Age-dependent mutational loads in human tRNA genes are tumor-specific and result in chimeric tRNA sequences that could disrupt the genetic code

Marina Murillo, Marina Salvadores, Aina Vaquer Picó, et al.

Genome Res. published online April 15, 2026
Access the most recent version at doi:10.1101/gr.281022.125

Age-dependent mutational loads in human tRNA genes are tumor-specific and result in chimeric tRNA sequences that could disrupt the genetic code.

Marina Murillo-Recio[1], Marina Salvadores[1,2], Aina Vaquer Picó[1], Lina Tsapanou[1], Adrián Gabriel Torres[1], Fran Supek[1,2,3*], Lluís Ribas de Pouplana[1,3*]

1. Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Barcelona, Catalonia, 08028, Spain.

2. Biotech Research & Innovation Centre (BRIC), University of Copenhagen, 2200 Copenhagen N, Denmark.

3. Catalan Institution for Research and Advanced Studies, Barcelona, Catalonia, 08010, Spain.

* To whom correspondence may be addressed.

Abstract.

Transfer RNA genes (tDNAs) are essential genomic elements that safeguard translational fidelity. Using the Telomere-to-Telomere version of the human genome we have mapped the position of human tDNAs and analyzed their individual transcriptional activities. Then we have characterized, at single base resolution, the impact of somatic mutations in human tDNAs and its relationship to the transcriptional status of each gene. We confirm that tDNAs are hotspots for somatic mutagenesis, and show that they display mutational loads that are directly proportional to their transcription rates. Highly transcribed tDNAs in tumors or healthy tissues accumulate mutations at rates up to nine-fold higher than highly transcribed protein-coding genes. Mutational loads at tDNAs are tumor-specific, and increase with patient age. Mutations at structurally conserved tRNA positions appear to be under negative selection. Anticodon nucleotides crucial for decoding frequently acquire somatic mutations, readily generating chimeric tRNA species potentially capable of systematically introducing amino acid substitutions across the proteome. Our results reveal a previously unrecognized source of somatic heterogeneity in human cancer and aging tissues that may directly impact upon translation efficiency and fidelity, and cause cell-specific proteostasis degeneration.

Keywords: tRNAs, tDNAs, somatic mutagenesis, transcription associated mutagenesis, mutational signatures, APOBEC3, mistranslation, cancer

Introduction.

The translational machinery of cells is finely tuned to maintain proteostasis, a key factor in cellular survival and proliferation (Gingold et al. 2014; Goodarzi et al. 2016; Torrent et al. 2018). Transfer RNAs (tRNAs) are central to translation, pairing their anticodons to codons in messenger RNAs and delivering the corresponding amino acids during translation. The human genome encodes over 600 tRNA-coding genes (tDNAs) (Chan and Lowe 2016), which are transcribed by RNA polymerase III (Pol III) via the recognition of two internal promoter regions that are highly conserved among all tDNAs (Fig. 1A) (Galli et al. 1981; Giege et al. 1998). All active tRNAs share a common three-dimensional structure required for their function in the ribosome, but each tRNA must present idiosyncratic recognition motifs essential to preserve the faithful translation of the genetic code (Fig. 1A) (Giege et al. 1998; Berg and Brandl 2021), which takes place when tRNAs are aminoacylated at their 3' end by aminoacyl-tRNA synthetases (ARS) (Giege et al. 1998; Ibba and Soll 2000). The identity elements used for accurate aminoacylation are idiosyncratic to each tRNA-ARS cognate pair and, in most cases, include the bases at the anticodon (Giege et al. 1998; Giege and Eriani 2023). However, some ARS do not interact with the anticodon sequence of their cognate tRNAs, and mutations at these anticodons can lead to chimeric tRNAs that cause amino acid substitutions throughout the proteome (Geslain et al. 2010; Parisien et al. 2013; Reverendo et al. 2014; Santos et al. 2018; Pinzaru and Tavazoie 2023).

