# A comprehensive rRNA variation atlas in health and disease

Daphna Rothschild[1]*, Teodorus Theo Susanto[1]*, Jeffrey P. Spence[1], Naomi R. Genuth[1,2], Nasa Sinnott-Armstrong[1], Jonathan K. Pritchard[1,2]§, Maria Barna[1]§

[1] Department of Genetics, Stanford University, Stanford, CA, 94305, USA
[2] Department of Biology, Stanford University, Stanford, CA, 94305, USA
* Co-first author
§ Co-senior author. Email: pritch@stanford.edu (JKP), mbarna@stanford.edu (MB).

## Summary

The hundreds of copies of ribosomal-DNA genes are the dark-matter of the human genome as it is unknown whether they possess sequence variation that forms different types of ribosomes. Here, we have overcome the technical hurdle of long-read sequencing of full-length ribosomal-RNA (rRNA) and developed an efficient algorithm for rRNA-variant detection. We discovered hundreds of variants that are not silent but are incorporated into translating ribosomes. These include tens of abundant variants within functionally important domains of the ribosome. Strikingly, variants assemble into distinct ribosome subtypes encoded on different chromosomes. With this first atlas of expressed rRNA-variants, we discover the impact of rRNA variation on health and disease. Across human tissues, we observe tissue-specific variant expression in endoderm/ectoderm derived tissues. In cancer, low abundant rRNA-variants become highly expressed. Together, this study provides a curated atlas for exploring rRNA variation and functionally links ribosome variation to tissue-specific biology and cancer.

## Introduction

The ribosome is a complex, ancient machine responsible for all protein synthesis, which has a core ribosomal RNA (rRNA) structure that is conserved across all kingdoms of life. However, when comparing the human and prokaryotic ribosome, a dramatic change occurred during the course of evolution, which increased the overall size of the ribosome. This is due in part to the insertion of blocks of sequences called expansion segments (ESs), as they “expand” the eukaryotic rRNA relative

to that of prokaryotes [1]. Interestingly, ESs vary greatly in length and sequence both within and among different species, and very little is known about their function [2–5]. These ESs are mostly solvent exposed, with highly flexible parts that allow for interactions with additional proteins or RNAs [6–8]. In humans, rRNA genes are present in hundreds of ribosomal DNA (rDNA) copies in tandem arrays that are spread across multiple chromosomal loci [9]. These high rDNA copy numbers are thought to be necessary to produce millions of ribosomes in each cell. Nevertheless, these hundreds of rDNA copies may allow for sequence variation between copies, which was first noted in mice and humans almost 50 years ago [10]. A recent important study has shown that about a quarter of positions in rDNA bear sequence variants [11]. Interestingly, the majority of these differences can be attributed to ES variability at the level of sequence variation in the form of indels and substitutions. Moreover, there have been a few examples of rRNA single nucleotide variants (SNVs) in non-human organisms [11–14]. Nevertheless, it is still an open question how abundant these expressed rRNA variants are in the context of a translating ribosome and if such variability is also linked to human physiology.

Until now, characterization of variants in genomic rDNA and expressed rRNA remained an outstanding challenge since previous studies used short-read sequencing, which has many limitations. One problem is that GC-rich sequences are highly depleted from short-read sequencing, and rRNA genes are extremely GC-rich, with certain regions including ESs having >80% GC on average [15–17]. This bias makes it difficult to reliably estimate the abundances of variants. Additionally, the assembly of longer rDNA regions has also remained a major problem with short read sequencing, given their high copy numbers, as combinatorially it is unknown which variants reside in the same rDNA copy. In order to compare variations between full length rDNA copies, new computational method development is also necessary to facilitate multiple sequence alignment, which is a hard computational problem [18]. Lastly, rRNA is highly structured with high GC-content and technologies that would enable long read sequences of full length 18S and 28S rRNA have not been developed (18S having about 1,800 bp and 28S having about 5,000 bp). As a result, previous studies mainly focused on copy number variations. Most interestingly, rDNA copy number was shown to decrease both with age and in cancer [19–23]. Recently, long read sequencing has successfully transformed the ability to analyze long repetitive regions in the human genome. Notably, this has enabled the successful complete assembly of a human genome by the Telomere-2-Telomere (T2T) consortium, including the positioning of rDNA copies in the five acrocentric chromosomes [24].

In this study, we address outstanding gaps in our understanding of rDNA variation. We developed novel experimental and computational strategies to comprehensively analyze rDNA and rRNA variants. We have overcome the challenges of long-read rRNA sequencing, which enabled us to characterize transcribed full length human 18S and 28S rRNA from actively translating ribosomes. Surprisingly, only a fraction of rRNA variants are expressed at high abundance. We discovered extensive 18S and 28S diversity including novel ribosome subtypes, which have multiple linked sequence variants, whose rRNA sequences are positioned on different chromosomes. We compiled this resource as the first coherent atlas of 18S and 28S gene variations. We have made this resource available in standard formats with detailed usage examples. We further use this atlas to find major changes in relative abundance of rRNA variants across different tissues. Moreover, we show that the rRNA variants can be detected in biopsies and that specific rRNA variants are at higher frequencies in some types of cancer. Together, these results suggest that ribosomes with unique sequence variation may be used to modulate different cellular programs underlying human physiology and disease.

# Results

## An atlas of 18S and 28S human rDNA variants validated in rRNA of translating ribosomes

Previously, a quarter of rDNA positions were reported to bear sequence variants [11]. Yet, with the exception of several variants, it is unknown if variants are transcribed and assembled into functionally active ribosomes. Moreover, the absolute abundance of any variant within the cell remains completely unknown. Due to technological barriers from using short read sequencing, nucleotide variants were not aggregated beyond Single Nucleotide Variants (SNVs), and it is unknown if variants are grouped to higher levels of structural organization units such as whole gene variants. With no reference of different rRNA subtypes, studies performing RNA sequencing discard the majority of rRNA reads. These rRNA variants remain completely ignored, which limits our understanding of the contribution of rRNA to human physiology and disease.

In this study we have characterized all human 18S and 28S rRNA full length variants from a human embryonic stem cell line (H7-hESC). Sequencing of rRNA has been historically technically

challenging [17,25] and we therefore have optimized long read sequencing. Here in order to extract rRNA from translationally active ribosomes, we first employed sucrose gradient fractionation wherein ribosomes can be separated into free ribosomal subunits, and translationally active ribosomes, which contain one or more bound ribosomes. We extracted RNA from translating ribosomes-containing fractions (**Figure S1**)**,** performed reverse transcription, and then sequenced complete 18S and 28S rRNA by HiFi accurate long read sequencing (**Figure 1A**, **Methods**). Here, we obtained 128,801 sequences of the 18S and 28,853 sequences of the 28S rRNAs from translating ribosomes. Since the reverse transcription may introduce errors at modified rRNA sites, we also sequenced the rDNA, and only included rRNA variants found in both rDNA and rRNA of H7-hESC in the atlas of variants (**Methods**).

In order to identify all positions with sequence variants, we developed a computational method which we term Reference Gap Alignment (RGA). Importantly, multiple sequence alignment (MSA) is computationally challenging and current MSA methods are heuristic and do not offer optimal sequence alignment. Here, we do not perform MSA. Instead, we extract all variants at a given position with respect to a common reference in a non-heuristic method as follows (**Figure 1B, S2**):

1. We classified sequences as either 18S or 28S followed by Needleman–Wunsch global sequence alignment [27] of each H7-hESC sequence to one RNA45S5 reference (either 18S or 28S based on read classification) [28].
2. We created a reference sequence that aligns to all other sequences that we call a gap-aligned reference. This gap-aligned reference has the same sequence as the reference but at each nucleotide position we extended a gap at the size of the maximal gap found by the global sequence alignment to all sequences. Importantly, this gap-aligned reference allows straightforward comparison among all sequences without requiring computationally expensive all-by-all pairwise sequence alignments.
3. We aligned all H7-hESC sequences to the gap-aligned reference using the previous global alignment with additional extended gaps at reference positions.
4. Lastly, we extracted all variants at a given position, across all aligned sequences.

In agreement with previous studies [11,17], we found that rDNA is highly variable (**Figure 2A-B**). Yet studies ignore rDNA variations as they are usually not thought as being expressed. An outstanding question is whether this rDNA is transcribed into function rRNA that is part of the translating

ribosome. In this regard, we find that rDNA variants are also found in functionally translating ribosomes as they are present in polysome fractions (**Figure S1**). Remarkably, we found 256 SNVs in the 18S and 1338 SNVs in the 28S, corresponding to about one variant every four rRNA positions (**Figure S3A-B**). Most variants (59%) are in ES regions (**Figure S3A-B**, ES/non-ES regions are annotated). Most importantly, with full coverage of the underlying rDNA and transcribed rRNA copies, we can accurately measure the frequency of each variant. The vast majority of rRNA variants are at low abundance, with only one variant from the 18S and 41 variants from the 28S having frequencies above 10% (**Figure 2C-D, Figure S3** annotated positions, **S3C** minor allele frequency with dashed like marking at 10% frequency).

