Rare Codons Not for Limiting Translation Speed in Mammals

Xiao-Yan Fan[1,2#], Tie-Gang Meng[1#], Teng Zhang[2#], Yu-Feng Wang[1#], Si-Tong Zhou[1#], Gui Zhang[1], Zhen Tang[3], Shi-Min Wang[1], Tao Jing[1], Lei Guo[1], Lei-Ning Chen[1], Sen Li[1], Juan Chen[1], Rui-Bao Su[1], Meng-Wen Hu[1], Jun-Yu Ma[1], Li-Zhang Zeng[3*], Qing-Yuan Sun[1*], Xiang-Hong Ou[1*], Shi-Ming Luo[1*]

[1]Guangzhou Key Laboratory of Metabolic Diseases and Reproductive Health, Guangdong-Hong Kong Metabolism & Reproduction Joint Laboratory, Reproductive Medicine Center, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, 510317, China.

[2]State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock (R2BGL), College of Life Sciences, Inner Mongolia University, Hohhot, 010070, China.

[3]Key Laboratory of Brain, Cognition and Education Science, Ministry of Education, South China Normal University, Guangzhou 510631, China.

[#]These authors contributed equally.

[*]e-mail: zenglzh@scnu.edu.cn; sunqy@gd2h.org.cn; ouxh@gd2h.org.cn; luoshm@gd2h.org.cn

Abstract

The prevailing view holds that rare codons function as evolutionarily conserved modulators of translation kinetics, a concept that has provided a critical framework for interpreting synonymous mutations in disease and guiding coding sequence optimization in gene therapy. Our findings, however, challenge this premise by demonstrating that rare codons are primarily passive byproducts of genome stability constraints, rather than adaptive regulators of translation. Supporting this, genomic analyses in human and mouse reveal that rare codons, predominantly ending in CpG or TpA dinucleotides, reflect genome-wide sequence biases rather than selective pressures for translational control. Furthermore, interrogation of tRNA abundance, via both genomic copy number and direct cellular quantification, shows no consistent correlation between codon rarity and cognate tRNA scarcity. Crucially, direct

experimental assessment via cytoplasmic microinjection of engineered mRNAs into mouse embryos demonstrates that even a substantial artificial increase in rare codon load has a negligible impact on protein synthesis kinetics. Consequently, we conclude that rare codons cannot be regarded as an independent factor limiting translation rates in mammalian cells. This insight demands a reassessment of the role of synonymous mutations in disease and expands the design principles for mRNA-based therapeutics by alleviating the presumed constraint of rare codon avoidance.

Keywords: Rare codons, codon optimization, translation kinetics, synonymous mutations, mRNA therapeutics, genomic constraints, tRNA abundance.

Introduction

The protein's function is fundamentally determined by its amino acid sequence, yet the degeneracy of the genetic code allows for a vast number of nucleotide sequences to encode the identical protein (*1-3*). With the exception of methionine and tryptophan, each amino acid is encoded by multiple synonymous codons, ranging from two to six per residue (*2*). However, this synonymy is not functionally silent: synonymous codon choice can influence gene expression through effects on transcription, mRNA stability, and translation efficiency (*4-6*). Understanding the functional consequences of synonymous codon usage, which are often manifested as synonymous single-nucleotide polymorphisms (synSNPs), is therefore critical for elucidating their roles in health and diseases such as cystic fibrosis and certain cancers (*7, 8*).

A central feature of synonymous codon usage is codon usage bias, the non-random preference for certain codons over others. Since the 1980s, it has been well established that genomes exhibit systematic codon preferences (*9-12*). In humans, for example, the leucine codon CTG is used more than five times as frequently as CTA, with relative frequencies of approximately 3.96 and 0.72, respectively (*13, 14*). A similarly pronounced (~4.4-fold) difference exists for the serine codons AGC and TCG, which have relative frequencies of 1.95 and 0.44 (*13, 14*). This biased distribution is thought to reflect a mechanism for fine-tuning translational efficiency (*6, 15, 16*).

A cornerstone of this paradigm is the specific belief that rare codons, those used infrequently within a genome, serve as evolutionarily conserved modulators of translation kinetics by slowing ribosomal elongation (*1, 17*). Although evidence from heterologous expression systems (e.g., expression of human genes in bacteria or yeast) supports a role for codon rarity in regulating protein expression (*18-20*), direct evidence that rare codons intrinsically slow translation in native eukaryotic contexts remains limited. Critically, the strong influence of local codon context and mRNA secondary structure on translation efficiency (*2, 21*), makes it difficult to determine whether rare codons are genuine drivers or merely passive correlates of altered translation dynamics .

To address this gap, we conducted the integrated genomic, bioinformatic, and experimental analyses. By analyzing genomic codon distributions and corresponding tRNA expression profiles, we propose that rare codons represent passive evolutionary byproducts rather than adaptively selected modulators of translation speed. Experimental evidence further demonstrates that even a substantial increasing rare codon frequency does not significantly affect translation kinetics, supporting this hypothesis.

Results

Genomic and Bioinformatics Evidence Supports Rare Codons as Passive Byproducts of Sequence Constraint

Protein translation is governed by multiple factors, with mRNA secondary structure playing a pivotal role. Such structural features can be influenced by complex base-pairing interactions, sometimes irrespective of nucleotide positioning. Therefore, we reasoned that individual gene- or mRNA-specific studies focusing on rare codons are likely confounded by local sequence context, and that systems-level 'omics' analyses of codon usage would offer more reliable insights into the relationship between codon selection and translational efficiency than correlations derived from individual mRNAs.

Moving beyond conventional definitions based solely on a codon's overall occurrence frequency (<1%) (*3*), we propose that a codon's rarity should also be defined relative to the usage of its synonymous partners. Using the Matched Annotation from NCBI and EMBL-EBI (MANE) transcript set for human and the protein-coding gene dataset for mouse from NCBI, we recomputed genome-wide codon frequencies across human and mouse coding sequences (CDSs) (Supplemental Table 1). A codon was classified as rare only if its overall frequency was below 1% and its usage was substantially lower (typically by less than 50%) than the most frequent synonymous codon.

Applying these criteria, we identified ten rare codons: ACG, CCG, GCG, TCG, ATA, CTA, GTA, TTA, CGA, and CGT (Fig. 1A), which are highly conserved between human and mouse (Fig. 1B and Supplemental Fig. 1). Except for CGA, which occur

at approximately half the frequency of its dominant synonyms, the remaining rare codons are generally used at less than 40% of the frequency of their most prevalent counterparts. Interestingly, with the exception of CGA and CGT, all identified rare codons end with the dinucleotides 'CG' or 'AT'. Given the established role of CpG islands in genomic stability and transcriptional regulation through methylation-mediated processes (*22, 23*), our observations raise the possibility that these codons may be passive byproducts of genome evolutionary constraints rather than actively selected for translational modulation.

A genome-wide scan of codon context revealed that, among the 64 triplet codons, the ten rare codons in human (GCG, ACG, CGT, CGA, TCG, CCG, GTA, CTA, ATA, TTA) rank 63rd, 62nd, 61st, 60th, 59th, 58th, 49nd, 46th, 21rd, and 9th, and in mouse (GCG, CGA, CCG, TCG, ACG, CGT, GTA, CTA, ATA, TTA) rank 64th, 63rd, 60th, 59th, 58th, 57th, 51st, 41st, 27th, and 23rd, respectively, in genomic distribution frequency (rank 1 = most frequent) (Fig. 1C and Supplemental Table 2). The collective rarity of these identified rare codons across the genome further supports the hypothesis that their occurrence in CDSs may be constrained by genomic context rather than positive selection for translational control.

Notably, in human, ATA and TTA exhibit higher genomic frequency ranks (21st and 9th) than their classification as rare in CDSs might suggest. Similarly, in mouse, their genomic ranks (27th and 23rd) are also substantially higher than expected for rare codons. This apparent contradiction prompted us to ask whether their rarity is relative, stemming from exceptionally high usage of their synonymous family members. We therefore reasoned that evaluating their usage within the context of their synonymous families would provide clarity. This analysis revealed a general pattern: all 'TA'-ending rare codons are used at a lower frequency in CDSs relative to their genomic abundance, whereas 'CG'-ending codons (including CGA and CGT) are used more frequently (Fig. 1D and Supplemental Table 3). When normalized to genomic abundance, 'CG'-ending rare codons show nearly two-fold increased usage in CDSs, while 'TA'-ending codons exhibit approximately 50% reduction (Fig. 1E). This pattern is consistent with the known, broader bias in coding sequences toward

codons ending in C or G, indicating that the behavior of rare codons follows this general principle (*24, 25*).

