Landes Bioscience

www.landesbioscience.com

Supplemental Material to:

Marc Parisien, Xiaoyun Wang, and Tao Pan

Diversity of human tRNA genes from the 1000-genomes project

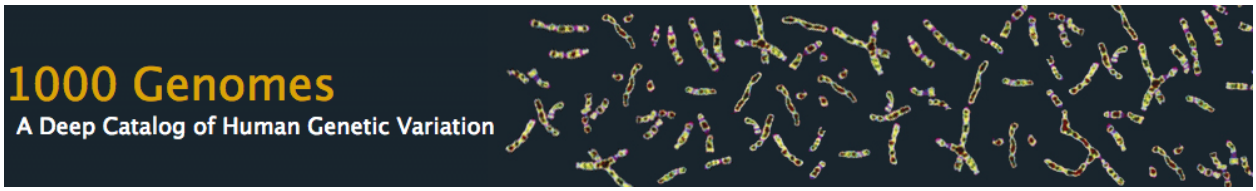
2013; 10(12)
http://dx.doi.org/10.4161/rna.27361

www.landesbioscience.com/journals/rnabiology/article/27361/

A

B

C



Overview: sequence a large number of people in order to provide a comprehensive resource on human genetic variations.

Goal: perform deep sequencing at a depth of about 4x coverage – not enough to reconstruct each individual's genomes, but sufficient to find most genetic variants that have frequencies of at least 1% in the studied populations.

Means: extract DNA from a pool of cells from a given individual, fragment the DNA into smaller pieces, then send these DNA fragments to deep sequencing. Because of the sharp decrease in cost of sequencing it is now possible to sequence many individuals.

D

FASTQ_FILE	MD5	RUN_ID	STUDY_ID	STUDY_NAME	CENTER_NAME	SUBMISSION_ID	SUBMISSION_DATE	SAMPLE_ID
SAMPLE_NAME	POPULATION	EXPERIMENT_ID	INSTRUMENT_PLATFORM	INSTRUMENT_MODEL	LIBRARY_NAME	RUN_NAME		
RUN_BLOCK_NAME	INSERT_SIZE	LIBRARY_LAYOUT	PAIRED_FASTQ	WITHDRAWN	WITHDRAWN_DATE	COMMENT	READ_COUNT	
BASE_COUNT	ANALYSIS_GROUP							
data/NA19238/sequence_read/ERR000018.filt.fastq.gz		3b092ef1661e2a8ff85050e01242707d				ERR000018		SRP000032
1000Genomes Project Pilot 2	BGI	ERA000013	2008-08-14 00:00:00		SRS000212	NA19238 YRI		ERX000014
ILLUMINA	Illumina Genome Analyzer		HU1000RADCAASE	BGI-FC307N0AAXX	0	SINGLE		0
9280498 334097928	high coverage							
data/NA19238/sequence_read/ERR000019.filt.fastq.gz		fcb89b0a755773872f1b073d0a518e0a				ERR000019		SRP000032
1000Genomes Project Pilot 2	BGI	ERA000013	2008-08-14 00:00:00		SRS000212	NA19238 YRI		ERX000014
ILLUMINA	Illumina Genome Analyzer		HU1000RADCAASE	BGI-FC307AWAAXX	0	SINGLE		0

E

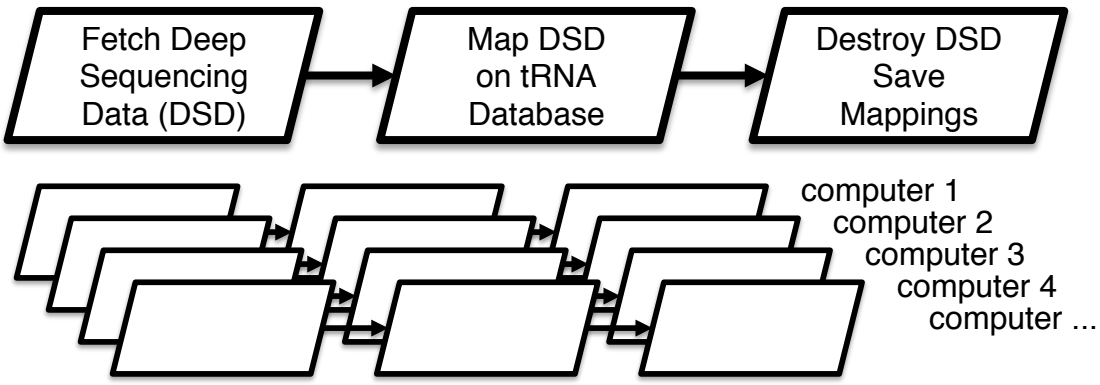


Figure S1. *Overview of the tRNA isodecoder diversity project.* **(A)** We employ the genomic tRNA database. **(B)** That database conveniently provides the DNA sequences and the secondary structures of most Human tRNA isodecoders and accepters. Shown here is a snapshot of the $tRNA^{LEU}$ data as presented on the web site. **(C)** We also employ the data collected by the 1000 genomes project. **(D)** Table from the 1000 genomes project describing the deep sequencing data (taken from ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/sequence.index). The panel shows the first two records that pertain to individual NA19238. **(E)** Computational tasks performed by a given computer for each entry in the table above.

