sufficient to support this conclusion.

> *(3) The number 3^9 is not immediately intuitive. It would be helpful if the authors also stated that this corresponds to approximately 20,000 possible non-standard genetic codes.*

We have revised the text to state both the exact number and the approximate value: 3^9 = 19,683, approximately 20,000 possible non-standard genetic codes.

> *(4) The rationale for using the three cost parameters (PR, MV, and HI) should be explained in greater detail. Because these parameters are central to the manuscript, a citation alone is not sufficient. A concise explanation of their biological relevance would improve the clarity and accessibility of the study.*

We agree that the biological relevance of the three cost parameters should be explained more clearly. In the revised manuscript, we have added a concise explanation of why polar requirement (PR), molecular volume (MV), and hydropathy index (HI) were used.

These parameters were selected because they have been widely used in theoretical studies of genetic code optimality and represent distinct physicochemical aspects of amino acid substitutions. PR reflects polarity-related interactions and has been a classical metric in error minimization analyses of the genetic code. MV represents side-chain size and steric volume, which could influence packing and structural stability in proteins. HI reflects hydrophobicity, which is closely related to protein folding and hydrophobic core formation. We have also clarified that these metrics are simplified descriptors and do not capture residue-specific structural or functional context, which we now discuss as a limitation of the study.

“PR reflects polarity-related interactions of amino acids and has been used as a classical measure of amino acid similarity in error minimization analyses. MV represents side-chain size and steric volume, which could affect protein packing and structural stability, whereas HI reflects hydrophobicity, which could be closely related to protein folding or hydrophobic core formation.”

> *(5) In Figure 3, the experimental framework would be easier to follow if the authors included a schematic and data for one representative non-SGC, explicitly illustrating how it differs from the near-SGC with respect to each of the three cost measures.*

We agree that showing one representative non-SGC would make the experimental framework and cost calculation more intuitive.

In the revised manuscript, we added a new panel to Fig. 3 comparing the near-SGC with a representative non-SGC. We selected the PR_{max} code as the representative example because it clearly illustrates how reassignment of vacant codon boxes can increase one mutational cost metric relative to the near-SGC. In this panel, we first show the codon assignment schemes of the near-SGC and PR_{max} code in the same genetic-code format used in Fig. 1. We then show the corresponding heatmap representations for the three physicochemical properties used in the cost calculation: polar requirement, molecular volume, and hydropathy index. The $Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$ values are shown for each code.

This new panel illustrates how changes in codon assignment are translated into different physicochemical cost landscapes and clarifies how the representative non-SGC differs from the near-SGC with respect to each of the three cost measures.

“To make the design of non-SGCs more explicit, we show one representative non-SGC together with the near-SGC in Fig. 3B. This comparison illustrates how assignment of Ala, Ser, or Leu to the vacant codon boxes changes the three mutational cost metrics, $Cost_{PR}$, $Cost_{MV}$, and $Cost_{HI}$.”

> *(6) In line 329, the phrase "similar pattern" is ambiguous and should be explained more explicitly.*

We have revised the ambiguous phrase “similar pattern” to describe the observation more explicitly. Specifically, we now state that the relative differences in GAL activity among

genetic codes observed in the low-mutation library were broadly retained in the high-mutation library, although overall activity decreased.

"For the high-mutation library, GAL activity decreased overall, while the relative differences in activity among genetic codes observed in the low-mutation library were broadly retained."

> *(7) Figure S7 appears to be an important control for the experiments shown in Figure 5, and I recommend moving it to the main figures.*

We thank the reviewer for this helpful suggestion. We agree that the HiBiT-based quantification of GAL protein amount is an important control for interpreting the GAL activity measurements in Fig. 5, and we appreciate the recommendation to increase its visibility. This analysis shows that the amount of C-terminally completed GAL products was broadly comparable among genetic codes, indicating that the large differences in GAL activity were not primarily attributable to differences in total translated protein amount.

After careful consideration, we have opted to retain this analysis in the supplementary figures because the main focus of Fig. 5 is the relationship between mutational cost and mutation-induced activity loss, quantified by the high-/low-mutation activity ratio. The HiBiT experiment addresses a related but distinct question: whether differences in absolute GAL activity among genetic codes can be explained by differences in protein abundance, and we felt that including it in the main figures might shift the emphasis away from the central message of Fig. 5. Nevertheless, we have added a clear reference to Figure 4–figure supplement 1 in the main text and the figure legend to ensure that readers are directed to this control when interpreting Fig. 5.

https://doi.org/10.7554/eLife.111164.2.sa0