Given that protein-coding sequences constitute only ~1.1–1.3% of the mammalian genome, codon usage within CDSs is unlikely to significantly influence overall genomic distribution. Instead, the observed genomic patterns, where the rarity of codon triplets mirrors their low frequency in the whole genome, argue against positive selection for translational modulation. These results collectively indicate that codon usage, including the occurrence of rare codons, is primarily shaped by genome-wide evolutionary and structural constraints rather than by adaptive optimization for translation. Consequently, the established mechanistic link between rare codons and tRNA-mediated ribosomal pausing requires critical re-examination.

tRNA Abundance Argues Against Rare Codons as a General Mechanism for Translation Attenuation

According to the prevailing model, rare codons modulate translation kinetics primarily through cognate tRNA abundance, with tRNA scarcity inducing ribosomal pausing to facilitate co-translational folding and regulate protein yield (*26*). To evaluate this hypothesis, we analyzed cognate tRNA availability for each codon in human and mouse using data from the Genomic tRNA Database (gtrnadb.ucsc.edu).

The results indicate that whether a codon is rare does not consistently correlate with its cognate tRNA copy number. Even after excluding codons decoded by wobble pairing with zero-copy tRNAs, tRNA copy number did not consistently correspond to codon rarity (Fig. 2A and Supplemental Table 4). For example, tRNAs cognate to the rare arginine codons CGT and CGA were among the most abundant. Moreover, for most other rare codons, tRNA copy numbers were not significantly lower than those of non-rare synonymous codons, except when compared to the single most abundant counterpart. This partial and inconsistent correlation indicates that the relationship between tRNA availability and codon usage frequency is not universal. Thus, while tRNA abundance may contribute to the effect of some rare codons, it cannot be the sole or primary determinant of translation efficiency across all such codons.

Furthermore, given potential variations in transcription and turnover, whether

genomic tRNA copy number serves as a reliable proxy for functional cellular abundance requires careful examination. Therefore, direct measurement of cellular tRNA levels is indispensable for conclusively evaluating the role of tRNA availability in rare codon-mediated regulation.

Despite advances in omics technologies, comprehensive tRNA profiling remains challenging due to their short length, high sequence similarity, extensive modifications, and complex secondary structures (*27*). Recent application of modification-induced misincorporation tRNA sequencing (mim-tRNAseq) to hiPSCs, HEK293T, and K562 cells enabled precise quantification (*28*). Re-analysis of this data revealed that, except for TCG (Ser; decoded by tRNA-CGA) and TTA (Leu; tRNA-TAA), tRNAs cognate to other rare codons generally showed intermediate abundance among their synonymous groups (Fig. 2B and Supplemental Table 5). Notably, the tRNAs cognate to the rare arginine codon CGT and the rare isoleucine codon ATA were, in fact, the most abundant within their respective synonymous families (Fig. 2C). This pattern remained consistent across stem cell and cancer lines (Fig. 2D), directly challenging the premise that translation slowing is a general consequence of cognate tRNA scarcity for rare codons. Having found that neither genomic distribution nor tRNA availability supports an adaptive regulatory role, we next asked the most direct question: do rare codons, even at high density, intrinsically limit translation rates in a living eukaryotic cell?

Direct Experimental Assessment Reveals Minimal Impact of Rare Codon Load on Translation Kinetics

To definitively and experimentally validate whether rare codons act as autonomous determinants of translation speed, we employed a falsifiability-based approach: if substantially increasing rare codon frequency yields no measurable effect on translation, it would provide direct evidence against their regulatory role.

Prior to experimental perturbation, we first assessed the natural usage of rare codons. Genomic analysis revealed that, in native coding sequences, rare codons are employed in fewer than 60% of the available synonymous positions and account for less than 30% of the total CDS length. The mean and median values further demonstrate their

limited usage: rare codons constitute only 6.6% (median 6.3%) of human and 6.5% (median 6.2%) of mouse CDS length, and are used at only 12.6% (median 12.2%) and 12.3% (median 11.9%) of substitutable synonymous positions, respectively (Fig. 3A and B, Supplemental Table 6 and 7). Building upon this baseline, we then designed mRNA constructs encoding the bright green and red fluorescent proteins mNeonGreen and mScarlet3-H, respectively. In particular, mNeonGreen matures exceptionally fast (in less than 10 minutes) (*29*), making it ideally suited for real-time monitoring of protein synthesis due to its rapid and consistent fluorescent kinetics. For each protein, we generated two distinct variants: a native sequence and a synonymous version in which approximately 80% of substitutable codons were replaced with their rare counterparts. This substitution strategy increased the rare codon content from 1.2% to 32.0% across the entire CDS for mScarlet3, and similarly from 4.9% to 32.3% for mNeonGreen. When considering only the substitutable codon positions, the utilization rate of rare codons increased dramatically from 2.86% to 79.1% for mScarlet3 and from 12.5% to 82.7% for mNeonGreen (Fig. 3C, Supplemental 2A and Supplemental Table 8).

Furthermore, unlike conventional plasmid-based transfection or transformation, which introduces asynchrony and heterogeneity, we microinjected *in vitro*-transcribed engineered mRNA directly into the cytoplasm of mouse embryos. This method circumvents variable plasmid uptake and transcription, ensuring synchronous translation initiation across a homogeneous cell population and providing a clean, physiologically relevant system for precise measurement. Crucially, measurements of the mNeonGreen and mScarlet3-H translation dynamics revealed no significant difference between the increased rare-codon and native versions (Fig. 3D, E, Supplemental 2B, C, and Supplemental Vedio1, 2).

These results demonstrate that even a substantial artificial increase in rare codon load has negligible effects on translation efficiency, providing direct experimental evidence that rare codons are not intrinsic limiters of translation speed in this mammalian system.

Discussion

Our integrated analysis leads to a revised paradigm: rare codons in mammalian genomes are primarily passive byproducts of genome evolutionary constraints, not adaptive translators of protein synthesis efficiency. This conclusion is supported by the congruence of rare codon distribution with genomic sequence biases, the lack of consistent correlation with tRNA scarcity, and most definitively, experimentally saturating mRNAs with rare codons fails to alter translational kinetics or protein production, in sharp contrast to the predictions of the prevailing model.

The foundational support for codon-centric regulation originated largely from seminal studies in bacterial systems (*30-33*). A classic example is the heterologous expression of mammalian genes in bacteria, where codons common in the donor organism are often rare in the host, frequently leading to reduced protein yield or misfolding. These defects are typically rescued by synonymously optimizing the gene sequence to match bacterial codon preferences (*30, 31, 34*). A critical caveat, however, is the fundamental mechanistic divergence in translation between prokaryotes and eukaryotes (*35, 36*). This divergence results in a different rate-limiting step, enabling a much faster elongation velocity in prokaryotes (10–20 aa/s) compared to eukaryotes (3–8 aa/s), a disparity of approximately 2- to 6-fold (*37, 38*). In eukaryotes, translation initiation constitutes a complex, rate-limiting process governed by the assembly of numerous initiation factors, contrasting sharply with the more streamlined bacterial system (*39-41*). Consequently, while codon usage robustly influences elongation kinetics and co-translational folding in bacteria, its regulatory impact in eukaryotes is likely attenuated or functionally distinct. This is because the slow initiation rate may overshadow codon-mediated effects during elongation as the primary determinant of overall translation efficiency.

Our model does not preclude that codon choice can influence gene expression, but it redirects the primary mechanistic focus. The effects of synonymous variants, including those introducing rare codons, may manifest predominantly through impacts on transcription, mRNA stability, or splicing processes sensitive to nucleotide sequence, rather than through translation elongation per se. Future investigations

should therefore prioritize the context of rare codons, such as their clustering or positioning within structural motifs, to understand their potential roles in these alternative regulatory layers.

By repositioning rare codons from active regulators to genomic bystanders, this model resolves some long-standing empirical and mechanistic contradictions, such as the inconsistency of Ribo-seq data, which shows poor reproducibility in ribosomal pausing sites and weak correlations between codon usage and elongation rates (*42, 43*). Furthermore, this paradigm shift opens new avenues for interpreting genetic variation and designing synthetic genes.

