Different types of secondary information in the genetic code

RICHARD J. MARAIA1,2 and JAMES R. IBEN1

1Intramural Research Program on Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

ABSTRACT

Whole-genome and functional analyses suggest a wealth of secondary or auxiliary genetic information (AGI) within the redundancy component of the genetic code. Although there are multiple aspects of biased codon use, we focus on two types of auxiliary information: codon-specific translational pauses that can be used by particular proteins toward their unique folding and biased codon patterns shared by groups of functionally related mRNAs with coordinate regulation. AGI is important to genetics in general and to human disease; here, we consider influences of its three major components, biased codon use itself, variations in the tRNAome, and anticodon modifications that distinguish synonymous decoding. AGI is plastic and can be used by different species to different extents, with tissue-specificity and in stress responses. Because AGI is species-specific, it is important to consider codon-sensitive experiments when using heterologous systems; for this we focus on the tRNA anticodon loop modification enzyme, CDKAL1, and its link to type 2 diabetes. Newly uncovered tRNAome variability among humans suggests roles in penetrance and as a genetic modifier and disease modifier. Development of experimental and bioinformatics methods are needed to uncover additional means of auxiliary genetic information.

Keywords: anticodon; codons; silent mutation; tRNA modification

INTRODUCTION

Paucity of understanding of the relationship between genes and the development of intricate phenotypes is accentuated by the fact that the number of protein-coding genes does not correlate with organismal complexity (Claverie 2001). Biological complexity may be better explained by considering numbers of gene-regulation networks, including alternative splicing and polyadenylation, rather than gene count (Szathmary et al. 2001). Here, we consider a layer of genetic information that resides in the synonymous codon component of the code and its functional interface with the genome-specific complement of tRNAs (the tRNAome) and the anticodon modifications that distinguish synonymous decoding.

The efficiency at which a codon is translated is determined by the amount of cognate tRNA activity available, which is in turn determined to a large extent by the tRNA gene copy number content (Gouy and Gautier 1982; Ikemura 1985; Dong et al. 1996; Berg and Kurland 1997; Percudani et al. 1997; Duret 2000; Kanaya et al. 2001; Tuller et al. 2010; Novoa et al. 2012), whereas fidelity is impacted by competition with noncognate tRNAs (Kramer and Farabaugh 2007; Kramer et al. 2010; Reynolds et al. 2010; Lamichhane et al. 2013a). If all codons were equally distributed among all mRNAs, all tRNAs were present and equally active at equimolar abundance, and high fidelity prevailed, there would be little if any consequence of synonymous or rare codon use, except on mRNA structure, stability, and other features such as splicing and transcription factor-binding sites (Chamary and Hurst 2005; Chamary et al. 2006, and references therein; Parmley et al. 2007; Gu et al. 2010; Goodman et al. 2013; Stergachis et al. 2013). However, evidence from bacteria and other microbes indicate that biased codon use is linked to translational efficiency and growth rate (e.g., Sharp et al. 2005). It is well known that a group of functionally related Escherichia coli mRNAs (for ribosomal proteins) share similar codon bias relative to other mRNAs (Ikemura 1981). Subsequent examination of 102 bacterial species with genomes of varying G + C content revealed that the faster growing ones contain higher numbers of tRNA genes but of fewer anticodon species, consistent with optimization, for growth, of a small subset of the presumed best-matched codons and anticodons (Rocha 2004). These and other analyses of

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

RNA 20:977–984; Published by Cold Spring Harbor Laboratory Press for the RNA Society

977
growing microbes support a view that biased codon selection reflects a balance of translational speed and accuracy, whereas mutational drift is a nonselective force accounting for codon variation among different genes of a genome (Bulmer 1991; Drummond et al. 2005; Wallace et al. 2013). However, optimization for growth alone may not be a most beneficial application of codon bias in higher eukaryotes (Akashi 1995; Kanaya et al. 2001; Sharp et al. 2005; Elhaik et al. 2009; Parmley and Huynen 2009). For this review, we wish to keep open the possibility that a component of codon use bias in higher eukaryotes may impact traits other than translational efficiency per se or growth (Urrutia and Hurst 2001; references therein).

Overall genome-wide codon use frequencies reflect averages consistent with species-specific overall G + C content (Duret 2002); and as noted above, differences in codon bias among mRNAs are known. Also reported was that bacteria and yeast partition codon bias to two functional mRNA groups, high abundance (for ribosomal proteins) and others (Bennetzen and Hall 1982; Forsburg 1994).

Recent examinations of whole genomes suggest a more elaborate landscape of codon bias variability that defined an additional mRNA group (Begley et al. 2007). A set of yeast mRNAs that encode proteins involved in a DNA damage response pathway share a unique pattern of codon bias that includes highly unequal distributions of certain synonymous codons (Begley et al. 2007). Additional groups of codon bias-related mRNAs involved in other specialized pathways have since been uncovered (Bauer et al. 2012; Fernández-Vázquez et al. 2013). These identify different groups of functionally related mRNAs linked by different patterns of biased codon use, and suggest that others remain to be discovered. The recently identified codon bias-related mRNA groups were uncovered because of their links to tRNA anticodon modification enzymes (Begley et al. 2007; Bauer et al. 2012; Fernández-Vázquez et al. 2013; Lamichhane et al. 2013a). Thus, there is secondary information in these codons in these mRNAs. Codon biases vary among species, suggesting that codon-related mRNA groupings may also fluctuate, representing a potential basis of a system of species-specific auxiliary genetic information (AGI).

