Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A common human anticodon polymorphism in an alanine-like tRNA gene predicts alanine incorporation at valine codons

Davis W. Cheng

dcheng@rockybiotech.org

Rocky Institute of Biotechnology

Research Article

Keywords: transfer RNA, anticodon, aminoacyl-tRNA synthetase, identity element, mistranslation, human population genomics

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11626981/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A common human anticodon polymorphism in an alanine-like tRNA gene predicts alanine incorporation at valine codons

Davis W. Cheng

Rocky Institute of Biotechnology, Clovis, CA 93619, USA

ORCID 0009-0001-1306-8923

Abstract

Transfer RNA identity is determined by elements recognized by aminoacyl-tRNA synthetases. In alanine, serine, and leucine tRNAs the principal identity elements lie outside the anticodon, so an anticodon substitution may uncouple aminoacylation specificity from codon recognition. Screening 431 human tRNA genes identified one candidate: tRNA-Val-AAC-6-1 on chromosome 6, which retains G3:U70, the major alanine identity element, but carries a valine anticodon. Sequence and isotype-model assignment support alanine-like identity: 88.9% maximum identity to alanine genes against 64.4% to valine genes. The anticodon polymorphism rs1233572 (chr6:28,735,467 T>C) distinguishes AGC, specifying alanine on about 61% of sampled chromosomes, from AAC, specifying valine on approximately 39% and ranging across nine cohorts from 13% in African ancestry groups to 60% in Finnish participants. Four primate species support AGC as ancestral. Under pooled Hardy–Weinberg expectations, the estimated AGC/AGC, AGC/AAC, and AAC/AAC frequencies are 37.2%, 47.6%, and 15.2%. The variant alters anticodon position 35, which separates the alanine and valine anticodon families in the analyzed gene set and is a principal valyl-tRNA synthetase determinant, while every alanine determinant is left intact and the retained acceptor-stem G:U pairs oppose valine charging. The AAC allele is therefore predicted to be partitioned between two synthetases rather than captured by one, and that partition sets the rate of alanine delivery to valine codons. Across the queried genes, common anticodon variation was confined to loci where it creates no conflict between the amino acid accepted and the synthetase's requirement.

Keywords transfer RNA · anticodon · aminoacyl-tRNA synthetase · identity element · mistranslation · human population genomics

Introduction

Accurate translation requires each transfer RNA to carry the amino acid specified by the codons it recognizes. Aminoacyl-tRNA synthetases recognize identity elements distributed throughout the tRNA. For many tRNA families, these elements include the anticodon, coupling aminoacylation specificity to codon recognition. Anticodon substitutions can disrupt this coupling when the principal aminoacylation identity elements lie elsewhere in the molecule.

Alanine, serine, and leucine tRNAs have prominent identity elements outside the anticodon. The acceptor-stem G3:U70 pair is a major determinant of alanine identity: introducing it into unrelated tRNAs can confer alanine acceptance (Hou and Schimmel 1988), whereas replacing it with G3:C70 disrupts recognition in the experimental system examined by Miller et al. (1991). Seryl- and leucyl-tRNA synthetases recognize structural features that include the long type-II variable arm. An anticodon substitution that preserves these identity elements may retain ancestral aminoacylation specificity while altering codon recognition. However, the synthetase corresponding to the new anticodon may also recognize the altered tRNA. Such substitutions could therefore generate substrates for competing synthetases, with the relative extent of aminoacylation by each enzyme influencing the translational consequence.

Anticodon substitutions occur repeatedly during evolution. In yeast, deletion of a tRNA gene can be compensated for by anticodon mutation at another locus (Yona et al. 2013). Anticodon shifts have also been documented across primates, flies, nematodes, yeasts, and enterobacteria, with gene redundancy proposed as a contributing factor (Rogers and Griffiths-Jones 2014). Such shifts have been cataloged as interspecies differences; the locus examined here appears in that survey as human_509. Targeted resequencing of 605 human tRNA genes in 84 individuals established that tRNA variation is far more extensive than genome-wide surveys had indicated, and identified 18 anticodon variants, several predicted to mistranslate (Berg et al. 2019). In humans, naturally occurring anticodon variants have been shown to cause mistranslation: alanine tRNA variants promote alanine incorporation at glutamate, valine, and threonine codons (Tennakoon et al. 2025), and a serine tRNA variant carried by approximately 2% of individuals promotes serine incorporation at phenylalanine codons in cells from healthy donors (Rozik et al. 2026).

Two questions organize this study. The first is mechanistic: how can a tRNA gene change the codons it reads without changing the amino acid it delivers? The second is evolutionary: what determines which loci are free to make that change? Here I used an identity-element-based screen to identify human tRNA genes whose predicted aminoacylation specificity differs from their anticodon assignment. The screen identified one non-alanine-annotated gene carrying G3:U70. Although this locus was previously classified as an alanine-to-valine anticodon shift, population data indicate that both anticodon states remain common in humans, so the transition can be examined while it is still segregating rather than after it has fixed. I then asked what distinguishes the loci at which anticodon variation reaches appreciable frequency from those at which it does not. The derived allele is a candidate for mistranslation rather than an experimentally established mistranslating variant.

Results

A non-alanine-annotated human tRNA gene retains the alanine identity element

I examined 431 human tRNA genes, comprising 429 genes in the GtRNAdb high-confidence set for GRCh38 and two genes excluded from that set because of isotype mismatches. The third acceptor-stem base pair was extracted from each predicted secondary structure. All 38 alanine genes carried G3:U70, providing complete recovery of the positive control. Among non-alanine isotypes, only tRNA-Val-AAC-6-1, located at chr6:28,735,429–28,735,500 on the minus strand, carried this pair (Figure 3a). The gene contains two consecutive G:U pairs, G3:U70 and G4:U69, in standard tRNA numbering.

Extending the scan to all 619 human tRNA predictions, including pseudogenes, identified 12 non-alanine genes carrying G3:U70. Eleven were pseudogenes or low-scoring predictions whose body-model assignments agreed with their anticodons. Two secondary-filtered loci in the same cluster, tRX-Asp-NNN-1-1 and tRNA-Und-NNN-1-1, had alanine-like sequences but were not classified as functional candidates.

The sequence screen did not identify a comparable candidate using the serine or leucine variable-arm criterion. Mature serine and leucine tRNAs had median lengths of 82 and 83 nucleotides, respectively, compared with 73 nucleotides for the other isotypes. Using a 79-nucleotide screening threshold, no gene outside the serine and leucine families was identified as carrying a type-II variable arm. This screen does not exclude mischarging mechanisms involving other identity elements.

The locus has an alanine-like tRNA sequence and a valine anticodon

Four lines of evidence support assignment of the locus to the alanine family (Table 1). Maximum pairwise sequence identity was 88.9% to an alanine gene, compared with 64.4% to the closest valine gene, and the ten closest sequence matches were all alanine genes. Removing the anticodon triplet increased the between-family difference from 24.5 to 25.6 percentage points for maximum identity and from 26.4 to 28.3 percentage points for median identity, indicating that the resemblance was not attributable to the anticodon. Four alanine genes occur within 25 kb of the locus. Consistent with these observations, tRNAscan-SE 2.0 (Chan et al. 2021) assigned a score of 88.1 against the alanine isotype model, compared with an overall score of 69.9, and flagged the valine anticodon assignment as an isotype mismatch.

Table 1. Evidence supporting an alanine-family assignment for tRNA-Val-AAC-6-1. Sequence identity was calculated as the number of matching alignment columns divided by the global alignment length, using mature sequences from the GtRNAdb high-confidence set (38 alanine and 27 valine genes). For masked comparisons, the anticodon triplet was removed from each sequence before realignment, and identity was calculated from the resulting alignment. IPD denotes the tRNAscan-SE isotype prediction difference; a negative value indicates disagreement between the isotype model and the anticodon assignment.