Next we focused on the high abundance SNVs (> 10% rRNA frequency). In the 18S gene, a single variant appears at high rRNA frequency but has low frequency in rDNA (**Figure 2C**). This variant is at position 807 and is part of the es6s region. This ES is in the binding site for important eukaryotic translation initiation factors eIF3 and eIF4G [29–31]. In contrast, for the 28S high abundance variants, there is very good agreement between variant frequencies in rRNA and rDNA (**Figure 2D**, R=0.93 Pearson correlation, correlating all variants (colored in red)). Notably, most of these variants are single nucleotide insertions, located in 5 ES regions (es7l, es15l, es27l, es31l and es39l), and 3 non-ES helices (28S:h11, h68 and h76) (**Figure 2E**). While the function of these ESs is largely unknown, a growing body of research supports various roles in translation regulation. es27l has been shown to be important for control of translation fidelity and binding of ribosome-associated proteins for several processes, such as initiator methionine cleavage from the nascent polypeptide [7] or acetylation of nascent polypeptides [6]. Moreover, es39l interacts with the Signal Recognition Particle (SRP), which identifies the signal sequence on nascent polypeptides emerging from the translating ribosome [32]. Interestingly, the most abundant variant in the non-ES helices, at position 60 in 28S:h11, a G to A substitution is unique to humans. The alternative allele, A, is the reference allele for other mammals including chimpanzees [33,34].

For most positions, the RNA45S5 reference allele is the major allele in the sequenced H7-hESC line (**Figure 2E**, reference allele in gray, **Extended Data 1-2 SNV atlas**). However, at 2 sites that belong to es15l region and another 2 sites that belong to es27l region, the 28S RNA45S5 reference allele is less abundant than other variants (**Figure 2E**, gray for the reference allele and red for variants). Importantly, the comparison between rRNA and rDNA revealed global high agreement between a

variant's copy number in rDNA and its rRNA expression levels. Additionally, most variants have low abundance. However, these variant frequencies change in the context of human physiology and disease (as we later analyze here).

rRNA genes are organized into domains with helix-like structures, including the ES regions (**Figure S3, S4, Table S1-2** for region annotation). Here, using full length rRNA and rDNA sequencing we bin variants to different structural units of the ribosome and created an atlas at four resolutions (1) nucleotide position (as described earlier), (2) helix, (3) ES and non-ES intervening regions, and (4) complete 18S and 28S gene-resolution with multiple linked sequence variants where we later describe to be separated into subtypes.

## 28S variants assemble to genomically distinct ribosome subtypes

Using long read sequencing and accurate variant calling we next analyzed 18S and 28S in their entirety to address the outstanding question of whether rDNA variations lead to the formation of different ribosome subtypes. For the 18S, variants appear at low frequency and therefore there is not much evidence for 18S subtypes. For the 28S, we found high agreement between rRNA and rDNA variant frequency so we first asked which rDNA variants are co-expressed on a given gene copy. To do this, we calculated the correlation coefficient (Pearson's $r^2$) between positions across all 28S H7-hESC rDNA sequences. This is analogous to measuring the linkage disequilibrium (LD) coefficient in population genetics, though across paralogous copies within a single genome rather than across individuals in a population. Notably, we found low global LD structure between highly abundant rDNA variants (**Figure 3A** showing LD for rDNA positions with found rRNA frequency > 10%), perhaps indicating high rates of nonallelic gene conversion across rDNA copies. The highest LD ($r^2 > 0.2$) was found within three ES regions, es15l, es27l and es39l. However for es7l region, all variants were in low LD. Comparing different regions, we found LD between four regions: 28S:h11, es15l, es27l and es39l, where in each region we identified a position with higher linkage to the other three regions (**Figure 3A**, four positions are annotated, position 60 being 28S:h11, with higher linkage to other positions). By considering the variants at this subset of positions, we found a total of 21 different haplotypes both in rDNA and rRNA that we consider as candidate 28S subtype variants (**Figure S5**).

We next asked if our atlas variants replicate in other datasets, and assessed their linkage. We ran our computational pipeline on two long-read DNA resources: (1) the fully assembled genome from the T2T with 219 rDNA copies with their chromosome location [24], and (2) the Genome In A Bottle (GIAB) dataset with HiFi long-read genomic DNA sequencing on two families and a cell line [35]. In the T2T we discovered that although there are only 35 reported rDNA SNVs, the high frequency rDNA SNVs in H7-hESC also appeared in high frequency in the T2T rDNA (**Figure S6A**, R=0.74 Pearson correlation). In the GIAB samples, like in the H7-hESC, we found hundreds of SNVs with high agreement between their frequencies and H7-hESC frequencies (**Figure S6B-D**). Next, we analyzed the LD of the same positions in the T2T and GIAB, and found once again that es7l positions have low LD, and es15l, es27l and es39l have relatively higher LD within each region (**Figure S7**). Moreover, 3 out of the 4 positions with higher linkage to other positions in the H7-hESC, had higher LD in the GIAB dataset (**Figure 3B**). Since these three variants are linked to variants at other positions (due to higher LD to other variants) we define the haplotypes formed by positions 60, 3513 and 4913, belonging to regions 28S:h11, es27l and es39l respectively, as different 28S subtypes (**Figure 3C**, **Extended Data 7 28S atlas**).

We next asked whether different 28S subtypes are spatially separated in the genome. For the H7-hESC we have 386 complete 28S rDNA sequences and in the GIAB dataset we randomly subsampled each GIAB sample to 386 complete 28S rDNA sequences. For these datasets, we do not know rDNA-chromosome positioning. Notably, by comparison of 28S rDNA sequence similarities we detected distinct 28S sequence groups in both hESC and GIAB (**Figure 3D**, Principal Coordinate Analysis, PCoA, of Bray-Curtis dissimilarities between 28S sequences [36], **Methods**). Notably, the different clusters in PCoA space match different 28S subtypes. Specifically, we observed that 28S sequences of a given subtype are more similar to one another in their entire sequence compared to 28S rRNAs of other subtypes (**Figure 3D**). In the T2T assembly, rDNA copies have chromosome coordinates, which enables us to measure 28S subtype presence at the five acrocentric chromosomes. Remarkably, we discovered that 28S subtypes are largely chromosome specific (**Figure 3E**). Taken together, this supports that 28S subtypes are genomically separated.

In conclusion, we present the first atlas of functional human rRNA variants at different resolutions, from SNVs to complete gene subtypes (**Extended Data 1-7**, **Methods**). Importantly, this atlas is an

essential resource for measuring the abundance of rRNA variants and their association across various biological conditions which we further analyze here.

## Quantifying rRNA variants relative abundances in expression data.

With this rRNA atlas, we next aimed to investigate how rRNA composition changes in different human conditions. Importantly, our atlas is a novel dataset of rRNA variants that enables us to measure rRNA variant composition in any sequencing dataset. Moreover, this resource is available in standard FASTA and VCF formats (**Extended Data**), which are useful for quantifying rRNA variant relative abundances, including in short read data without *de novo* assembly of variants. We aggregated SNVs to regional ES/non-ES regional variants as an atlas reference, which is well suited for mapping short-reads (**Figure 4A, Figure S8**, **Table S2**, **Extended Data 5-6, Methods** for usage instructions).