While the fundamental blueprint for health and disease is encoded in the genome, the ability to fully decipher this information is still a central challenge. Advances in omics technologies have begun to address this gap, enabling the systematic elucidation of how genetic variations influence physiological and pathological states. In particular, the role of single-nucleotide variants (SNVs) in human health and disease has been increasingly clarified (*8, 44*). Among these, however, the functional impact of synonymous variants, which represent the most abundant class of SNVs, is often complicated by contradictory evidence (*45-47*). Our finding that rare codons exert minimal influence on translation efficiency offers a key insight that helps resolve some of these contradictions.

Furthermore, mRNA therapy, which operates by instructing the body's own cells to produce proteins capable of preventing or treating a broad spectrum of diseases, from infectious diseases to cancers, is currently undergoing rapid development (*48, 49*). A central objective in its development is the optimization of mRNA sequences to maximize protein yield while minimizing the amount of mRNA required. Since rare codons have traditionally been thought to constrain translational efficiency, current mRNA optimization strategies generally prioritize the avoidance of such codons (*5, 50*). Our results challenge this rationale, demonstrating that rare codons do not pose a significant limitation. Importantly, as mRNA-based therapeutics bypass the constraints of genome stability and transcription regulation in cell, the optimization focus can be directed specifically toward enhancing translation efficiency and mRNA

stability. This finding substantially expands the design possibilities for therapeutic mRNAs, enabling the engineering of molecules with superior *in vivo* stability, translational efficiency and fidelity, and low immunogenicity. Notably, for recombinant protein production in non-mammalian systems such as bacteria and engineered yeasts, where the impact of rare codon usage is well-documented, the conventional codon optimization strategies, including rare codon avoidance, remain fully applicable and effective.

Collectively, our study demonstrates that rare codons in eukaryotes primarily arise as passive byproducts of genomic and evolutionary constraints, rather than serving as direct regulators of protein translation efficiency. This revised view not only consolidates previously conflicting genomic and experimental data but also provides a clearer foundation for future research into the complex interplay between genome architecture, RNA biology, and phenotypic expression.

Materials and Methods

CDSs Acquisition and Processing

To obtain a high-confidence set of human protein-coding sequences (CDS), we utilized the Matched Annotation from NCBI and EMBL-EBI (MANE) dataset (Release 1.4) (*51*), downloaded from https://ftp.ncbi.nlm.nih.gov/refseq/MANE/MANE_human/release_1.4/. This dataset provides a manually curated, genome-wide set of representative transcripts with a one-to-one correspondence between NCBI RefSeq and Ensembl gene annotations, ensuring high-quality, well-supported sequences. A custom Python script was employed to parse the FASTA files and extract the canonical CDS for each transcript. The script processed the sequence records, extracting the CDS based on annotated coordinates and filtering for valid, unique coding sequences. From the initial dataset, this procedure yielded a final, non-redundant set of 19,352 high-quality CDS entries for subsequent analysis.

To establish a non-redundant set of CDSs for the mouse (Mus musculus) genome, transcripts were sourced from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/datasets/gene/). A custom Python script was then employed to select a single representative isoform per gene using a hierarchical prioritization strategy. The algorithm first grouped all transcripts by gene identifier, parsed from filenames. For each gene, it preferentially selected the longest experimentally validated transcript (NM_ accession) to prioritize high-confidence annotations. In the absence of an NM_ isoform, the longest predicted model (XM_ accession) was chosen. This approach, which uses file size as a proxy for CDS length to favor the most comprehensive sequence, effectively generated a canonical transcript set by systematically resolving isoform redundancy. Ultimately, from an initial pool of 98,888 CDS sequences, the filtration process yielded a final set of 22,410 canonical transcripts, ensuring each gene was represented by its most reliable and complete protein-coding sequence for subsequent analysis.

Genome Sequence Acquisition and Processing

Complete, telomere-to-telomere (T2T) genome assemblies were obtained for both

human and mouse from the NCBI database. For human (Homo sapiens), we used the T2T-CHM13v2.0 assembly (accession GCF_009914755.1), downloaded from: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/009/914/755/GCF_009914755.1_T2T-CHM13v2.0/GCF_009914755.1_T2T-CHM13v2.0_genomic.fna.gz For mouse (Mus musculus), we employed the T2T_mhaESC_v1.0 assembly (accession GCA_050437135.1), available at: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/050/437/135/GCA_050437135.1_T2T_mhaESC_v1.0/GCA_050437135.1_T2T_mhaESC_v1.0_genomic.fna.gz These complete, gap-free genome assemblies provide the most comprehensive reference sequences currently available for their respective species. The genomic data were processed using a standardized pipeline that included quality control checks, sequence validation, and annotation mapping to ensure data integrity prior to downstream analyses.

Codon Frequency Analysis in CDSs and Tri-Nucleotide Frequency Analysis in Genome

The CDS-based analysis extracts codon usage patterns from protein-coding regions. Sequences are processed in strict three-base increments starting from the first nucleotide position, maintaining the correct reading frame throughout. Each CDS sequence is divided into non-overlapping triplets corresponding to actual codons, ensuring that every codon is counted exactly once in its functional context. This approach reflects the biological reality of translation where ribosomes read mRNA sequences in consecutive, non-overlapping groups of three nucleotides.

In contrast, the genome-wide analysis scans entire chromosome sequences using a sliding window approach where every possible three-base combination is counted regardless of reading frame position. Each nucleotide serves as the starting point for a triplet count, meaning that the analysis effectively performs three separate scans: one starting from each of the first three nucleotide positions. This results in overlapping counts where the same genomic region contributes to multiple triplet occurrences, providing a comprehensive background frequency without reading frame constraints.

The fundamental difference lies in the counting strategy: CDS analysis uses

frame-specific, non-overlapping codon counting that mirrors biological translation, while genomic analysis employs frame-agnostic, overlapping triplet counting that captures all possible three-base combinations across the entire genome.

Source of tRNA Data

Cognate tRNA gene copy numbers for human and mouse were obtained from the Genomic tRNA Database (GtRNAdb; http://gtrnadb.ucsc.edu/), which provides comprehensive tRNA gene predictions generated by tRNAscan-SE analysis of complete genomes.

Experimentally measured tRNA abundances in hiPSCs, HEK293T, and K562 cells were derived from Supplementary Table 1 of the study by Behrens et al. (*28*), which employed mim-tRNAseq for high-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes.

Plasmid Construction and mRNA Preparation

Gene sequences for mNeonGreen and mScarlet3 were obtained from VectorBuilder. To generate high rare-codon variants, we redesigned the coding sequences by substituting alternative codons with their rare counterparts. Importantly, to circumvent potential adverse effects on transcription, mRNA stability, or translation that could arise from consecutive rare codons, we deliberately avoided introducing stretches of adjacent rare codons. Consequently, not all substitutable codons were replaced. For mNeonGreen, 86 of 104 substitutable codons were replaced, increasing the utilization rate of rare codons at substitutable positions from 12.5% to 82.7%. For mScarlet3, 83 of 105 substitutable codons were replaced, increasing this rate from 2.86% to 79.1% (see Supplemental Table 8 for the detailed sequences). Notably, even with this conservative design approach, the resulting rare codon utilization rate in both engineered mRNAs substantially exceeds the maximum rate observed at substitutable positions in any endogenous human or mouse coding sequence, which is less than 57.14% (Supplemental Tables 6 and 7).

All sequences were synthesized and cloned into the pAZ-RNA-14 vector (RN2640) by Suzhou Genewiz Biotechnology Co. Ltd. (Suzhou, China). Constructs contained defined 5'-UTR and 3'-UTR/polyA sequences (see Supplemental Table 8). Plasmids

were linearized with BspQI, and mRNAs were synthesized *in vitro* using T7 RNA polymerase, followed by purification and lyophilization. mRNA powders were stored at -80° C until use.

mRNA Microinjection and Live-cell Fluorescence Microscopy

All animal studies in this project were approved by the Ethics Committee of Guangdong Second Provincial General Hospital (Approval No. 2025-DW-KZ-121-01). Female and male C57BL/6J mice (8 weeks old; Beijing Huafukang Bioscience, Beijing, China) were housed in the animal facility of Guangdong Second Provincial General Hospital under a 12-hour light/dark cycle. Female mice were superovulated by intraperitoneal injection of 10 IU PMSG (Ningbo Animal Hormone Factory), followed by 10 IU hCG (Ningbo Animal Hormone Factory, Ningbo, China) 48 hours later, and then paired with males. Fertilized embryos were collected approximately 30 hours after hCG injection.

