Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/4/2/eaao0665/DC1

Supplementary Materials for

Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression

Matthew M. Parks, Chad M. Kurylo, Randall A. Dass, Linda Bojmar, David Lyden, C. Theresa Vincent, Scott C. Blanchard

Published 28 February 2018, *Sci. Adv.* **4**, eaao0665 (2018)
DOI: 10.1126/sciadv.aao0665

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/4/2/eaao0665/DC1)

- table S4 (Microsoft Excel format). Human individuals with variants in intersubunit bridges.
- table S5 (Microsoft Excel format). rRNA variants whose AFs are stratified by populations.
- table S7 (Microsoft Excel format). Common rRNA variant positions in mouse and human.
  table S8 (Microsoft Excel format). Log intensity of proteins detected from quantitative mass spectrometry analysis.

Supplementary Figures

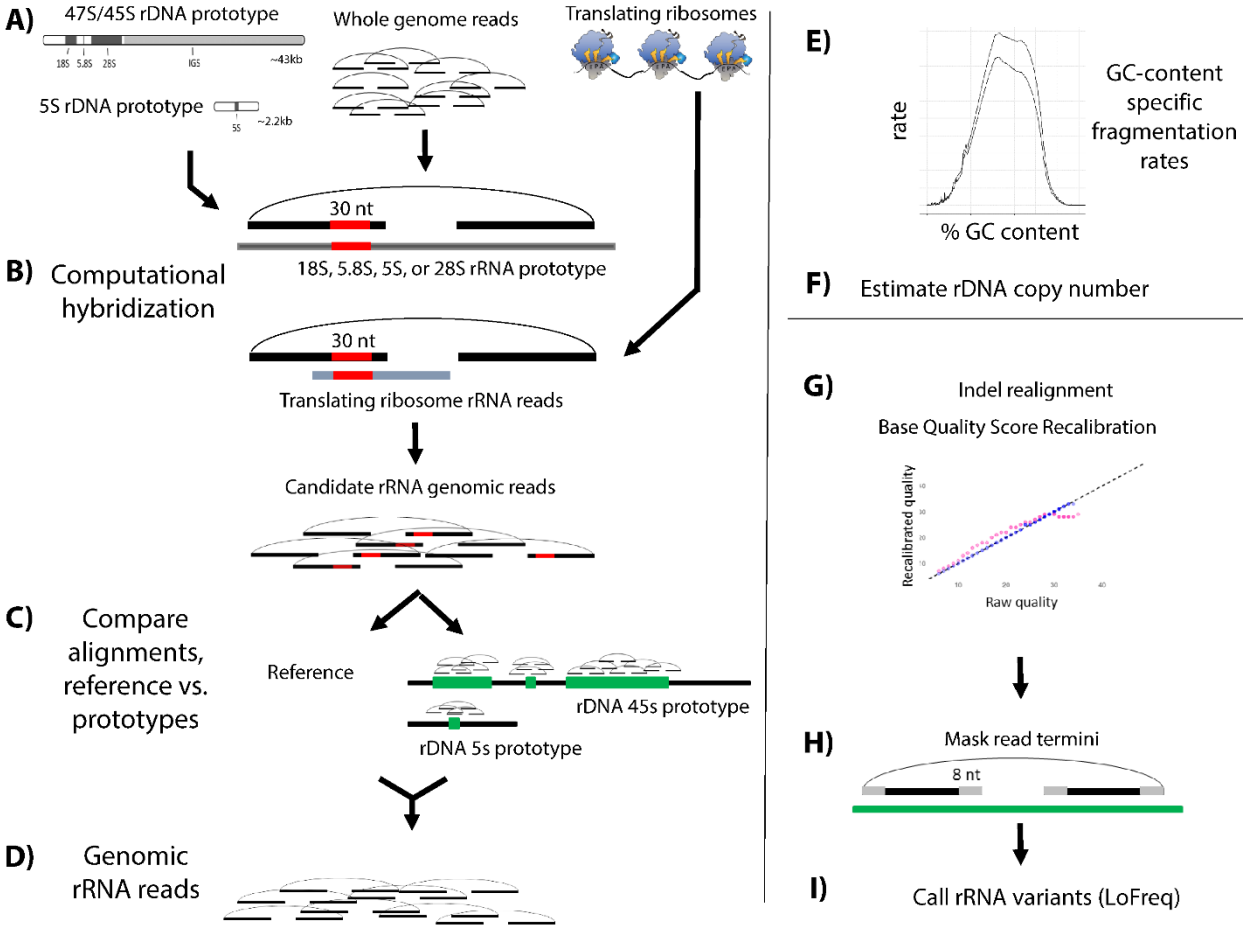

fig. S1. Bioinformatics strategy for rRNA copy number estimation and variant discovery. A) The rDNA prototype and RNA-seq reads from actively translating ribosomes are used to identify whole genome sequencing (WGS) reads putatively generated from rDNA regions by **B)** computational hybridization, wherein paired-end reads are selected if at least one of the mates contains a contiguous stretch of 30 nt of perfect identity to the prototype or any RNA-seq read. **C)** Candidate rRNA WGS reads are then aligned to the reference genome and the rDNA prototype, separately, and **D)** only paired-end reads which do not have better alignments to the reference are retained. **E)** Sample-specific biases in read depth