Figure S1

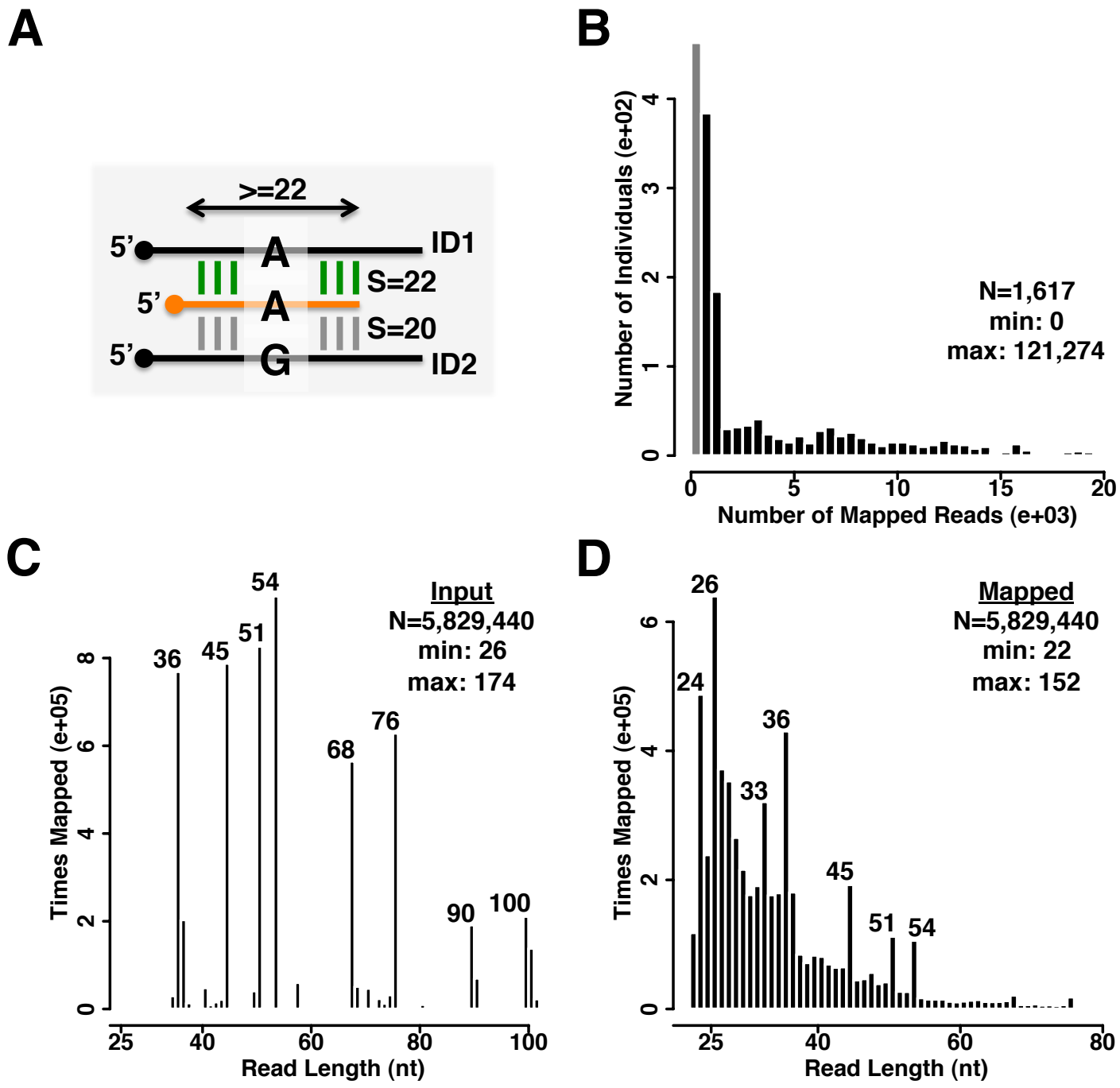


Figure S2. *Deep-sequencing reads mapping on the tRNA database.* **(A)** Mapping strategy. A deep-sequencing read (orange) can be aligned (vertical bars) to two tRNA isodecoders (ID1 and ID2) that differ at one nucleotide position (A/G) on a (minimum) length of 22 nucleotides. The score S of an alignment is computed as the number of matches minus the number of mismatches. The read is assigned to ID1 (green) as opposed to ID2 (grey) since the alignment has a better score with ID1 (22-0=22) versus ID2 (21-1=20). This way, a read aligned with a mismatch hasn't found any better tRNA isoacceptor or isodecoder sequence match. The reference tRNA sequences feature their introns. **(B)** Distribution of the number of mapped deep-sequencing reads per individual. About 70% of individuals have more than 500 reads mapped (black vertical bars). **(C)** Distribution of the mapped deep-sequencing reads length as found in the data files. Selected peaks are identified. **(D)** Distribution of the mapped deep-sequencing reads length. Selected peaks are identified. Pearson's R^2=0.008 for the correlation between mapped read length and the length of the input read.