Lyophilized mRNA was reconstituted in nuclease-free water to a stock concentration of 1 μg/μL, aliquoted into 1 μL volumes, and stored at −80° C until use. For microinjection, mRNA stock was diluted to 300 μg/μL in nuclease-free water. Embryos were injected in room temperature medium using a Piezo-driven microinjection system (PiezoXpert, Eppendorf). Immediately after injection, embryos were transferred to a Tokai Hit stage-top incubation chamber mounted on a Leica Stellaris 8 Falcon Flim Microscope, maintained at 37℃ and 5% CO_2. Fluorescence imaging was performed at 5-minute intervals for 1 hour for real-time monitoring of protein synthesis.

Data Analysis

Fluorescence signal quantification was performed using the proprietary image analysis toolkit integrated with the Leica SP8 FALCON confocal microscope system. For each experimental condition, quantification was conducted across three independent biological replicates. Statistical analysis was performed using a two-tailed unpaired Student’s t-test. Data are presented as mean ± standard error of the mean (s.e.m.), with P-values used to assess the significance of observed

differences between experimental groups.

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2022YFC2702203), the National Natural Science Foundation of China (32571304, 82301878) and the Talents' Plan Foundation of Guangdong Second Provincial General Hospital (2024D004).

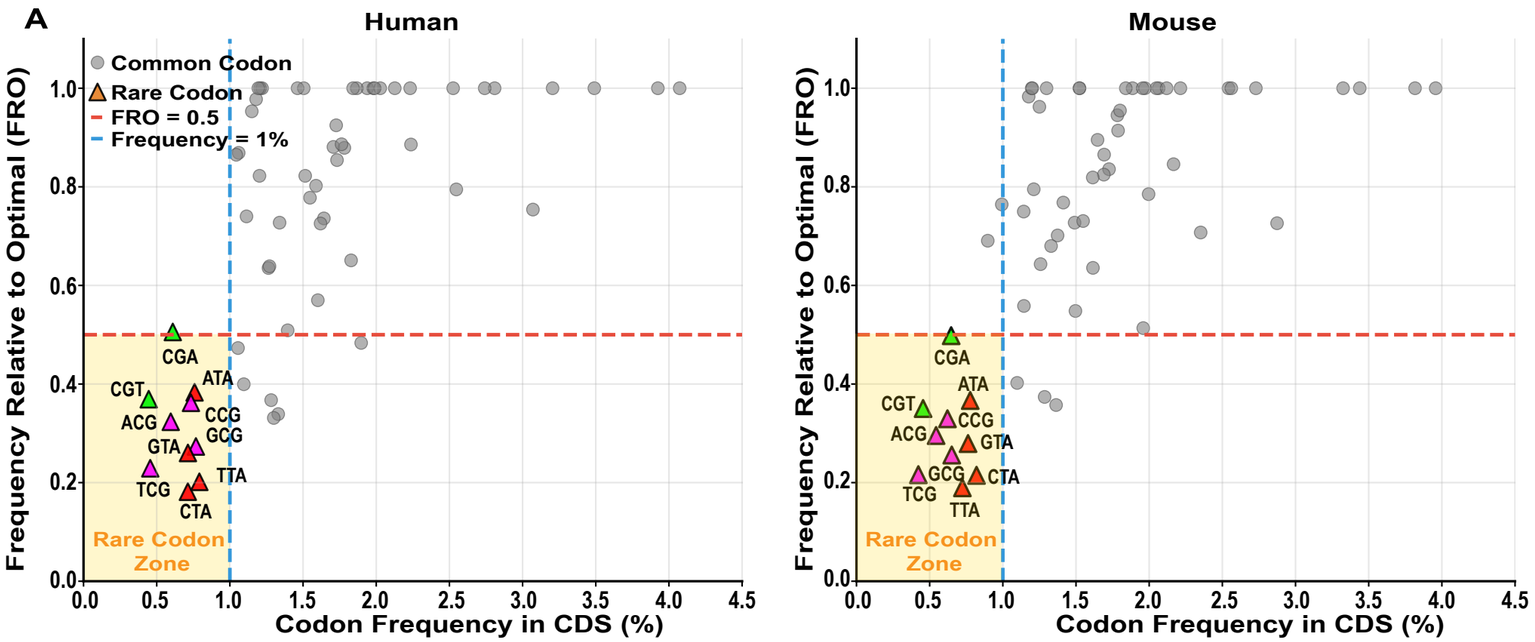
References

1. G. Hanson, J. Coller, Codon optimality, bias and usage in translation and mRNA decay. *Nature reviews. Molecular cell biology* **19**, 20-30 (2018).
2. G. Cannarozzi *et al.*, A role for codon order in translation dynamics. *Cell* **141**, 355-367 (2010).
3. W. Ding *et al.*, Rare codon recoding for efficient noncanonical amino acid incorporation in mammalian cells. *Science* **384**, 1134-1142 (2024).
4. J. V. Chamary, L. D. Hurst, Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome biology* **6**, R75 (2005).
5. H. Zhang *et al.*, Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* **621**, 396-403 (2023).
6. Y. Liu, Q. Yang, F. Zhao, Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annual review of biochemistry* **90**, 375-401 (2021).
7. N. W. Schroder, R. R. Schumann, Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *The Lancet. Infectious diseases* **5**, 156-164 (2005).
8. J. Zhang, W. Qian, Functional synonymous mutations and their evolutionary consequences. *Nature reviews. Genetics* **26**, 789-804 (2025).
9. T. Ikemura, Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. *Journal of molecular biology* **158**, 573-597 (1982).
10. R. Grantham, C. Gautier, M. Gouy, R. Mercier, A. Pave, Codon catalog usage and the genome hypothesis. *Nucleic acids research* **8**, r49-r62 (1980).
11. R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, R. Mercier, Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic acids research* **9**, r43-74 (1981).
12. T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular biology and evolution* **2**, 13-34 (1985).
13. B. P. Thornlow *et al.*, Predicting transfer RNA gene activity from sequence and genome context. *Genome research* **30**, 85-94 (2020).
14. P. P. Chan, T. M. Lowe, tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods in molecular biology* **1962**, 1-14 (2019).
15. R. Hershberg, D. A. Petrov, Selection on codon bias. *Annual review of genetics* **42**, 287-299 (2008).
16. Q. Wu, A. A. Bazzini, Translation and mRNA Stability Control. *Annual review of biochemistry* **92**, 227-245 (2023).
17. T. E. Quax, N. J. Claassens, D. Soll, J. van der Oost, Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular cell* **59**, 149-161 (2015).
18. C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression. *Trends in biotechnology* **22**, 346-353 (2004).
19. C. H. Kim, Y. Oh, T. H. Lee, Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. *Gene* **199**, 293-301 (1997).