Feature	tRNA-Val-AAC-6-1	Expected of an alanine gene	Expected of a valine gene
Acceptor-stem pair 3:70	G:U (plus a second G:U at 4:69)	G:U	G:C
Reference anticodon (GRCh38)	AAC	AGC, CGC or TGC	AAC or CAC
Highest identity to the family	88.9%	—	64.4%
Median identity to the family	79.9%	—	53.4%
Highest identity, anticodon excluded	89.9%	—	64.3%
Ten most similar human tRNA genes	All 10 are alanine genes	alanine	valine
tRNA genes within 25 kb	4 alanine genes	—	—
tRNAscan-SE isotype model	Ala, score 88.1 (overall 69.9); IPD −43.10, isotype mismatch	Ala	Val
Mature length and variable arm	72 nt, short arm	72–73 nt, short	72–73 nt, short

The locus contains potential recognition determinants for two synthetases

Because alanyl-tRNA synthetase (AlaRS) and valyl-tRNA synthetase (ValRS) recognize different regions of the tRNA, I compared the locus with both families at positions reported to contribute to either aminoacylation identity (Table 2). The discriminator nucleotide was invariant across the 38 alanine and 27 valine genes: all 65 carried A73. This nucleotide is favorable for AlaRS recognition (Shi and Schimmel 1991) and is the wild-type discriminator in the ValRS system examined by Horowitz et al. (1999), in which G73 acts as a negative determinant. Position 73 therefore does not distinguish the two families in the analyzed human gene set and is compatible with recognition by both enzymes at this locus.

The families differed at the acceptor stem and anticodon. All 38 alanine genes carried G3:U70, whereas all 27 valine genes carried U3:A70; the candidate locus carried G3:U70 and G4:U69. At the anticodon, all 38 alanine genes carried G35 and all 27 valine genes carried A35, while position 36 was C in both families. Position 35 therefore distinguished the alanine and valine anticodon families throughout the analyzed gene set and was the position altered by

rs1233572. The two alleles consequently differ at a principal determinant of valine identity while retaining the alanine-associated determinants examined here.

Table 2. AlaRS and ValRS identity determinants at tRNA-Val-AAC-6-1. Family columns show the nucleotide or base pair observed among 38 alanine and 27 valine genes in the high-confidence set; these features were invariant within each family unless otherwise indicated. Determinant assignments are based on experiments in Escherichia coli and Thermus thermophilus (Hou and Schimmel 1988; Tamura et al. 1991; Liu et al. 1997; Horowitz et al. 1999; Fukai et al. 2003). Human ValRS identity elements have not been determined directly; the predicted consequences are extrapolations from these experimental systems.

Determinant	Read by	Alanine genes (n = 38)	Valine genes (n = 27)	tRNA-Val-AAC-6-1	Predicted consequence at this locus
Acceptor-stem pair 3:70	AlaRS (positive); ValRS (negative)	G:U	U:A	G:U	Alanine charging; valine charging reduced
Acceptor-stem pair 4:69	ValRS (negative)	G:U	U:A	G:U	Valine charging further reduced
Discriminator 73	AlaRS and ValRS	A	A	A	Favorable for both; does not discriminate
Anticodon position 34	Decoding (inosine substrate when A)	A, C or U	A, C or U	A	Inosine predicted; reads GUU, GUC, GUA
Anticodon position 35	ValRS (principal)	G	A	A or G (rs1233572)	A confers valine recognition; G does not
Anticodon position 36	ValRS (principal)	C	C	C	Present on both alleles

Variants at the alanine identity element itself are rare

A variant disrupting the alanine identity element while preserving the anticodon would represent a complementary configuration: a tRNA retaining alanine-codon recognition but potentially impaired in alanine aminoacylation. Querying gnomAD v4 at both nucleotides of G3:U70 in all 38 alanine genes identified 72 variants across 24 genes (Supplementary Table S9). None had a frequency of at least 1%. Three reached 0.1%, eight reached 0.01%, and the most frequent, a position-3 substitution in tRNA-Ala-TGC-5-1, reached 0.28%. Although four genes carried anticodon variants above 1%, this comparison does not establish differential constraint. The screen queried 76 identity-element positions and 1,293 anticodon positions; applying the observed rate of common anticodon variation to the smaller set yields an expected count of 0.24 common identity-element variants. The analysis therefore provides a descriptive bound rather than evidence of asymmetric constraint. No common variant disrupting G3:U70 was identified among the surveyed alanine genes, whereas anticodon variation identified potential discordance between decoding specificity and aminoacylation identity.

The anticodon is a common polymorphism

Human tRNA genes exhibit exceptionally high mutation rates (Thornlow et al. 2018). Querying gnomAD v4 at the anticodon triplets of the selected genes returned results for 429 loci; queries failed for two loci on an unplaced contig, which were excluded rather than recorded as variant-negative. The successful queries yielded 856 single-nucleotide variants across 347 genes. Of these, 606 were classified as altering the predicted assignment under the standard genetic code. One of the two selenocysteine records falls in that set; because the selenocysteine anticodon decodes UGA through a dedicated pathway rather than through the standard code, both selenocysteine records were excluded, leaving 605 identity-changing variants, of which nine reached an allele frequency of 0.1%. The most frequent identity-changing variant was located in tRNA-Val-AAC-6-1 (Figure 3b).

The variant rs1233572 (chr6:28,735,467 T>C) alters the middle anticodon nucleotide. Because the gene is on the minus strand, the genomic C allele specifies the AGC alanine anticodon, whereas the genomic T allele specifies the AAC valine anticodon. Their frequencies were approximately 61% and 39%, respectively; T is the GRCh38 reference allele. Both alleles retain G3:U70 because the substitution occurs in the anticodon rather than the acceptor stem (Figure 1). If adenosine at position 34 is modified to inosine by the ADAT2–ADAT3 heterodimer (Schaub and Keller 2002), the corresponding anticodons are expected to recognize U-, C-, and A-ending codons within the alanine or valine codon family. Modification and decoding by this particular tRNA have not been established experimentally.

The AAC allele was common, and its frequency differed substantially among ancestry groups (Figure 2a). Frequencies were corroborated across nine cohorts spanning three continents (Figure 2b), and the variant had no gnomAD quality flags. Pooling ancestry groups produced an approximately 8% excess of homozygotes, a pattern potentially attributable to population structure or mapping artifacts involving similar tRNA paralogs. Summing expectations calculated within ancestry groups yielded 30,517 observed versus 30,294 expected homozygotes, a ratio of 1.007. This attenuation is consistent with the Wahlund effect, although it does not independently exclude mapping artifacts.

Both anticodon states were common (Table 3). Under Hardy–Weinberg expectations calculated from pooled allele frequencies, an estimated 37.2% of individuals would carry two AGC copies, 47.6% one AGC and one AAC copy, and 15.2% two AAC copies; 62.8% would carry at least one AAC copy. Within ancestry groups, the expected AAC/AAC frequency ranged from 1.6% in African ancestry groups to 16.9% in Admixed American and Ashkenazi Jewish groups, 26.1% in non-Finnish Europeans, and 36.4% in Finnish participants. These are model-based genotype expectations rather than observed proportions of individuals. The two anticodon states therefore correspond to potentially common genotype classes.

Table 3. Allele frequencies and expected genotype frequencies of rs1233572 by ancestry group. Allele frequencies were obtained from gnomAD v4 genomes. Genotype percentages are Hardy–Weinberg expectations calculated from the allele frequencies in each row. AGC specifies alanine and AAC specifies valine. Groups with fewer than 1,000 sampled chromosomes are omitted. The pooled row provides a theoretical expectation based on pooled allele frequencies and does not represent observed genotype frequencies.

Ancestry group	Chromosomes	AGC (Ala)	AAC (Val)	AGC/AGC	AGC/AAC	AAC/AAC
Finnish	10,546	0.396	0.604	15.7%	47.9%	36.4%
Non-Finnish European	67,966	0.489	0.511	23.9%	50.0%	26.1%
South Asian	4,820	0.587	0.413	34.5%	48.5%	17.0%
Admixed American	15,272	0.589	0.411	34.6%	48.4%	16.9%
Ashkenazi Jewish	3,468	0.589	0.411	34.7%	48.4%	16.9%
Remaining	2,110	0.621	0.379	38.5%	47.1%	14.4%
East Asian	5,154	0.628	0.372	39.4%	46.7%	13.9%
African/African-American	41,522	0.872	0.128	76.0%	22.3%	1.6%
All sampled chromosomes	152,060	0.610	0.390	37.2%	47.6%	15.2%

Two entries in the ranked variant list corresponded to variants characterized experimentally by other groups, and their frequencies were consistent with published estimates. The frequency of rs1041881 in tRNA-Ala-AGC-6-1 was 6.59%, compared with 6.5% reported by Tennakoon et al. (2025), and the frequency of the tRNA-Ser-AGA-2-3 variant was 1.9%, compared with approximately 2% reported by Rozik et al. (2026). Recovery of these published estimates supports the consistency of the frequency extraction.

