tRNA gene copy number variation in humans

James R. Iben and **Richard J. Maraia**[1],[*]
Intramural Research Program on Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

The human tRNAome consists of more than 500 interspersed tRNA genes comprising 51 anticodon families of largely unequal copy number. We examined tRNA gene copy number variation (tgCNV) in six individuals; two kindreds of two parents and a child, using high coverage whole genome sequence data. Such differences may be important because translation of some mRNAs is sensitive to the relative amounts of tRNAs and because tRNA competition determines translational efficiency vs. fidelity and production of native vs. misfolded proteins. We identified several tRNA gene clusters with CNV, which in some cases were part of larger iterations. In addition there was an isolated tRNALysCUU gene that was absent as a homozygous deletion in one of the parents. When assessed by semiquantitative PCR in 98 DNA samples representing a wide variety of ethnicities, this allele was found deleted in hetero- or homozygosity in all groups at ~50% frequency. This is the first report of copy number variation of human tRNA genes. We conclude that tgCNV exists at significant levels among individual humans and discuss the results in terms of genetic diversity and prior genome wide association studies (GWAS) that suggest the importance of the ratio of tRNALys isoacceptors in Type-2 diabetes.

Keywords

1. Introduction

Copy number variation (CNV) of genomic segments is a source of genetic differences among individuals, in many cases associated with complex disease phenotypes (Fanciulli et al., 2010; Zhang et al., 2009). More than 500 tRNA genes are interspersed on all of the human chromosomes (except Y chromosome), similar to short Alu retroposon elements but on a smaller scale. While classical CNV usually refers to reiterations of otherwise single copy loci, as described here the meaning for tRNA genes is somewhat different because they are multi-gene families of near-identical copies dispersed to multiple loci on multiple chromosomes.

A wide diversity exists in the tRNA gene contents of species (Chan and Lowe, 2009). Even closely related species show significant differences in tRNA gene copy numbers that are discordant with relatively minimal differences in other genomic features such as size, gene number, gene size and intron positions (Iben and Maraia, 2012). While tRNA gene content

---

[*]Corresponding author at: 31 Center Drive, Bld. 31, Rm 2A25, Bethesda, MD 20892-2426. Tel.: +1 301 402 3567. maraiar@mail.nih.gov (R.J. Maraia).
[1]Commissioned Corps, U.S. Public Health Service.

Conflict of interest
The authors note no conflict of interest.

is generally matched to codon use, even in higher eukaryotes (Duret, 2000; Gouy and Gautier, 1982; Grosjean and Fiers, 1982; Ikemura, 1981; Novoa et al., 2012; Pechmann and Frydman, 2013; Sharp et al., 1986), changes in the tRNAome that accompany speciation can appear to offset the match in some cases (Iben and Maraia, 2012) while in others they appear to restore balance (Higgs and Ran, 2008). In other cases variance in the fractional content of isoacceptors among highly related species was balanced by differential amplification of tRNA genes that can wobble to decode synonymous codons (Aldinger et al., 2012; Iben and Maraia, 2012). These observations support the idea that match of tRNAome composition and codon use is generally conserved, presumably to maintain fidelity and/or efficiency of translation (Lamichhane et al., 2013).

Mismatches of codon use and tRNAs can affect translation efficiency, fidelity, and/or protein folding via multiple mechanisms, in some cases as part of stress response programs or cell cycle control (Bauer et al., 2012; Begley et al., 2007; Berg and Kurland, 1997; Esberg et al., 2006; Fedyunin et al., 2012; Grosjean and Fiers, 1982; Hilterbrand et al., 2012; Jansen et al., 2003; Kimchi-Sarfaty et al., 2007; Lamichhane et al., 2013; Letzring et al., 2010; Patil et al., 2012; Pechmann and Frydman, 2013; Plotkin and Kudla, 2011; Xu et al., 2013; Zhou et al., 2013). Alterations of tRNA anticodon modifications result in phenotypes attributable to distorted translation of mRNAs over-enriched in cognate codons, that can be reversed by providing extra copies of the tRNA genes (Bauer et al., 2012; Begley et al., 2007; Esberg et al., 2006; Lamichhane et al., 2013; Wei et al., 2011; Xu et al., 2013; Zhou et al., 2013). Anticodon modification enzymes can be activated by stress, affecting dynamic control of codon-specific translation (Chan et al., 2010; Paredes et al., 2012). Cancers have been linked to tRNA overexpression (Pavon-Eternod et al., 2009) and anticodon hypermodification (Kuchino and Borek, 1978; Kuchino et al., 1982). These observations support a model in which some key mRNAs are sensitive to alterations of specific tRNA abundance or anticodon modification, in a manner relevant to phenotype.

The cumulative observations lead to the possibility that the degree of match between tRNA gene copy number and cognate codon use frequencies of individual genes may be a determinant of gene expression patterns. Yet tRNA gene CNV in individual yeast strains of the same species propagated in different laboratories does occur, in one case upon selective pressure and in other cases with no apparent selection, detectable by analysis of Illumina/Solexa sequence read coverage depth (Iben and Maraia, 2012; Iben et al., 2011).

We previously observed spontaneous amplification of tRNA genes in the fission yeast, *Schizosaccharomyces pombe* from whole genome Illumina/Solexa sequence datasets of sufficient read coverage (Iben and Maraia, 2012; Iben et al., 2011). The data suggest that tRNA gene copy number variation (tgCNV) may be widespread although it has not been examined much and not at all in humans. As alluded to above, tgCNV may affect translation of specific mRNAs (Brackley et al., 2011). Depending on the identity of the affected tRNA(s) and functions of the affected mRNAs, phenotypic output might be significant in some cases. Accordingly, tRNAome variation represents an unexamined layer of genetic diversity among individuals. This led us to question if tgCNV could be observed in humans. Here we evaluated available human genome-wide high coverage sequence datasets at all predicted tRNA gene loci and indeed found multiple instances of tgCNV among six individuals.

2. Materials and methods

All whole genome sequence data for the six individuals and their DNA (used here for PCR verification) are freely available through the 1000 Genomes Project. The genomic DNA for these samples are available through the Coriell Institute (12878, 12892, 119238, 19239,

19240); in one case the DNA was not available but the fibroblasts were so we prepared the DNA for PCR analysis (12891) (also obtained from Coriell Institute).

2.1. Sequence read depth at tRNA gene loci

Whole genome high coverage sequence datasets were identified and obtained from the 1000 Genomes Project database (Genomes Project et al., 2010). At the time of this study only 6 deposited genomes were identified as high coverage and genome-wide, comprised of two trio families: one of European-American ethnicity and one of Nigerian origin. The locations of all the tRNA genes were extracted for these genomes using the previously compiled data of the Genomic tRNA Database (Chan and Lowe, 2009). Aligned sequence data in indexed BAM format and the mapped read depth at the tRNA loci were extracted along with an additional 1 kb to either side of each tRNA gene. Read depth for the tRNA sequences was obtained by averaging across the tRNA region (Supplemental Table 1, columns G–L).

2.2. Overall total genome-wide read coverage

Using Samtools (Li et al., 2009), a pileup of per base coverage levels was derived for each dataset either over all bases of the common chromosomes 1–22, or of Chr-X alone. For each sample, a number of bases exhibiting each coverage value were summed and the resulting profile was plotted. Average coverage is calculated as the average base coverage for the main portion of the Gaussian curve (0–100 bp); the relative coverages among the six individuals determined are consistent with the relative coverages reported by the 1000 Genomes Project (not shown).