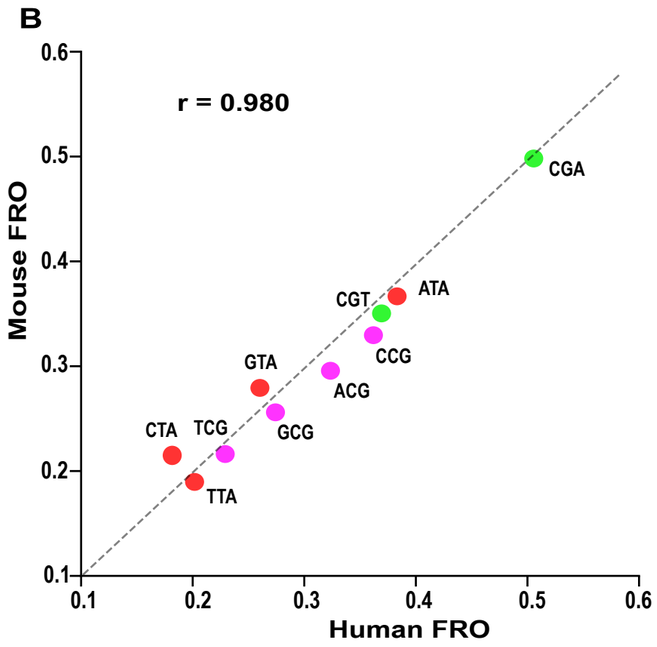
20. X. Hu, Q. Shi, T. Yang, G. Jackowski, Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in Escherichia coli. *Protein expression and purification* **7**, 289-293 (1996).
21. K. Leppek *et al.*, Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nature communications* **13**, 1536 (2022).
22. V. R. Ramirez-Carrozzi *et al.*, A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-128 (2009).
23. W. Reik, Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432 (2007).
24. J. Li, J. Zhou, Y. Wu, S. Yang, D. Tian, GC-Content of Synonymous Codons Profoundly Influences Amino Acid Usage. *G3* **5**, 2027-2036 (2015).
25. M. Nabiyouni, A. Prakash, A. Fedorov, Vertebrate codon bias indicates a highly GC-rich ancestral genome. *Gene* **519**, 113-119 (2013).
26. E. M. Novoa, L. Ribas de Pouplana, Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in genetics : TIG* **28**, 574-581 (2012).
27. M. C. Lucas *et al.*, Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nature biotechnology* **42**, 72-86 (2024).
28. A. Behrens, G. Rodschinka, D. D. Nedialkova, High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Molecular cell* **81**, 1802-1815 e1807 (2021).
29. N. C. Shaner *et al.*, A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nature methods* **10**, 407-409 (2013).
30. S. L. Martin, B. Vrhovski, A. S. Weiss, Total synthesis and expression in Escherichia coli of a gene encoding human tropoelastin. *Gene* **154**, 159-166 (1995).
31. B. Q. Wang, L. Lei, Z. F. Burton, Importance of codon preference for production of human RAP74 and reconstitution of the RAP30/74 complex. *Protein expression and purification* **5**, 476-485 (1994).
32. E. Goldman, A. H. Rosenberg, G. Zubay, F. W. Studier, Consecutive low-usage leucine codons block translation only when near the 5' end of a message in Escherichia coli. *Journal of molecular biology* **245**, 467-473 (1995).
33. J. F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current opinion in biotechnology* **6**, 494-500 (1995).
34. G. L. Rosano, E. A. Ceccarelli, Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial cell factories* **8**, 41 (2009).
35. M. Kozak, Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37 (2005).
36. M. Kozak, Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187-208 (1999).
37. A. Prabhakar, J. Choi, J. Wang, A. Petrov, J. D. Puglisi, Dynamic basis of fidelity and speed in translation: Coordinated multistep mechanisms of elongation and termination. *Protein science : a publication of the Protein Society* **26**, 1352-1362 (2017).
38. E. Siller, D. C. DeZwaan, J. F. Anderson, B. C. Freeman, J. M. Barral, Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *Journal of molecular biology* **396**, 1310-1318 (2010).

39. T. Tuller *et al.*, An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344-354 (2010).
40. J. Brito Querido, I. Diaz-Lopez, V. Ramakrishnan, The molecular basis of translation initiation and its regulation in eukaryotes. *Nature reviews. Molecular cell biology* **25**, 168-186 (2024).
41. N. Sonenberg, A. G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745 (2009).
42. C. H. Yu *et al.*, Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular cell* **59**, 744-754 (2015).
43. C. A. Charneski, L. D. Hurst, Positively charged residues are the major determinants of ribosomal velocity. *PLoS biology* **11**, e1001508 (2013).
44. Z. E. Sauna, C. Kimchi-Sarfaty, Understanding the contribution of synonymous mutations to human disease. *Nature reviews. Genetics* **12**, 683-691 (2011).
45. R. S. Dhindsa *et al.*, A minimal role for synonymous variation in human disease. *American journal of human genetics* **109**, 2105-2109 (2022).
46. L. Kruglyak *et al.*, Insufficient evidence for non-neutrality of synonymous mutations. *Nature* **616**, E8-E9 (2023).
47. X. Shen, S. Song, C. Li, J. Zhang, Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725-731 (2022).
48. Y. Wang *et al.*, mRNA vaccine: a potential therapeutic strategy. *Molecular cancer* **20**, 33 (2021).
49. C. Liu *et al.*, mRNA-based cancer therapeutics. *Nature reviews. Cancer* **23**, 526-543 (2023).
50. Y. Li *et al.*, Deep generative optimization of mRNA codon sequences for enhanced mRNA translation and therapeutic efficacy. *Nature communications* **16**, 9957 (2025).
51. J. Morales *et al.*, A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* **604**, 310-315 (2022).

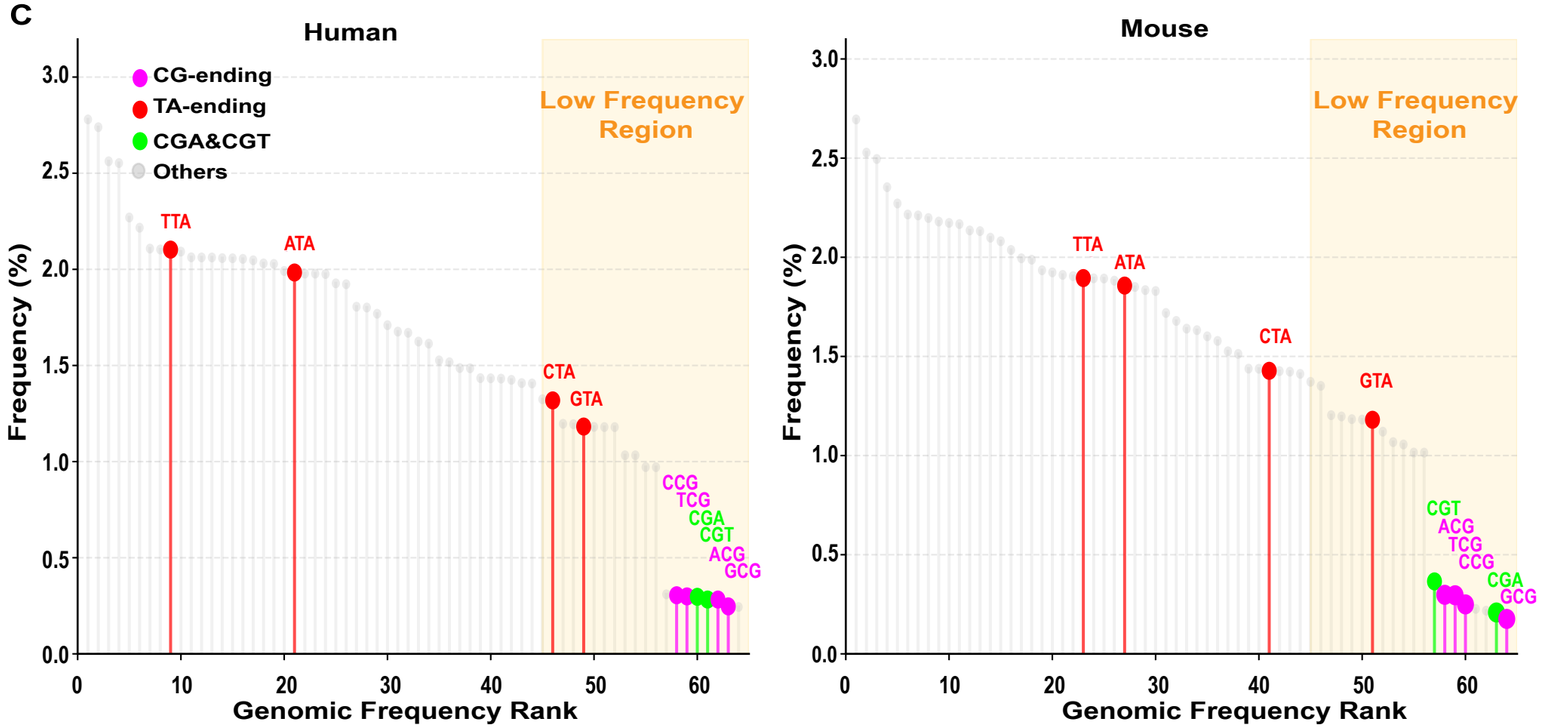
A


Human
Mouse
Common Codon
Rare Codon
FRO = 0.5
Frequency = 1%
Frequency Relative to Optimal (FRO)
Codon Frequency in CDS (%)
Rare Codon Zone
CGA
CGT
ATA
CCG
ACG
GCG
GTA
TCG
TTA
CTA