The valine anticodon is derived in the human lineage

Whole-genome pairwise alignments mapped the human locus to a single orthologous region in each of three primate species: chimpanzee chr6:28,977,923, gorilla chr6:29,366,916, and macaque chr4:141,512,719. Extraction and reverse complementation yielded full-length 72-nucleotide tRNA sequences differing from the human sequence at four to six positions. All three contained AGC and retained G3:U70 (Figure 1a). These sequences support an ancestral alanine anticodon and a derived valine anticodon in humans, consistent with a G-to-A substitution at anticodon position 35.

An initial annotation-based comparison did not recover these orthologs because it searched for genes annotated as valine tRNAs with alanine-like sequences. Direct whole-genome alignment resolved the discrepancy: the corresponding primate genes were present and annotated as alanine tRNAs.

Orangutan was examined as an additional outgroup and carried C at the aligned variant position, consistent with the alanine anticodon state observed in chimpanzee, gorilla, and macaque. This polarity agrees with the assignment by Rogers and Griffiths-Jones (2014), who classified the locus as an alanine-to-valine alloacceptor shift using synteny across six primate genomes, including alanine-anticodon orthologs in chimpanzee, orangutan, and marmoset. The present sequence comparisons corroborate that assignment. Reference assemblies alone cannot establish whether the derived state is fixed within humans; the population data demonstrate that both states remain segregating.

Common anticodon variants occur at a small subset of cataloged shift loci

The persistence of both alleles motivated an assessment of variation at other cataloged anticodon-shift loci. Twenty-six shift sets reported by Rogers and Griffiths-Jones (2014) included human genes, comprising 59 human tRNA genes on primary chromosomes. Mapping GRCh37 coordinates to GRCh38 and matching them to the current gene set resolved 51 loci, all concordant with the historical isotype and anticodon annotations. Two exclusion criteria were applied. Eight genes did not map successfully or fell outside an annotated gene. These loci occurred in the 1q21 duplication block and the MHC, where the assembly changed between builds; 41 GRCh38 genes with matching isotypes and anticodons were identified as candidates within those intervals, none carrying an anticodon variant above 0.053%. A further 12 mapped loci were secondary-filtered predictions outside the selected 431-gene set. They were not tested and were excluded rather than assigned zero variants. The resulting comparison included 39 shift loci and 390 other genes from the successfully queried set. Of the 39 shift loci, 32 carried at least one recorded anticodon variant, although most variants were rare (Table 4). Eleven loci had a variant reaching 0.01%, and two had a variant reaching 1%: tRNA-Val-AAC-6-1 and tRNA-Ala-AGC-6-1. The latter carried a 6.6% variant specifying a glycine-decoding anticodon characterized by Tennakoon et al. (2025). Four of the 429 queried genes carried a common anticodon variant; two occurred among the shift loci, which comprised approximately 9% of the queried set.

The comparison yielded an odds ratio of 10.5 (2 of 39 versus 2 of 390; one-tailed Fisher exact $P = 0.043$). Because this result depends on two observations, it should not be interpreted as established enrichment. The result was sensitive to the frequency threshold: $P = 0.19$ at 0.1% and $P = 0.009$ at 0.01%. Regional clustering is a potential confounder: 143 of the 429 queried genes were located in the extended MHC, including 12 of the 39 shift loci and three of the four genes with common anticodon variants. Restricting the comparison to selected MHC genes yielded 2 of 12 shift loci versus 1 of 131 other genes ($P = 0.019$). This restricted comparison does not establish independence from regional confounding, and no multiple-testing-adjusted claim is made. Descriptively, common anticodon variation occurred at two tested shift loci. Both carried variants potentially affecting the relationship between aminoacylation specificity and codon recognition; mistranslation has been demonstrated for one, whereas the other is examined here.

Table 4. Standing anticodon variation at cataloged anticodon-shift loci. Shift loci are human members of the primate shift sets reported by Rogers and Griffiths-Jones (2014) that map to GRCh38 and fall within the 429 genes successfully queried in gnomAD v4. Twelve

additional mapped loci outside that set were not queried and were excluded rather than classified as variant-negative. P values were calculated using one-tailed Fisher exact tests. The final row restricts both groups to genes within the extended MHC (chr6:25.7–33.4 Mb).

Criterion	Shift loci	Other queried genes	Odds ratio	P
Anticodon variant at AF ≥ 1%	2 / 39	2 / 390	10.5	0.043
Anticodon variant at AF ≥ 0.1%	2 / 39	7 / 390	2.96	0.19
Anticodon variant at AF ≥ 0.01%	11 / 39	47 / 390	2.87	0.0086
Any anticodon variant recorded	32 / 39	—	—	—
AF ≥ 1%, extended MHC only	2 / 12	1 / 131	26.0	0.019

Population differentiation and linkage disequilibrium constrain interpretation

Allele-frequency differences among ancestry groups prompted an assessment of population differentiation. Across seven gnomAD ancestry groups with sufficient counts, the descriptive differentiation index for rs1233572 was 0.077. This index summarizes variation in allele frequency among groups; it is not a selection test, and no inference regarding selection is drawn from it. Whether selection has acted at this locus remains unresolved and would require additional analyses, such as haplotype-based tests using phased genotypes.

The locus lies within the extended major histocompatibility complex (MHC), where linkage disequilibrium across HLA and non-HLA genes complicates causal-variant identification (de Bakker et al. 2006). No association for rs1233572 itself was identified in the queried GWAS Catalog release. The analysis identified 260 linked variants at $r^2 \geq 0.6$ and 527 associations spanning 241 traits within the surrounding 300-kb interval, approximately five times the genome-wide association density. These regional associations cannot be attributed to rs1233572 on the basis of linkage alone.

Sparse sequencing evidence supports transcription of the AAC allele

Expression must be interpreted in relation to the anticodon genotype: a cell line homozygous for AGC lacks an AAC allele at this locus. Reads spanning the anticodon identify the expressed anticodon state but do not establish genomic zygosity. Eight human mim-tRNAseq samples (Behrens et al. 2021), totaling approximately 35 million reads, were screened. Each read was assigned to its best match among the candidate locus, represented with a wildcard anticodon, and all human alanine and valine tRNA genes, requiring an edit-distance advantage of at least two over the second-best match.

Two reads in HEK293T replicate 1 and one in a heat-stressed HEK293T sample matched the locus; all three contained AAC. No AAC reads were assigned in K562 or iPSC samples. A single K562 read contained GGC, corresponding to neither allele and providing an indication of background assignment noise. Three decoy sequences, each differing by at least four edits from every human tRNA reference, captured no reads in any sample. Both reads from HEK293T replicate 1 showed an adenosine-to-guanosine mismatch at position 58, consistent with a reverse-transcription signature of m^1A58 and with modified tRNA rather than genomic DNA contamination.

These reads support transcription of the AAC allele in the sampled HEK293T cells but do not establish genomic zygosity or permit a reliable estimate of transcript abundance. Three assigned reads across the samples, compared with 1,404 reads for the closest paralog in one sample, indicate low representation in the analyzed libraries. Detection was inconsistent across replicates with similar sequencing depth.

Discussion

Variation at anticodon position 35 distinguishes two common predicted decoding states of an alanine-like human tRNA. Both alleles retain G3:U70, G4:U69, and A73 and are therefore predicted to remain compatible with AlaRS recognition despite their different anticodons. The variant instead alters a principal determinant recognized by ValRS. The AAC tRNA could consequently accept alanine from AlaRS or valine from ValRS, with the aminoacylated product determining the amino acid delivered to valine codons. Under pooled Hardy–Weinberg expectations, approximately 15.2% of individuals would be AAC homozygotes. The decoding range depends on anticodon modification: inosine formation at position 34 would be expected to permit recognition of GUU, GUC, and GUA. Among the 431 screened genes, this was the only non-alanine-annotated locus carrying G3:U70. Neither aminoacylation specificity nor alanine misincorporation was measured in this study.