Here, we used the publicly available Genotype-Tissue Expression (GTEx) short-read RNA-seq dataset to test if rRNA variant frequencies are associated with human tissue biology. Previous studies comparing mRNA across tissues in the GTEx dataset found tissue-specific, including brain-specific, gene expression [37,38]. Here we analyzed 2,618 samples from 332 individuals and 44 tissues from GTEx and asked if rRNA variants differ in their expression in these tissues (**Figure 4A**). We hypothesized that the rRNA regions, which were found here earlier to have high abundant variants in hESC (**Figure 2E**, namely, regions es6s from the 18S and h11, es7l, es15l, es27l, h76, es31l and es39l from the 28S), may be important for tissue development. Surprisingly, some of the highly abundant hESC regional variants were not highly abundant in GTEx (**Table S3**, highest regional variants expressed in GTEx). Nonetheless, the most abundant variant in GTEx in each of these regions significantly differed in expression in many tissues (**Figure 4B-I, Figure S9-12**, and **Table S4**: P-value < 0.05 FDR corrected Mann-Whitney U rank sum test; **Table S5**: P-value < 0.05 FDR corrected MANOVA test). Notably, when comparing the regional variant expression across tissues, we observe significant difference between tissues derived from the ectoderm and endoderm germ layers (**Figure 4J-Q, Table S6** FDR corrected ranksums test comparing regional variant median relative abundances of ectoderm-derived tissues in blue and endoderm-derived tissues in red). Most of the ectoderm derived tissues belong to brain tissues and most endoderm derived tissues are digestive-system tissues (**Figure 4A** endoderm and ectoderm derived tissues are labeled). Next, we asked if rRNA subtypes associate with tissues and analyzed the newest nanopore long-read

RNA-seq GTEx dataset with 88 samples from 12 tissues [39]. In this dataset, however, we found only 213 complete rRNA reads, 208 reads being from fibroblasts (**Table S7**). While this dataset is not suitable for comparing the 28S subtypes across tissues, we found the same 28S subtypes in the fibroblasts (**Table S7**). Taken together, our results support major changes in the expression of rRNA variants across tissues which could be a result of differential expression of 28S subtypes.

Lastly, we asked if changes in the expression of rRNA variants are associated with cancer. For this, we used 10,030 samples of short-read RNA-seq with clinical phenotypes from The Cancer Genome Atlas (TCGA) [40]. When comparing cancer types, we found distinct expression patterns of rRNA regional variants across cancers (**Figure S13-18**). This observation is in agreement with our finding that rRNA variants are differentially expressed across tissues (**Figure 4**). To test if rRNA variants are cancer-specific, we compared cancer biopsies to control biopsies from the same tissues. Strikingly, we identified specific rRNA regional variants which are found in low abundance in control biopsies and are significantly differentially expressed in cancer biopsies for 11 cancers (**Figure 5**, **Table S8** for alternative allele regional variant abundances**, Table S9,** P-value < 0.05 after FDR correction, Mann-Whitney U rank sum test). Thus even low abundant variants hold immense importance as disease biomarkers.

We conclude that our atlas enables direct measuring of rRNA variants changes in expression data. Moreover, we showed that atlas variants are functional in translating ribosomes and that they are differentially expressed across tissues and cancer-types.

# Discussion

With hundreds of rDNA copy numbers in the genome and poor understanding of their sequence variation, rDNA and other repetitive genomic regions could be considered the ‘dark-matter’ of the genome. Neither CRISPR genome editing nor assembly of rDNA genes with short-read sequencing is currently possible with these high copy numbers. However, long-read sequencing has enabled the final assembly of these highly repetitive regions. Here, using high quality long read sequencing of rDNA, and rRNA of actively translating ribosomes, we were able to curate an atlas of over a thousand rRNA sequence variants from an H7-hESC line. By analyzing full length 18S and 28S rRNA, we built an atlas at different resolutions from nucleotide position variants to 28S gene-level subtypes as a

useful resource for studying rRNA variations, and composition across biological conditions (**Extended data 1-5**).

Utilizing the first fully assembled human genome, from the T2T Consortium [24], we have discovered chromosome associated rDNA subtypes. It is an interesting finding in light of the prevailing understanding in the field about concerted evolution of rDNA units [41–43]. Specifically, it has been hypothesized that a tandem array of rDNA genes is generated from unequal crossing over between sister chromatids [44–46], leading to homogenization of rDNA units within a single chromosomal locus. It is possible that chromosomally distinct rDNA subtypes are important to maintain functionally different subtypes from being lost. Additionally, spatial separation of rDNA haplotypes might enable regulation of their expression at the chromosome level, through allelic inactivation of rDNA loci or inactivation of nucleolar organizer regions (NORs) in the distal junction [47–49].

With the creation of this comprehensive characterization of functional human rRNA variants, we have enabled greater understanding of the often neglected yet ubiquitous rRNA sequencing data. Our study offers the community a new resource to study rRNA in various biological systems. By utilizing publicly available datasets from GTEx and TCGA, we have found tissue and cancer-specific expression of human rRNA variants, which suggest regulation of ribosome composition in both normal and diseased physiology. In GTEx, we identified tissue-specific rRNA composition. It is yet to be discovered if the differential expression of these variants comes from different 28S subtypes. We have analyzed the GTEx long read dataset [39], however very few full length rRNA reads were found, highlighting the importance of our method in optimizing full length rRNA reverse transcription for 28S subtype analyses.

It is possible that human rRNA variants fine-tune regulation of mRNA translation, based on differential association with ribosome-associated proteins, mRNA transcripts, or other factors. We observed interesting differential expression of rRNA variants between tissues beloging to ectoderm and endoderm lineages. This pattern might hint at specialized functions of different ribosome subtypes. Long-lived cells associated with the nervous system might require ribosome subtypes that emphasize translation fidelity over speed, as compared to rapidly dividing cells in the digestive tract for example that require constant replacement given harsh local environments. Indeed, our lab and others have previously shown that es27l plays a role in translation fidelity through association with

ribosome-associated proteins [7,50,51]. Such interactions that trade speed over fidelity might be fine-tuned by the expression of different rRNA variants.

Finally, in cancer we discovered that some low abundant rRNA variants in control biopsies were elevated in cancer biopsies. Future work is needed to understand whether they promote oncogenic ribosome activity and how they are regulated. These results are in agreement with the growing appreciation that ribosomes in cancer may have different rRNA modifications and activities [52–56]. Therefore our results provide another layer of ribosome specificity wherein cancer cells might deploy a particular rRNA variant that is more compatible with their cellular fitness. Importantly, we found that specific rRNA variants may be used as biomarkers for disease. Notably, 5-fluorouracil, a common chemotherapy drug, was recently shown to incorporate into rRNA and promote drug resistance by changing mRNA translation [57]. Together, it may be that drugs directly target specific rRNA variants and further examination would be needed to test whether they should be used for cancer specific therapy.

Together, our results provide the first atlas to distinguish different types of ribosomes and to link them to different cellular programs, including those underlying human health and disease.

# Acknowledgments

We thank Xiangling Meng, Craig Kerr, Ali Wilkening and Michael Montgomery for help with early experiments that were not later followed in this study. We thank the Barna and Pritchard group members for discussions.

# Funding

MB is supported by the New York Stem Cell Foundation and the National Institutes of Health grant R01HD086634. JKP is supported by RO1 HG008140. DR is supported by ALTF 1042-2019 EMBO and LT000218/2020-L HFSP postdoctoral scholarships. TTS is supported by a National Science Scholarship (PhD) from the Agency for Science, Technology and Research. MB is a NYSCF Robertson Investigator.


# Authors contributions

DR conceived the project, designed experimental and computational analyzes, conducted all computational analyses, interpreted the results, and wrote the manuscript. TTS designed and conducted all experiments, interpreted the results, and wrote the manuscript. JPS helped in computational analyses. NG helped in experimental data collection. NSA helped interpret GTEx data, MB and JKP conceived and directed the project and analyses, designed the analyses, interpreted the results and wrote the manuscript.

# Competing interests

The authors declare no competing interests.

# Data and materials availability

The atlas is available as Extended Data to this publication.
H7-hESC raw rDNA and rRNA sequencing data is available under BioProject ID PRJNA926787.