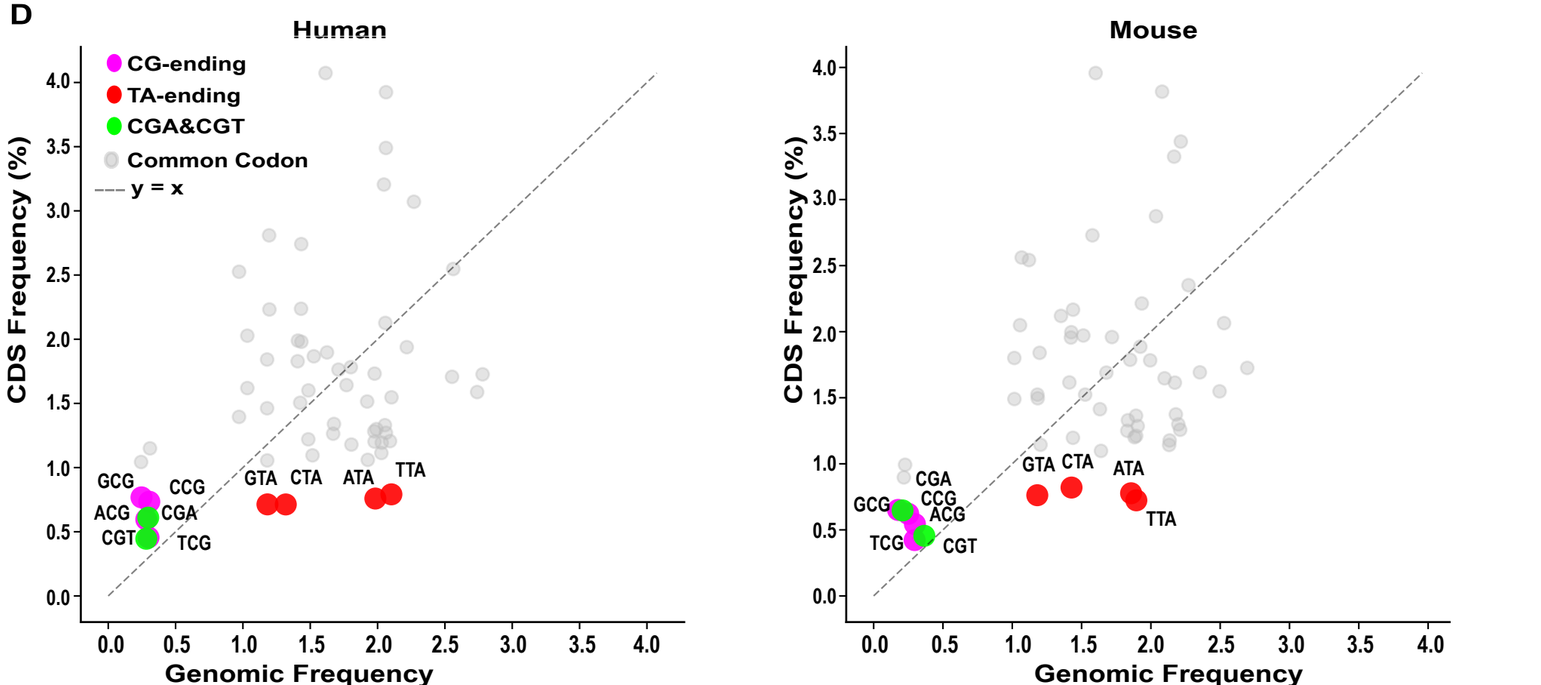
B


r = 0.980
Mouse FRO
Human FRO
CGA
CGT
ATA
CCG
GTA
ACG
GCG
CTA
TCG
TTA

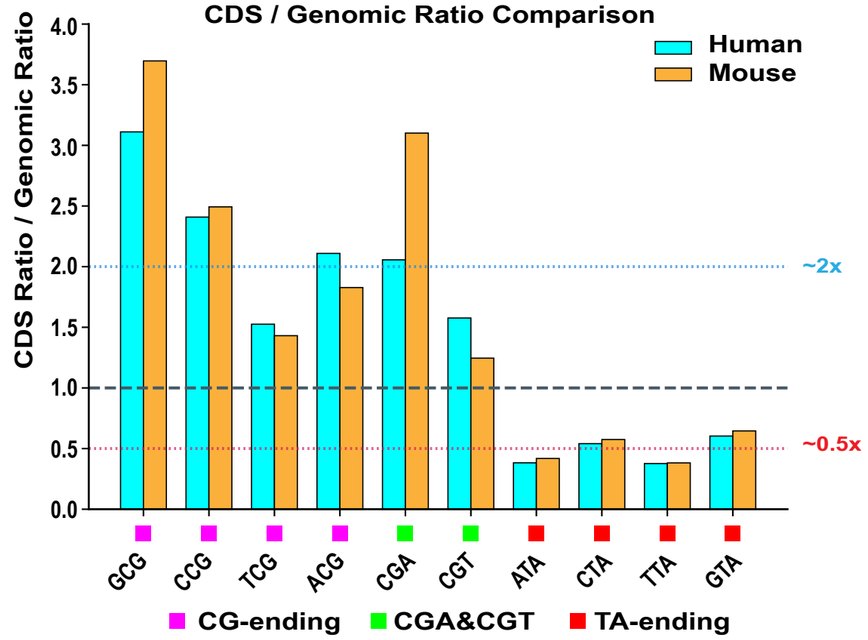
C


Human
Mouse
CG-ending
TA-ending
CGA&CGT
Others
Low Frequency Region
Frequency (%)
Genomic Frequency Rank
TTA
ATA
CTA
GTA
CCG
TCG
CGA
CGT
ACG
GCG

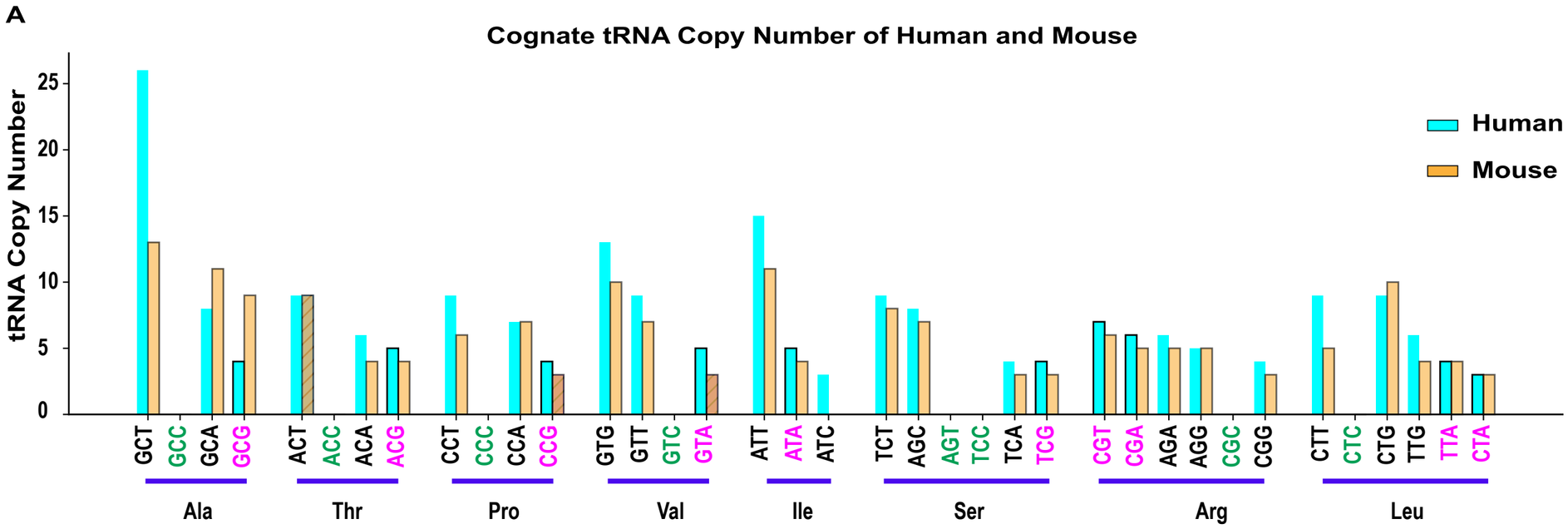
D


Human
Mouse
CG-ending
TA-ending
CGA&CGT
Common Codon
y = x
CDS Frequency (%)
Genomic Frequency
GCG
CCG
ACG
CGA
CGT
TCG
GTA
CTA
ATA
TTA