A kinetic partitioning model for aminoacylation at this locus

The determinants summarized in Table 2 provide a basis for considering recognition by both synthetases. AlaRS does not require anticodon recognition. An acceptor-stem G:U pair can confer alanine acceptance on tRNAs with cysteine or phenylalanine anticodons (Hou and Schimmel 1988), and the same pair governs recognition of human alanine tRNA by human AlaRS (Hou and Schimmel 1989). Acceptor-stem minihelices lacking an anticodon arm can also undergo alanylation (Francklyn and Schimmel 1989). Structural studies indicate discrimination of approximately two orders of magnitude in kcat and show that G:U-dependent geometry is propagated along the acceptor stem to the 3′ terminus (Naganuma et al. 2014). Both alleles at this locus retain G3:U70, G4:U69, and A73; rs1233572 does not alter these positions. Both are therefore predicted to remain compatible with alanylation.

ValRS recognizes a different set of determinants. Its principal anticodon determinants are A35 and C36 (Horowitz et al. 1999). In the enzyme–tRNA complex, these bases are stacked and inserted into a synthetase pocket that forms base-specific hydrogen bonds (Fukai et al. 2003). The AAC allele retains both bases and is therefore a candidate ValRS substrate on the basis of its anticodon, whereas the AGC allele lacks A35.

Acceptor-stem features may influence the relative contributions of the two synthetases. G:U pairs in the acceptor helix can act as negative determinants of ValRS recognition. Introducing G:U at the conserved 4:69 position of tRNA-Val reduces aminoacylation efficiency, primarily through an increase in KM; G:U at 3:70 produces a smaller but significant reduction. Conversion of tRNA-Ala or tRNA-Phe into an efficient valine acceptor required removal of the acceptor-stem G:U pair in addition to introduction of a valine anticodon (Liu et al. 1997). The candidate locus retains G:U at both positions. A related experimental precedent was reported by Tamura et al. (1991): a G35→A35 substitution in Escherichia coli tRNA-Ala, with its acceptor-stem pairs retained, produced 18% valine charging, whereas near-complete valine charging required additional replacement of those pairs. The anticodon change associated with rs1233572 is analogous to that substitution, although its effect in the human tRNA context remains unmeasured.

These observations motivate a kinetic partitioning model in which the AAC tRNA can be aminoacylated by both synthetases. Alanylation is hypothesized to be favored because the alanine-associated determinants remain intact, whereas the valine-associated anticodon determinants occur together with acceptor-stem features that may reduce ValRS recognition. The biological consequence would depend on the fraction of candidate tRNAs aminoacylated with alanine and their contribution to decoding at valine codons. Candidate expression and the composition of the competing isoacceptor pool would influence that contribution. The relevant competing pool depends on the codon: with inosine at position 34 the AAC isoacceptor reads GUU, GUC, and GUA, so at GUU and GUC the only competing human genes are the ten Val-AAC genes, whereas at GUA the five Val-TAC genes also contribute; GUG is read by the

Val-TAC and Val-CAC isoacceptors and is not decoded by this locus. The locus is one of ten human Val-AAC genes; equal expression would imply an approximately 10% contribution by gene copy number, but copy number alone does not establish its actual contribution to the functional tRNA pool. In the AGC state, the locus would instead constitute a twenty-seventh Ala-AGC gene. These quantities are experimentally measurable, but none has been measured for this locus. Whether partitioning in human cells resembles the 18% valine charging reported in the bacterial system is unknown. The bacterial experiments provide a mechanistic precedent rather than a quantitative estimate for humans.

Two aspects of translational quality control may influence the consequences of such partitioning. AlaRS and free-standing AlaX editing factors recognize inappropriate amino acids attached to alanine tRNA substrates rather than mismatches between the attached amino acid and the anticodon. Correctly alanylated tRNA may therefore escape these editing activities despite an altered anticodon. A precedent for correction by another synthetase is provided by alanine misacylation of threonine tRNA: the cognate synthetase performs cross-editing, whereas AlaRS does not edit the product (Chen et al. 2020). By analogy, ValRS cross-editing is a candidate corrective mechanism at this locus because its principal anticodon recognition determinants remain intact. Whether human ValRS deacylates the candidate alanyl-tRNA remains untested. Elongation-factor binding may provide a further influence. In the bacterial system studied by Schrader et al. (2011), affinity for elongation factor Tu depends on the esterified amino acid and three T-stem base pairs rather than the anticodon. An alanylated tRNA with an alanine-like body may therefore retain elongation-factor compatibility, although this extrapolation requires verification in the human system.

The frequency of the candidate allele distinguishes this locus from previously characterized mistranslating tRNA variants. The rs1041881 allele in tRNA-Ala-AGC-6-1 has a reported frequency of approximately 6.5% and directs alanine incorporation at glycine codons (Tennakoon et al. 2025). A serine tRNA variant associated with mistranslation in cells from healthy donors occurs in approximately 2% of individuals (Rozik et al. 2026). The AAC allele examined here occurs on approximately 39% of sampled chromosomes and on more than half of chromosomes in some European ancestry groups. Its allele frequency is approximately sixfold higher than that reported for rs1041881; estimates expressed per individual and per chromosome are not directly comparable.

Three findings limit interpretation. First, no selection test was performed: the differentiation index is descriptive, and allele frequency alone does not establish adaptation. Second, extensive MHC linkage disequilibrium prevents attribution of regional trait associations to this variant without additional evidence. Third, transcription is supported by only three assigned reads, with uncertainty regarding abundance and reproducibility; near-identical isodecoder sequences make unambiguous assignment of short reads to individual tRNA loci intrinsically difficult, and reference choice measurably affects the result (Pichot et al. 2021).

The representation of this locus may complicate its identification in screens organized by annotated tRNA family. Because GRCh38 carries the valine anticodon allele, the gene is cataloged among valine tRNAs and may be omitted from screens restricted to alanine annotations. An evolutionary catalog based on reference assemblies also cannot establish whether an inferred anticodon shift is fixed within a species. The targeted resequencing survey of Berg et al. (2019) illustrates the point concretely. That study applied the two screens that would be expected to recover this configuration, searching for anticodon variants in the families whose aminoacylation is anticodon-independent and for variants that create a G3:U70 pair, and reported candidates of both kinds. This locus falls between them: it is annotated as a valine gene, so it lies outside the isotype-restricted anticodon screen, and it already carries G3:U70 in the reference assembly, so no variant creates one. Screens built on either criterion alone will systematically miss loci of this type. These considerations support integrating tRNA identity-element screening with population variation

data. Among the tested human members of the Rogers and Griffiths-Jones (2014) catalog, one additional locus carried a common anticodon variant; its mistranslation potential has already been characterized.

Two complementary measurements would test the proposed mechanism. The first would identify the amino acid attached to the AAC tRNA. Charging-state sequencing indicates whether a tRNA is aminoacylated but generally does not identify the attached amino acid; an aminoacylation assay or direct analysis of aminoacyl-tRNA is therefore required. The second would test alanine incorporation at valine-coded positions in human proteins. Amino acid substitutions can be detected systematically in shotgun proteomes (Mordret et al. 2019), and a genotype-stratified comparison of AAC and AGC homozygotes could assess the predicted difference. Supplementary Data S7 presents sampling requirements under the stated assumptions. The occurrence of both genotypes in the population provides a potential basis for endogenous comparisons, complementing experiments using engineered constructs.

What constrains which loci can change a decoding assignment

The locus examined here suggests that the freedom of a tRNA gene to change its anticodon is set at three levels, and the present data speak to each. The first is recognition architecture. Where the cognate synthetase reads the anticodon, an anticodon substitution changes the amino acid delivered as well as the codon read, and the two properties move together; where the principal determinants lie elsewhere, the properties can separate. Among the 429 queried genes, four carried an anticodon variant above 1%, and the pattern across them is consistent: three occur at loci whose aminoacylation identity is determined outside the anticodon, and the fourth, in tRNA-Arg-CCT-5-1, changes CCT to TCT and so alters which arginine codons are read while still specifying arginine. None of the four disturbs the relationship between a tRNA and the synthetase that charges it: at the three decoupled loci the anticodon is not the recognition signal, and at the fourth the substitution is synonymous in amino-acid terms. With only four observations this is a pattern rather than a demonstration, but it is the pattern the recognition architecture predicts.