An independent variable is the tRNA complement, which can differ significantly in gene copy number not only among related species (Chan and Lowe 2009; Iben and Maraia 2012) but also in different strains of a single yeast species (Iben et al. 2011; Iben and Maraia 2012) and among different humans (Iben and Maraia 2013; Parisien et al. 2013). tRNAome variability can be a determinant of expression of specific mRNAs, at least in some microorganisms (Bulmer 1987). Most genomes lack genes for tRNAs whose anticodons can engage in perfect-match Watson:Crick (W:C) pairing to moderate- and high-use codons (Bollenbach et al. 2007; Itzkovitz and Alon 2007); in humans, there are thirteen such codons (Chan and Lowe 2009). This requires one tRNA to function at two or more codons, one that matches perfectly and another that must be wobble decoded. Intriguingly, certain anticodon modifications can differentially alter the specificity of a tRNA for its perfect match versus synonymous codon (Agris 2004, 2008; Begley et al. 2007; Maraia et al. 2008).

A notable use of codon bias appears to serve a conserved mechanism of slow translation at the very beginning of ORFs to optimize ribosome use and facilitate efficient translation (Cannarozzi et al. 2010; Fredrick and Ibba 2010; Tuller et al. 2010; Pechmann and Frydman 2013). Others have found that rare codons at the N-terminal regions of genes reflect reduced RNA structure rather than codon rarity itself (Gu et al. 2010; Goodman et al. 2013). In other cases, rare codons reflect avoidance of Shine-Dalgarno sequences within coding regions of bacterial genes (Li et al. 2012). Codons can be more significantly constrained by factor binding sites within coding exons, applicable to ~15% of human codons (Stergachis et al. 2013). Thus, there is much information in codon use, and flexibility is afforded by the redundancy component of the code (Weatheritt and Babu 2013).

This review focuses on two types of information derived from biased codon use, codon-induced translational pauses that are key toward unique folding pathways for a subset of susceptible proteins, of which we note only a few cases, and codon use bias patterns that are shared by functionally related mRNAs that contribute to coordinate regulation, for which a small but significant and increasing fraction is known. Auxiliary genetic information has wide-reaching implications, including for genetic disease and speciation, and new methods are needed to unveil it in full.

**Codon-tRNA imbalances and wobble decoding can affect protein-specific synthetic rate and folding**

Biased codon use in some microorganisms correlates with gene expressivity, yet codon-derived information beyond primary amino acid sequence had also been known to exist (Grantham et al. 1980; Gouy and Gautier 1982). Punctuation of mRNA sequence with low frequency codons or synonymous substitutions can be a determinant of protein folding (Burrows et al. 1987; Cortazzo et al. 2002; Kimchi-Sarfaty et al. 2007; Tsai et al. 2008; Kramer et al. 2009; Pechmann and Frydman 2013; Zhou et al. 2013). Although synonymous substitutions are silent with regard to amino acid identity, they can alter the rate of translation elongation and protein folding efficiency with significant effects on activity (Spencer et al. 2012). Elongation rate at a codon that is wobble decoded is slower than at its synonymous perfect-match W:C codon decoded by the same tRNA, and the difference is greater in humans than in worms (Stadler and Fire 2011).

The degree of codon use bias and imbalance with the tRNA complement can vary among species. When it does, some mRNAs from one are poorly translated in another due to a poor match of the AGI components. Widely known examples of both perspectives exist—introduction into *E. coli* of extra...
copies of specific tRNA genes to improve translation of eukaryotic mRNAs, and synonymous codon replacement in reporter genes derived from one species for better translation in another, e.g., “humanized” GFP (Zolotukhin et al. 1996) and CRE recombinase (Shimshek et al. 2002). However, codon optimization alone may not lead to high specific activity of the heterologous protein. Harmonization of codon content and context with the right balance of tRNAs that wobble decode versus W:C decode can better re-create the balance of elongation and pausing to enhance the efficiency of correct folding (Spencer et al. 2012). This indicates that interpretation of codon-sensitive experiments should consider the potential for species-specific effects of the model organism, as will be examined later in a specific case related to pre-proinsulin and Type 2 diabetes.