# References


1. Gerbi, S.A. (1996). Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. Ribosomal RNA—Structure, evolution, processing, and function in protein synthesis, 71–87.
2. Ramesh, M., and Woolford, J.L. (2016). Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. RNA *22*, 1153–1162. 10.1261/rna.056705.116.
3. Morgan, D.G., Ménétret, J.F., Radermacher, M., Neuhof, A., Akey, I.V., Rapoport, T.A., and Akey, C.W. (2000). A comparison of the yeast and rabbit 80 S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments. J. Mol. Biol. *301*, 301–321. 10.1006/jmbi.2000.3947.
4. van Nues, R.W., Venema, J., Planta, R.J., and Raué, H.A. (1997). Variable region V1 of Saccharomyces cerevisiae 18S rRNA participates in biogenesis and function of the small ribosomal subunit. Chromosoma *105*, 523–531. 10.1007/BF02510489.
5. Houge, G., Robaye, B., Eikhom, T.S., Golstein, J., Mellgren, G., Gjertsen, B.T., Lanotte, M., and Døskeland, S.O. (1995). Fine mapping of 28S rRNA sites specifically cleaved in cells undergoing apoptosis. Mol. Cell. Biol. *15*, 2051–2062. 10.1128/MCB.15.4.2051.
6. Knorr, A.G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., and Beckmann, R. (2019). Ribosome-NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. Nat. Struct. Mol. Biol. *26*, 35–39. 10.1038/s41594-018-0165-y.
7. Fujii, K., Susanto, T.T., Saurabh, S., and Barna, M. (2018). Decoding the function of expansion segments in ribosomes. Mol. Cell *72*, 1013-1020.e6. 10.1016/j.molcel.2018.11.023.
8. Leppek, K., Fujii, K., Quade, N., Susanto, T.T., Boehringer, D., Lenarčič, T., Xue, S., Genuth, N.R., Ban, N., and Barna, M. (2020). Gene- and Species-Specific Hox mRNA Translation by Ribosome Expansion Segments. Mol. Cell *80*, 980-995.e13. 10.1016/j.molcel.2020.10.023.
9. Henderson, A.S., Warburton, D., and Atwood, K.C. (1972). Location of ribosomal DNA in the human chromosome complement. Proc Natl Acad Sci USA *69*, 3394–3398. 10.1073/pnas.69.11.3394.
10. Arnheim, N., and Southern, E.M. (1977). Heterogeneity of the ribosomal genes in mice and men. Cell *11*, 363–370. 10.1016/0092-8674(77)90053-8.
11. Parks, M.M., Kurylo, C.M., Dass, R.A., Bojmar, L., Lyden, D., Vincent, C.T., and Blanchard, S.C. (2018). Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci. Adv. *4*, eaao0665. 10.1126/sciadv.aao0665.
12. Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the Arabidopsis NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. Nat. Commun. *12*, 387. 10.1038/s41467-020-20728-6.
13. McCutchan, T.F., de la Cruz, V.F., Lal, A.A., Gunderson, J.H., Elwood, H.J., and Sogin, M.L.

(1988). Primary sequences of two small subunit ribosomal RNA genes from Plasmodium falciparum. Mol. Biochem. Parasitol. *28*, 63–68. 10.1016/0166-6851(88)90181-8.

14. Locati, M.D., Pagano, J.F.B., Girard, G., Ensink, W.A., van Olst, M., van Leeuwen, S., Nehrdich, U., Spaink, H.P., Rauwerda, H., Jonker, M.J., et al. (2017). Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. RNA *23*, 1188–1199. 10.1261/rna.061515.117.

15. Clark, C.G., Tague, B.W., Ware, V.C., and Gerbi, S.A. (1984). Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res. *12*, 6197–6220. 10.1093/nar/12.15.6197.

16. Wakeman, J.A., and Maden, B.E. (1989). 28 S ribosomal RNA in vertebrates. Locations of large-scale features revealed by electron microscopy in relation to other features of the sequences. Biochem. J. *258*, 49–56. 10.1042/bj2580049.

17. Fan, W., Eklund, E., Sherman, R.M., Liu, H., Pitts, S., Ford, B., Rajeshkumar, N.V., and Laiho, M. (2022). Widespread genetic heterogeneity of human ribosomal RNA genes. RNA *28*, 478–492. 10.1261/rna.078925.121.

18. Carrillo, H., and Lipman, D. (1988). The multiple sequence alignment problem in biology. SIAM J. Appl. Math. *48*, 1073–1082. 10.1137/0148063.

19. Xu, B., Li, H., Perry, J.M., Singh, V.P., Unruh, J., Yu, Z., Zakari, M., McDowell, W., Li, L., and Gerton, J.L. (2017). Ribosomal DNA copy number loss and sequence variation in cancer. PLoS Genet. *13*, e1006771. 10.1371/journal.pgen.1006771.

20. Malinovskaya, E.M., Ershova, E.S., Golimbet, V.E., Porokhovnik, L.N., Lyapunova, N.A., Kutsev, S.I., Veiko, N.N., and Kostyuk, S.V. (2018). Copy number of human ribosomal genes with aging: unchanged mean, but narrowed range and decreased variance in elderly group. Front. Genet. *9*, 306. 10.3389/fgene.2018.00306.

21. Tiku, V., and Antebi, A. (2018). Nucleolar function in lifespan regulation. Trends Cell Biol. *28*, 662–672. 10.1016/j.tcb.2018.03.007.

22. Ganley, A.R.D., and Kobayashi, T. (2014). Ribosomal DNA and cellular senescence: new evidence supporting the connection between rDNA and aging. FEMS Yeast Res. *14*, 49–59. 10.1111/1567-1364.12133.

23. Ren, R., Deng, L., Xue, Y., Suzuki, K., Zhang, W., Yu, Y., Wu, J., Sun, L., Gong, X., Luan, H., et al. (2017). Visualization of aging-associated chromatin alterations with an engineered TALE system. Cell Res. *27*, 483–504. 10.1038/cr.2017.18.

24. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. 10.1126/science.abj6987.

25. Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, e72. 10.1093/nar/gks001.

26. Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V.R., Hunicke-Smith, S., Swamy, S., et al. (2013). Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. RNA *19*, 958–970. 10.1261/rna.039743.113.

27. Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453. 10.1016/0022-2836(70)90057-4.

28. Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, R.L., Yates, B., and Bruford, E.A. (2021). Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Res. *49*, D939–D946. 10.1093/nar/gkaa980.

29. Toribio, R., Díaz-López, I., and Ventoso, I. (2016). New insights into the topology of the scanning ribosome during translation initiation: Lessons from viruses. RNA Biol. *13*, 1223–1227. 10.1080/15476286.2016.1247146.

30. Yu, Y., Abaeva, I.S., Marintchev, A., Pestova, T.V., and Hellen, C.U.T. (2011). Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors. Nucleic Acids Res. *39*, 4851–4865. 10.1093/nar/gkr045.

31. Díaz-López, I., Toribio, R., Berlanga, J.J., and Ventoso, I. (2019). An mRNA-binding channel in the ES6S region of the translation 48S-PIC promotes RNA unwinding and scanning. eLife *8*. 10.7554/eLife.48246.

32. Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature *427*, 808–814. 10.1038/nature02342.

33. Qu, L.H., Nicoloso, M., and Bachellerie, J.P. (1988). Phylogenetic calibration of the 5' terminal domain of large rRNA achieved by determining twenty eucaryotic sequences. J. Mol. Evol. *28*, 113–124. 10.1007/BF02143502.

34. Qu, L.H., Nicoloso, M., and Bachellerie, J.P. (1991). A sequence dimorphism in a conserved domain of human 28S rRNA. Uneven distribution of variant genes among individuals. Differential expression in HeLa cells. Nucleic Acids Res. *19*, 1015–1019. 10.1093/nar/19.5.1015.

35. Zook, J.M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C.E., Alexander, N., et al. (2016). Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci. Data *3*, 160025. 10.1038/sdata.2016.25.

36. Bray, J.R., and Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol. Monogr. *27*, 325–349. 10.2307/1942268.

37. Taskesen, E., and Reinders, M.J.T. (2016). 2D Representation of Transcriptomes by t-SNE Exposes Relatedness between Human Tissues. PLoS ONE *11*, e0149853. 10.1371/journal.pone.0149853.

38. Donovan, M.K.R., D'Antonio-Chronowska, A., D'Antonio, M., and Frazer, K.A. (2020). Cellular

deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. Nat. Commun. *11*, 955. 10.1038/s41467-020-14561-0.

39. Glinos, D.A., Garborcauskas, G., Hoffman, P., Ehsan, N., Jiang, L., Gokden, A., Dai, X., Aguet, F., Brown, K.L., Garimella, K., et al. (2022). Transcriptome variation in human tissues revealed by long-read sequencing. Nature *608*, 353–359. 10.1038/s41586-022-05035-y.
40. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400-416.e11. 10.1016/j.cell.2018.02.052.
41. Brown, D.D., Wensink, P.C., and Jordan, E. (1972). A comparison of the ribosomal DNA's of Xenopus laevis and Xenopus mulleri: the evolution of tandem genes. J. Mol. Biol. *63*, 57–73. 10.1016/0022-2836(72)90521-9.
42. Nei, M., and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. *39*, 121–152. 10.1146/annurev.genet.39.073003.112240.
43. Eickbush, T.H., and Eickbush, D.G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics *175*, 477–485. 10.1534/genetics.107.071399.
44. Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. Trends Genet. *18*, 74–82. 10.1016/s0168-9525(02)02592-1.
45. Carvalho, C.M.B., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. *17*, 224–238. 10.1038/nrg.2015.25.
46. Haber, J.E. (2018). DNA repair: the search for homology. Bioessays *40*, e1700229. 10.1002/bies.201700229.
47. Schlesinger, S., Selig, S., Bergman, Y., and Cedar, H. (2009). Allelic inactivation of rDNA loci. Genes Dev. *23*, 2437–2447. 10.1101/gad.544509.
48. van Sluis, M., van Vuuren, C., Mangan, H., and McStay, B. (2020). NORs on human acrocentric chromosome p-arms are active by default and can associate with nucleoli independently of rDNA. Proc Natl Acad Sci USA *117*, 10368–10377. 10.1073/pnas.2001812117.
49. Grob, A., Colleran, C., and McStay, B. (2014). Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes Dev. *28*, 220–230. 10.1101/gad.234591.113.
50. Wild, K., Aleksić, M., Lapouge, K., Juaire, K.D., Flemming, D., Pfeffer, S., and Sinning, I. (2020). MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. Nat. Commun. *11*, 776. 10.1038/s41467-020-14603-7.
51. Shankar, V., Rauscher, R., Reuther, J., Gharib, W.H., Koch, M., and Polacek, N. (2020). rRNA expansion segment 27Lb modulates the factor recruitment capacity of the yeast ribosome and shapes the proteome. Nucleic Acids Res. *48*, 3244–3256. 10.1093/nar/gkaa003.
52. Babaian, A., Rothe, K., Girodat, D., Minia, I., Djondovic, S., Milek, M., Spencer Miko, S.E.,