The second level is local sequence context, which determines how completely a decoupled locus actually decouples. At tRNA-Val-AAC-6-1 the same acceptor-stem feature acts in both directions: G3:U70 and G4:U69 license alanylation, and acceptor-helix G:U pairs are simultaneously negative determinants for ValRS (Liu et al. 1997). The feature that permits the uncoupling therefore also biases the partition between the two synthetases. Decoupling in this sense is quantitative rather than binary, and the degree of it is a property of the individual locus rather than of the isotype.

The third level is dosage. A locus that contributes a small share of its isoacceptor pool can sustain a change in decoding with a correspondingly small effect on translation. This gene is one of ten human Val-AAC genes and was sparsely represented in the examined libraries. The present data do not, however, support redundancy as a general explanation: an exploratory comparison found per-gene frequencies of an anticodon variant above 0.01% of 0.14 in families with at least 20 genes and 0.11 in smaller families, which does not distinguish the hypothesis from the null.

One part of the question is beyond the reach of these data, and the limit is worth stating plainly. If decoupled loci are free to vary at the anticodon, the complementary prediction is that they are constrained at the identity element itself. No common variant disrupting G3:U70 was found among the 38 alanine genes. That observation is consistent with constraint but does not establish it: applying the observed rate of common anticodon variation to the 76 identity-element positions queried yields an expected count of 0.24, so the absence of such variants is what sampling alone would produce. Distinguishing constraint from sampling at these positions requires either a larger population sample or direct measurement of fitness effects, and the present analysis supplies a bound rather than an answer.

Identity-decoupling variants as a class. The locus can be considered within an operationally defined class of candidate identity-decoupling variants. Three criteria are proposed: aminoacylation specificity is not principally determined by the anticodon; the tRNA retains an identity element associated with that specificity; and an anticodon variant segregates in the population. Sequence annotation and population data can identify candidates, although functional consequences require experimental verification. Among the 429 successfully queried human genes, four carried a common anticodon variant, and three met the proposed criteria (Table 5). These were tRNA-Val-AAC-6-1 and tRNA-Ala-AGC-6-1, which retain G3:U70, and tRNA-Ser-AGA-2-3, which has the long variable arm associated with serine identity. The fourth, tRNA-Arg-CCT-5-1, does not meet the first criterion because arginyl-tRNA synthetase recognizes anticodon elements. Mistranslation has been demonstrated for two members (Tennakoon et al. 2025; Rozik et al. 2026); the remaining candidate is examined here. This framework groups candidates by shared sequence and recognition criteria without establishing their functional equivalence.

A potential contribution to protein variation between individuals. Alternative splicing and RNA editing diversify protein sequences by modifying the RNA template. Candidate identity-decoupling variants could contribute to protein variation through changes in the decoding machinery while leaving the transcript unchanged. At this locus, AAC and AGC homozygotes may differ in their tRNA contributions to valine-codon decoding; their expected frequencies under pooled Hardy–Weinberg assumptions are 15.2% and 37.2%, respectively. Whether these genotype differences produce measurable protein substitutions remains unknown. Applying identity-element screening to additional sequenced populations could identify other candidates for functional investigation.

Table 5. Four human tRNA genes carrying common anticodon variants. The table lists genes with an anticodon variant reaching an allele frequency of 1% among the 429 genes successfully queried in gnomAD v4. Anticodons are shown in the orientation of the gene. The identity-element column indicates a principal recognition determinant for the cognate synthetase. When aminoacylation identity is determined outside the anticodon, an anticodon substitution may alter codon recognition while preserving aminoacylation specificity. Three of the four variants meet the proposed criteria for candidate identity-decoupling variants.

Gene	Variant (GRCh38)	rsID	Anticodon	Amino acid specified	Allele frequency	Principal identity element	Anticodon read by synthetase
tRNA-Val-AAC-6-1	6-28735467-T-C	rs1233572	AAC → AGC	Val → Ala	0.610	G3:U70 (alanine)	no
tRNA-Arg-CCT-5-1	16-3193951-C-T	rs403863	CCT → TCT	Arg → Arg	0.268	anticodon (arginine)	yes
tRNA-Ala-AGC-6-1	6-28812110-C-G	rs1041881	AGC → ACC	Ala → Gly	0.066	G3:U70 (alanine)	no
tRNA-Ser-AGA-2-3	6-27495848-G-A	rs147439337	AGA → AAA	Ser → Phe	0.019	long variable arm (serine)	no

Biological significance of substoichiometric substitution. A mistranslating tRNA can alter a fraction of protein molecules, allowing substituted and unsubstituted forms to coexist within a cell. Under the model used here, the probability of at least one substitution increases with the number of susceptible codons (Table 6). For a protein containing 30 valine codons, the projected fraction of molecules with at least one alanine substitution ranges from 0.3% at a per-codon rate of 10^{-4} to approximately 79% at the rate reported for an overexpressed construct. The actual substitution rate must be measured to establish the extent of proteome involvement. The alanine–valine substitution pair has a BLOSUM62 score of zero (Tennakoon et al. 2025), although consequences depend on the structural and functional context of each site. Potential effects may be deleterious or beneficial under different conditions. Replacing valine with alanine removes two side-chain carbons and reduces packing volume at buried positions. In a well-characterized experimental system, cavity-creating substitutions destabilized folded proteins by 2.7–5.0 kcal mol^{-1}, with the energetic cost related to the volume of the cavity (Eriksson et al. 1992). Selection associated with mistranslation-induced misfolding has been proposed to constrain coding-sequence evolution,

particularly in neural tissue (Drummond and Wilke 2008). This provides a rationale for investigating tissue-specific consequences but does not establish a neural effect at this locus. Conversely, controlled translational ambiguity can generate heterogeneous protein populations and has been described as adaptive in several systems (Hoffman et al. 2017). The present data do not resolve which consequences, if any, apply here. A phenotype could be quantitative and genotype-dependent. The expected AAC/AAC and AGC/AGC frequencies of 15.2% and 37.2%, respectively, identify potentially common groups for investigation.

Table 6. Projected fraction of protein molecules carrying at least one alanine substitution. Values were calculated as $1 - (1 - r)^n$ for a protein containing n valine codons and a per-codon misincorporation rate r. These are projections at assumed rates, not measurements; no misincorporation rate was measured in this study. For context, reported missense error rates per codon range from 4×10^{-5} to 6.9×10^{-4} in yeast and from 10^{-4} to 10^{-3} in Escherichia coli (Kramer and Farabaugh 2007; Kramer et al. 2010). A rate of 5×10^{-2} was reported for an overexpressed alanine tRNA anticodon variant (Tennakoon et al. 2025). The corresponding rate for the AAC allele at this locus remains unknown.

Valine codons in the protein	$r = 10^{-4}$	$r = 10^{-3}$	$r = 5 \times 10^{-3}$	$r = 5 \times 10^{-2}$
10	0.1%	1.0%	4.9%	40.1%
30	0.3%	3.0%	14.0%	78.5%
60	0.6%	5.8%	26.0%	95.4%
100	1.0%	9.5%	39.4%	99.4%

A hypothesis for persistence of the AAC allele. The AAC allele occurs on approximately 39% of sampled chromosomes. The proposed partitioning model offers a possible explanation for how an allele with the potential to uncouple aminoacylation identity from codon recognition might persist at high frequency. If ValRS aminoacylates a substantial fraction of AAC transcripts, only part of the candidate tRNA pool would carry alanine to valine codons. Its overall contribution would also depend on expression relative to other valine-decoding tRNAs and on any corrective cross-editing activity. The locus is one of ten Val-AAC genes, but gene copy number alone does not bound its functional contribution. Its representation in the examined libraries was low, with three assigned reads across samples compared with 1,404 reads for its closest paralog in one sample. Together, limited expression, competing valine aminoacylation, and potential cross-editing could reduce misincorporation. Whether the resulting rate approaches the reported yeast missense-error range of 4×10^{-5} to 6.9×10^{-4} (Kramer et al. 2010) remains unknown. Such a rate would not, by itself, establish neutrality or explain the allele frequency. The model is a hypothesis based on observations in other systems, rather than a demonstrated mechanism. A rate above approximately 10^{-3} at valine codons in AAC homozygotes would challenge the proposed low-misincorporation scenario and motivate additional functional and evolutionary analyses. An exploratory comparison of the simpler redundancy hypothesis did not distinguish it from the null: per-gene frequencies of an anticodon variant above 0.01% were 0.14 in families with at least 20 genes and 0.11 in smaller families.