2.3. Robust estimation of tRNA copy number and variation

For each dataset, the averaged depth at any given tRNA locus was normalized against the overall whole genome average depth, tabulated for all genes (Supplemental Table 1, columns M–R). Average normalized genomic depth along with standard deviation was tabulated for each gene locus (Supplemental Table 1, columns S, T). Individual gene copy number variation was predicted at loci where normalized depth across samples exceeded a $1.0\times$ threshold. A total of 26 individual genes were seen to exceed this threshold (Table 2). The same was done for loci that exceeded a $0.5\times$ threshold, yielding 92 tRNA gene loci (Supplemental Table 2).

2.4. PCR validation of tgCNV at the Chr-11 tRNALysCUU

PCR primers flanking the Chr-11 tRNALysCUU gene: "437_Lys_CTTF" 5′-GGCCACAGGAGCTTCAAGTA-3′ and "437_Lys_CTTR" 5′-TGTGACTCAGGGGGCATAAT-3′ were used to amplify a ~400 bp product. As a control the Chr-16 tRNALysUUU gene, which exhibited the least variable at nearly uniform $1\times$ genomic average coverage across the 6 datasets (Supplemental Table 1), was amplified in the same duplex PCR reaction with primers: "C_193_Lys_CTTF" 5′-CGCAGGCGCTTCTTAGTATT-3′ and "C_193_Lys_CTTR" 5′-ACACACGGATCGGAGAACAC-3′ to produce a ~300 bp product. Duplex PCR was performed using Platinum Supermix (Invitrogen) in 50 µL, with template from either purchased genomic DNA where available (12878, 12892, 119238, 19239, 19240) or genomic DNA prepared from fibroblasts (12891) (DNA and cells obtained from Coriell Institute). Data not shown demonstrated that under the conditions used, 27 cycles yielded product quantities in the linear range. Thus, after 27 cycles, 5 µL aliquots of the PCRs were run on 1% agarose gel and visualized with ethidium bromide. The relative amounts of the ~300 and ~400 bp products shown in Fig. 3 were confirmed in triplicate experiments (not shown).

2.5. Assessment of Chr-11 tRNALysCUU gene deletion in diverse samples

A panel of 92 ethnically diverse genomic DNA samples was obtained from HPA (Health Protection Agency, Public Health England, product EDP-1, cat. # 07020701p). The PCR assay described above was scaled to 20 μL and run in triplicate for the 92 samples plus 3 original samples (12878, 12891, 12892) to serve as controls for genotypes +/−, −/−, and +/+, respectively. Analysis and quantitation of the relative amounts of ~300 and ~400 bp products were done on a WAVE column system (Transgenomic) and summarized in Table 3 which also indicates that distribution of the three genotypes among males and females according to the vendor provided ethnicity and gender information.

3. Results

Analysis of whole genome sequence read depth is useful for determining tgCNV in yeast (Iben and Maraia, 2012; Iben et al., 2011). Basically, differences in the number of Illumina/ Solexa sequence reads that map to different sequence tracts can reflect copy number differences if the overall genome read depth is uniform and of sufficiently high coverage. Previously mapped human sequence datasets were obtained for each of the six individuals for who existed moderately high coverage (16-fold) in the 1000 Genomes Project database (NA12878, NA12891, NA12892, NA19238, NA19239, NA19240) (Genomes Project et al., 2010). These had been annotated as two kindred trios each consisting of two parents and a daughter. One of the trio kindreds was from Utah with European ancestry (NA12891, NA12892, NA19239) and the other was from Nigeria (NA12878, NA19238, NA19240). These mapped sequence datasets were evaluated for read depth at the 506 human tRNA gene loci (as well as 3 each of selenocysteine and suppressor tRNAs) previously predicted for Build 37.1 of the human genome hg19 assembly. For this purpose we included for each locus the tRNA sequence itself plus 1 kb of DNA on each side. The 506 tRNA genes including five assigned to chromosome-X (Chr-X), collectively produce 51 tRNA anticodon families for the 20 standard amino acids that decode all of the 61 amino acid codons (Goodenbour and Pan, 2006). In the human genome, ten codons have no tRNA that match them directly by Watson:Crick base pairing and must rely on wobble decoding. Excluding these, the human tRNAome contains large inequalities in the gene copy numbers for its 51 tRNA anticodon families, ranging from 1 to 32, however this is not too dissimilar from some other mammals (Chan and Lowe, 2009).

3.1. Estimation of genomic copy number of tRNA genes

We first determined the average total read depth across the entire genome for each of the six datasets to serve as normalization for comparison and calibration so that we could derive tRNA gene counts per genome. The average total genome read depth for Chr-1 to Chr-22 was found to range from 17 to 34 fold coverage for the six samples, each with comparable uniform distribution around the median (Fig. 1A). We also determined the overall read depth coverage for Chr-X which reassuringly revealed that the two males in our sample set showed $0.5\times$ coverage relative to the four females which centered around $1\times$ (Fig. 1B).

Absolute copy number differences at the tRNA loci were tabulated for the six datasets (Supplemental Table 1, columns G–L). The tRNA gene loci copy numbers were also expressed as ratios relative to global average read depth (Supplemental Table 1, columns M–R). The copy number averages and standard deviations were determined for each tRNA gene locus (Supplemental Table 1, cols. S, T) which indicated that read depth across most tRNA gene loci in each of the six genomes was largely similar, at $\sim1\times$ relative to average genomic coverage. Specifically, 87% of all the tRNA gene loci in the 6 genomes were covered at read depths within $0.4\times$ of the genomic average depth, suggesting a high degree

of cross-individual reproducibility and overall constancy of their tRNAomes (Supplemental Fig. S2).

Human tRNA genes of the same anticodon family share significant identity throughout the tRNA sequence itself (Goodenbour and Pan, 2006). Accordingly, some tRNA sequence reads may be distributively "mapped" to other members of the same tRNA anticodon family in the genome assembly, which could therefore be a source of discordance with estimates based on flanking regions in some cases. Thus, we grouped all reads within the tRNA sequences into the 51 anticodon families (Table 1). We then plotted the read depths per genome for the six datasets (Fig. 2A). Cursory examination of Table 1 revealed good agreement both among the six samples and with copy numbers predicted from hg19 by tRNA-ScanSE; for example 1–2 copies for the tRNAs for Asn AUU and Tyr AUA anticodons, and ~30 copies for the Asn GUU and Ala AGC anticodons (Chan and Lowe, 2009). However, there was more variability among the six individuals for some tRNAs, e.g., Asn GUU, Cys GCA and Glu CUC anticodons, than for others. It is noteworthy that the direction of change, i.e., increase or decrease, differed among individuals for different tRNAs (Fig. 2A).

3.2. Copy number variation differs significantly among tRNA gene families

We more directly compared tRNA gene copy numbers derived from read depth from the six individuals with those predicted for the hg19 assembly (Fig. 2B). This confirmed good general agreement for most of the tRNA gene families with an overall correlation coefficient of R = 0.956, with few significant outliers (Fig. 2B). This analysis revealed additional information. There was more tgCNV for some tRNA families than for other tRNA families among the six individuals. However, this tgCNV was not uniform nor strictly correlated with copy number since the tRNA families Glu-CUC and Gly-GCC which were each predicted at   15 copies for hg19, showed a high degree of CNV whereas tRNA Ala-AGC and Cys GCA whose copy estimate of 29 and 30 compared well with the copy numbers in hg19, showed relatively less CNV (Table 1 and Fig. 2B). We noted that for the 18 tRNA gene families with copy numbers of   6, there was little apparent variation. For example, all five Arg tRNA families which comprise 28 genes matched hg19 well (Table 1). The analyses suggested that some tRNA genes were subjected to more CNV than others. The analysis indicated that while hg19 served as a very good indicator of tRNA gene number, there is variability for some tRNA gene families in the six individuals.