Figure S2

Homo sapiens CTS telomere maintenance complex component 1 (CTC1), transcript variant 2, non-coding RNA

```
NR_046431  CCCTATATTAAGATTTGAAGTAGACCCGGAAAGTTAGTGGCCGGTTAGCTCAGTTGGTTA 4680
ILE-AAT    ------------------------------------GGCCGGTTAGCTCAGTTGGTTA 22
                                               **********************

NR_046431  GAGCGTGGTGCTAATAACGCCAAGGTCGCGGGTTCGAACCCCGTACGGGCCAGTGGGTGG 4740
ILE-AAT    GAGCGTGGTGCTAATAACGCCAAGGTCGCGGGTTCGATCCCCGTACGGGCCA-------- 74
           ********************************** **************
```

Homo sapiens long intergenic non-protein coding RNA 324 (LINC00324), non-coding RNA

```
NR_026951  TCGGCTCGTTGGTCTAGGGGTATGATTCTCGCTTCGGGTGCGAGAGGTCCCGGGTTCAAA 660
ProAGG     ----------------------------------------AGAGGTCCCGGGTTCAAA 81
                                                   ******************

NR_026951  TCCCGGACGAGCCCTCCTTTACCTTTTACTGAGACAAGAGTGTCTTCAAGGAATAGGTTA 720
ProAGG     TCCCGGACGAGCCC---------------------------------------------- 95
           **************
```

Homo sapiens microRNA 3676 (MIR3676), microRNA

```
NR_037447  ---------------TTGGTTAAAGCGCCTGTCTAGTAAACAGGAGATCCTGGGTTCGAA 45
ThrAGT     AGCACCATGGCTTAGCTGGTTAAAGCACCTGTCTAGTAAACAGGAGATCCTGAGTTTCAA 60
                         **********.*************************.***  **

NR_037447  TCCCAGCGGTGCCTCCGTGTTTCCCCCACGCTTTTGCCAA 85
ThrAGT     TTCCAATGGTGCCT-------------------------- 74
           * ***. *******
```

Figure S3. *Transfer RNAs with significant sequence identity to Human RNA transcripts.*

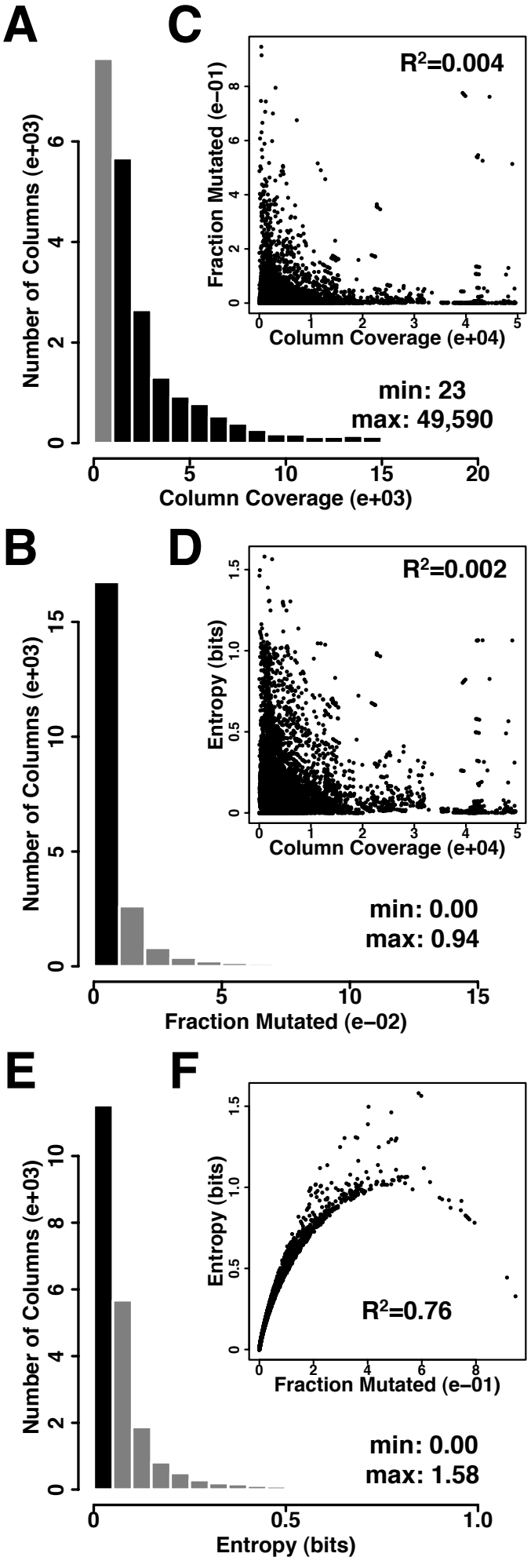


Figure S4. *Column coverage on the tRNA database.* A total of 21,785 columns are inspected for deep-sequencing coverage. **(A)** Distribution of column coverages. About 65% of columns have coverage of at least 1,000 (vertical black bars). **(B)** Distribution of mutation rates per columns. The rate is computed from the fraction of nucleotides in the column that do not correspond to the reference sequence. As much as 55% of columns have mutation rates below 0.5% (vertical black bar). **(C)** Scatter plot of column mutation rate with respect to the column's coverage. Pearson's R^2 indicates no correlation. **(D)** Scatter plot of column entropy rate with respect to the column's coverage. Pearson's R^2 indicates no correlation. **(E)** Distribution of sequence entropy per column. The sequence entropy of a column gives an indication on how many different symbols appear in a column and in what quantities. A column with equal shares of A, C, G and T's would have a perfect entropy of 2. About 53% of columns have entropy values less than 0.05 (vertical black bar). **(F)** Scatter plot of entropy rate of a column with respect to its mutation rate. Pearson's R^2 indicates a strong correlation. Indeed, a high entropy column requires a high mutation rate. The inverse is not necessarily true: a high mutation rate column could feature only one alternative symbol, therefore keeping the entropy low (arch-like plot). **(G)** Scatter plots of mutation rates versus column coverage: for the first 30 nucleotides of tRNAs with scores greater than 50 (left) and for the last 40 nucleotides of tRNAs with scores lower than 50 (right).