distribution according to GC-content are computed to **F)** estimate rDNA copy number from the identified rRNA WGS reads. **G)** WGS rRNA reads are aligned to the prototype rDNA, indel realignment and base quality score recalibration are performed via GATK, **H)** base qualities at terminal 8nt on both ends of the reads are manually reduced to Q0, and **I)** variants are called using LoFreq.

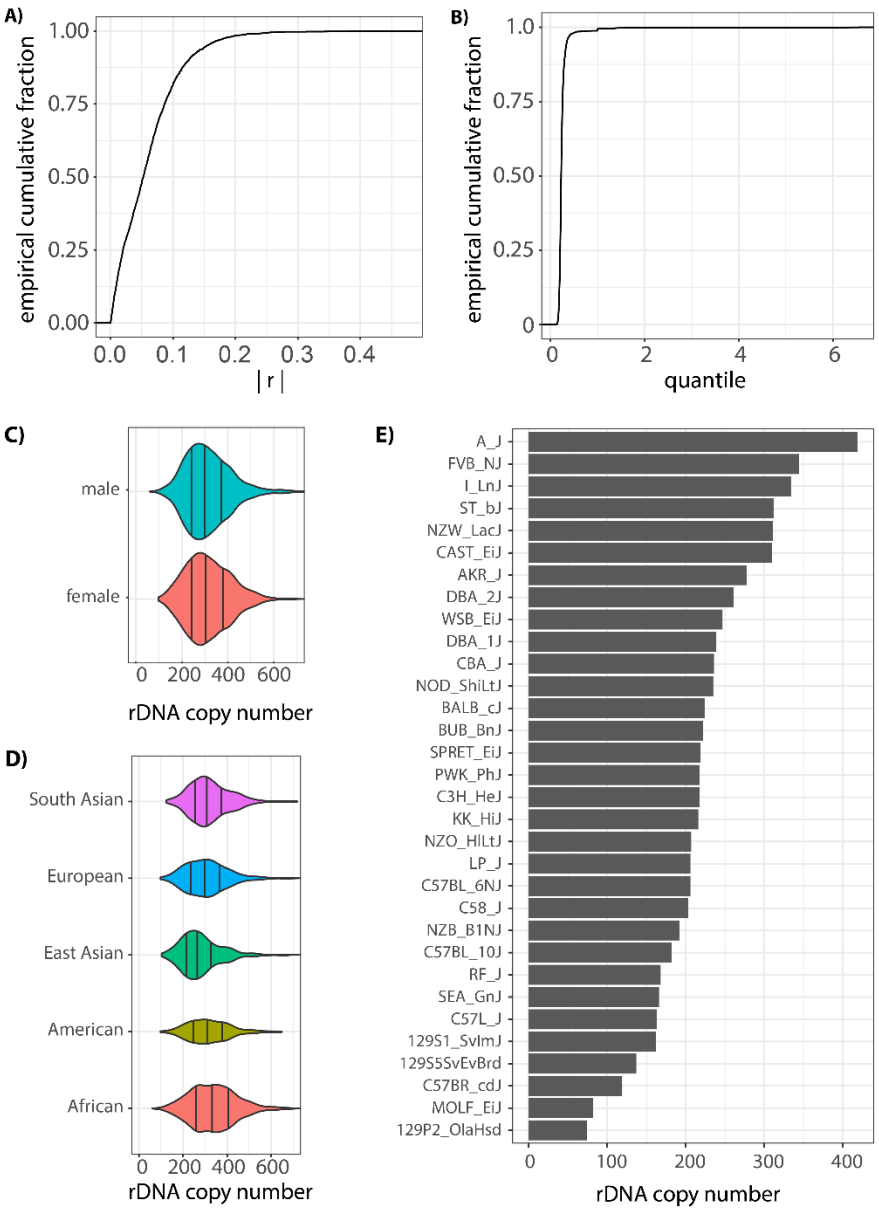

fig. S2. rDNA copy number estimation in human and mouse. A) Empirical cumulative distribution of the correlation coefficient for GC-content vs. predicted copy number of control regions across individuals. For each individual analyzed, 50,000 regions of width 4,000 nt of the reference genome that were not overlapping with regions with detected structural variation were randomly sampled. Copy number predictions for each sample were computed as described in **Methods**. The Pearson correlation coefficient *r* for correlation between predicted copy number and GC-content of the region was computed using the 50,000 randomly sampled regions, for each individual. The plot shows the empirical distribution of $|r|$ across individuals. **B)** Empirical cumulative distribution across individuals of the 90% quantile of absolute error in copy number estimate for the 50,000 randomly sampled regions of length 4,000 nt used for copy number estimation validation. **C,D)** Estimated rDNA copy number estimates per human individual from the 1000 Genomes Project, grouped by **C)** sex and **D)** continent. **E)** Estimated rDNA copy number per mouse strain from the Mouse Genomes Project.