Cotranslational protein folding appears to be more common in eukaryotes than bacteria and is related to functional definition of protein domains (Netzer and Hartl 1997). Other differences extend to vertebrates and reflect divergence from bacteria in translational mechanisms that are relevant to cotranslational folding (Oresic and Shalloway 1998; Oresic et al. 2003). In a most compelling ongoing study, human alleles of the multidrug resistance 1 (MDR1) gene that vary at a SNP for synonymous codons produce proteins of identical sequence but with distinct drug recognition profiles and stabilities, apparently due to structural differences (Kimchi-Sarfaty et al. 2007; Fung et al. 2014). Thus, synonymous codons can specify functionally alternate protein structures and are not always silent due to a variety of mechanisms (Chamary et al. 2006; Parmley et al. 2007; Sauna et al. 2007; Hunt et al. 2009; Brest et al. 2011; Hurst 2011; Plotkin and Kudla 2011; Stergachis et al. 2013). Synonymous substitutions can confer structural changes in mRNA that may affect ribosome pausing independent of cognate tRNA availability (Chamary and Hurst 2005; Nackley et al. 2006; Kudla et al. 2009), and a means to address this has been developed (Salari et al. 2013). At any rate, in considering this type of AGI, it is important to recognize that the same codon in a similar mRNA context may not render another protein susceptible to alternate folding nor be relevant to its function. Such effects may be dependent on the mRNA structure and stability and also on its translational robustness (Fig. 3 in Drummond et al. 2005).

In any case, selection of alternative folding outcomes for particular proteins may be determined evolutionarily by codon mutation or by changes in relative tRNA levels, even perhaps in a tissue-specific manner (Dittmar et al. 2006; Parmley and Huynen 2009) (see below) or by stress-induced anticodon modifications that alter tRNA specific activity. A challenge is to develop methods to identify such proteins whose translation is sensitive to AGI-mediated alternative outcomes of functional relevance.

An illustrative example of tRNA-, modification-, and codon-mediated AGI is the public health relationship between CDKAL1, a tRNA wobble base modification enzyme, and Type 2 diabetes (T2D). Multiple different SNP-containing Cdkal1 alleles were identified through genome-wide association studies (GWAS) as conferring relatively high risk of T2D. Unlike most tRNA modification enzymes, Cdkal1 has only one substrate, tRNA^{lys}{UUU}. It transfers a methythiol group to threonylcarbomoyl-N^2A37 (t^A37) in the anticodon loop of tRNA^{lys}{UUU} to form tRNA^{lys}{UUU}-ms^t^A37, which can decode either of the only two Lys codons, AAA and AAG. In contrast to this, its isoacceptor, tRNA^{lys}{CUU}, which does not contain ms^t^A37, can decode only its perfect-match codon, AAG. Pre-proinsulin is a rather short polypeptide that contains only two Lys codons, both AAG, one in the B chain and the other at the junction of the C peptide and A chain, immediately adjacent to codon mutations known to cause hyperproinsulinemia, a disease in which insulin is insufficiently processed prior to secretion (Dhanvantari et al. 2003; Stoy et al. 2010). The conditional knockout of cdkal1 in insulin-producing pancreatic β cells in mice indeed leads to development of T2D with proinsulin accumulation (Wei and Tomizawa 2010; Wei et al. 2011). According to a CDKAL1 model of T2D, hypomodification of tRNA^{lys}{UUU} compromises its ability to decode its wobble codon (AAG) in pre-proinsulin mRNA, resulting in misfolding, poor processing, and impaired secretion (Wei et al. 2011). It will be important to determine to what extent the associated dysfunction is due to CDKAL1 deficiency-mediated mistranslation of pre-proinsulin, i.e., with noncognate amino acid inserted at the AAG codon, or altered kinetics of cotranslational folding.

Most T2D-associated CDKAL1 alleles have mutations in intron 5 (Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research 2007) involved in alternative splicing of the only isoform specific to (the endoplasmic reticulum of) pancreatic β cells (Brambilla et al. 2012). Pre-proinsulin is expressed at very high levels in β cells and is presumably vulnerable also because the AAG codon is in a sensitive position, whereas other proteins may not be as susceptible. Although the mouse cdkal1 knockout allele was limited to pancreatic β cells, of which pre-proinsulin is a major translation product (Wei et al. 2011), the human cdkal1 alleles likely produce deficiencies, rather than ablation, with less severe effects.

This model reasonably fits with the subtle genetic influences expected to contribute to complex multigenic disease phenotypes such as T2D. As such, considerations of experimental model systems, including ones in which tRNA isoacceptors that differ in wobble decoding and anticodon modifications compete for key codons, are noteworthy. It should be emphasized that according to this model, the only two Lys codons in pre-proinsulin are exact matches for tRNA^{lys}{CUU}, which is not itself a substrate of CDKAL1. Therefore, in the presence of unaffected cognate tRNA^{lys}{CUU}, an observed alteration in pre-proinsulin translation that is due to hypomodification of tRNA^{lys}{UUU} suggests that the stoichiometric balance of these two tRNAs^{lys}, which both compete for the AAG codon, should be a
factor in the penetrance of CDKAL1 variants. In the human reference genome build hg19, there are 16 genes for tRNA\textsuperscript{UAA}UUU and 17 genes for tRNA\textsuperscript{UAG}CUU.