Conclusion

This study set out to ask how a tRNA gene can change the codons it reads without changing the amino acid it delivers, and what determines which loci are free to do so. On the first question, the answer at tRNA-Val-AAC-6-1 is that the two properties are read from different parts of the molecule by different enzymes. The polymorphism rs1233572 (chr6:28,735,467 T>C), a single-nucleotide substitution at the central anticodon position that distinguishes the alanine-specifying AGC anticodon from the valine-specifying AAC anticodon, alters position 35, which separates the alanine and valine anticodon families throughout the analyzed human gene set and is a principal determinant of ValRS recognition, while G3:U70, G4:U69 and A73 are present on both alleles and are untouched by the variant. The AAC allele is therefore predicted to be a substrate for both synthetases rather than for either alone, with the retained

acceptor-stem G:U pairs biasing the partition toward alanylation, and the resulting alanyl-tRNA is not a substrate for editing activities that check the amino acid rather than the anticodon. On the second question, the data indicate that such changes are confined to loci where they do not place the tRNA's amino acid in conflict with its synthetase, that the degree of uncoupling is set by local sequence context rather than by isotype, and that gene dosage modulates the consequence without, on these data, explaining which loci vary. What remains unmeasured is the quantity on which both answers turn: the partition between the two synthetases, and hence the rate of alanine delivery to valine codons. The identity-decoupling criteria give a way to find the other loci where the same division of labor is open to variation.

Methods

Gene sets and structures. tRNA predictions, secondary structures, and mature sequences for GRCh38 were obtained from GtRNAdb files dated 2019. The resource uses tRNAscan-SE 2.0 (Chan et al. 2021), including the isotype-specific covariance models and isotype prediction difference scores used here. The analysis set comprised 429 high-confidence genes and two genes annotated as isotype mismatches. Acceptor-stem pairs were identified by stack matching the tRNAscan-SE structure strings and are reported in standard tRNA numbering, in which the acceptor stem contains pairs 1:72 through 7:66. Raw sequence coordinates were distinguished from standard structural positions. The gene body ends at the discriminator nucleotide, position 73, and the mature CCA tail is added post-transcriptionally. Variable-arm screening used intron-free mature sequences; genomic lengths were not used because introns inflate apparent lengths and compromised recovery of the positive control.

Identity determinants. Acceptor-stem base pairs, anticodon nucleotides, and the discriminator nucleotide at position 73 were extracted from tRNAscan-SE secondary-structure strings for the high-confidence set by matching the leading and trailing acceptor-stem helices. The discriminator was identified as the single unpaired nucleotide immediately 3’ of the acceptor stem. The post-transcriptionally added CCA tail was absent from the gene body. Determinant assignments were obtained from the experimental literature cited in the text rather than inferred solely from these sequences.

Sequence similarity. Pairwise sequence identity between the candidate locus and all high-confidence alanine and valine tRNA genes was calculated from mature sequences using global Needleman–Wunsch alignment with a match score of 1, mismatch score of 0, gap-opening penalty of −2, and gap-extension penalty of −0.5. Identity was defined as the number of matching columns divided by the alignment length. For masked comparisons, the three anticodon nucleotides were removed from each sequence before realignment, and identity was calculated from the resulting alignment.

Population variation. Anticodon coordinates were queried in gnomAD v4 (Karczewski et al. 2020) through its GraphQL interface. Queries failed for two of the 431 loci, both on an unplaced contig; these loci were excluded from the variant analysis rather than recorded as variant-negative, so all gnomAD-based counts and enrichment tests use the 429 successfully queried genes as their denominator. For each variant, the amino acid specified by the alternative anticodon was determined after accounting for gene orientation. Hardy–Weinberg expectations were calculated within ancestry groups and summed across nine groups with sufficient counts: afr, nfe, eas, sas, amr, fin, asj, mid, and remaining. The ami group was excluded because of its small sample size. These groups differ from the seven used for the differentiation index, and a summed observed-to-expected ratio near unity does not constitute a formal test of within-group equilibrium. Reference bases were checked against GRCh38 using Ensembl. Additional cohort frequencies were obtained from the dbSNP record for rs1233572.

Identity-element variants. Genomic coordinates of both nucleotides in the third acceptor-stem pair were derived from the tRNAscan-SE structure strings for each of the 38 high-confidence alanine genes, accounting for gene orientation. Each nucleotide was queried in gnomAD v4 using a single-position interval. All returned variants with called genotypes are reported in Supplementary Table S9.

Primate orthology. LASTZ_NET pairwise whole-genome alignments were retrieved from Ensembl for human chr6:28,735,300–28,735,650 against chimpanzee, gorilla, and macaque. The human gene interval was mapped through each alignment, and the orthologous sequence was extracted and reverse-complemented. Anticodons and acceptor-stem pairs were then examined directly. Orangutan was queried only within a 13-nucleotide window centered on the variant position; it therefore contributed evidence at the aligned variant position rather than a full-length orthologous sequence.

Population differentiation. A descriptive differentiation index was calculated as the unweighted variance of allele frequencies across seven gnomAD ancestry groups divided by the heterozygosity corresponding to the pooled, sample-size-weighted allele frequency. This is neither the Hudson nor the Weir–Cockerham estimator and is reported as a descriptive index rather than a formal FST estimate. Comparisons with background sets of frequency-matched variants were attempted but are not reported. Across four sampling windows in the extended MHC and on chromosome 2, the median background index varied approximately fivefold, indicating sensitivity of the estimated percentile to background-window selection. The index is therefore reported without a background percentile and is not interpreted as a selection test.

Expression. Raw reads from eight human mim-tRNAseq samples in GSE152621 (Behrens et al. 2021) were obtained from the European Nucleotide Archive. Reads containing ATGCTT or its reverse complement were aligned by edit distance against the candidate locus with a wildcard anticodon, three decoys, and all human alanine and valine tRNA genes. A read was assigned when the best edit distance was no greater than 6 and was at least 2 lower than the second-best distance. Decoys were generated by substituting four positions outside the anticodon, and each was verified to differ by at least four edits from every human tRNA reference. Series GSE141436 and GSE272400 were excluded because untrimmed reads yielded only adapter matches and, after trimming, none of 3,000 sampled reads matched a human tRNA.

Anticodon-shift loci. Supplementary Files 2 and 3 from Rogers and Griffiths-Jones (2014) were obtained from the publisher and parsed to identify shift sets containing human genes. GRCh37 coordinates were mapped to GRCh38 using the Ensembl assembly-mapping endpoint and matched to GtRNAdb genes by nearest overlap with a 30-nucleotide tolerance. Isotype and anticodon concordance with the historical annotations was checked separately and confirmed for all 51 matches; the matching algorithm itself did not enforce concordance. Matched loci were cross-referenced against the anticodon-variant table. Loci outside the 431-gene selected set were classified as not tested and excluded rather than recorded as variant-negative. Enrichment was assessed using one-tailed Fisher exact tests against the other selected genes and was also examined within the extended MHC.

Associations. The full GWAS Catalog association release, containing 1,203,840 records, was searched by rsID for rs1233572 and its linked variants and by genomic position within chr6:28.60–28.90 Mb. Linked variants at $r^2 \geq 0.6$ were retrieved from Ensembl for the CEU, YRI, CHB, and FIN panels.

Software and assistance. Analyses were implemented in Python 3 using the requests, NumPy, and Matplotlib libraries. Sequence alignment used a custom global Needleman–Wunsch implementation with a match score of 1, mismatch score of 0, gap-opening penalty of −2, and gap-extension penalty of −0.5. Under the author's direction, Anthropic's

Claude was used to develop and execute analysis scripts, query the public databases listed above, and prepare draft text for author revision. OpenAI's ChatGPT was used for American scientific English editing. Further details are provided in the artificial intelligence disclosure. Analysis code and derived data tables accompany the manuscript as Supplementary Information.