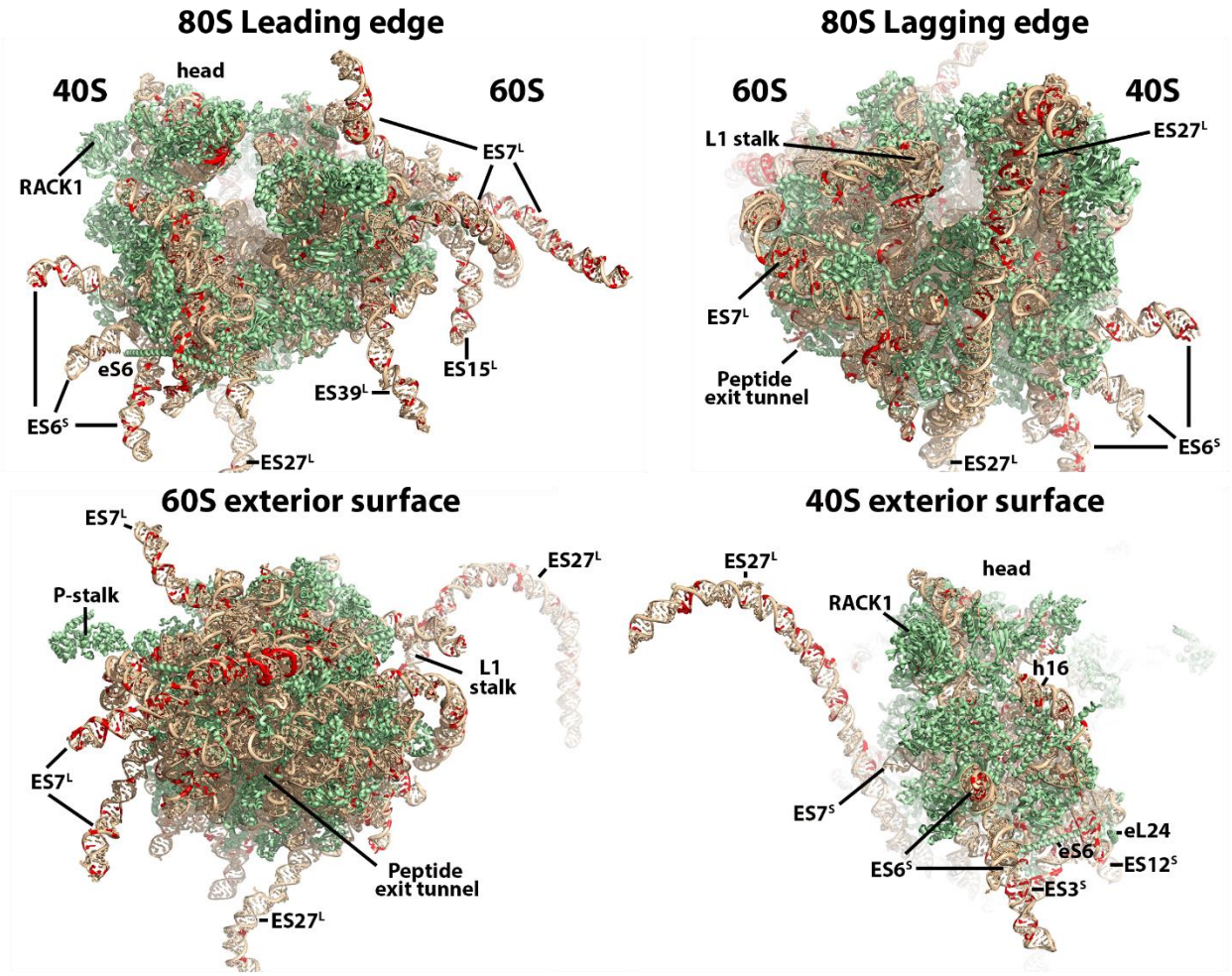

fig. S3. Tertiary structure of rRNA variants. High frequency rRNA variants (AF ≥ 20%) (red) on the tertiary structure of the ribosome, with associated proteins (green). Ribosome tertiary structures used were: PDB ID 4V6X(*49*).