However, the relative number of tRNA gene copies for tRNA\textsuperscript{UAA}UUU and tRNA\textsuperscript{UAG}CUU vary significantly among humans such that the ratios of the two tRNAs\textsuperscript{Lys} differ by as much as 23% among the six individuals examined (Iben and Maraia 2013). The CDKAL1 model of T2D suggests that this variability might be a source of relevant genetic diversity that could possibly affect penetrance of T2D (Iben and Maraia 2013).

With regard to use of model organisms to examine codon-related traits, we note that in mice used to study CDKAL1-T2D (Wei et al. 2011), there are 11 and 19 tRNA\textsuperscript{Lys}UUU and tRNA\textsuperscript{Lys}CUU genes, respectively (Chan and Lowe 2009), significantly different from human. Moreover, the study utilized Lys codon-reporters in \textit{Bacillus subtilis} (Wei et al. 2011), which contains four genes for tRNA\textsuperscript{Lys}UUU and none for tRNA\textsuperscript{Lys}CUU (Chan and Lowe 2009). Thus, it may be important when using model systems for disease analysis to consider the AGI components that are not conserved.

Sets of mRNAs linked by shared patterns of codon bias

A type of AGI of focus here is derived from patterns of codon bias shared by mRNAs that comprise functional groups, e.g., a set of yeast mRNAs involved in a stress response (Begley et al. 2007). The translational efficiencies of these mRNAs are coordinated by a modification on the wobble uridines (U34) of two tRNAs—tRNA\textsuperscript{Arg}UCU and tRNA\textsuperscript{Glu}UCU—which are cognate for the biased codons (Begley et al. 2007). The heavily biased use of one codon each for Arg and Glu, but not their synonymous codons, define a small group of DNA damage response mRNAs whose translational efficiency is controlled in part by Trm9, a tRNA methyltransferase. Trm9 transfers a methyl group as the last step of biosynthesis of 5-methylcarbonylmethyluridine (mcm5U34) and 5-methylcarbonylmethyl-2-thiouridine (mcm5s2U34) on tRNA\textsuperscript{Arg}UCU and tRNA\textsuperscript{Glu}UCU. In the absence of Trm9, the DNA damage response is compromised (Begley et al. 2007). Remarkably, deficiency of human ABH18, a Trm9 homolog, leads to hypomodification of the corresponding human tRNAs and impairs the human DNA damage response (Fu et al. 2010), although identification of the cognate codon-sensitive human mRNAs responsible remains to be demonstrated.

Investigations in yeast extend the link between codon-use patterns in functionally related mRNAs, specific wobble base modifications of the cognate tRNAs, and phenotype (Bauer et al. 2012; Fernández-Vázquez et al. 2013). Although wobble base U34 is the most diversely modified nucleotide in tRNA, the second most diversely modified is position 37, flanking the other side of the anticodon. A link between functionally related mRNAs, tRNA anticodon modification, and phenotype was extended to isopentenyl-A\textsuperscript{6}-adenosine of A37 (i6A37) in fission yeast, most prominently for the high abundance subset of serine codons (Lamichhane et al. 2013a).

Bacterial mutants of the i6A37 modification enzyme, MiaA, sensitize the σ factor RpoS coding sequence (and maybe others) enriched in codons decoded by MiaA-modified tRNAs, in this case a subset of Leu codons, to translational insufficiency (Thompson and Gottesman 2014).

tRNAomes are highly variable and plausibly contribute to trait penetrance

Phylogenetic evidence indicates that the fractional tRNA gene content of the tRNAome, including isoacceptor ratios, can vary significantly among related species and even among individual humans (Marck and Grosjean 2002; Iben and Maraia 2012, 2013; Parisien et al. 2013). The subsets of tRNA isoacceptors that are substrates for anticodon modifications can also vary even when the modification enzymes themselves are conserved (Lamichhane et al. 2011, 2013b). Relative numbers of tRNA isoacceptor genes is markedly different in related \textit{Schizosaccharomyces} species, and among individual persons, particularly relevant to wobble decoding (Iben and Maraia 2012, 2013). A wide variability in \textit{Schizosaccharomyces} tRNAomes exists despite similar genome size and remarkable conservation of other features of the genomes (Iben and Maraia 2012). In addition to variation in tRNA gene number, there was a striking reorganization of the fractional content of isoacceptor tRNA genes in these species. Lack of any genes for tRNA\textsuperscript{Ala}AGC is specific to \textit{Schizosaccharomyces japonicas}, and this was accompanied by amplification of the tRNA\textsuperscript{Ala}UGC gene necessitating wobble decoding of GCU codons not required in the other species (Iben and Maraia 2012). Gene numbers for tRNA\textsuperscript{Ser}UGA, tRNA\textsuperscript{Thr}UCC, and tRNA\textsuperscript{Val}UAC have similarly diverged among \textit{Schizosaccharomyces}, all encoding U34, the most versatile of the wobble bases and most deterministic of synonymous decoding (Iben and Maraia 2012). Variability of these highly related tRNAomes presumably reflects their readiness to evolve (Yona et al. 2013; Bloom-Ackermann et al. 2014).