3.3. A tRNA gene cluster on Chr-1 is a source of tgCNV

Of the 506 tRNA gene loci examined here, only 26 exceeded a 1×-fold difference in read depth between any two of the six individuals (Table 2). Upon further examination, 17 of these 26 loci mapped to a ~28 kb stretch on Chr-1 (161413094–161440995) of hg19 (Table 2, lines 10–26). These 17 tRNA gene loci are represented as four imperfect repeats of a cluster of five tRNA genes; Glu-CTC, Gly-TCC, Asp-GTC, Leu-CAG and Gly-GCC, found in ~7 kb reiterations in the hg19 assembly (Supplemental Fig. S1). This suggests that these ~7 kb repeats or parts thereof were subjected to different degrees of iteration in the six individuals. These can be described in more detail and in genome-wide context by focusing on tRNA Gly-GCC. While the total number of tRNA Gly-GCC genes in hg19 is 15, only 6 of these map to Chr-1 (Chan and Lowe, 2009) (Table 1). The six genomes harbor 14–29 tRNA Gly-GCC genes and their increases relative to hg19 appear to be limited to imperfect duplication of one or more of the ~7 kb repeats on Chr-1 (Table 2). The same is true for Gly-UCC, Glu-CUC and the other members of the Chr-1 cluster but to lesser extents (bold font in Fig. 2B).

An additional 6 of the 26 loci in Table 1 were on Chr-1, although these were located 12 Mb away from the nearest ~7 kb cluster. These include a pair each of Val-CAC and Asn-GTT tRNA genes. Thus, accounting for the repeats of the tRNA genes on the ~7 kb regions and other close proximity isodecoder copies on Chr-1, 11 of the 506 hg19 tRNA gene loci were detected as exhibiting significant CNV in the six individuals (Table 1).

We also found 92 additional loci with significant differences of 0.5×-fold among some of the six individual datasets, potentially reflecting a source of tgCNV attributable to haploid states in the diploid DNA-derived sequence data samples (Supplemental Table 2).

3.4. A deletion allele of a tRNALysCUU gene-containing locus on Chr-11

Of specific note was a single locus encoding tRNALysCUU (Chr-11:51359900–51359972), which fully lacked coverage in one of the six samples (Table 2, line 29, NA12891). Two other samples showed 1× coverage at this tRNA gene (NA12892, NA19239) and the three other samples showed 0.5× coverage (NA12878, NA19238, NA19240) relative to their genomic averages (Table 2). Of all the genes examined, only this one lacked coverage entirely in an individual, suggesting deletion in NA12891 of both chromosomal copies.

The read depth coverage analysis of this Chr-11 tRNALysCUU gene locus fit with a Mendelian pattern of inheritance in the kindred with the predicted homozygous deletion, because the mother had twice as much read depth coverage of this tRNALysCUU locus as the daughter while it was entirely absent for the father (Table 2, last line). We note that this is the only tRNALysCUU gene on Chr-11; the other tRNALysCUU genes are distributed on eight other chromosomes (Chan and Lowe, 2009). Further analysis of the sequence data for this sample revealed the absence of flanking DNA coverage on both sides of this tRNALysCUU totaling ~21 kb, spanning Chr-11:51341115–51363134 (not shown). Upon literature search we found that these boundaries roughly coincide with deletions described in earlier global CNV studies, although homozygous deletion was not noted (Ju et al., 2010; Matsuzaki et al., 2009; Shaikh et al., 2009). No features other than this single tRNALysCUU gene are annotated within this ~21 kb region in RefSeq or Ensembl. The next closest annotated feature is a gene for olfactory receptor 4A5 (OR4A5), located ~50 kb downstream of the deletion.

Interestingly, the tRNALysCUU from this locus is a perfect Watson: Crick decoder of the lysine AAG codon whose altered translation in pre-proinsulin is implicated in development of Type-2 diabetes (T2D) in mice that lack the CDKAL1 gene previously implicated as a high quality risk indicator for T2D by genome wide association studies (GWAS) (Wei et al., 2011). Given the potential implications of this ifnding and the possibility that it may be useful as an independent factor, we wanted to establish a genomic PCR assay that could be used to validate the quantitative variation of this locus in the six individuals.

3.5. Validation of copy number variation of the Chr-11 tRNALysCUU

We obtained DNA samples available through the 1000 Genomes Project for the six individuals, and established a semiquantitative duplex PCR assay. We amplified the Chr-11 tRNALysCUU gene and an internal control tRNA gene determined to be invariant among the six individuals, in the same PCR reactions. The PCR products were ~300 bp for the invariant gene and ~400 bp for the tRNALysCUU gene, visualized after electrophoresis on 1% agarose gel (Fig. 3). The relative amounts of the tRNALysCUU gene PCR products fit well with the read depth data for the six samples (Fig. 3, see under the lanes) and were reproducible upon triplicate analyses (not shown). After quantitation it seemed clear from lanes 1–3 of Fig. 3 that the Utah kindred daughter was heterozygous (lane 1), the father had a homozygous deletion (lane 2), and the mother was homozygous for the presence of the

tRNALysCUU gene (lane 3). The read depth and PCR data also fit for this tRNALysCUU gene in the Nigerian kindred represented in lanes 4–6.

3.6. The deletion allele of the Chr-11 tRNALysCUU is widespread

Heterozygosity for the tRNALysCUU deletion allele in the Utah residents and Nigerians suggested that it might be widespread in the human population. We investigated this using the duplex PCR assay in ethnically diverse samples. Ninety-two samples of genomic DNAs were obtained as an ethnically diverse panel. The PCR assay was run in triplicate for these samples along with 3 of the original six samples as controls for each genotype and the products were fractionated and quantitated using a WAVE column system. We found that the deletion allele of the Chr-11 tRNALysCUU gene occurs in members of each ethnic group on the panel, represented in both genders, with an overall frequency of ~50%, as summarized in Table 3.

4. Discussion

4.1. tRNA gene copy number variation is readily observable in humans

Eukaryotic tRNA genes are short, self-contained transcription units that are interspersed throughout the genome, and dispersed among the chromosomes. As noted in the Introduction, all of the tRNA genes evaluated here are members of one or another of the 51 multigene families.

Hg19 is a composite assembly of DNA obtained from ~13 donors (Editorial, 2010). The tRNAscan-SE algorithm finds tRNA genes in genomes with high accuracy (Lowe and Eddy, 1997). Application of tRNAscan-SE to genome assemblies has led to the appreciation that the tRNAome is a highly variable component of both closely related and distantly related genomes (Chan and Lowe, 2009).

For the present work, we sought to evaluate quantitative representation of the 506 tRNA genes in hg19 in each of six individuals. Briefly, we analyzed sequence read depth across all tRNA gene loci normalized against whole genome coverage depth for the six persons. The normalized tRNA gene read depth was taken to estimate copy number per genome, on the same scale as for hg19 (Chan and Lowe, 2009) (Supplemental Table 1). This revealed tgCNV among humans. By this approach, tgCNV reflects whether a member of any tRNA multigene family is present or absent in heterozygous or homozygous form, or as an extra copy(s).