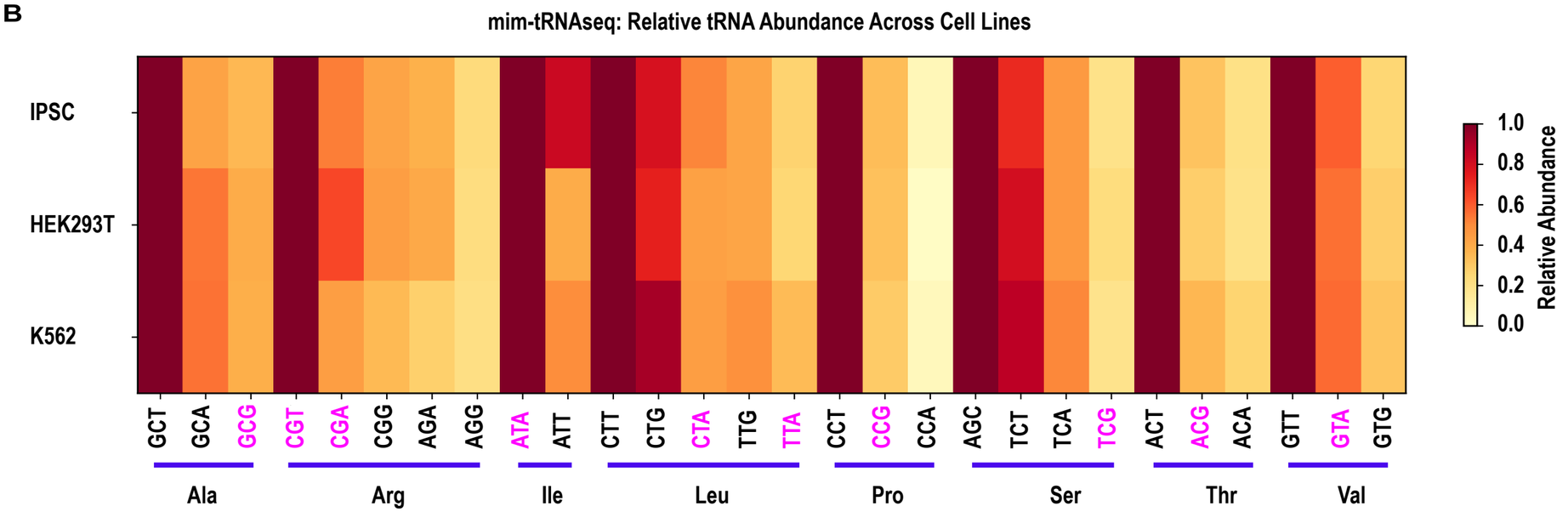
E


CDS / Genomic Ratio Comparison
Human
Mouse
CDS Ratio / Genomic Ratio
~2x
~0.5x
GCG
CCG
TCG
ACG
CGA
CGT
ATA
CTA
TTA
GTA
CG-ending
CGA&CGT
TA-ending

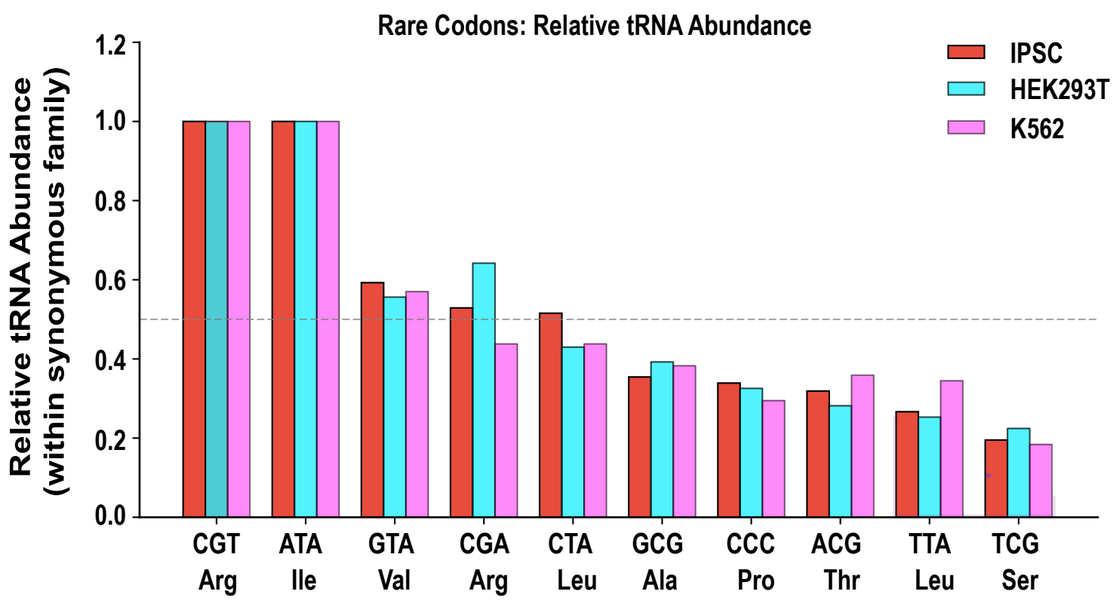
A


Cognate tRNA Copy Number of Human and Mouse
tRNA Copy Number
0
5
10
15
20
25
Human
Mouse
GCT GCC GCA GCG
ACT ACC ACA ACG
CCT CCC CCA CCG
GTG GTT GTC GTA
ATT ATA ATC
TCT AGC AGT TCC TCA TCG
CGT CGA AGA AGG CGC CGG
CTT CTC CTG TTG TTA CTA
Ala
Thr
Pro
Val
Ile
Ser
Arg
Leu

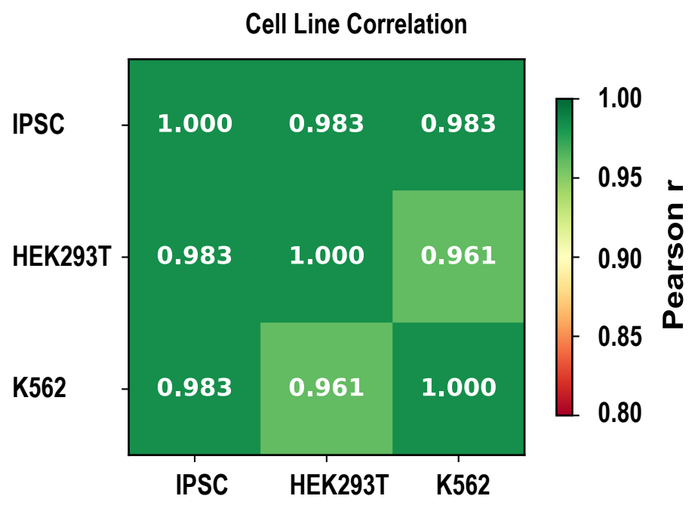
B


mim-tRNAseq: Relative tRNA Abundance Across Cell Lines
IPSC
HEK293T
K562
GCT GCA GCG CGT CGA CGG AGA AGG ATA ATT CTT CTG CTA TTG TTA CCT CCG CCA AGC TCT TCA TCG ACT ACG ACA GTT GTA GTG
Ala
Arg
Ile
Leu
Pro
Ser
Thr
Val
Relative Abundance
1.0
0.8
0.6
0.4
0.2
0.0

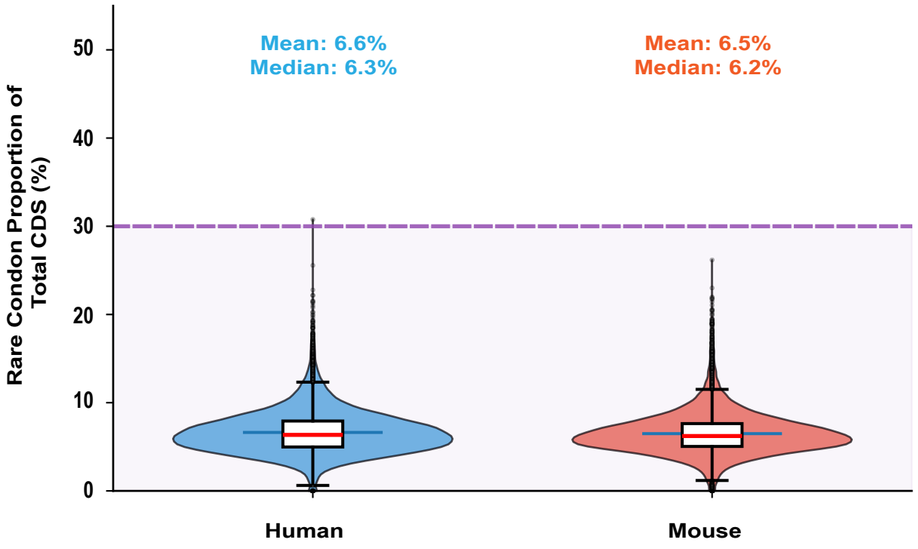
C


Rare Codons: Relative tRNA Abundance
Relative tRNA Abundance (within synonymous family)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
IPSC
HEK293T
K562
CGT Arg
ATA Ile
GTA Val
CGA Arg
CTA Leu
GCG Ala
CCC Pro
ACG Thr
TTA Leu
TCG Ser