Figures

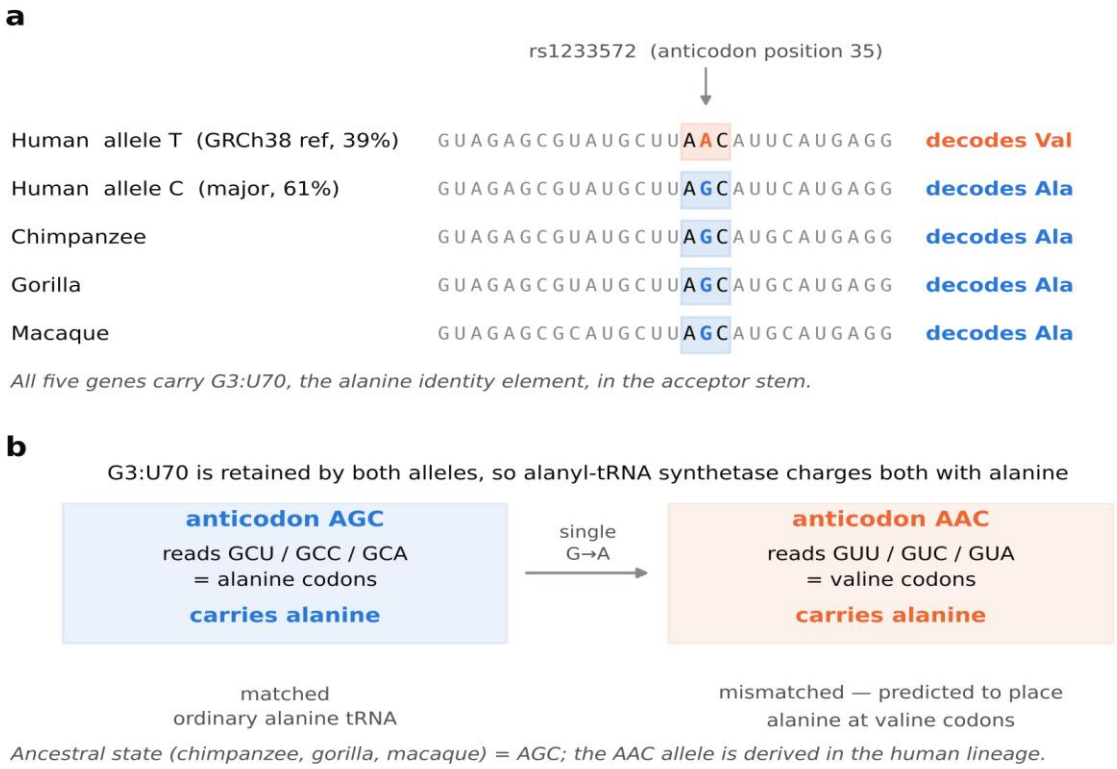


Figure 1. The rs1233572 polymorphism and the ancestral state of the locus. (a) Anticodon-loop sequences of the two human alleles and orthologous genes from three primate species. All five sequences retain G3:U70, the major alanine identity element. (b) Proposed mechanism: retention of G3:U70 predicts alanine aminoacylation of both alleles, whereas their anticodons specify different predicted codon-recognition patterns. Aminoacylation and mistranslation have not been experimentally demonstrated for this locus.

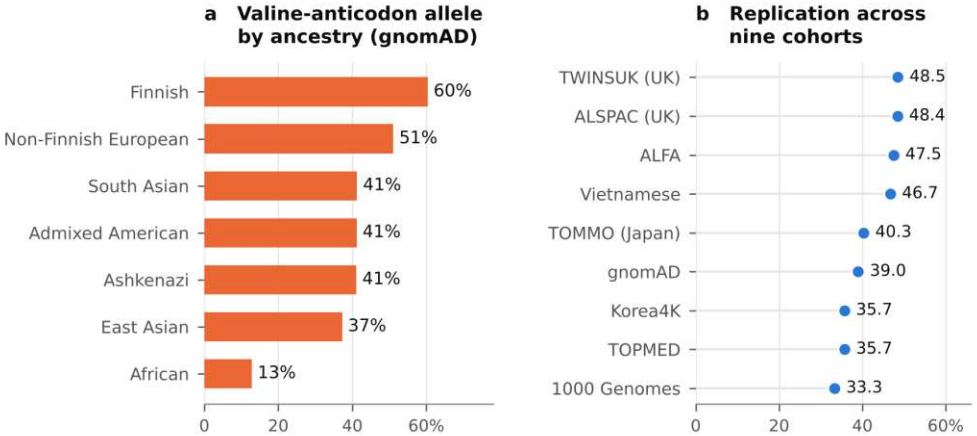


Figure 2. Population distribution of the valine-anticodon allele. (a) Allele frequency across gnomAD ancestry groups. (b) Allele frequency across nine cohorts.

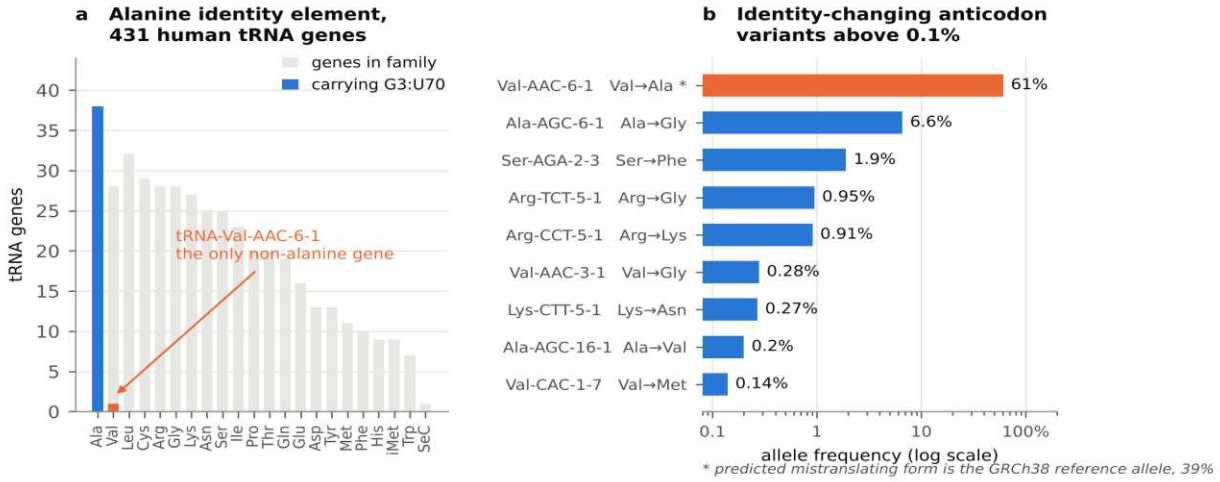


Figure 3. Distribution of G3:U70 and anticodon variants in the screened human tRNA genes. (a) Genes carrying G3:U70 by isotype among 431 screened genes: all 38 alanine genes, one valine-annotated gene, and none of the other screened isotypes. (b) Anticodon variants that alter the predicted assignment under the standard genetic code and reach an allele frequency of at least 0.1%. Labels and frequencies follow the gnomAD reference-to-alternative convention: arrows indicate the amino acid specified by the GRCh38 reference anticodon followed by that specified by the alternative anticodon, and bars show alternative-allele frequencies. At rs1233572, the alternative C allele specifies the ancestral AGC anticodon and occurs at approximately 61%; the candidate mistranslating form is the reference T allele, which specifies the derived AAC anticodon and occurs at approximately 39%. In this asterisk-marked row, the plotted allele is therefore the ancestral alanine-specifying allele rather than the candidate mistranslating allele.

Acknowledgments

The author thanks the maintainers of GtRNAdb, gnomAD, Ensembl, the GWAS Catalog, and the European Nucleotide Archive for providing public access to the data used in this study.

Statements and declarations

Funding

This study received no external funding. No grants, contracts, or other support from public, commercial, or nonprofit agencies were used for its design, conduct, analysis, or reporting.

Competing interests

The author declares no financial or nonfinancial interests, either currently or during the preceding three years, that are relevant to this article.

Ethics approval and consent

Not applicable. This study analyzed aggregate, de-identified data from public reference and population resources and publicly available sequencing datasets. No participants were recruited, no human or animal samples were collected, and no individual-level genotypes were accessed. Ethics approval and informed consent were therefore not required.