While each tRNA gene locus may be considered individually, given that many copies of identical tRNA sequences exist in the genome it was informative to consider tRNA anticodon identities collectively. Accordingly there was good agreement with the tRNA gene numbers in hg19 but with a few exceptions (Fig. 2B). For a majority of these exceptions, hg19 revealed lower counts than our analysis especially for a cluster of tRNA genes on Chr-1 that reside within larger repeat units (Supplemental Fig. S1). Such discrepancy probably reflects the limitations of building an assembly in the vicinity of tandem repeats. For the tRNA genes clustered on Chr-1, each of the six individuals had higher copy numbers than represented by hg19 (e.g., Gln, Glu, Gln, Fig. 2B). These Chr-1 tRNA gene clusters appear to be subject to a significant amount of CNV among individuals.

However, we found that variation was not due only to tRNA gene clusters. An example is an isolated tRNALysCUU gene on Chr-11 that completely lacked read coverage in one individual in one of the kindreds. Our estimated copy numbers suggested a heterozygous deletion of this tRNA gene in other individuals consistent with Mendelian inheritance in both kindreds. Genomic PCR analysis confirmed the absence of this tRNA gene in the DNA

of the identified individual, and further indicated its heterozygous or homozygous presence in the others, in good agreement with read depth estimates. Using PCR we extended analysis of the Chr-11 tRNALysCUU gene to a larger ethnically diverse group of genomes. This revealed that a tRNALysCUU gene deletion allele is present in every ethnicity examined, at an overall frequency of about 50%.

4.2. tgCNV and implications for gene expression and genetic-based disease

Relative amounts of tRNAs are important because their competition for cognate and non-cognate codons determines translational efficiency vs. fidelity as well as propensity to produce native vs. misfolded proteins (Reynolds et al., 2010). As the molecular mechanisms of some diseases attributed to protein misfolding are poorly understood, tgCNV could potentially contribute. As noted, perturbation of decoding of a Lys AAG codon in pre-proinsulin was implicated in development of T2D in CDKAL1-deficient mice, experimentally validating GWAS results (Wei et al., 2011). CDKAL1 is a tRNA methylthiotransferase that modifies threonylcarbamoyl-adenosine (t6A) at position A37 in the anticodon loop of one tRNA species, tRNALysUUU, to ms2t6A37. There are two codons for lysine, AAG and AAA, and in humans there are near equal numbers of tRNALysCUU and tRNALysUUU genes to decode them (Table 1). Curiously, tRNALysUUU can decode both its cognate codon AAA and by wobble, the AAG codon, whereas tRNALysCUU can decode only its cognate codon, AAG. Pre-proinsulin mRNA has two lysine codons, both AAG, one in the B chain region and the other at the junction of the C peptide and A chain, immediately adjacent to codons whose mutation cause hyperproinsulinemia, a disease in which insulin is insufficiently processed prior to secretion (Dhanvantari et al., 2003; Stoy et al., 2010). According to a CDKAL1 model of T2D, hypomodification of tRNALysUUU alters its ability to compete with tRNALysCUU at its wobble codon AAG which results in misfolding, poor processing and impaired secretion of insulin (Wei et al., 2011). Such a model plausibly fits with the subtle genetic influences expected to contribute to complex multigenic phenotypes such as T2D.

The prevalence of the deletion allele of the Chr-11 tRNALysUUU gene uncovered here in combination with the importance of tRNA competition in protein folding homeostasis, and the CDKAL1 T2D model of impaired AAG decoding, promoted us to consider more generally variances of tRNALysCUU and tRNALysUUU gene numbers and other isoacceptor pairs among the six individuals examined. Each row of Table 4 lists the ratios of gene copy numbers of two tRNA isodecoders for 8 different amino acids. For each pair of these isodecoders one of the tRNAs can wobble to the other codon, similar to that for tRNALys**C**UU and tRNALys**U**UU. The last column of Table 4 compiles the range of variance of the ratios of the two isodecoders among the six individuals. The individuals with the least and greatest variance of tRNALys**C**UU and tRNALys**U**UU genes differ in their ratios by 23%. For the two Arg tRNAs the range is 40%, and for the Tyr pair it is 47%. We suggest the possibility that such differences may contribute to phenotypic differences among individuals, especially if combined with other genetic differences in protein coding genes, including synonymous substitutions.

It should be noted that after this study was completed, we learned that a region of mouse distal chromosome 1 (Chr 1) that corresponds to human Chr 1q21–q23 includes quantitative trait loci for a diverse set of traits (Mozhui et al., 2008). This region encodes a large number of genes including ~20 aminoacyl-tRNA synthetases (aaRSs) in addition to a tRNA gene cluster. In addition to their role in catalyzing the production of aminoacyl-tRNAs used for protein synthesis during mRNA translation, several diverse nontranslational functions of aaRSs have been discovered (reviewed in Guo and Schimmel, 2013). In addition, tRNAs can feedback regulate their aaRSs (Ryckelynck et al., 2005). Therefore, tgCNV may have

impact not only on mRNA translation directly but also indirectly, via effects on the aaRSs (Mozhui et al., 2008).

Another perspective of the potential effects of tgCNV is with regard to classical CNV of large genomic regions more generally (Fanciulli et al., 2010; Zhang et al., 2009). Since tRNA genes are distributed on all chromosomes the likelihood that a duplication or other amplification of any large genomic segment will contain a tRNA gene(s) increases with size of the affected region. Depending on the tRNA species and the copy number of its anticodon family, it is conceivable that it might contribute to an associated phenotype.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Abbreviations

tRNA	transfer RNA
CNV	copy number variation
tgCNV	tRNA gene copy number variation
PCR	polymerase chain reaction
T2D	type-2 diabetes
Chr	chromosome
bp	base pair(s)
GWAS	genome wide association studies

References

Aldinger CA, Leisinger AK, Gaston KW, Limbach PA, Igloi GL. The absence of A-to-I editing in the anticodon of plant cytoplasmic tRNA (Arg) ACG demands a relaxation of the wobble decoding rules. RNA Biol. 2012; 9:1239–1246. [PubMed: 22922796]

Bauer F, et al. Translational control of cell division by elongator. Cell Rep. 2012; 1:424–433. [PubMed: 22768388]

Begley U, et al. Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol Cell. 2007; 28:860–870. [PubMed: 18082610]

Berg OG, Kurland CG. Growth rate-optimised tRNA abundance and codon usage. J Mol Biol. 1997; 270:544–550. [PubMed: 9245585]

Brackley CA, Romano MC, Thiel M. The dynamics of supply and demand in mRNA translation. PLoS Comput Biol. 2011; 7:e1002203. [PubMed: 22022250]

Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 2009; 37:D93–D97. [PubMed: 18984615]

Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet. 2010; 6:e1001247. [PubMed: 21187895]

Dhanvantari S, et al. Disruption of a receptor-mediated mechanism for intracellular sorting of proinsulin in familial hyperproinsulinemia. Mol Endocrinol. 2003; 17:1856–1867. [PubMed: 12829804]

Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 2000; 16:287–289. [PubMed: 10858656]

Editorial. E pluribus unum. Nat Methods. 2010; 7:331. [PubMed: 20440876]

Esberg A, Huang B, Johansson MJ, Bystrom AS. Elevated levels of two tRNA species bypass the requirement for elongator complex in transcription and exocytosis. Mol Cell. 2006; 24:139–148. [PubMed: 17018299]