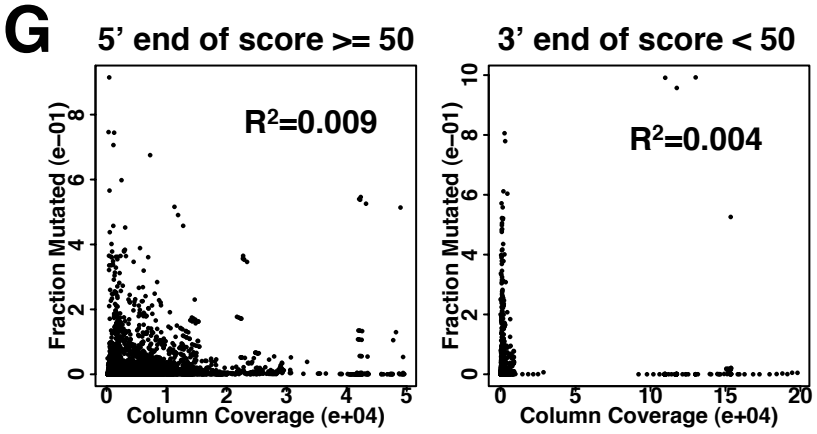


Figure S4

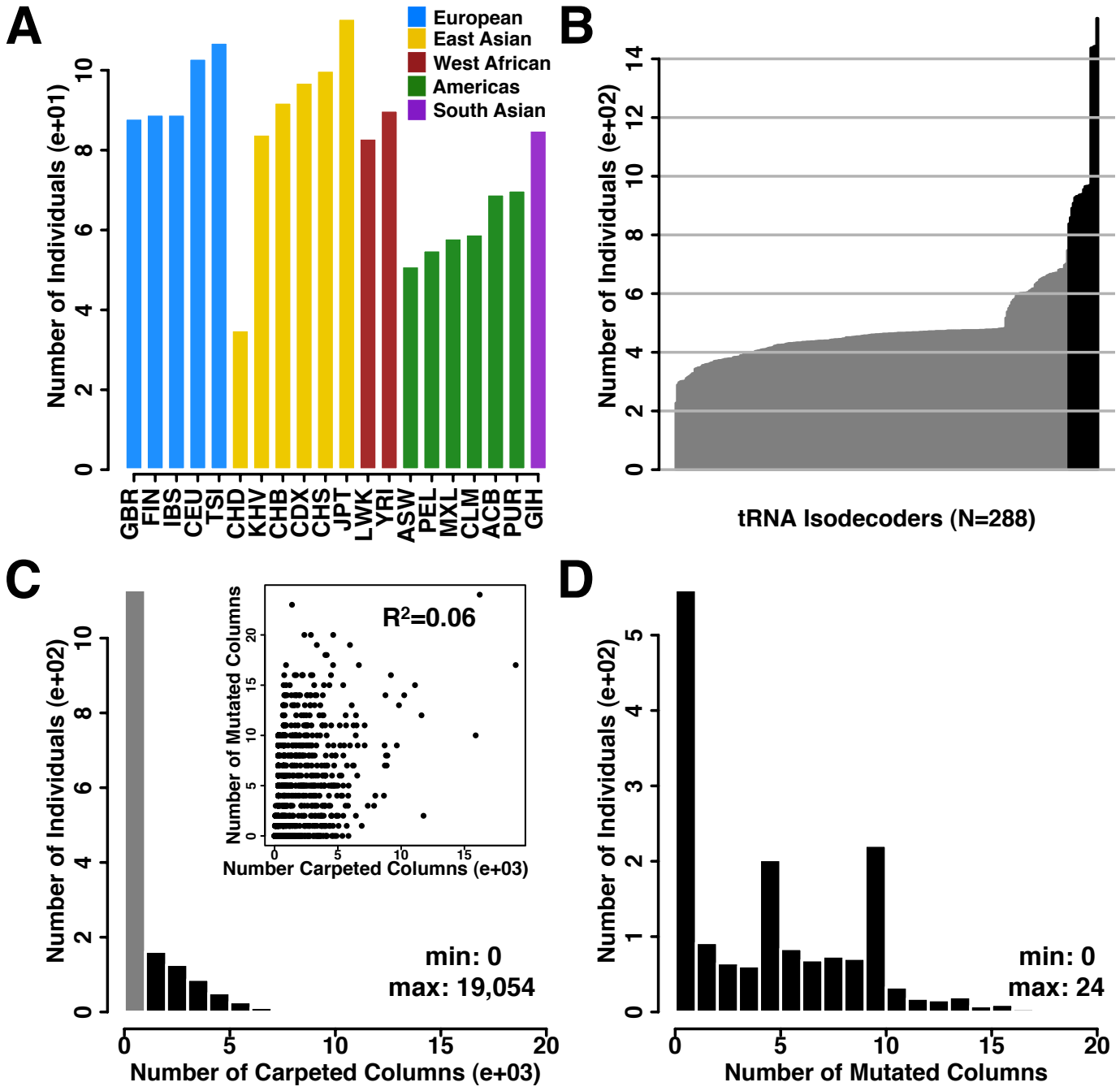


Figure S5. *Coverage per individual, for N=1,617 individuals.* **(A)** Number of individuals per population code. **(B)** Number of individuals per tRNA isodecoders. An individual is counted for an isodecoder only if it has some deep-sequencing data that maps on that isodecoder. There are 288 isodecoders with tRNA-Scan scores >= 50. Only 20 isodecoders are present in at least half of the individuals (vertical black bars). **(C)** Distribution of carpeted columns, those with good coverage (>=10). Only 30% of individuals have more than 1,000 columns with good coverage among the total 21,785 (vertical black bars). Pearson's R^2=0.06 indicates that there is no correlation between the number of carpeted columns and their mutated status (inset). **(D)** Distribution of mutated columns.

Figure S5

A

```
TRNA_0336:T:11   17  TRNA-IleTAT  Homo_sapiens_chr19.trna10-IleTAT
>>>>>>>*.>>>>*___**__*<<<<.>>>>>_*|||_____________________<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
          T
GCTCCAGTGGCGCAATCGGTtAGCGCGCGGTACTTATAtgacagtgcgagcggagcaATGCCGAGGtTGTGAGTTCGATCCTCACCTGGAGCA
```

B

```
TRNA_0319:T:26   21  TRNA-IleAAT  Homo_sapiens_chr6.trna163-IleAAT
>>>>>>>*.>>>>*___**__*<<<<.>>>>>_*|||__<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
                         T
GGCCGGTTAGCTCAGTTGGTtAGAGCGTGGTGCTAATAACGCTAAGGtCGCGGGTTCGATCCCCGTACTGGCCA
```

C

```
TRNA_0353:A:13   22  TRNA-LeuCAA  Homo_sapiens_chr1.trna58-LeuCAA
>>>>>>>*.>>>>*__**___*<<<<.>>>>>_*|||________________________<<<<<_______________>>>>>***_*__<<<<<<<<<<<<*
            A
GTCAGGATGGCCGAGTGGTctAAGGCGCCAGACTCAAGgtaagcaccttgcctgcgggctTTCTGGTCTCCGGATGGAGGCGTGGGTTCGAATCCCACTTCTGACA
```

D

```
TRNA_0051:G:40   30  TRNA-ArgACG  Homo_sapiens_chr6.trna8-ArgACG
>>>>>>>*.>>>>*__**__*<<<<.>>>>>_*|||__<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