The relative abundances of and competition among tRNAs determine translational efficiency, pausing, and miscoding, the latter associated with misfolding and protein aggregation (Drummond et al. 2005; Plant et al. 2007; Reynolds et al. 2010). Thus, for certain mRNAs, two different tRNA complements might produce differential folding. In higher eukaryotes, tRNA complements are tissue specific (Dittmar et al. 2006), consistent with evidence of cell type-specific transcription of different tRNA genes (Barski et al. 2010; Canella et al. 2010; for review, see White 2011; Canella et al. 2012) and possibly relevant to a human codon bias pattern found mostly in tissue-specific mRNAs (Parmley and Huynen 2009). A set of codon-biased mRNAs that match the tissue-specific tRNA complement has been proposed as contributing to a breast cancer phenotype (Pavon-Eternod et al. 2009).
The complement of human tRNA genes and their sequence diversity vary among individuals (Iben and Maraia 2013; Parisien et al. 2013). As noted above for Lys synonymous codon decoding of pre-proinsulin, the ratio of gene numbers for isodecoder pairs may be relevant to trait penetrance (Iben and Maraia 2013). There are seven isodecoder tRNA gene pairs for which one can wobble to the other codon in humans whose ratios differ from 18% to 55% among the six individuals examined (Table 4 in Iben and Maraia 2013). Thus, tRNAome variability comprises significant genetic variation in humans with potential consequences to penetrance and thereby should be considered as a genetic modifier of some traits and, in some cases, as a specific disease modifier.

The anticodon modification component of AGI

Although tRNA abundance often reflects tRNA gene dosage, specific activity can be affected by tRNA modifications, especially to the anticodon loop. Some anticodon modifications appear to differentially affect the tRNA’s activity for its exact match cognate versus wobble codon (Agris 2004, 2008; Begley et al. 2007; Maraia et al. 2008). Increasing evidence indicates that some anticodon modification enzymes can be activated by stress, affecting dynamic control of codon-specific translation (Chan et al. 2010, 2012; Paredes et al. 2012; Dedon and Begley 2014).

tRNA position 37 is part of the “extended anticodon” (Yarus 1982; Agris 2008), whose modified base identity covaries with identity at wobble position 34 (Yarus 1982; Agris et al. 2007). As noted above, the link of functionally related sets of mRNAs extends to tRNA position 37, at least for i6A37, in yeast and possibly in bacteria (Lamichhane et al. 2013a; Thompson and Gottesman 2014). A prevalent modification in bacteria is iA37 on tRNAs for UNN codons—Ser, Tyr, Leu, Phe, Cys, and Trp—and this is further modified in many bacterial species. iA37 is not further modified in eu-karyotes, and it is also less prevalent, widely absent on tRNAs Phe and Leu, and variable on some others. tRNA\textsubscript{Cys} carries iA37 in 	extit{Saccharomyces cerevisiae} but tRNA\textsubscript{Trp} does not, whereas the exact opposite is true for 	extit{Schizosaccharomyces pombe} (Lamichhane et al. 2011). Although this is explained for tRNA\textsubscript{Cys} by the fact that 	extit{S. pombe} tRNA\textsubscript{Cys} carries G at 37 rather than A as in 	extit{S. cerevisiae}, the discrepancy for tRNA\textsubscript{Trp} is of greater biological interest. The 	extit{S. pombe} tRNA isopentenyltransferase (Tit1) can modify tRNA\textsubscript{Trp} from either 	extit{S. cerevisiae} or 	extit{S. pombe}, whereas the 	extit{S. cerevisiae} homolog, Mod5, is restricted in substrate recognition and cannot modify tRNA\textsubscript{Trp} from either source (Lamichhane et al. 2011). Although human TRIT1 is also less restricted, the subset of human tRNAs that carry iA37 differs from either yeast (Table 1; Lamichhane et al. 2013b).

Recent evidence from 	extit{S. pombe} indicate that the presence of iA37 increases fidelity at cognate but decreases fidelity at noncognate codons (Lamichhane et al. 2013a). Perhaps more significantly, iA37 increases the specific decoding activity of the tRNAs to which it is attached by nearly fourfold (Lamichhane et al. 2013a). Accordingly, most iA37 occurs on abundant tRNAs that decode very highly used codons enriched in the most abundant mRNAs, those for ribosome subunits, translation factors, and energy enzymes (Lamichhane et al. 2013a). Highly abundant mRNAs are biased toward highly used codons, and these include those decoded by iA37-containing tRNAs (Lamichhane et al. 2013a). Thus, iA37 promotes translational efficiency of the most highly produced proteins in 	extit{S. pombe}.