Fanciulli M, Petretto E, Aitman TJ. Gene copy number variation and common human disease. Clin Genet. 2010; 77:201–213. [PubMed: 20002459]

Fedyunin I, Lehnhardt L, Bohmer N, Kaufmann P, Zhang G, Ignatova Z. tRNA concentration fine tunes protein solubility. FEBS Lett. 2012; 586:3336–3340. [PubMed: 22819830]

Genomes Project C etal. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]

Goodenbour JM, Pan T. Diversity of tRNA genes in eukaryotes. Nucleic Acids Res. 2006; 34:6137–6146. [PubMed: 17088292]

Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 1982; 10:7055–7074. [PubMed: 6760125]

Grosjean H, Fiers W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene. 1982; 18:199–209. [PubMed: 6751939]

Guo M, Schimmel P. Essential nontranslational functions of tRNA synthetases. Nat Chem Biol. 2013; 9:145–153. [PubMed: 23416400]

Higgs PG, Ran W. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. Mol Biol Evol. 2008; 25:2279–2291. [PubMed: 18687657]

Hilterbrand A, Saelens J, Putonti C. CBDB: the codon bias database. BMC Bioinforma. 2012; 13:62.

Iben JR, Maraia RJ. Yeast tRNAomics: tRNA gene copy number variation and codon use provide bioinformatics evidence of a new wobble pair in a eukaryote. RNA. 2012; 18:1358–1372. [PubMed: 22586155]

Iben JR, et al. Comparative whole genome sequencing reveals phenotypic tRNA gene duplication in spontaneous *Schizosaccharomyces pombe* La mutants. Nucleic Acids Res. 2011; 39:4728–4742. [PubMed: 21317186]

Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol. 1981; 151:389–409. [PubMed: 6175758]

Jansen R, Bussemaker HJ, Gerstein M. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. Nucleic Acids Res. 2003; 31:2242–2251. [PubMed: 12682375]

Ju YS, et al. Reference-unbiased copy number variant analysis using CGH micro-arrays. Nucleic Acids Res. 2010; 38:e190. [PubMed: 20802225]

Kimchi-Sarfaty C, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science. 2007; 315:525–528. [PubMed: 17185560]

Kuchino Y, Borek E. Tumour-specific phenylalanine tRNA contains two supernumerary methylated bases. Nature. 1978; 271:126–129. [PubMed: 202873]

Kuchino Y, Borek E, Grunberger D, Mushinski JF, Nishimura S. Changes of post-transcriptional modification of wye base in tumor-specific tRNAPhe. Nucleic Acids Res. 1982; 10:6421–6432. [PubMed: 6924749]

Lamichhane TN, et al. Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast. Mol Cell Biol. 2013; 33:2918–2929. [PubMed: 23716598]

Letzring DP, Dean KM, Grayhack EJ. Control of translation efficiency in yeast by codon–anticodon interactions. RNA. 2010; 16:2516–2528. [PubMed: 20971810]

Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25:2078–2079.

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25:955–964. [PubMed: 9023104]

Matsuzaki H, Wang PH, Hu J, Rava R, Fu GK. High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians. Genome Biol. 2009; 10:R125. [PubMed: 19900272]

Mozhui K, Ciobanu DC, Schikorski T, Wang X, Lu L, Williams RW. Dissection of a QTL hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. PLoS Genet. 2008; 4:e1000260. [PubMed: 19008955]

Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. A role for tRNA modifications in genome structure and codon usage. Cell. 2012; 149:202–213. [PubMed: 22464330]

Paredes JA, et al. Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC Biol. 2012; 10:55. [PubMed: 22715922]

Patil A, et al. Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. Cell Cycle. 2012; 11:3656–3665. [PubMed: 22935709]

Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res. 2009; 37:7268–7280. [PubMed: 19783824]

Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol. 2013; 20:237–243. [PubMed: 23262490]

Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011; 12:32–42. [PubMed: 21102527]

Reynolds NM, Lazazzera BA, Ibba M. Cellular mechanisms that control mistranslation. Nat Rev Microbiol. 2010; 8:849–856. [PubMed: 21079633]

Ryckelynck M, Giege R, Frugier M. tRNAs and tRNA mimics as cornerstones of aminoacyl-tRNA synthetase regulations. Biochimie. 2005; 87:835–845. [PubMed: 15925436]

Shaikh TH, et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. 2009; 19:1682–1690. [PubMed: 19592680]

Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986; 14:5125–5143. [PubMed: 3526280]

Stoy J, Steiner DF, Park SY, Ye H, Philipson LH, Bell GI. Clinical and molecular genetics of neonatal diabetes due to mutations in the insulin gene. Rev Endocr Metab Disord. 2010; 11:205–215. [PubMed: 20938745]

Wei F, et al. Deficit of tRNALys modification by Cdkal1 causes the development of type 2 diabetes in mice. J Clin Invest. 2011; 121:3598–3608. [PubMed: 21841312]

Xu Y, Ma P, Shah P, Rokas A, Liu Y, Johnson CH. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature. 2013; 495:116–120. [PubMed: 23417065]

Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 2009; 10:451–481. [PubMed: 19715442]

Zhou M, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. 2013; 495:111–115. [PubMed: 23417067]