Wieden, H.-J., Landthaler, M., Morin, G.B., et al. (2020). Loss of m1acp3Ψ Ribosomal RNA Modification Is a Major Feature of Cancer. Cell Rep. *31*, 107611. 10.1016/j.celrep.2020.107611.

53. Jansson, M.D., Häfner, S.J., Altinel, K., Tehler, D., Krogh, N., Jakobsen, E., Andersen, J.V., Andersen, K.L., Schoof, E.M., Ménard, P., et al. (2021). Regulation of translation by site-specific ribosomal RNA methylation. Nat. Struct. Mol. Biol. *28*, 889–899. 10.1038/s41594-021-00669-4.

54. Sulima, S.O., Hofman, I.J.F., De Keersmaecker, K., and Dinman, J.D. (2017). How ribosomes translate cancer. Cancer Discov. *7*, 1069–1087. 10.1158/2159-8290.CD-17-0550.

55. McMahon, M., Contreras, A., Holm, M., Uechi, T., Forester, C.M., Pang, X., Jackson, C., Calvert, M.E., Chen, B., Quigley, D.A., et al. (2019). A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS. eLife *8*. 10.7554/eLife.48847.

56. Larson, M.H., Pan, W., Kim, H.J., Mauntz, R.E., Stuart, S.M., Pimentel, M., Zhou, Y., Knudsgaard, P., Demas, V., Aravanis, A.M., et al. (2021). A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. Nat. Commun. *12*, 2357. 10.1038/s41467-021-22444-1.

57. Therizols, G., Bash-Imam, Z., Panthu, B., Machon, C., Vincent, A., Ripoll, J., Nait-Slimane, S., Chalabi-Dchar, M., Gaucherot, A., Garcia, M., et al. (2022). Alteration of ribosome function upon 5-fluorouracil treatment favors cancer cell drug-tolerance. Nat. Commun. *13*, 173. 10.1038/s41467-021-27847-8.