Intriguingly, although iA37 is found in all kingdoms of life, it is used variably on tRNAs that decode the full set or various subsets of UNN codons (Table 1). Accordingly, the translational attributes of iA37 may be partitioned to

| RNA anticodon | E. coli | S. pombe | S. cerevisiae | Homo sapiens |
|---------------|---------|----------|---------------|--------------|
| Ser A\textsuperscript{AGA} | —\textsuperscript{b} | i\textsuperscript{A37} | i\textsuperscript{A37} | i\textsuperscript{A37} |
| Ser CGA | A\textsuperscript{37}\textsuperscript{c} | — | — | — |
| Ser CGA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | i\textsuperscript{A37} | i\textsuperscript{A37} | i\textsuperscript{A37} |
| Ser UGA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | i\textsuperscript{A37} | i\textsuperscript{A37} | i\textsuperscript{A37} |
| Tyr GUU | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | i\textsuperscript{A37} | i\textsuperscript{A37} | G37\textsuperscript{d} |
| Tyr AUU | — | — | — | G37\textsuperscript{d} |
| Cys GCA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | G37 | i\textsuperscript{A37} | G37 |
| Cys ACA | — | — | — | — |
| Ser UCA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | — | i\textsuperscript{A37} |
| Trp CCA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | A\textsuperscript{37}\textsuperscript{c} | i\textsuperscript{A37} | G37 |
| Phe GAA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | G37 | G37 | G37 |
| Phe GAA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | G37 | G37 | G37 |
| Leu CAA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | G37 | G37 | G37 |
| Leu UAA | ms\textsuperscript{2}\textsuperscript{i}A\textsuperscript{37} | G37 | G37 | A37\textsuperscript{d} |

Bacterial tRNAs as reported for 	extit{E. coli} in the tRNA data base, tRNAdb 2009 (Jühlhing et al. 2009); for the other three species see Lamichhane et al. (2011, 2013a,b). The profiles of mitochondrial tRNAs with iA37 in these three eukaryotes differ from their cytosolic profiles and from each other (Lamichhane et al. 2011, 2013a,b).

The anticodon sequence shown (positions 34–36) reflects the DNA-encoded nucleotide rather than the actual identity after modification, e.g., inosine is found at position 34 in many tRNAs, ms\textsubscript{2} AGA.

\textsuperscript{a}The anticodon sequence shown (positions 34–36) reflects the DNA-encoded nucleotide rather than the actual identity after modification, e.g., inosine is found at position 34 in many tRNAs, ms\textsubscript{2} AGA.

\textsuperscript{b}The “—” reflects absence of any genes encoding a tRNA with the anticodon listed, as per the genomic tRNA database (Chan and Lowe 2009).

\textsuperscript{c}The iA37 is present and is part of the AAA sequence at positions 36–38 required by the tRNA isopentenyltransferases, but it is unmodified as reported in Jühlhing et al. (2009) or lacks iA37 as reported in Lamichhane et al. (2011, 2013a,b).

\textsuperscript{d}G37 is encoded in this tRNA anticodon species in some organisms.

\textsuperscript{e}Modification of this tRNA was found to be only 40%–50%, whereas the other human cytosolic iA37-tRNAs were ~95% modified (Lamichhane et al. 2013b).

\textsuperscript{f}Most human tRNA genes for Leu CAA and UAA encode G at position 37, but the few that encode A37 appear to produce tRNA that lacks the iA37 modification (Lamichhane et al. 2013b).
different mRNAs in different species. Thus, although the genetic code is “universal,” translation occurs locally, in this case imposed by tRNAs that fit the substrate preferences of the isopentenyltransferase at hand. A case in point is that the profile of human iA37-tRNAs differs from yeast, and the most cognate codon-enriched mRNAs are not represented by those for ribosomal proteins and related factors as they are in yeast (JR Iben and RJ Maraia, unpubl.).

In summary, several studies indicate that the translation of some mRNAs will be more sensitive to certain anticodon modifications than others, dependent on codon use bias. By selection in their distinct genetic backgrounds of codon modifications than others, dependent on codon use of some mRNAs will be more sensitive to certain anti-ed by those for ribosomal proteins and related factors as they are in yeast (JR Iben and RJ Maraia, unpubl.).

Nonrandom codon use by the mRNAome—a resource of auxiliary information?

As reviewed above, the exomes of microorganisms contain groups of functionally related mRNAs that share specific patterns of synonymous codon bias and are linked or “keyed” to specific subsets of tRNAs and their anticodon modifications (Begley et al. 2007; Bauer et al. 2012; Fernández-Vázquez et al. 2013; Lamichhane et al. 2013a; Thompson and Gottesman 2014). An important question is whether the same may be true in mammals and humans.

Codon bias in humans is more complex than in microorganisms; for a comprehensive review of and relevant perspective on synonymous codon bias, one should see Chamary et al. (2006). For example, the human genome, like in other mammals, is a mosaic of isochore patterns of alternating low and high GC contents, and this has significant effects on codon bias (Kanaya et al. 2001; Chamary et al. 2006; Elhaik et al. 2009).