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.11.049.
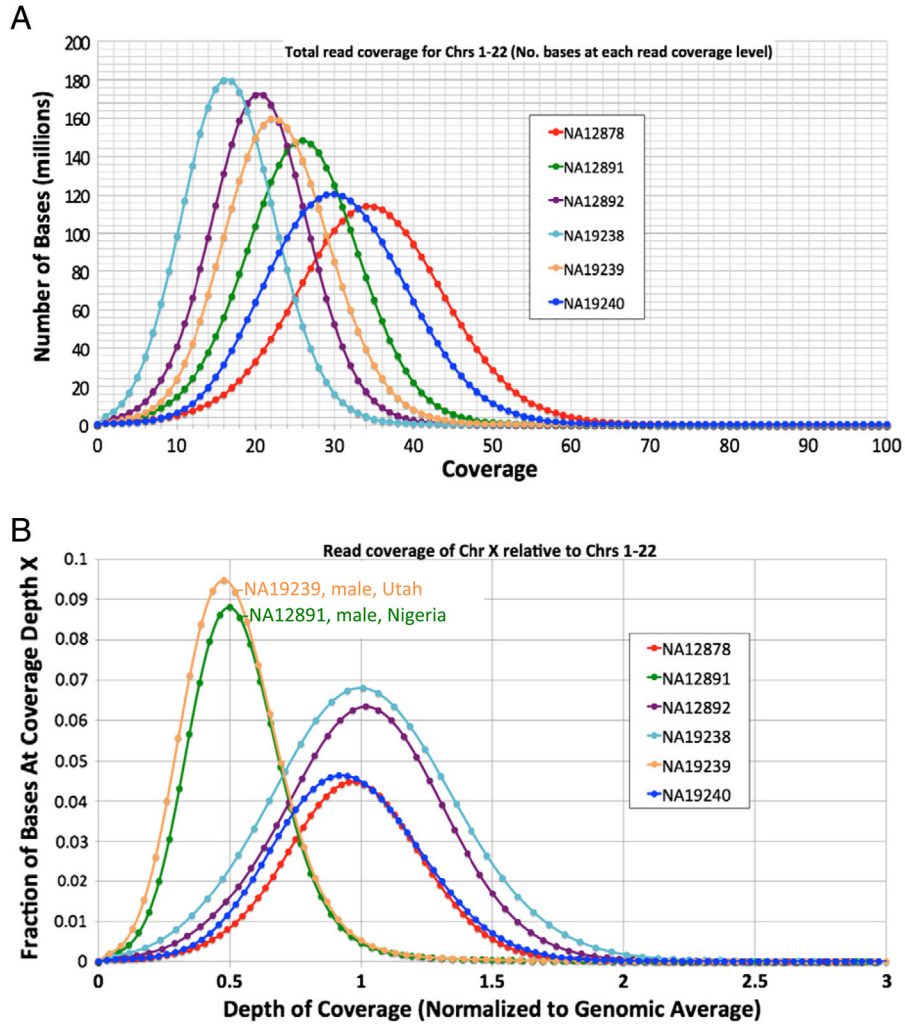
Fig. 1.
Read coverage distributions among the six individuals examined here. A) Totals for chromosomes 1–22. Relatively high coverage, e.g., greater than 40-fold for NA19238 and 60-fold for NA19240, likely reflect repeat sequences. B) Read coverage on chromosome-X relative to total genomic (Chr 1–22) average read depth; coverage centers at 0.5-fold for the two males as annotated, but centers around 1-fold for the four other samples, females.
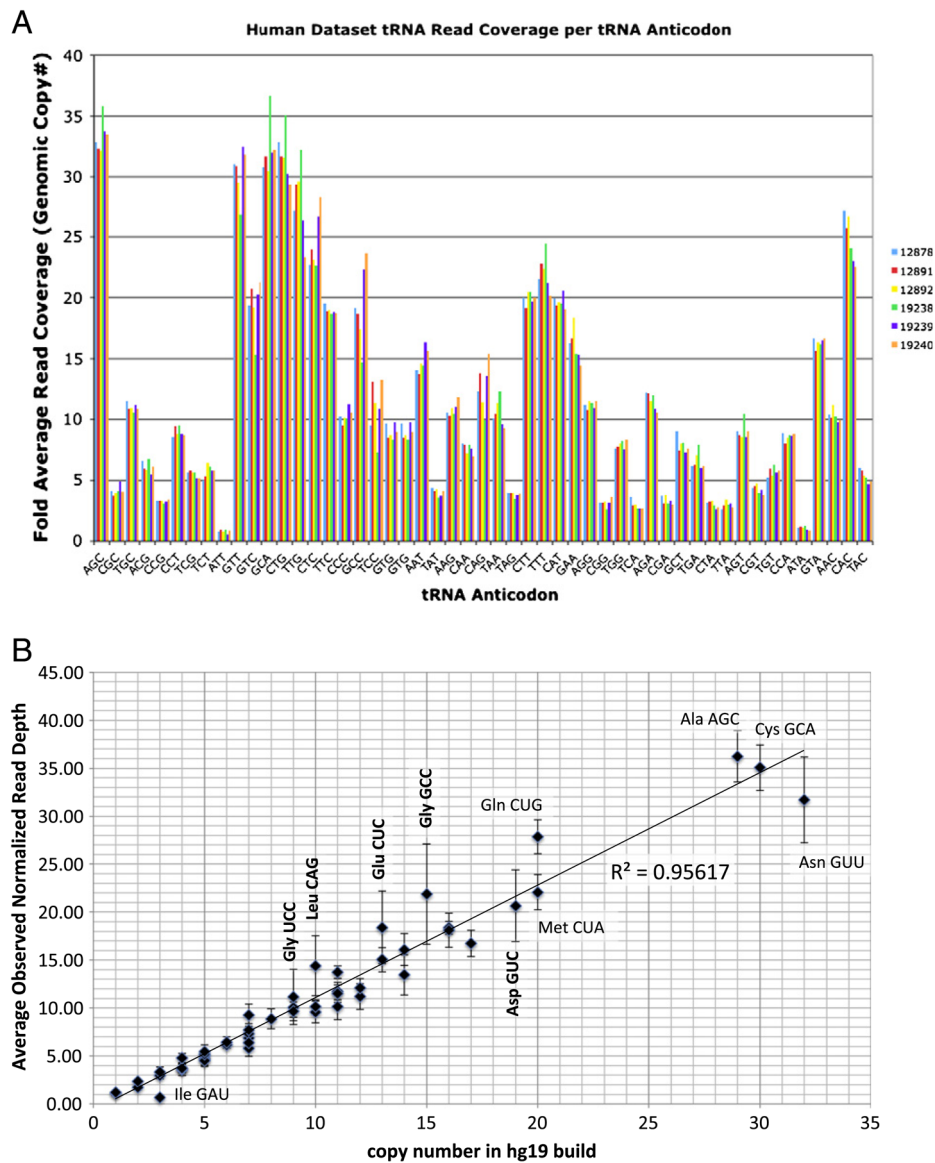
Fig. 2.
tRNA gene copy number estimates. A)Normalized tRNA gene read depth was summed for each of the 51 tRNA anticodon species (Table 1) and plotted for each individual according to the color code to the right. This reflects largely reproducible gene copy numbers per anticodon with few notable exceptions (see text). B) Estimated gene counts from (A, y-axis), plotted against tRNA gene counts per anticodon as predicted by tRNA-ScanSE for the hg19 assembly (3). Some of the tRNA genes specifically referred to in the text are annotated. Error bars represent standard deviation of estimated gene counts across the six individuals of the study.
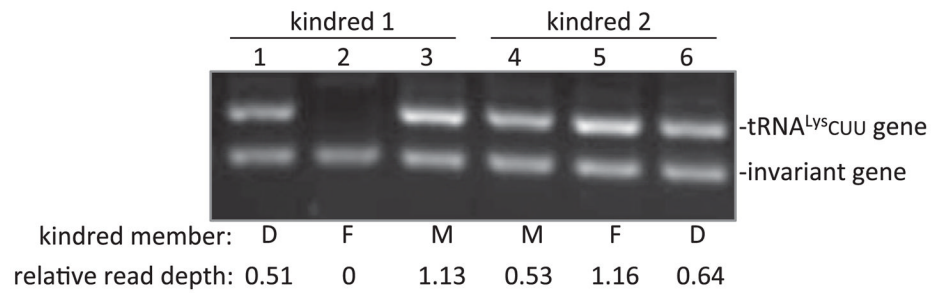
Fig. 3.
Validation of copy number variation of the Chr-11 tRNA^{Lys}CUU gene. Duplex PCR of the Chr-11 tRNALysCUU gene locus and an invariant control tRNA gene locus as indicated to the right of the bands for each of the 6 individuals in the two kindreds. An ethidium bromide stained agarose gel is shown. The kindred members are indicated below the lanes: D = daughter, F = father, M = mother. The relative read depths at the Chr-11 tRNALysCUU gene locus derived from Table 1 are also shown. For this locus, values above 1.0 reflect the diploid state of the Chr-11 tRNALysCUU gene, the values 0.43–0.54 reflect the haploid state and the value of 0 reflects a homozygous deletion.