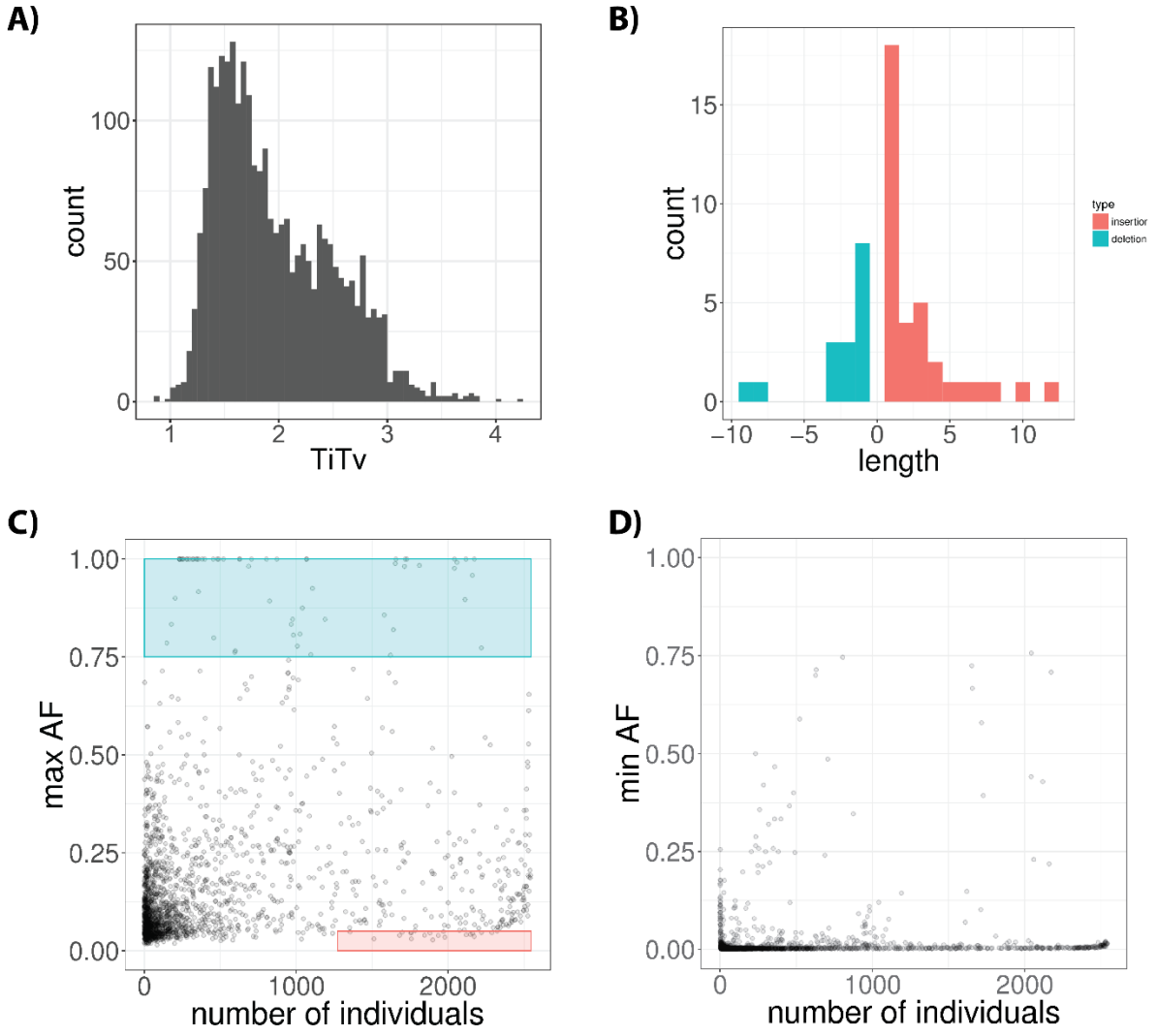

fig. S4. Properties of detected rDNA variants detected in rRNA genes in human. A) Transition/Transversion (Ti/Tv) ratio per individual from 1000 Genomes Project. **B)** Length distribution of detected rRNA indel variants. **C)** rRNA variant maximum genomic allele frequency across individuals vs number of individuals with the variant. Regions of extreme penetrance to number of individuals relationship are highlighted red and blue. **D)** rRNA variant minimum genomic allele frequency across individuals vs number of individuals with the variant.