Although synonymous mutations had been considered as neutrally evolving, it has become apparent that there are several selective forces that act on them, including at the level of mRNA structure, stability, and splicing, and moreover that many synonymous SNPs define disease-associated alleles (Chamary et al. 2006, Table 1). Yet there may be selective forces on synonymous codon bias in higher eukaryotes that remain unknown (for review, see Chamary et al. 2006). Further, it has been suggested that the weak selection for some synonymous mutations may be strengthened by interaction among synonymous mutations, via synergistic effects (Akashi 1995; Chamary et al. 2006). It seems to us that this leaves open the possibility that some component of codon bias operates in higher eukaryotes through a network of connections of synonymous codons at different sites, i.e., in different genes, consistent with the functional linkage of mRNAs through shared codon bias. Tests of whether these presumed connections exist are needed. A question is whether groups of mRNAs with shared patterns of specific synonymous codon bias might exist in higher eukaryotes, and if so are they functionally related? If so, this would represent a potential source of AGI that would be important, in conjunction with tRNAome variance and anticodon modification dynamics, toward better understanding of health and disease and their genetic modifiers.

ACKNOWLEDGMENTS

We thank Robert Crouch, Dolph Hatfield, Mark M ortin, and Harold Burgess for critical reading of this work at various stages of its development; T. Begley, N. Blewett, R. Crouch, and the PGD Seminar Series members for discussion; R. Crouch, T. Pederson, and P. Schimmel for encouragement; and anonymous reviewers for insightful and helpful comments. This work was supported by the Intramural Research Program on Genomics of Differentiation in the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

REFERENCES

Agris PF. 2004. Decoding the genome: a modified view. Nucleic Acids Res 32: 223–238.
Agris PF. 2008. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep 9: 629–635.
Agris PF, Vendeix FA, Graham WD. 2007. tRNA’s wobble decoding of the genome: 40 years of modification. J Mol Biol 366: 1–13.
Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in Drosophila DNA. Genetics 139: 1067–1076.
Barski A, Chepelev I, Liko D, Cuddapah S, Fleming AB, Birch J, Cui K, White RJ, Zhao K. 2010. Pol II and its associated epigenetic marks are present at Pol III–transcribed noncoding RNA genes. Nat Struct Mol Biol 17: 629–634.
Bauer F, Matsuyama A, Candiacci J, Dieu M, Scheliga J, Wolf DA, Yoshida M, Hermand D. 2012. Translational control of cell division by elongator. Cell Rep 1: 424–433.
Begley U, Dyavaiah M, Patil A, Rooney JP, Direnzo D, Young CM, Agris PF. 2008. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep 9: 629–635.
Begley U, Dyavaiah M, Patil A, Rooney JP, Direnzo D, Young CM, Agris PF. 2004. Decoding the genome: a modified view. Nucleic Acids Res 32: 223–238.
Berg OG, Kurland CG. 1997. Growth rate-optimised tRNA abundance optimization of the genetic code. J Mol Biol 270: 544–550.
Bloom-Ackermann Z, Navon S, Ginsgold H, Towers R, Pilpel Y, Dahan O. 2014. A comprehensive tRNA deletion library unravels the genetic architecture of the tRNA pool. PLoS Genet 10: e1004084.
Bollenbach T, Vetsigian K, Kishony R. 2007. Evolution and multilevel optimization of the genetic code. Genome Res 17: 401–404.
Brambillasca S, Altkru¨ger A, Colombo SF, Friederich A, Eickelmann P, Mark M, Borgese N, Solimena M. 2012. CDK5 regulatory subunit-associated protein 1-like 1 (CDKAL1) is a tail-anchored protein in the endoplasmic reticulum (ER) of insulinoma cells. J Biol Chem 287: 41808–41819.
Brest P, Lapuquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, Mari B, Babry P, Mosnier JP, Hebuterne X, et al. 2011. A synonymous variant in IRGM alters a binding site for mIR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn’s disease. Nat Genet 43: 242–245.
Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325: 728–730.
Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129: 897–907.
Lamichhane TN, Blewett NH, Maraia RJ. 2011. Plasticity and diversity of tRNA anticodon determinants of substrate recognition by eukaryotic A37 isopentenyltransferases. RNA 17: 1846–1857.

Lamichhane TN, Blewett NH, Cherkasova VA, Crawford AK, Iken JR, Farabaugh PJ, Begley TJ, Maraia RJ. 2013a. Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast. Mol Cell Biol 33: 2918–2929.

Lamichhane TN, Mattijsen S, Maraia RJ. 2013b. Human cells have a limited set of tRNA anticodon loop substrates of the tRNA isopentenyltransferase TRIT1 tumor suppressor. Mol Cell Biol 33: 4900–4908.

Li GW, Oh E, Weissman JS. 2012. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484: 538–541.

Maraia RJ, Blewett NH, Bayfield MA. 2008. It’s a mod mod tRNA world. Nat Chem Biol 4: 162–164.

Mark C, Grossjean H. 2002. tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticonodon-sparing strategies and domain-specific features. RNA 8: 1189–1232.

Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Dietchenko L. 2006. Human tRNA-COL-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314: 1930–1933.

Netzer WF, Harl FL. 1997. Recombination of protein domains facilitates co-translational folding in eukaryotes. Nature 388: 343–349.

Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. 2009. Methylation of the 1000-genomes project.

Oresic M, Shalloway D. 1998. Specific correlations between relative synonymous codon usage and protein secondary structure. J Mol Biol 281: 31–48.