Table 1

Estimated tRNA gene count per anticodon family. For each of the 6 individuals, normalized average read depth was summed for each anticodon to serve as an estimate of genomic tRNA gene content. Gene count values over the six individuals are averaged and compared to the counts predicted by tRNA-ScanSE for the hg19 assembly.

| | | Daughter | Father | Mother | Mother | Father | Daughter | | | | |
| | | Estimated gene count | | | | | | | | | |
AA	Anticodon	12878	12891	12892	19238	19239	19240	AVERAGE	STDEV	tRNA-ScanSE Count	Difference
Ala	AGC	37.56	33.54	34.25	34.04	37.78	40.19	36.23	2.68	29	7.23
Ala	CGC	4.91	4.02	4.35	4.08	5.69	5.03	4.68	0.65	5	−0.32
Ala	UGC	11.11	9.44	9.47	8.63	10.75	11.00	10.07	1.02	9	1.07
Arg	ACG	7.83	6.43	6.38	6.64	6.34	7.58	6.87	0.66	7	−0.13
Arg	CCG	3.95	3.59	3.68	3.06	3.72	4.26	3.71	0.40	4	−0.29
Arg	CCU	5.55	5.21	4.52	4.05	5.22	5.48	5.01	0.59	5	0.01
Arg	UCG	6.69	6.32	6.15	5.57	5.93	6.40	6.18	0.39	6	0.18
Arg	UCU	5.93	5.77	7.00	5.98	6.66	7.16	6.42	0.60	6	0.42
Asn	AUU	1.59	1.69	1.65	1.65	1.49	2.19	1.71	0.25	2	−0.29
Asn	GUU	33.85	30.70	29.91	24.19	34.92	36.62	31.70	4.47	32	−0.30
Asp	GUC	21.69	21.50	19.86	13.84	22.10	24.94	20.66	3.72	19	1.66
Cys	GCA	35.28	33.02	32.10	35.14	35.93	38.86	35.06	2.37	30	5.06
Gln	CUG	31.21	27.41	27.28	26.17	26.80	28.26	27.85	1.79	20	7.85
Gln	UUG	14.35	13.04	13.27	14.67	13.69	13.36	13.73	0.64	11	2.73
Glu	CUC	17.43	17.33	16.20	13.77	21.27	24.37	18.39	3.80	13	5.39
Glu	UUC	16.53	14.81	15.34	12.82	14.82	15.92	15.04	1.27	13	2.04
Gly	CCC	9.64	8.48	8.78	7.70	10.61	10.43	9.27	1.15	7	2.27
Gly	GCC	22.68	20.20	19.01	14.39	25.68	29.28	21.87	5.23	15	6.87
Gly	UCC	10.02	13.08	11.34	6.48	11.06	14.93	11.15	2.87	9	2.15
His	GUG	11.48	9.21	9.54	8.20	11.26	11.08	10.13	1.34	11	−0.87
Ile	AAU	13.21	11.84	12.53	11.12	16.04	15.93	13.44	2.09	14	−0.56
Ile	GAU	0.87	0.99	0.74	0.40	0.38	0.61	0.66	0.25	3	−2.34
Ile	UAU	5.13	4.47	4.65	3.47	4.31	5.13	4.52	0.62	5	−0.48
Leu	AAG	11.33	10.27	10.71	9.61	11.84	13.45	11.20	1.35	12	−0.80

		Daughter	Father	Mother	Mother	Father	Daughter				
		Estimated gene count									
AA	Anticodon	12878	12891	12892	19238	19239	19240	AVERAGE	STDEV	tRNA-ScanSE Count	Difference
Leu	CAA	8.03	7.36	6.74	6.63	7.58	7.53	7.31	0.54	7	0.31
Leu	CAG	14.54	14.99	12.52	9.74	15.60	19.04	14.41	3.12	10	4.41
Leu	UAA	8.02	6.98	8.50	8.05	6.88	7.86	7.71	0.64	7	0.71
Leu	UAG	3.24	2.96	2.93	2.32	2.96	3.37	2.96	0.36	3	-0.04
Lys	CUU	17.89	15.27	16.88	14.94	17.12	18.38	16.75	1.38	17	-0.25
Lys	UUU	19.32	18.71	17.96	17.34	18.31	18.49	18.36	0.67	16	2.36
Met	CAU	23.56	21.00	21.47	19.19	23.67	23.54	22.07	1.83	20	2.07
Phe	GAA	12.91	11.63	13.26	10.48	11.95	12.24	12.08	0.99	12	0.08
Pro	AGG	10.20	8.97	9.51	7.84	9.79	11.16	9.58	1.13	10	-0.42
Pro	CGG	3.79	3.48	3.52	2.56	3.64	4.54	3.59	0.64	4	-0.41
Pro	UGG	5.87	5.44	5.70	4.95	5.49	7.43	5.81	0.85	7	-1.19
SeC	UCA	4.32	3.20	3.33	2.64	3.14	3.31	3.32	0.55	3	0.32
Ser	AGA	13.45	11.80	11.28	10.89	11.23	11.88	11.76	0.91	11	0.76
Ser	CGA	4.38	3.34	4.15	3.02	3.80	3.71	3.73	0.50	4	-0.27
Ser	GCU	10.70	8.05	8.74	7.91	8.42	9.41	8.87	1.04	8	0.87
Ser	UGA	5.10	4.57	4.87	5.61	5.40	5.45	5.17	0.40	5	0.17
Sup	CUA	1.12	1.49	1.21	0.81	0.90	1.12	1.11	0.24	1	0.11
Sup	UUA	2.27	2.39	2.49	2.38	2.35	2.26	2.36	0.08	2	0.36
Thr	AGU	10.68	9.44	9.40	10.32	9.92	11.15	10.15	0.70	10	0.15
Thr	CGU	5.14	4.94	5.17	3.87	4.82	4.71	4.77	0.48	4	0.77
Thr	UGU	6.16	6.46	5.92	6.17	6.50	7.12	6.39	0.42	7	-0.61
Trp	CCA	10.55	8.66	9.24	8.58	9.98	10.94	9.66	0.99	9	0.66
Tyr	AUA	1.32	1.31	1.20	1.28	1.10	1.05	1.21	0.11	1	0.21
Tyr	GUA	17.48	14.75	15.71	13.84	16.46	18.32	16.09	1.68	14	2.09
Val	AAC	12.28	10.98	12.20	10.05	11.26	12.30	11.51	0.91	11	0.51
Val	CAC	19.68	17.45	17.90	15.11	18.50	20.07	18.12	1.78	16	2.12
Val	UAC	6.16	5.85	5.47	4.67	4.61	6.01	5.46	0.68	5	0.46

Table 2

Identification of tRNA genes with predicted CNV. For each of the 506 tRNA loci (Sup Table 1), only those demonstrating a normalized read depth difference of greater than 1× relative to the genomic average depth for the six individuals are listed. Highlighted in bold blue are 5 tRNAs within a 7 kb sequence present in multiple imperfect copies on the hg19 assembly; light blue font highlight other genes on Chr-1 (see text). Highlighted in red is a single tRNA gene lacking any mapped reads in one individual.