# Figures

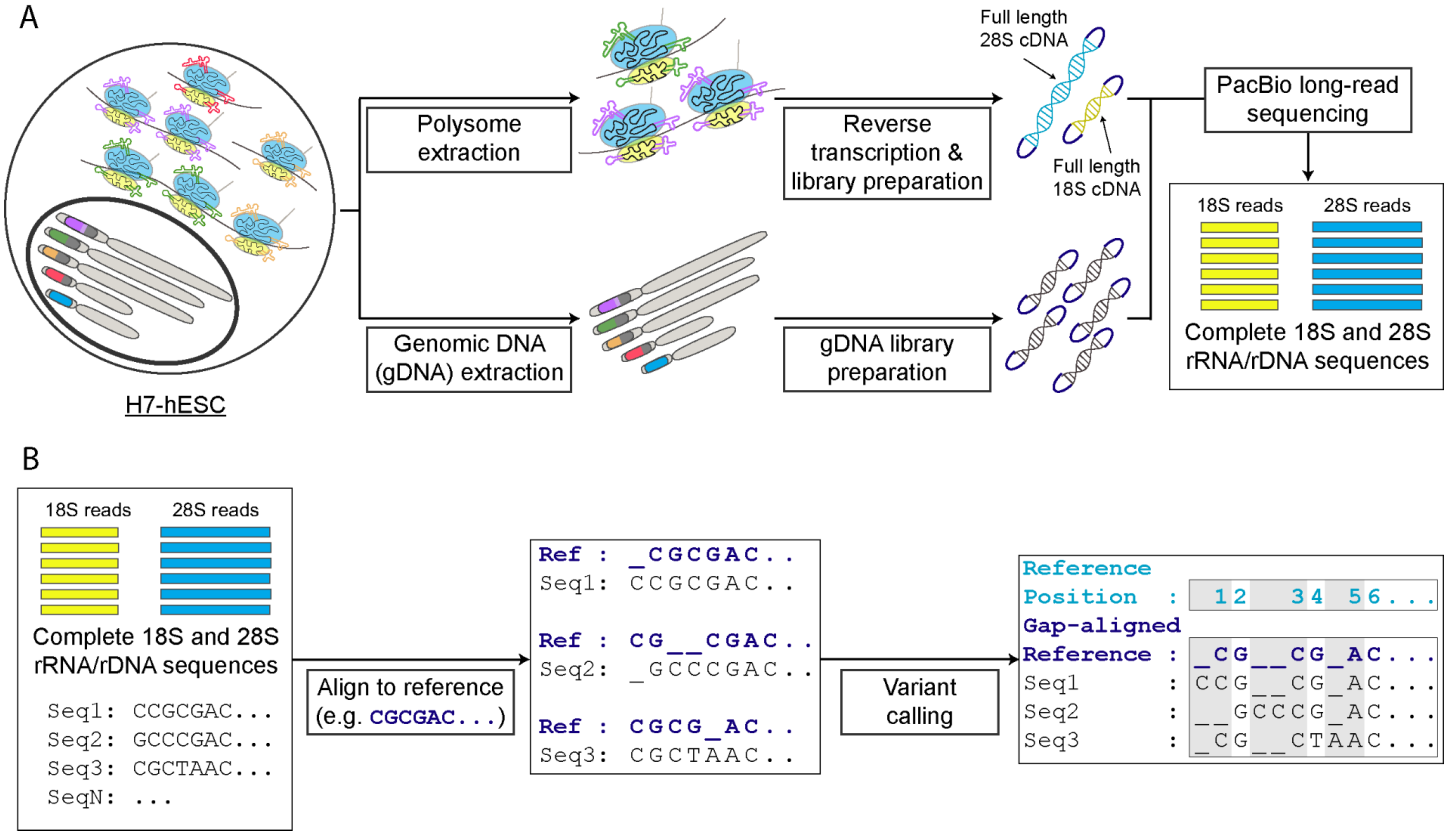



**Figure 1. Graphical Illustration of the rRNA atlas variant extraction pipeline.**

A. 18S and 28S rDNA / rRNA sequence extraction pipeline from H7-hESC

B. Variant extraction pipeline that starts with 18S and 28S rRNA / rDNA sequences from (A)

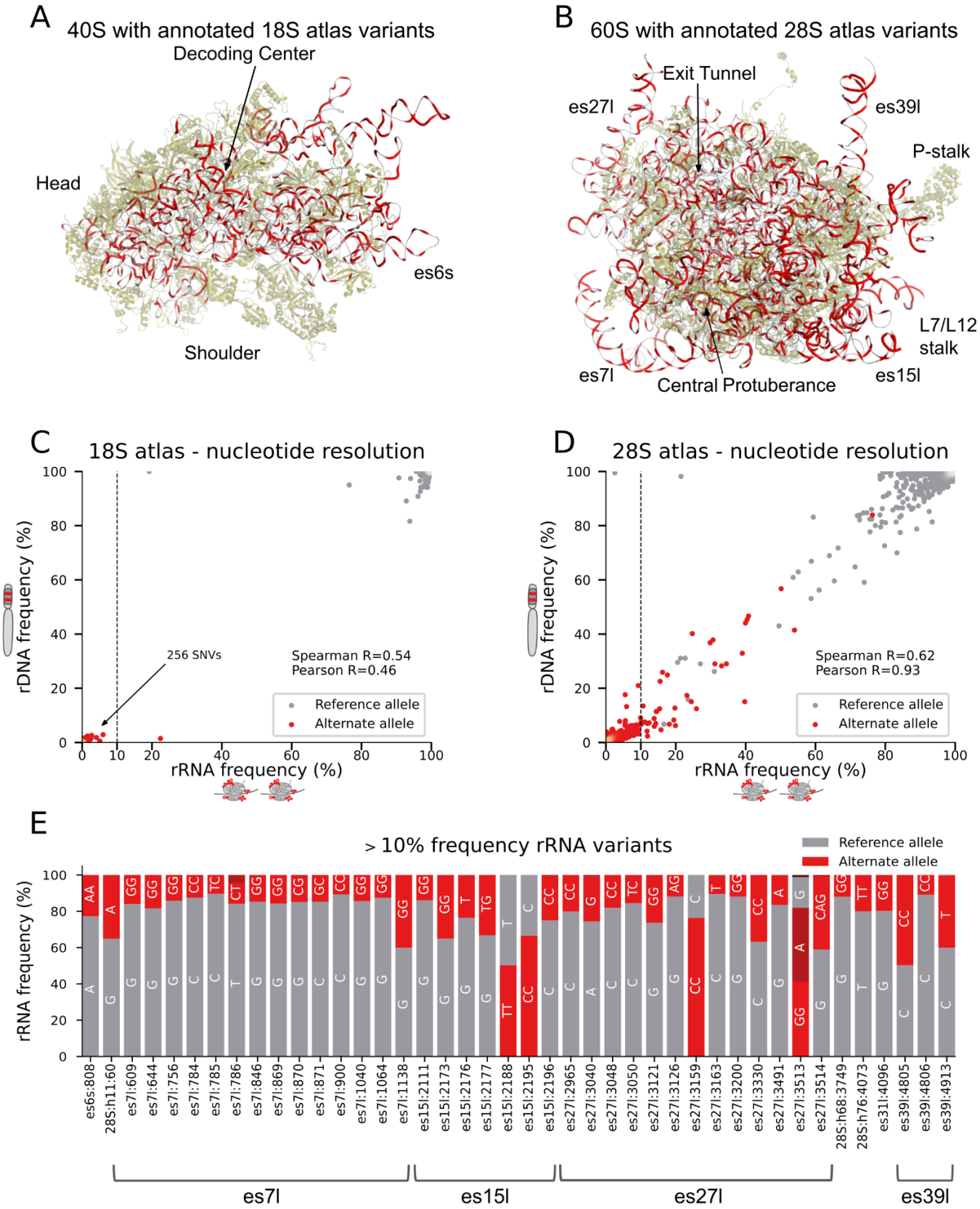
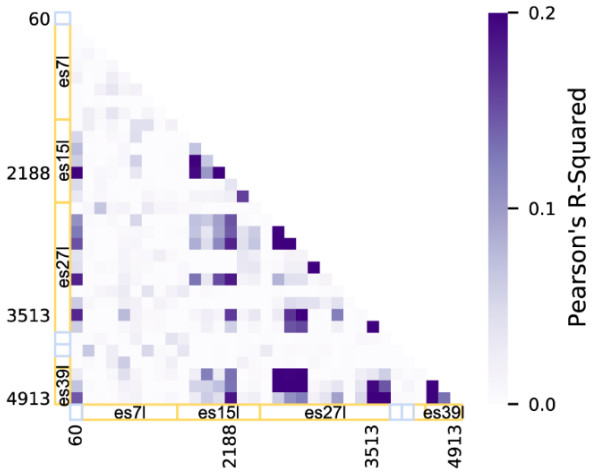

A
40S with annotated 18S atlas variants
Decoding Center
Head
es6s
Shoulder
B
60S with annotated 28S atlas variants
Exit Tunnel
es27l
es39l
P-stalk
L7/L12
stalk
es7l
Central Protuberance
es15l
C
18S atlas - nucleotide resolution
rDNA frequency (%)
rRNA frequency (%)
256 SNVs
Spearman R=0.54
Pearson R=0.46
Reference allele
Alternate allele
D
28S atlas - nucleotide resolution
rDNA frequency (%)
rRNA frequency (%)
Spearman R=0.62
Pearson R=0.93
Reference allele
Alternate allele
E
> 10% frequency rRNA variants
Reference allele
Alternate allele
rRNA frequency (%)
es6s:808
28S:h11:60
es7l:609
es7l:644
es7l:756
es7l:784
es7l:785
es7l:786
es7l:846
es7l:869
es7l:870
es7l:871
es7l:900
es7l:1040
es7l:1064
es7l:1138
es15l:2111
es15l:2173
es15l:2176
es15l:2177
es15l:2188
es15l:2195
es15l:2196
es27l:2965
es27l:3040
es27l:3048
es27l:3050
es27l:3121
es27l:3126
es27l:3159
es27l:3163
es27l:3200
es27l:3330
es27l:3491
es27l:3513
es27l:3514
28S:h68:3749
28S:h76:4073
es31l:4096
es39l:4805
es39l:4806
es39l:4913
es7l
es15l
es27l
es39l

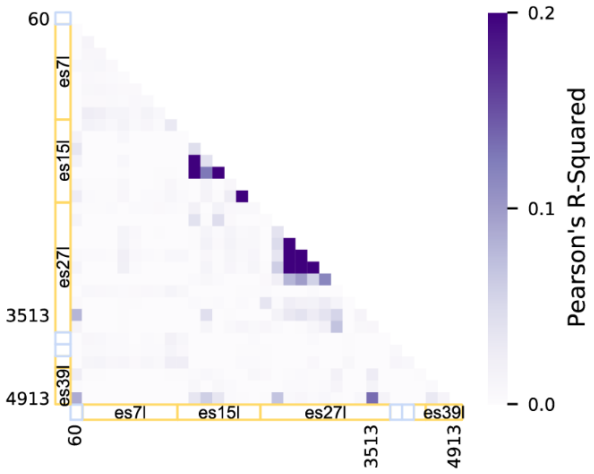
**Figure 2. 18S and 28S rRNA atlas variants and high correlation between rDNA and rRNA variant frequency.**

A. 40S 3D structure with 18S atlas SNVs. Ribosomal proteins are presented in semitransparent green, 18S rRNA in gray. SNVs variants are highlighted in red.
B. Same as (A) for the 60S and the 28S atlas SNVs.
C. Scatter plot of 18S rRNA frequency (x-axis) and rDNA frequency (y-axis) for reference and alternate alleles. Alternate alleles are marked in red and reference alleles are marked in gray. A dashed black line indicates rRNA frequency equal to 10%. Spearman and Pearson correlations for rRNA frequency and rDNA frequency between alternate alleles alone are presented (calculated on variants, red dots alone).
D. Same as (C) for the 28S.
E. Stacked bar plots of allele frequencies at positions with rRNA SNVs with frequency >10%. The nucleotide sequence matching the alleles are indicated inside the bar plots for variants with >10% allele frequency. The reference allele is indicated in gray and alternate alleles are indicated in red colors.

A



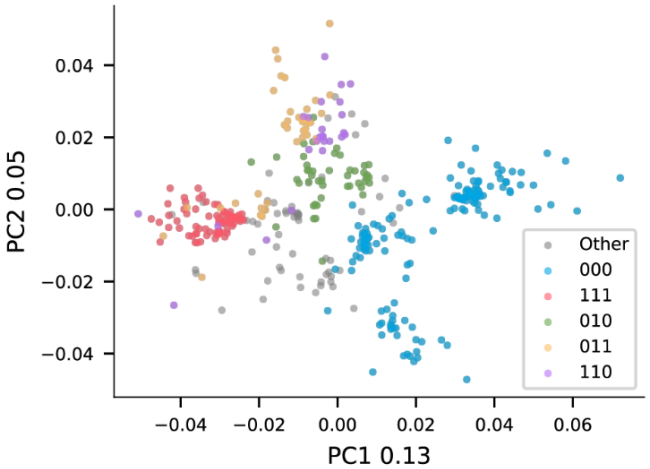

B



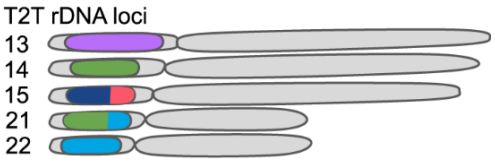

C

| 28S position | Digit | Nucleotide |
|---|---|---|
| **60** | 0 | G |
| | 1 | A |
| **3513** | 0 | A |
| | 1 | GG |
| | 2 | G |
| **4913** | 0 | C |
| | 1 | T |