                                       G
GGGCCAGTGGCGCAATGGAtAACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGGCTGGCTCG
```

E

```
TRNA_0158:G:27   43  TRNA-CysGCA  Homo_sapiens_chr14.trna8-CysGCA
>>>>>>>*.>>>>*__**_*<<<<.>>>>>_*|||__<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
                          G
GGGGGTATAGCTCAGGGGTAGAGCATTTGACTGCAGATCAAGAGGtCCCCGGTTCAAATCCGGGTGCCCCCT
```

F

```
TRNA_0356:C:102 435  TRNA-LeuCAA  Homo_sapiens_chr6.trna140-LeuCAA
>>>>>>>*.>>>>*__**___*<<<<.>>>>>_*|||__________________________<<<<<_______________>>>>>***_*__<<<<<<<<<<<<*
                                                                                                     C
GTCAGGATGGCCGAGTGGTctAAGGCGCCAGACTCAAGcttactgcttcctgtgttcgggtcTTCTGGTCTCCGTATGGAGGCGTGGGTTCGAATCCCACTTCTGACA
```

G

```
TRNA_0356:C:16  712  TRNA-LeuCAA  Homo_sapiens_chr6.trna140-LeuCAA
>>>>>>>*.>>>>*__**___*<<<<.>>>>>_*|||__________________________<<<<<_______________>>>>>***_*__<<<<<<<<<<<<*
               C
GTCAGGATGGCCGAGTGGTctAAGGCGCCAGACTCAAGcttactgcttcctgtgttcgggtcTTCTGGTCTCCGTATGGAGGCGTGGGTTCGAATCCCACTTCTGACA
                          TGCG

>>>>>>>..>>>...........<<<.>>>>>.......<<<<<.>>>>...<<<<..>>>>>.......<<<<<<<<<<<<.
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr6.trna7-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna67-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna42-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna40-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna38-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna36-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTCCTGACA Homo_sapiens_chr1.trna34-LeuCAG Sc: 77.22
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTTCTGACA Homo_sapiens_chr16.trna26-LeuCAG Sc: 75.90
GTCAGGATGGCCGAGCGGTctAAGGCGCTGCGTTCAGGTCGCAGTCTCCCCTGGAGGCGTGGGTTCGAATCCCACTTCTGACA Homo_sapiens_chr16.trna17-LeuCAG Sc: 75.90
```