| Chrom# | tRNA# | Start | End | ID | AC | Read coverage | | | | | | Normalized to average coverage | | | | | |
						Daughter 12878	Father 12891	Mother 12892	Mother 19238	Father 19239	Daughter 19240	Daughter 12878	Father 12891	Mother 12892	Mother 19238	Father 19239	Daughter 19240
1	133	16872504	16872434	Gly	CCC	51.17	27.79	35.07	23.80	56.42	53.73	1.48	1.07	1.70	1.41	2.45	1.76
1	2	17053780	17053850	Gly	CCC	82.87	61.90	43.94	34.24	59.61	101.35	2.39	2.38	2.13	2.03	2.59	3.31
1	6	17201958	17202031	Asn	GTT	62.96	39.65	36.50	26.11	69.38	72.51	1.82	1.53	1.77	1.55	3.02	2.37
1	7	17216172	17216245	Asn	GTT	28.88	26.50	18.77	9.27	43.62	46.24	0.83	1.02	0.91	0.55	1.90	1.51
1	99	149294736	149294666	Val	CAC	72.77	36.99	35.62	11.32	36.70	32.20	2.10	1.42	1.73	0.67	1.60	1.05
1	98	149298627	149298555	Val	CAC	67.12	42.79	29.70	37.62	57.85	88.93	1.94	1.65	1.44	2.23	2.51	2.90
1	35	161413094	161413164	Gly	GCC	80.25	55.76	42.52	21.00	76.55	127.46	2.32	2.15	2.06	1.24	3.33	4.16
1	80	161417089	161417018	Glu	CTC	62.01	39.50	35.07	21.11	68.92	110.29	1.79	1.52	1.70	1.25	3.00	3.60
1	79	161417446	161417375	Gly	TCC	46.26	50.81	30.44	13.00	29.07	63.28	1.34	1.96	1.47	0.77	1.26	2.07
1	36	161418741	161418823	Leu	CAG	57.98	44.49	28.42	17.57	44.13	81.29	1.67	1.71	1.38	1.04	1.92	2.66
1	37	161420467	161420537	Gly	GCC	79.94	64.41	41.35	26.15	88.82	138.58	2.31	2.48	2.00	1.55	3.86	4.53
1	**77**	**161424469**	**161424398**	**Glu**	**CTC**	**70.81**	**62.40**	**34.89**	**24.33**	**64.63**	**108.25**	**2.05**	**2.40**	**1.69**	**1.44**	**2.81**	**3.54**
1	**76**	**161424827**	**161424756**	**Gly**	**TCC**	**38.89**	**52.82**	**25.13**	**13.06**	**46.99**	**81.03**	**1.12**	**2.03**	**1.22**	**0.77**	**2.04**	**2.65**
1	**75**	**161425485**	**161425414**	**Asp**	**GTC**	**50.71**	**30.13**	**27.50**	**12.64**	**43.88**	**69.49**	**1.46**	**1.16**	**1.33**	**0.75**	**1.91**	**2.27**
1	**38**	**161426122**	**161426204**	**Leu**	**CAG**	**48.67**	**48.93**	**30.19**	**16.65**	**43.35**	**75.72**	**1.41**	**1.88**	**1.46**	**0.99**	**1.88**	**2.47**
1	**39**	**161427898**	**161427968**	**Gly**	**GCC**	**99.93**	**60.79**	**39.24**	**27.35**	**72.75**	**125.73**	**2.89**	**2.34**	**1.90**	**1.62**	**3.16**	**4.11**
1	74	161431880	161431809	Glu	CTC	68.65	49.94	33.69	19.56	68.29	93.11	1.98	1.92	1.63	1.16	2.97	3.04
1	73	161432237	161432166	Gly	TCC	39.88	54.56	32.46	12.15	33.25	51.44	1.15	2.10	1.57	0.72	1.45	1.68
1	72	161432895	161432824	Asp	GTC	47.06	43.51	25.42	6.56	34.78	54.07	1.36	1.67	1.23	0.39	1.51	1.77
1	41	161435258	161435328	Gly	GCC	90.77	56.76	41.85	20.20	83.52	141.86	2.62	2.18	2.03	1.20	3.63	4.63
1	71	161439260	161439189	Glu	CTC	65.78	46.93	37.19	22.93	60.68	122.96	1.90	1.81	1.80	1.36	2.64	4.02
1	70	161439618	161439547	Gly	TCC	48.10	47.75	32.74	12.56	40.96	73.71	1.39	1.84	1.59	0.74	1.78	2.41

| | | | | | | Read coverage | | | | | | Normalized to average coverage | | | | | |
| | | | | | | Daughter | Father | Mother | Mother | Father | Daughter | Daughter | Father | Mother | Mother | Father | Daughter |
Chrom#	tRNA#	Start	End	ID	AC	12878	12891	12892	19238	19239	19240	12878	12891	12892	19238	19239	19240
1	42	161440913	161440995	Leu	CAG	61.17	53.93	32.53	14.88	48.07	83.36	1.77	2.08	1.58	0.88	2.09	2.72
6	163	26745328	26745255	Ile	AAT	19.36	13.72	13.36	14.70	61.05	58.49	0.56	0.53	0.65	0.87	2.65	1.91
6	25	26751918	26751990	Ala	AGC	24.38	17.36	19.18	25.07	53.41	69.23	0.70	0.67	0.93	1.48	2.32	2.26
11	2	51359900	51359972	Lys	CTT	17.58	0.00	23.30	8.95	26.77	19.56	0.51	0.00	1.13	0.53	1.16	0.64

Table 3

Assessment of the Chr-11 tRNALysCUU gene status (+/+, +/–, –/–) in ethnically diverse DNA samples as determined by a high throughput version of the PCR assay shown in Fig. 3.

Ethnicity	Males	Females	+/+ M	+/– M	–/– M	+/+ F	+/– F	–/– F	Notes
GI = German/Italian	1	0	0	0	1	0	0	0	
AA = Australian Aborigine	6	3	3	1	2	0	3	0	
BA = Black African	3	4	1	1	1	1	3	0	Includes 1000 Genomes Project stocks
AJ = Ashkenazi Jewish	5	0	1	1	3	0	0	0	
FC = French Caucasian	2	3	0	2	0	0	2	1	
FCn = French Canadian	0	0	0	0	0	0	0	0	
SAI = South American Indian	5	3	0	3	2	0	0	3	
T = Thai	6	4	2	3	1	0	2	2	
O = Oriental	10	1	3	2	5	1	0	0	1 unknown gender +/–
J = Japanese	11	4	7	2	2	0	3	1	
F = French	4	2	0	3	1	0	1	1	
C = Caucasian	3	7	1	1	1	2	4	1	Includes 1000 Genomes Project stocks
I = Italian	7	3	1	5	1	1	1	1	
– = blank well control	0	0	0	0	0	0	0	0	
Sums	63	34	19	24	20	5	19	10	
			+/+				+/–	–/–	
Totals			24				44	30	
Ratios			24.49%				44.90%	30.61%	

Table 4

Relative ratios of tRNA gene copy numbers for selected tRNA isoacceptor pairs in the six individuals. Variance reflects the range of difference between the values for the two individuals with the highest and lowest ratios.

ID			HG19 tRNA-SCAN			HG19 normalized ratio between isodecoder pairs									
AA	Anticodon 1	Anticodon 2	Count1	Count2	Ratio	HG19	12878	12891	12892	19238	19239	19240	STDEV	Variance
Arg	UCU	CCU	6	5	1.20	1	0.89	0.92	1.29	1.23	1.06	1.09	0.16	40%
Asn	GUU	AUU	32	2	16.00	1	1.33	1.13	1.13	0.92	1.47	1.05	0.20	55%
Gln	UUG	CUG	11	20	0.55	1	0.84	0.87	0.88	1.02	0.93	0.86	0.07	18%
Glu	UUC	CUC	13	13	1.00	1	0.95	0.85	0.95	0.93	0.70	0.65	0.13	29%
Leu	UAA	CAA	7	7	1.00	1	1.00	0.95	1.26	1.21	0.91	1.04	0.14	35%
Lys	UUU	CUU	16	17	0.94	1	1.15	1.30	1.13	1.23	1.14	1.07	0.08	23%
Tyr	GUA	AUA	14	1	14.00	1	0.94	0.81	0.93	0.77	1.07	1.24	0.18	47%