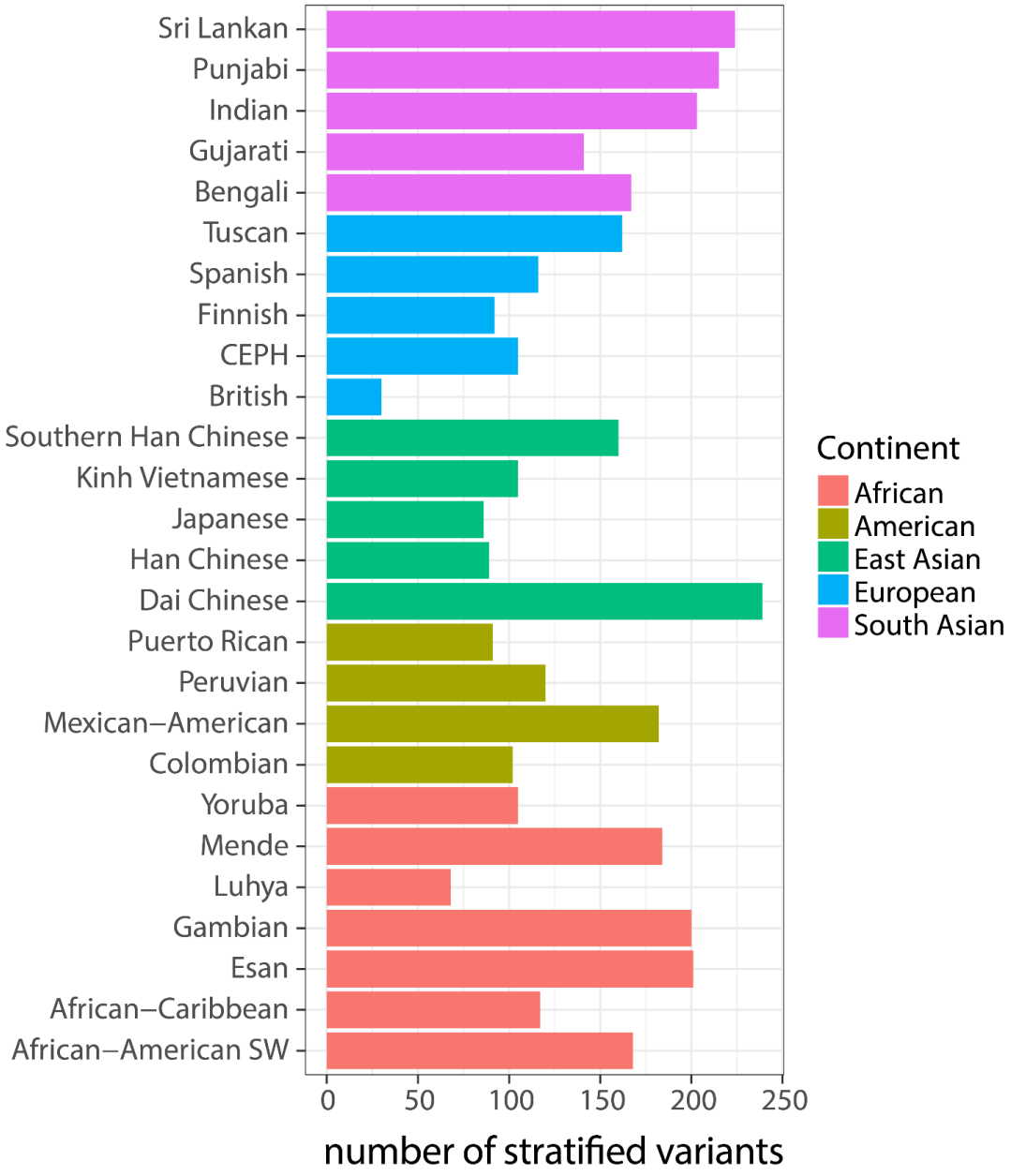

fig. S5. The number of variants stratified by each population. Here, a variant is considered stratified by population *x* if there exists another population *y* such that the allele frequency of the variant is stratified populations *x* and *y*.