H

```
TRNA_0570:A:26  185  TRNA-TyrGTA  Homo_sapiens_chr8.trna4-TyrGTA
>>>>>>>*.>>>>*___**_*<<<<.>>>>>_*|||_____________________<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
                         A
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA

pairwise comparisons:
>>>>>>>..>>>>........<<<<.>>>>>.........................<<<<<.....>>>>>.......<<<<<<<<<<<<.
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA Homo_sapiens_chr8.trna4-TyrGTA Sc: 73.04
     *    *                ^       *                *       *         *              *
CCTTCAATAGTTCAGCTGGTAGAGCAGAGGACTATAGctacttcctcagtaggagacGTCCTTAGGtTGCTGGTTCGATTCCAGCTTGAAGGA Homo_sapiens_chr2.trna14-TyrATA Sc: 55.93

>>>>>>>..>>>>........<<<<.>>>>>.........................<<<<<.....>>>>>.......<<<<<<<<....<<<<.
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCG....AAGGA Homo_sapiens_chr8.trna4-TyrGTA Sc: 73.04
* ****         **        ^****       ********************   ***     *     *  *  *  *  * **********
TCAATTATAGCTCAGT.GGTAGAGCATTTAACTGTAG....................ATCAAGAGGtCCCTGGATCAACTCTGGGTGCccccTTTAA Homo_sapiens_chr7.trna24-TyrGTA Sc: 33.50

>>>>>>>..>>>>..........<<<<.>>>>>.........................<<<<<........>>>>>.......<<<<<<<<<<<<.
CCTTCGATAGCTCAGCTGGT..AGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAG...GtCGCTGGTTCGATTCCGGCTCGAAGGA Homo_sapiens_chr8.trna4-TyrGTA Sc: 73.04
******* *  *  * *   **  * *^ ** *     ******************** * ***  *** *****     *     ***********
GTCAGTGTTGCACAAC.GGTtaAGTGAAG.AGGCTGTAA...................ACC..CAGACTG.GATGGGTTCAATTCCCATCTCTGCCG Homo_sapiens_chr15.trna5-TyrGTA Sc: 24.21
```

Figure S6

I

```
TRNA_0570:A:6    71  TRNA-TyrGTA  Homo_sapiens_chr8.trna4-TyrGTA
>>>>>>>*.>>>>*___**_*<<<<.>>>>>_*|||______________________<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
     A
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA

pairwise comparisons:
>>>>>>>..>>>>.........<<<<.>>>>>...........................<<<<<....>>>>>.......<<<<<<<<<<<<.
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA Homo_sapiens_chr8.trna4-TyrGTA Sc: 73.04
     ^    *              *       *               *       *         *              *   *
CCTTCAATAGTTCAGCTGGTAGAGCAGAGGACTATAGctacttcctcagtaggagacGTCCTTAGGtTGCTGGTTCGATTCCAGCTTGAAGGA Homo_sapiens_chr2.trna14-TyrATA Sc: 55.93

>>>>>>>..>>>>........<<<<.>>>>>...........................<<<<<....>>>>>........<<<<<<<<<<<<.
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGctacttcctcagcaggagacATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA Homo_sapiens_chr8.trna4-TyrGTA Sc: 73.04
*    ^                               * * *** * **********  *      **       *       *  *  *  **
TCTTCAATAGCTCAGCTGGTAGAGCGGAGGACTGTAGgtgcacgcccgtggcc....ATTCTTAGG.TGCTGGTTTGATTCCGACTTGGAGAG Homo_sapiens_chr8.trna12-TyrGTA Sc: 46.11
```

J

```
TRNA_0450:C:25   60  TRNA-PheGAA  Homo_sapiens_chr6.trna56-PheGAA
>>>>>>>*.>>>>*___**_*<<<<.>>>>>_*|||_____<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
                        C
GCCGAAATAGCTCAATTGGGAGAGTGTTAGACTGAAGatcTTCTGCAGGtCTCTGGTTCAATTCCGGGTTTCGACA

>>>>>>>..>>>>........<<<<.>>>>>..........<<<.<<......>>>>>.......<<<<<<<<<<<<.
GCCGAAATAGCTCAATTGGGAGAGTGTTAGACTGAAGatcTTCT.GCAGGt.CTCTGGTTCAATTCCGGGTTTCGACA Homo_sapiens_chr6.trna56-PheGAA Sc: 53.54
              *         ^            ***
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna96-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna109-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr19.trna14-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr13.trna1-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr12.trna11-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr11.trna15-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr11.trna13-PheGAA Sc: 82.45
GCCGAGATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna106-PheGAA Sc: 82.13
GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACCGAAG...ATCTtAAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna103-PheGAA Sc: 67.94
GCCAAAATTGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCACCA Homo_sapiens_chr6.trna72-PheGAA Sc: 59.98
GCTGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAG...ATCT.TAAAGttCCCTGGTTCAACCCTGGGTTTCAGCC Homo_sapiens_chr6.trna112-PheGAA Sc: 56.32
GCCAAAATAGCTCAGCTGGGAGAGCATTAGACTGAAG...ATCT.AAAGGt.CTCTGGTTTGATCCTGGGTTTCAGAA Homo_sapiens_chr6.trna88-PheGAA Sc: 43.56
GCTGAGATAGCTCGGTTGGGAGGGCATCAGCCTGAAG...ATCT.AAAGGa.GACTGGTTCAATTCTGGGTTTTGGCA Homo_sapiens_chr6.trna116-PheGAA Sc: 27.39
```