D




E




| Chromosome | Haplotype | T2T copy number |
|---|---|---|
| **Chr13** | 110 | 73 |
| | 000 | 1 |
| | 100 | 1 |
| | 120 | 1 |
| **Chr14** | 010 | 14 |
| **Chr15** | 100 | 33 |
| | 111 | 16 |
| | 011 | 1 |
| **Chr21** | 010 | 39 |
| | 000 | 16 |
| | 100 | 1 |
| **Chr22** | 000 | 21 |

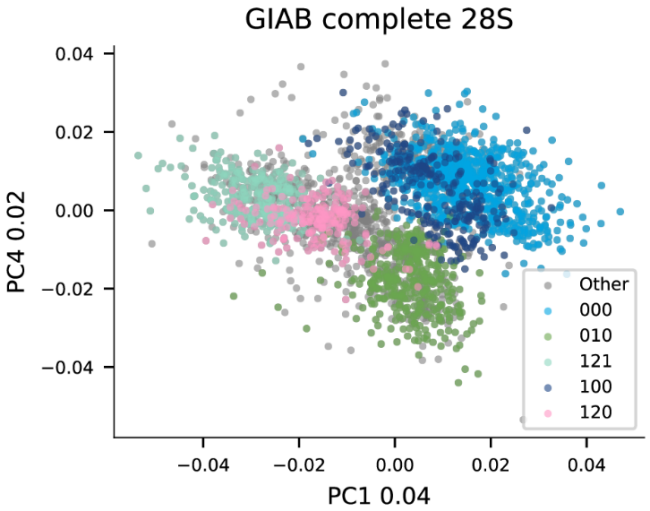


**Figure 3. 28S subtypes found by haplotype analysis**

A. Correlation coefficient (Pearson's $r^2$) heatmap between positions across H7-hESC 28S rDNA with variant frequency >10%. X-axis and Y-axis are annotated by regions. Helix regions are annotated by light blue and ES regions are annotated by yellow. Individual positions with higher $r^2$ between regions are also indicated.
B. Same as (A) for the Genome In A Bottle (GIAB) dataset.
C. Haplotype digit code to variant sequence conversion at the 3 positions with higher $r^2$ in (A) and (B)
D. Bray-Curtis Principal Coordinate Analysis (PCoA) of 386 H7-hESC 28S rDNA sequences (upper panel) and 386 28S rDNA sequences from each GIAB sample (lower panel). Each dot is a complete 28S rDNA sequence. The colors correspond to coloring an rDNA sequence by its 3 position haplotype described in (C). Numbers in the X and Y labels represent the PCoA explained variance.
E. Telomere-to-telomere haplotype distribution across the 5 acrocentric chromosomes. The rDNA acrocentric arms are presented in a schematic cartoon with proportions of rDNA haplotypes in different colors as found in the matching table below. Haplotypes match the 3 position haplotypes in (C). We indicate the rDNA copy number of each haplotype in every chromosome.

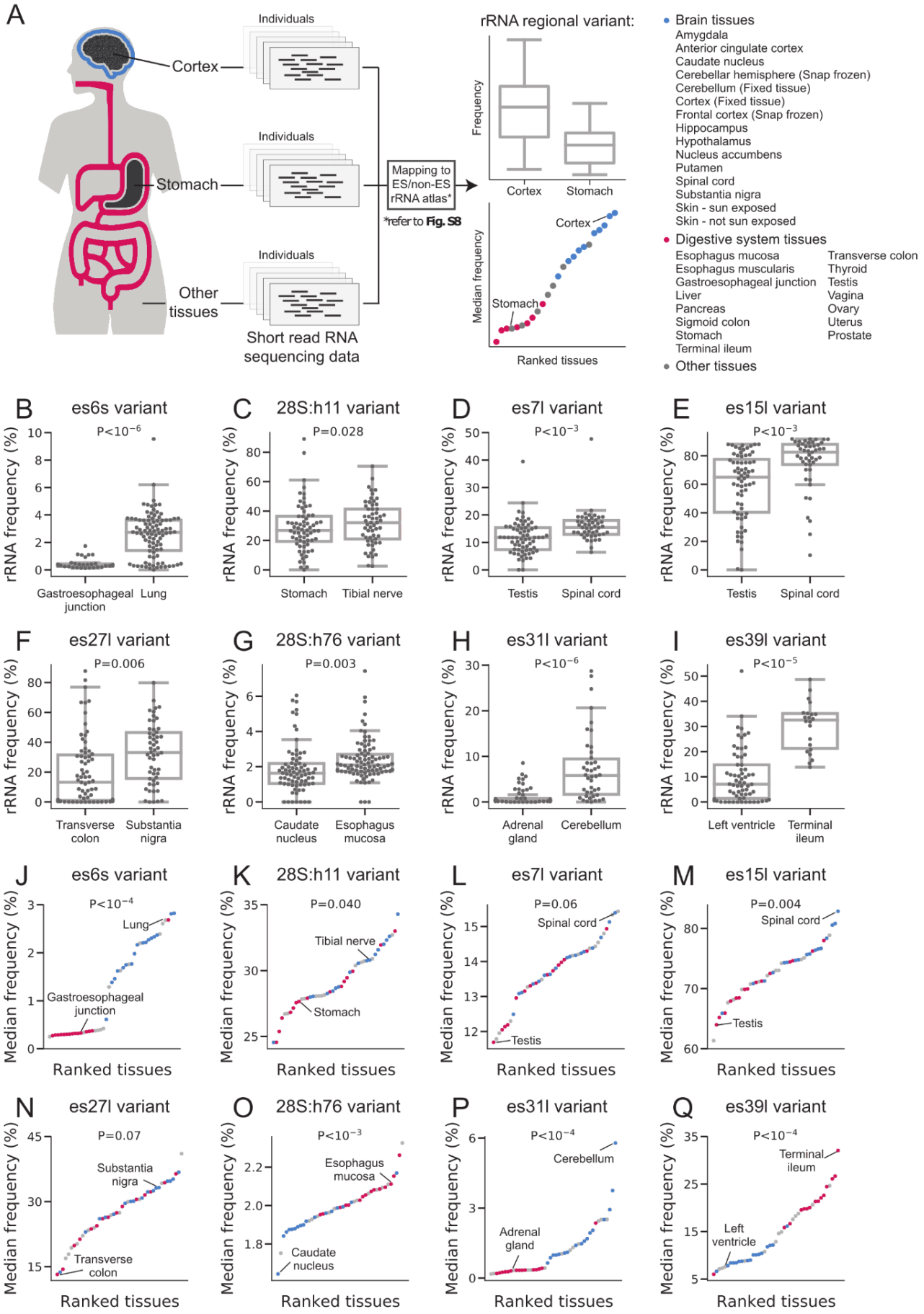

A
Cortex
Stomach
Other tissues
Individuals
Short read RNA sequencing data
Mapping to ES/non-ES rRNA atlas*
*refer to Fig. S8
rRNA regional variant:
Frequency
Cortex
Stomach
Median frequency
Ranked tissues
Brain tissues
Amygdala
Anterior cingulate cortex
Caudate nucleus
Cerebellar hemisphere (Snap frozen)
Cerebellum (Fixed tissue)
Cortex (Fixed tissue)
Frontal cortex (Snap frozen)
Hippocampus
Hypothalamus
Nucleus accumbens
Putamen
Spinal cord
Substantia nigra
Skin - sun exposed
Skin - not sun exposed
Digestive system tissues
Esophagus mucosa
Esophagus muscularis
Gastroesophageal junction
Liver
Pancreas
Sigmoid colon
Stomach
Terminal ileum
Transverse colon
Thyroid
Testis
Vagina
Ovary
Uterus
Prostate
Other tissues
B es6s variant P<10^-6 rRNA frequency (%) Gastroesophageal junction Lung
C 28S:h11 variant P=0.028 rRNA frequency (%) Stomach Tibial nerve
D es7l variant P<10^-3 rRNA frequency (%) Testis Spinal cord
E es15l variant P<10^-3 rRNA frequency (%) Testis Spinal cord
F es27l variant P=0.006 rRNA frequency (%) Transverse colon Substantia nigra
G 28S:h76 variant P=0.003 rRNA frequency (%) Caudate nucleus Esophagus mucosa
H es31l variant P<10^-6 rRNA frequency (%) Adrenal gland Cerebellum
I es39l variant P<10^-5 rRNA frequency (%) Left ventricle Terminal ileum
J es6s variant P<10^-4 Lung Gastroesophageal junction Median frequency (%) Ranked tissues
K 28S:h11 variant P=0.040 Tibial nerve Stomach Median frequency (%) Ranked tissues
L es7l variant P=0.06 Spinal cord Testis Median frequency (%) Ranked tissues
M es15l variant P=0.004 Spinal cord Testis Median frequency (%) Ranked tissues
N es27l variant P=0.07 Substantia nigra Transverse colon Median frequency (%) Ranked tissues
O 28S:h76 variant P<10^-3 Esophagus mucosa Caudate nucleus Median frequency (%) Ranked tissues
P es31l variant P<10^-4 Cerebellum Adrenal gland Median frequency (%) Ranked tissues
Q es39l variant P<10^-4 Terminal ileum Left ventricle Median frequency (%) Ranked tissues