Oresic M, Dehn M, Korenbloom D, Shalloway D. 2003. Tracing specific synonymous codon–secondary structure correlations through evolution. J Mol Evol 56: 473–484.

Paredes JA, Carreto I, Simões J, Bezzera AR, Gomes AC, Santamaria R, Kapushesky M, Moura GR, Santos MA. 2012. Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC Biol 10: 55.

Parisien M, Wang X, Pan T. 2013. Diversity of human tRNA genes from the 1000-genomes project. RNA Biol 10: 1853–1867.

Parnsey KL, Huynen MA. 2009. Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. PLoS Genet 5: e1000548.

Parnsey KL, Urrutia AO, Potrebowski L, Kaessmann H, Hurst LD. 2007. Splicing and the evolution of proteins in mammals. PLoS Biol 5: e14.

Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. 2009. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res 37: 7268–7280.

Pechmann S, Frydman J. 2013. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol 20: 237–243.

Percudani R, Favesi A, Ottonello S. 1997. Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. J Mol Biol 272: 327–330.

Plant EP, Nguyen P, Russ JR, Pittman YR, Nguyen T, Quesinberry JT, Kinzy TG, Dinman JD. 2007. Differentiating between near- and non-cognate codons in Saccharomyces cerevisiae. PLoS One 2: e517.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12: 32–42.

Purvis IJ, Bettany AJ, Santiago TC, Coggins JR, Duncan K, Eason R, Brown AJ. 1987. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. J Mol Biol 193: 413–417.

Reynolds NM, Lazazerra BA, Ibaa M. 2010. Cellular mechanisms that control mistranslation. Nat Rev Microbiol 8: 849–856.

Rocha EP. 2004. Codon usage bias from tRNA’s point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res 14: 2279–2286.

Salari R, Kimchi-Sarfaty C, Gottesman MM, Przytycka TM. 2013. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. Nucleic Acids Res 41: 44–53.

Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM. 2007. The sounds of silence: synonymous mutations affect function. Pharmacogenomics 8: 527–532.

Sharp PM, Bailes E, Grocock RJ, Peden JF, Sackett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res 33: 1141–1153.

Shimshesk DR, Kim J, Hubner MR, Speigel DJ, Buchholz F, Casanova E, Stewart AF, Seeburg PH, Sørensen R. 2002. Codon-improved Cre recombinase (ICre) expression in the mouse. Genesis 32: 19–26.

Spencer PS, Siller E, Anderson JF, Barral JM. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol 422: 328–335.

Stadler M, Fire A. 2011. Wobble base-pairing slows in vivo translation elongation in metazoans. RNA 17: 2063–2073.

Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akery JM, et al. 2013. Exonic transcription factor binding directs codon choice and affects protein evolution. Science 342: 1367–1372.

Stoy J, Steiner DF, Park SY, Ye H, Philipson LH, Bell GI. 2010. Clinical and molecular genetics of neonatal diabetes due to mutations in the insulin gene. Rev Endod Metab Disord 11: 205–215.

Szathmary E, Jordan F, Pal C. 2001. Molecular biology and evolution. Can genes explain biological complexity? Science 292: 1315–1316.

Thompson KM, Gottesman S. 2014. The MiaA tRNA modification enzyme is necessary for robust RpoS expression in Escherichia coli. J Biol Chem 196: 754–761.

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R. 2008. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J Mol Biol 383: 281–291.

Tuller T, Carmi A, Vestigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141: 344–354.

Urrutia AO, Hurst LD. 2001. Codon usage bias covariates with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 159: 1191–1199.

Wallace EW, Airoldi EM, Drummond DA. 2013. Estimating selection on synonymous codon usage from noisy experimental data. Mol Biol Evol 30: 1438–1453.

Weatheritt RJ, Babu MM. 2013. The hidden codes that shape protein evolution. Science 342: 1325–1326.

Wei FY, Tomizawa K. 2010. Functional loss of Cdkal1, a novel tRNA modification enzyme, causes the development of type 2 diabetes. Endocr J 58: 819–825.

Wei F, Suzuki T, Watanabe S, Kimura S, Kaitsuka T, Fujimura A, Matsui H, Atta M, Michiue H, Fontecave M, et al. 2011. Deficit of tRNA3G modification by Cdkal1 causes the development of type 2 diabetes in mice. J Clin Invest 121: 3598–3608.

White RJ. 2011. Transcription by RNA polymerase III: more complex than we thought. Nat Rev Genet 12: 459–463.

Yarus M. 1982. Translational efficiency of transfer RNA’s: uses of an extended anticodon. Science 218: 646–652.

Yona AH, Bloom-Ackermann Z, Frumkin I, Hansson-Smith V, Charpak-Amikam Y, Feng Q, Boeke JD, Dahan O, Pilpel Y. 2013. tRNA genes rapidly change in evolution to meet novel translational demands. Elife 2: e01339.

Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y. 2013. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature 495: 111–115.

Zolotukhin S, Potter M, Hauswirth WW, Guy J, Muzyczka N. 1996. A “humanized” green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J Virol 70: 4646–4654.