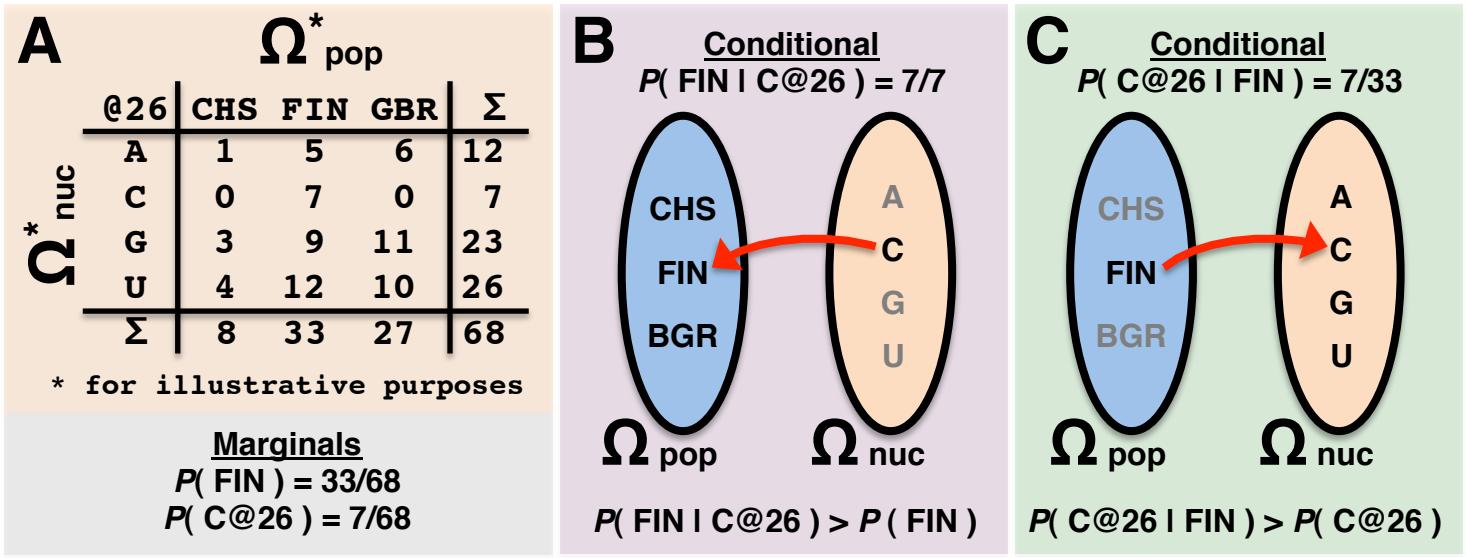
K

```
TRNA_0450:G:15   33  TRNA-PheGAA  Homo_sapiens_chr6.trna56-PheGAA
>>>>>>>*.>>>>*___**_*<<<<.>>>>>_*|||_____<<<<<_____>>>>>***_*__<<<<<<<<<<<<*
              G
GCCGAAATAGCTCAATTGGGAGAGTGTTAGACTGAAGatcTTCTGCAGGtCTCTGGTTCAATTCCGGGTTTCGACA

pairwise comparison:
>.>>>>>..>>>>.........<<<<..>>>>..........<<<.<.......>>>>>.......<<<<<<<<<<.<.
GCCGAAATAGCTCAATTGGG.AGAGTGTTAGACTGAAGatcTTCT.GCAGGt.CTCTGGTTCAATTCCGGGTTTCGACA Homo_sapiens_chr6.trna56-PheGAA Sc: 53.54
              ^          C            ***
>>>>>>>..>>>>.........<<<<.>>>>>..........<<<.<<......>>>>>.......<<<<<<<<<<<<.
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna96-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna109-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr19.trna14-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr13.trna1-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr12.trna11-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCGGCA Homo_sapiens_chr11.trna15-PheGAA Sc: 84.19
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr11.trna13-PheGAA Sc: 82.45
GCCGAAATAGCTCAGTTGGG.AGAGCGTTAGACCGAAG...ATCTtAAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna103-PheGAA Sc: 67.94
GCCGAGATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCAATCCCGGGTTTCGGCA Homo_sapiens_chr6.trna106-PheGAA Sc: 82.13
GCCAAAATTGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.AAAGGt.CCCTGGTTCGATCCCGGGTTTCACCA Homo_sapiens_chr6.trna72-PheGAA Sc: 59.98
GCTGAAATAGCTCAGTTGGG.AGAGCGTTAGACTGAAG...ATCT.TAAAGttCCCTGGTTCAACCCTGGGTTTCAGCC Homo_sapiens_chr6.trna112-PheGAA Sc: 56.32
GCTGAGATAGCTCGGTTGGG.AGGGCATCAGCCTGAAG...ATCT.AAAGGa.GACTGGTTCAATTCTGGGTTTTGGCA Homo_sapiens_chr6.trna116-PheGAA Sc: 27.39
GCCAAAATAGCTCAGCTGGG.AGAGCATTAGACTGAAG...ATCT.AAAGGt.CTCTGGTTTGATCCTGGGTTTCAGAA Homo_sapiens_chr6.trna88-PheGAA Sc: 43.56

>.>>>>>..>>>>.........<<<<..>>>>..........<<<.<.......>>>>>.......<<<<<<<<<<.<.
GCCGAAATAGCTCAATTGGG.AGAGTGTTAGACTGAAGatcTTCT.GCAGGt.CTCTGGTTCAATTCCGGGTTTCGACA Homo_sapiens_chr6.trna56-PheGAA Sc: 53.54
              ^*    *                 ***
GCCAAAATAGCTCAGCTGGG.AGAGTATTAGGTTGAAG...ATAC.AAAGTt.CCTTGGCTCAATCCAGAGTTTGGGGG Homo_sapiens_chr8.trna9-PheGAA Sc: 20.84
TGCATGGTTGTCTAGT.GGCtAGGATTCGGTGCTGAAA...GAGC.CACGG..CCCGGGTTCGATTCCCGGTCAGGGAA Homo_sapiens_chr1.trna100-PheGAA Sc: 32.35
TGCATGGTTGTCTAGT.GGCtAGGATTCGGTGCTGAAA...GCGT.CACGG..CCCGGGTTCGATTCCCGGTCAGGGAA Homo_sapiens_chr1.trna92-PheGAA Sc: 32.31
```