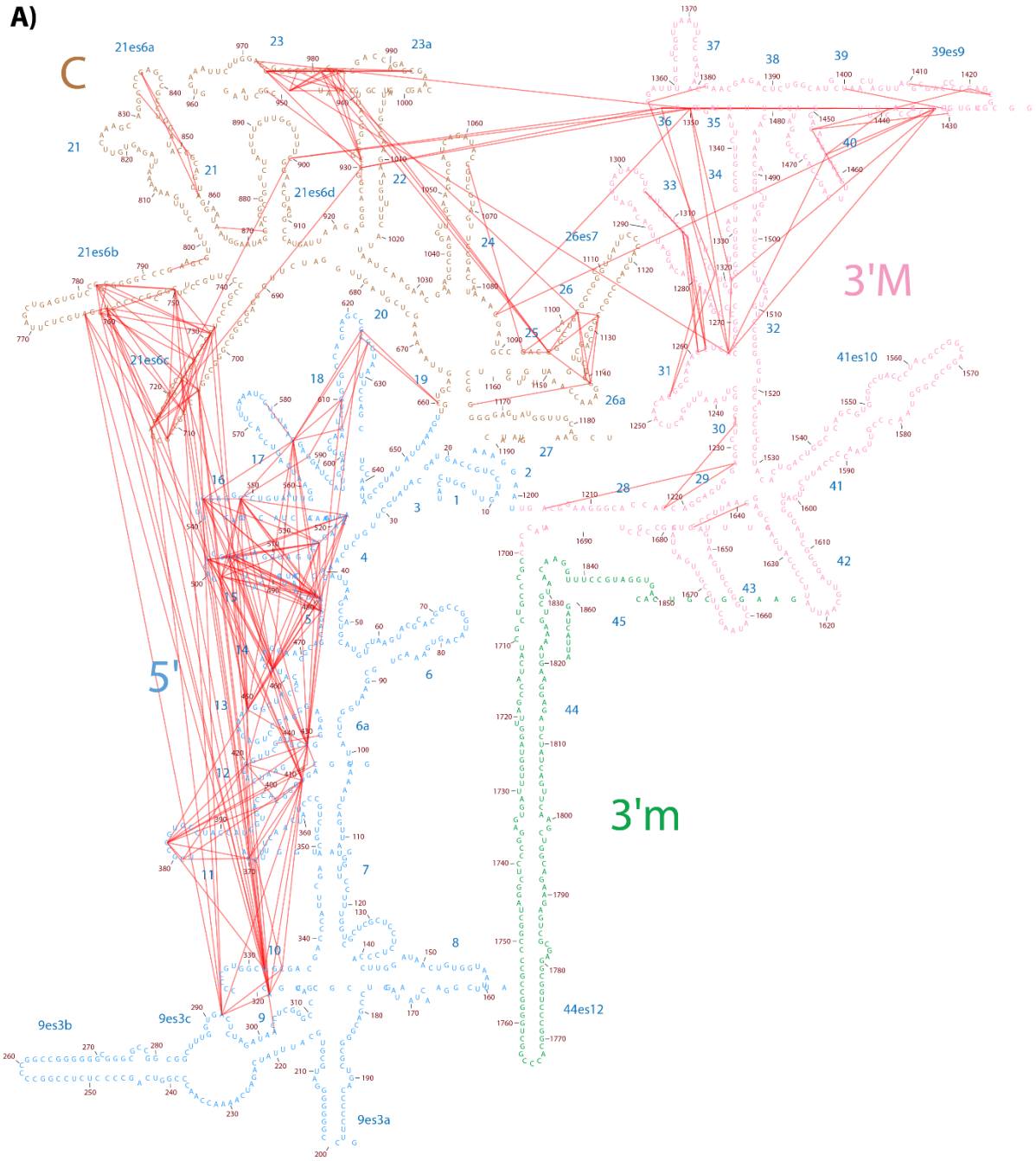

fig. S6. rRNA variants with correlated AF in human. Secondary structure diagram of the 18S rRNA showing intra-individual genomic AF across human individuals from the 1000 Genomes Project and are less than 500 nt apart. Nucleotides are color-coded by domain, as indicated. Red lines link pairs of rRNA variants with highly correlated ($|r| \geq 0.75$) intra-individual AF. Diagrams were made using Ribovision.

B)