**Figure 4. rRNA variant expression is tissue specific and differs between tissues derived from the ectoderm and endoderm lineages**

A. Schematic representation of the GTEx analysis focusing on Cortex versus Stomach comparison. We map rRNA reads from different samples to rRNA regional variants (See **Figure S8** for ES/non-ES atlas and **Methods** for usage). Per rRNA regional variant, we illustrate variant expression comparison between tissues: Upper panel with a box plot comparing rRNA regional expression in Cortex and Stomach samples. Bottom panel shows the median rRNA regional expression across all tissues. Cortex and Stomach are annotated, and all ectoderm and endoderm tissues are highlighted in blue/red marks.
B. Box plot comparing es6s rRNA regional variant, ES:es6s:12_d1_r31, expression in Lung and Gastroesophageal Junction samples.
C. Same as (B) for region 28S:h11 (pre-es5l), ES:pre_es5l:1_d119_r4743, highlighting Stomach and Tibial nerve.
D. Same as (B) for region es7l, ES:es7l:7_d1_r39, highlighting Testis and Brain Spinal cord.
E. Same as (B) for region es15l, ES:es15l:19_d1_r17, highlighting Testis and Brain Spinal cord.
F. Same as (B) for region es27l, ES:es27l:1_d5_r280, highlighting Transverse colon and Brain Substantia nigra.
G. Same as (B) for region h76 (pre-es31l), ES:pre_es31l:1_d21_r5146, highlighting Esophagus mucosa and Brain Caudate nucleus.
H. Same as (B) for region es31l, ES:es31l:2_d1_r488, highlighting Adrenal gland and Brain Cerebellum.
I. Same as (B) for region es39l, ES:es39l:11_d2_r45, highlighting Terminal ileum and Heart Left ventricle.
J. Scatter plot showing median frequency of es6s regional variant, ES:es6s:12_d1_r31, across all tissues. Tissues derived from ectoderm are marked in blue, tissues derived from endoderm are marked in red, other tissues in gray. The Lung and Gastroesophageal Junction that were shown in panel (B) are annotated with a line.
K. Same as (J) for region 28S:h11 (pre-es5l), ES:pre_es5l:1_d119_r4743, the Stomach and Tibial nerve that were shown in panel (C) are annotated with a line.
L. Same as (J) for region es7l, ES:es7l:7_d1_r39, the Testis and Brain Spinal cord that were shown in panel (D) are annotated with a line.

M. Same as (J) for region es15l, ES:es15l:19_d1_r17, the Testis and Brain Spinal cord that were shown in panel (E) are annotated with a line.

N. Same as (J) for region es27l, ES:es27l:1_d5_r280, the Transverse colon and Brain Substantia nigra that were shown in panel (F) are annotated with a line.

O. Same as (J) for region h76 (pre-es31l), ES:pre_es31l:1_d21_r5146, the Esophagus mucosa and Brain Caudate nucleus that were shown in panel (G) are annotated with a line.

P. Same as (J) for region es31l, ES:es31l:2_d1_r488, the Adrenal gland and Brain Cerebellum that were shown in panel (H) are annotated with a line.

Q. Same as (J) for region es39l, ES:es39l:11_d2_r45, the Terminal ileum and Heart Left ventricle that were shown in panel (I) are annotated with a line.

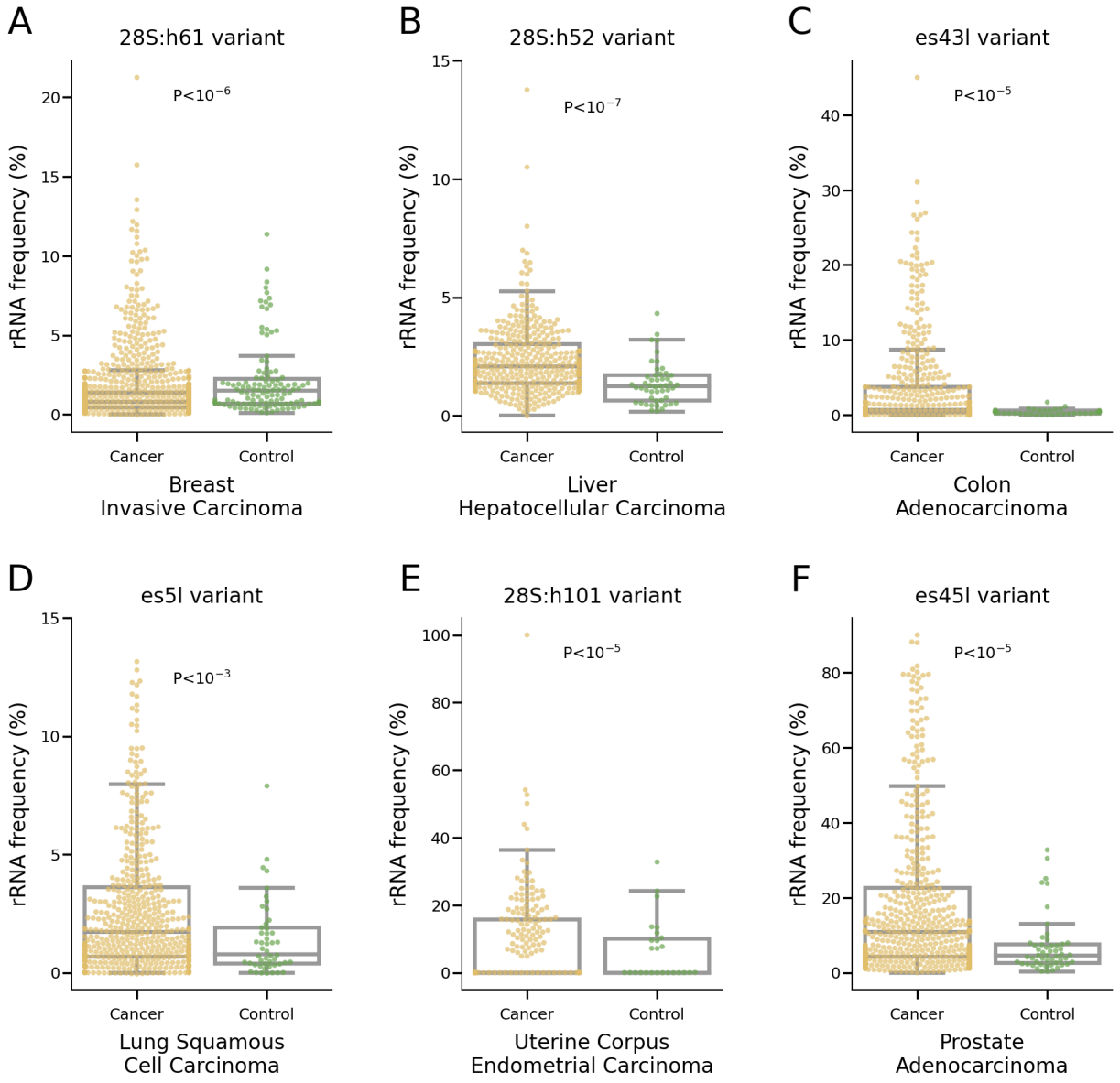



**Figure 5. Cancer-specific rRNA variant expression**

A. Box plot showing the rRNA frequency of the top expressed, alternate allele regional variant, of 28S:h61 (pre-es27l), ES:pre_es27l:11_d1_r27, across TCGA cancer and control samples for Breast Invasive Carcinoma.
B. Same as (A) for region 28S:h52 (pre-es20l), ES:pre_es20l:1_d1_r348, in Liver Hepatocellular Carcinoma.
C. Same as (A) for region es43l, ES:es43l:2_d9_r420, in Colon Adenocarcinoma.
D. Same as (A) for region es5l, ES:es5l:2_d4_r495, in Lung Squamous Cell Carcinoma.

E. Same as (A) for region h101 (post-es41l), ES:post_es41l:2_d2_r168, in Uterine Corpus Endometrial Carcinoma.

F. Same as (A) for region es45l, ES:es45l:4_d1_r82, in Prostate Adenocarcinoma.