Figure S6. *Assessment of the 11 new tRNA isodecoders.* The assessment is unambiguous when no other tRNA isodecoders feature the altered nucleotide identity at the indicated position in the new isodecoder (A-F). When other isodecoders can be found to feature the altered nucleotide identity, it suffices to find sequence mismatches between the new isodecoder sequence with the existing ones near the altered site (G-K).



D

P(POP | N @ i)				tRNA	ID
POP	N	i	P		
PUR	U	23	1.0	ArgUCU	chr9.trna4
PUR	U	11	1.0	IleUAU	chr6.trna55
GBR	G	20	1.0	TyrGUA	chr6.trna14
GBR	A	18	1.0	IleUAU	chr19.trna10
FIN*	C	26	1.0	TyrGUA	chr14.trna18
CEU	U	17	1.0	IleUAU	chr2.trna5
CEU	A	24	1.0	SerGCU	chr6.trna123
CDX	U	11	0.7	IleUAU	chr2.trna5

Figure S7. *Conditional probability analysis for tRNA alleles.* (A) Consider a simplified contingency table where we collect occurrences of all four nucleotides Ω_{nuc} for three population codes Ω_{pop} at the 26th column of a given tRNA isodecoder sequence alignment. From the table, we can compute marginal probabilities, like *P (FIN)*, the probability of an individual to be of the FIN population group, and *P(C@26)*, the probability to observe a C at column 26 (grey area). (B) Now we can ask if there's any gain for an element in a sample space Ω given the knowledge of an element in the other sample space; a conditional probability. For instance, we can ask what's the probability of an individual to be of the FIN population code given that it has a C at column 26, which we write as *P(FIN|C@26)*. The space of ancestries Ω_{pop} and the space of nucleotide identities Ω_{nuc} are laid side-by-side: The red arrow suggests a two-step process: FIN ancestry is estimated given the knowledge of nucleotide identity at position 26. Here, the conditional probability *P(FIN|C@26)* is higher than the marginal *P (FIN)* (magenta area). (C) Likewise, we can ask if there's any gain in knowing the population code of an individual on the nucleotide distribution. Here too, the conditional probability *P(C@26|FIN)* is higher than the marginal *P(C@26)* (green area). (D) Probability table conditioned on nucleotide identity. The table shows the probability *P* of ancestry POP given the knowledge of nucleotide identity *N* at position *i*. All entries have *P* (POP)<<*P*(POP|N@i), that is, the knowledge of nucleotide identity increases significantly the knowledge about ancestry. The entry with a star corresponds to the case depicted in panel B, with actual numbers from the 1,000 genomes project. The probability is rounded up to a value of 1.0 which makes it almost certain that a person would be of Finnish ancestry given we observe a C at position 26 of tRNA Tyr^{GUA}. Probability tables conditioned on ancestry can be found in the main figures.

POP code	Description
European Ancestry (Euro)	
CEU	Northern and Western European
TSI	Toscani in Italia
GBR	British from England and Scotland
FIN	Finnish from Finland
IBS	Iberian in Spain
East Asian Ancestry (EaAs)	
CHB	Han Chinese in Beijing, China
JPT	Japanese in Tokyo, Japan
CHS	Han Chinese South
CDX	Chinese Dai in Xishuangbanna
KHV	Kinh in Ho Chi Minh City, Vietnam
CHD	Chinese in Denver, Colorado
West African Ancestry (WeAf)	
YRI	Yoruba in Ibadan, Nigeria
LWK	Luhya in Webuye, Kenya
Americas (Amer)	
ASW	African Ancestry in Southwest US
ACB	African Caribbean in Barbados
MXL	Mexican Ancestry in Los Angeles, CA
PUR	Puerto Rican in Puerto Rico
CLM	Colombian in Medellin, Colombia
PEL	Peruvian in Lima, Peru
South Asian Ancestry (SoAs)	
GIH	Gujarati Indian in Houston, TX

Table S1. *Population codes found in this study and their ancestry.*