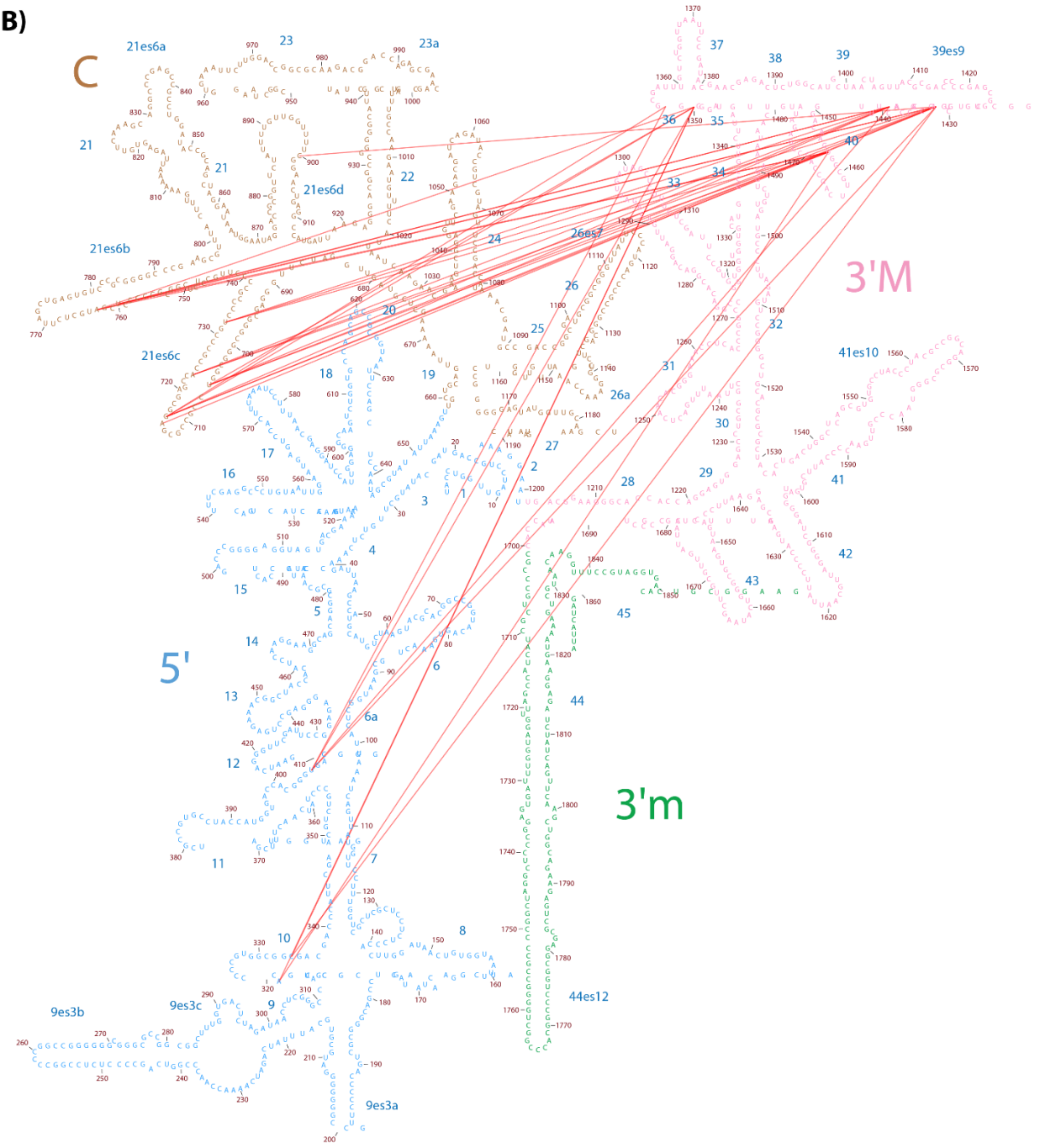

fig. S6. rRNA variants with correlated AF in human. Secondary structure diagram of the 18S rRNA showing intra-individual genomic AF across human individuals from the 1000 Genomes Project that are at least 500 nt apart in primary sequence. Nucleotides are color-coded by domain, as indicated. Red lines link pairs of rRNA variants with highly correlated ($|r| \geq 0.75$) intra-individual AF. Diagrams were made using Ribovision.