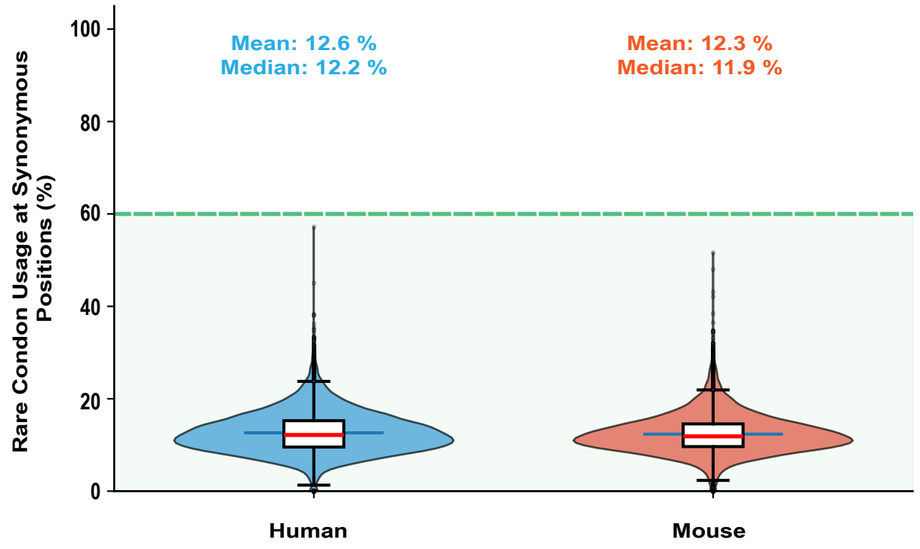
D


Cell Line Correlation
IPSC
HEK293T
K562
1.000
0.983
0.983
0.983
1.000
0.961
0.983
0.961
1.000
IPSC
HEK293T
K562
Pearson r
1.00
0.95
0.90
0.85
0.80

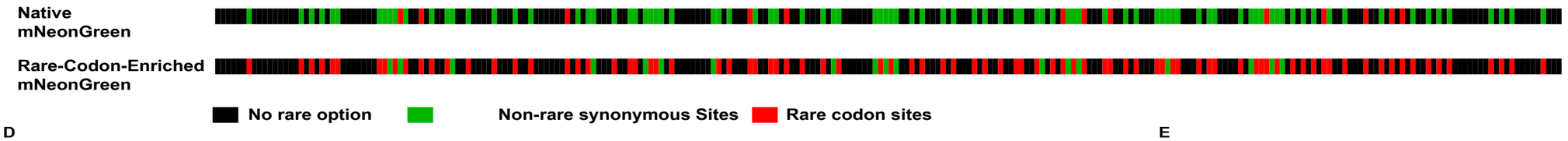
A


Mean: 6.6%
Median: 6.3%
Mean: 6.5%
Median: 6.2%
Rare Condon Proportion of Total CDS (%)
0
10
20
30
40
50
Human
Mouse

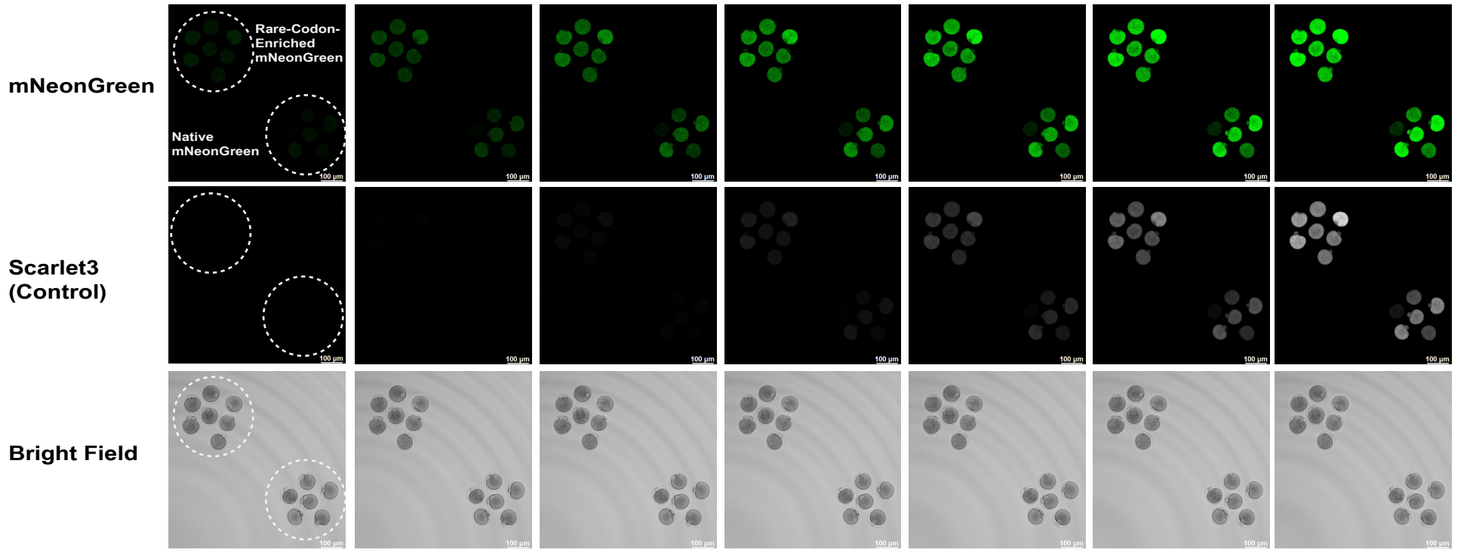
B


Mean: 12.6 %
Median: 12.2 %
Mean: 12.3 %
Median: 11.9 %
Rare Condon Usage at Synonymous Positions (%)
0
20
40
60
80
100
Human
Mouse

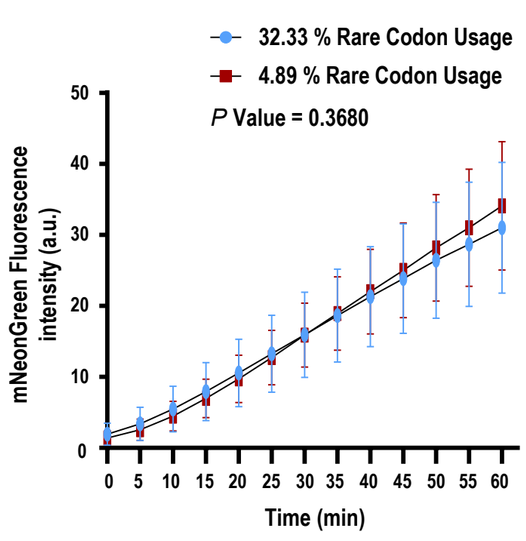
C


Native mNeonGreen
Rare-Codon-Enriched mNeonGreen
No rare option
Non-rare synonymous Sites
Rare codon sites

D


0 Min
10 Min
20 Min
30 Min
40 Min
50 Min
60 Min
mNeonGreen
Rare-Codon-Enriched mNeonGreen
Native mNeonGreen
Scarlet3 (Control)
Bright Field
100 µm

E


32.33 % Rare Codon Usage
4.89 % Rare Codon Usage
P Value = 0.3680
mNeonGreen Fluorescence intensity (a.u.)
0
10
20
30
40
50
0 5 10 15 20 25 30 35 40 45 50 55 60
Time (min)