Data availability

All primary data are publicly available. tRNA gene sets, secondary structures, and mature sequences were obtained from GtRNAdb (https://gtrnadb.ucsc.edu/). Population allele frequencies were obtained from gnomAD v4 (https://gnomad.broadinstitute.org/) and dbSNP. Reference sequences and pairwise whole-genome alignments were obtained from Ensembl (https://rest.ensembl.org/). Trait associations were obtained from the GWAS Catalog full

release (https://www.ebi.ac.uk/gwas/). Sequencing reads are available in NCBI GEO series GSE152621. The focal variant is rs1233572 at chr6:28,735,467 (GRCh38); tRNA-Val-AAC-6-1 is located at chr6:28,735,429–28,735,500 on the minus strand. No new sequence data were generated. Derived data are provided in Supplementary Tables S1–S3, S8, and S9 and Supplementary Data S4–S7. Table S3 lists all 856 analyzed anticodon variants with classification flags. Of these, 606 carry the identity-change flag, including 31 that specify stop codons. One of the two flagged selenocysteine records is among the 606; both are excluded from the standard-code comparison because the selenocysteine anticodon is decoded through a dedicated pathway, leaving 605. Supplementary Tables are numbered as follows: S1, acceptor-stem pair 3 across the screened genes; S2, raw gnomAD anticodon-variant records with callability flags; S3, classified anticodon variants; S8, standing variation at the Rogers and Griffiths-Jones shift loci; S9, variants at the alanine identity element. Two records with no called genotypes (AC = 0, AN = 0) are marked as not callable in Table S2 and excluded from analysis. Analysis code is supplied with the submission and will be deposited in Zenodo with a citable DOI upon acceptance.

Materials and code availability

All analysis scripts used to generate the reported results, figures, and tables are provided as Supplementary Information to support reproduction of the analyses using the cited public resources. No new materials, reagents, or cell lines were generated.

Author contributions

D.W.C. is the sole author. He conceived the study, formulated the hypotheses, designed and directed the analyses, interpreted the results, wrote the manuscript, and verified the reported data.

Use of artificial intelligence

The author conceived the research question, formulated the hypotheses, wrote the original manuscript, and substantially revised it in five subsequent rounds. He directed and revised the analytical strategy, independently analyzed the data, and identified and corrected substantive computational errors during review, including an incorrect enrichment denominator and sequence-identity values. These analyses were recomputed and corrected before submission. The author used Anthropic's Claude to develop and execute analysis scripts, query the public databases described in the Methods, assemble figures and tables, and check the manuscript for internal consistency. OpenAI's ChatGPT was later used to revise the final version of manuscript for American scientific English, including grammar, terminology, clarity, and scientific phrasing. The author reviewed, revised and verified the reported values against their sources. Analysis code accompanies the submission to support independent reproduction using public data. The author takes full responsibility for the content, interpretation, and conclusions.

References

Behrens A, Rodschinka G, Nedialkova DD (2021) High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Mol Cell 81:1802–1815. https://doi.org/10.1016/j.molcel.2021.01.028

Berg MD, Giguere DJ, Dron JS, Lant JT, Genereaux J, Liao C, Wang J, Robinson JF, Gloor GB, Hegele RA, O'Donoghue P, Brandl CJ (2019) Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. RNA Biol 16:1574–1585. https://doi.org/10.1080/15476286.2019.1646079

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res 49:9077–9096. https://doi.org/10.1093/nar/gkab688

Chan PP, Lowe TM GtRNAdb: Genomic tRNA Database. https://gtrnadb.ucsc.edu/. Accessed 31 Oct 2026

Chen M, Kuhle B, Diedrich J, Liu Z, Moresco JJ, Yates JR 3rd, Pan T, Yang XL (2020) Cross-editing by a tRNA synthetase allows vertebrates to abundantly express mischargeable tRNA without causing mistranslation. Nucleic Acids Res 48:6445–6457. https://doi.org/10.1093/nar/gkaa469

de Bakker PIW, McVean G, Sabeti PC et al (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 38:1166–1172. https://doi.org/10.1038/ng1885

Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352. https://doi.org/10.1016/j.cell.2008.05.042

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183. https://doi.org/10.1126/science.1553543

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

Fukai S, Nureki O, Sekine S, Shimada A, Vassylyev DG, Yokoyama S (2003) Mechanism of molecular interactions for tRNA-Val recognition by valyl-tRNA synthetase. RNA 9:100–111. https://doi.org/10.1261/rna.2760703

Hoffman KS, O'Donoghue P, Brandl CJ (2017) Mistranslation: from adaptations to applications. Biochim Biophys Acta Gen Subj 1861:3070–3080. https://doi.org/10.1016/j.bbagen.2017.01.031

Horowitz J, Chu WC, Derrick WB, Liu JC, Liu M, Yue D (1999) Synthetase recognition determinants of E. coli valine transfer RNA. Biochemistry 38:7737–7746. https://doi.org/10.1021/bi990490b

Hou YM, Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 333:140–145. https://doi.org/10.1038/333140a0

Hou YM, Schimmel P (1989) Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. Biochemistry 28:6800–6804. https://doi.org/10.1021/bi00443a003

Karczewski KJ, Francioli LC, Tiao G et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581:434–443. https://doi.org/10.1038/s41586-020-2308-7

Kramer EB, Farabaugh PJ (2007) The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. RNA 13:87–96. https://doi.org/10.1261/rna.294907

Kramer EB, Vallabhaneni H, Mayer LM, Farabaugh PJ (2010) A comprehensive analysis of translational missense errors in the yeast Saccharomyces cerevisiae. RNA 16:1797–1808. https://doi.org/10.1261/rna.2201210

Liu M, Chu WC, Liu JC, Horowitz J (1997) Role of acceptor stem conformation in tRNA-Val recognition by its cognate synthetase. Nucleic Acids Res 25:4883–4890. https://doi.org/10.1093/nar/25.24.4883

Miller WT, Hou YM, Schimmel P (1991) Mutant aminoacyl-tRNA synthetase that compensates for a mutation in the major identity determinant of its tRNA. Biochemistry 30:2635–2641. https://doi.org/10.1021/bi00224a011

Mordret E, Dahan O, Asraf O et al (2019) Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors. Mol Cell 75:427–441. https://doi.org/10.1016/j.molcel.2019.06.041

Naganuma M, Sekine S, Chong YE, Guo M, Yang XL, Gamper H, Hou YM, Schimmel P, Yokoyama S (2014) The selective tRNA aminoacylation mechanism based on a single G·U pair. Nature 510:507–511. https://doi.org/10.1038/nature13440

Pichot F, Marchand V, Helm M, Motorin Y (2021) Non-redundant tRNA reference sequences for deep sequencing analysis of tRNA abundance and epitranscriptomic RNA modifications. Genes 12:81. https://doi.org/10.3390/genes12010081

Rogers HH, Griffiths-Jones S (2014) tRNA anticodon shifts in eukaryotic genomes. RNA 20:269–281. https://doi.org/10.1261/rna.041681.113

Rozik P, Moore H, Lant JT et al (2026) Mistranslation from an endogenous tRNA variant in human pan-genome cell lines. Nucleic Acids Res 54(5). https://doi.org/10.1093/nar/gkag224

Schaub M, Keller W (2002) RNA editing by adenosine deaminases generates RNA and protein diversity. Biochimie 84:791–803. https://doi.org/10.1016/s0300-9084(02)01446-3

Schrader JM, Chapman SJ, Uhlenbeck OC (2011) Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. Proc Natl Acad Sci U S A 108:5215–5220. https://doi.org/10.1073/pnas.1102128108

Shi JP, Schimmel P (1991) Aminoacylation of alanine minihelices. Discrimination of a single atom. J Biol Chem 266:2705–2708. https://doi.org/10.1016/S0021-9258(18)49902-2

Tamura K, Himeno H, Asahara H, Hasegawa T, Shimizu M (1991) Identity determinants of E. coli tRNA-Val. Biochem Biophys Res Commun 177:619–623. https://doi.org/10.1016/0006-291x(91)91833-x

Tennakoon R, Bily TMI, Hasan F, Hoffman KS, O'Donoghue P (2025) Natural human tRNA-Ala anticodon variants mistranslate the genetic code. RNA 31:791–806. https://doi.org/10.1261/rna.080450.125

Thornlow BP, Hough J, Roger JM et al (2018) Transfer RNA genes experience exceptionally elevated mutation rates. Proc Natl Acad Sci U S A 115:8996–9001. https://doi.org/10.1073/pnas.1801240115

Yona AH, Bloom-Ackermann Z, Frumkin I et al (2013) tRNA genes rapidly change in evolution to meet novel translational demands. eLife 2:e01339. https://doi.org/10.7554/eLife.01339

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Cheng2026SupplementaryFiles.zip