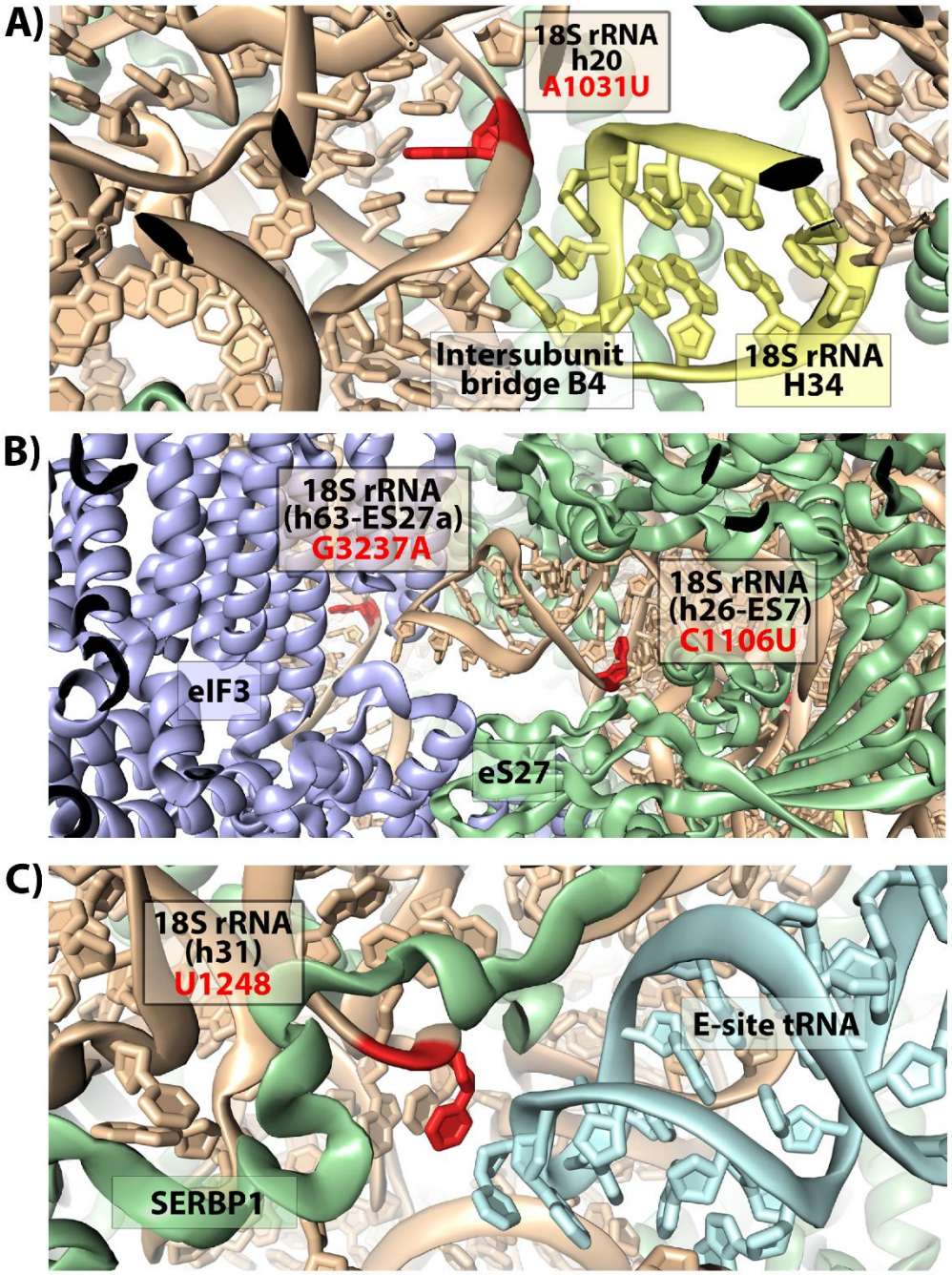

fig. S7. rRNA variants that are differentially expressed between mouse tissues that localize to known functional centers of the ribosome. Structural models based on **A)** PDB ID 5LZS, **B)** eIF3 structures from EMD-3056-3058 modeled onto PDB ID 5LZS, and **C)** PDB ID 4V6X.

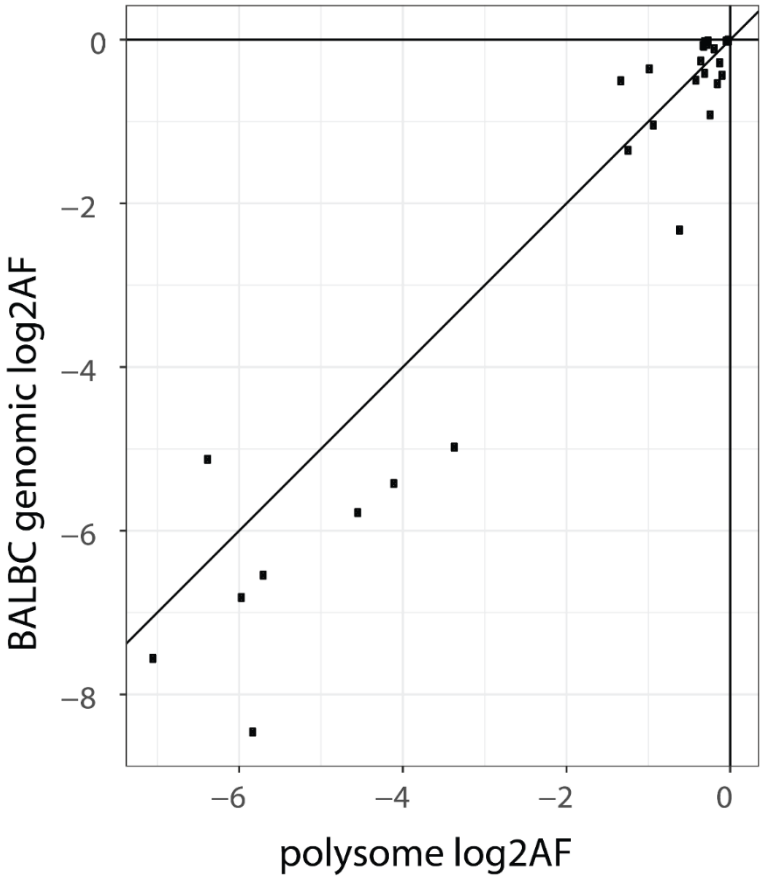

fig. S8. Genomic AF and polysome expression. Correlation of log2 allele frequency of 30 rRNA variants that are found in both actively translating ribosomes from polysome isolates in NMuMG cells and in the genome for any of the 32 strains analyzed from the Mouse Genomes Project. For the genomic allele frequency (AF), the median genomic AF amongst strains with the variant was used.

Supplementary Tables

table S1. Population stratification of rDNA copy number (V_{st} > 0.2). Size gives the number of individuals in each population.

table S2. Detected rRNA indel variants.

table S3. Detected rDNA variants in human that occur at positions of rRNA known to be modified in eukaryotes.

table S4. Human individuals with variants in intersubunit bridges.

table S5. rRNA variants whose AFs are stratified by populations.

table S6. Summary of rRNA variants detected in mouse strains from the Mouse Genomes Project.

	SNP		deletion		insertion		Total	
	variant alleles	positions	variant alleles	positions	variant alleles	positions	variant alleles	positions
18S	38	38	2	2	3	3	43	43
28S	184	180	8	8	21	17	213	205
5.8S	24	23	1	1	0	0	25	24
5S	4	4	0	0	0	0	4	4
Total	250	245	11	11	24	20	285	276

table S7. Common rRNA variant positions in mouse and human.

table S8. Log intensity of proteins detected from quantitative mass spectrometry analysis.