Alterations in tRNA populations are linked to numerous human diseases (Orellana et al. 2022). For example, increased tRNA abundance is a feature of cancer cells, which adapt their tRNA populations to optimize the translation of specific genetic programs (Goodarzi et al. 2016; Zhang et al. 2018; Gupta et al. 2022; Pinzaru and Tavazoie 2023). Quantitative and qualitative alterations in tRNA populations -including tRNA-derived fragments (tRFs)- are widespread in human cancers (Huang et al. 2018; Cabrelle et al. 2024), where they have been linked to aberrant signaling and translational reprogramming in several cancer types, including lung and breast cancer (Skeparnias et al. 2020;

Kwon et al. 2021). Specific tRNA isoacceptors and post-transcriptional base modifications regulate cancer cell survival and influence metastatic potential by controlling the translation of proliferation-related genes (Earnest-Noble et al. 2022; Garcia-Vilchez et al. 2023). Moreover, metabolic and therapeutic stresses such as chemotherapy induce codon-biased aberrant protein production through altered tRNA function and decoding. Thus, variations in tRNA populations reshape the cancer proteome and support tumor adaptation and therapy resistance (Kochavi et al. 2024; Wernaart et al. 2024; Yang et al. 2024).

Somatic mutagenesis causes changes in the DNA sequence of somatic cells, drives the emergence of cancers, and is an important contributor to the aging process (Vijg and Dong 2020). The nature and dynamics of human somatic mutagenesis has been studied mostly in protein-coding genes, and its impact upon non-coding RNA genes is less understood. The activity of members of the APOBEC3 subfamily of cytidine deaminase enzymes is an important contributor to somatic mutagenesis (Taylor et al. 2013; Mas-Ponte and Supek 2020; Langenbucher et al. 2021; Sanchez et al. 2024). APOBEC3 enzymes function, primarily, as an innate immunity mechanism to defend against viruses and mobile genetic elements but, in some cancers, APOBEC3A and APOBEC3B paralogs can generate a high mutational burden (Burns et al. 2013; Roberts et al. 2013; Nik-Zainal et al. 2014; Supek and Lehner 2017; Petljak et al. 2022). APOBECs can access only single-stranded nucleic acids, and are known to act upon single-stranded DNA (ssDNA) intermediates during DNA repair, as well as ssDNA segments within structured, stem-loop DNA (Roberts et al. 2013; Buisson et al. 2019; Mas-Ponte and Supek 2020).

Studies in bacteria and yeast have reported that non-protein-coding genes can accumulate mutations also caused by the activity of APOBEC3 enzymes. In an *E. coli* model, mutational impact of heterologous expressed APOBEC3B was highest in genes transcribed by Pol III, such as tDNAs (Sakhtemani et al. 2019). Mutational studies in a yeast model also suggested that human APOBEC3B

overexpression causes severe DNA damage in Pol III-transcribed genes (Saini et al. 2017), suggesting a transcription-related mechanism. In human germline cells, mutations in nuclear tDNAs are subject to strong purifying selection (Thornlow et al. 2018; Seplyarskiy et al. 2023), but analysis of tRNA sequences obtained from the plasma of healthy human volunteers revealed significant levels of human heterogeneity (Parisien et al. 2013; Berg et al. 2019). Human tDNAs are known to accumulate somatic mutations at high rates, and APOBEC has been identified as a major contributor to this phenomenon (Sakhtemani et al. 2019). However, a detailed analysis of tDNA somatic mutations, including their distribution at single-base resolution and their potential physiological impact in cancer and aging is lacking.

Here we have mapped all tDNAs in the Telomere-to-Telomere (T2T-CHM13) version of the human genome (Fig. 1B), and used this information to study somatic mutagenesis at these loci (Hoyt et al. 2022; Nurk et al. 2022). Combining our data with POLR3D ChIP-seq and 'biotin-capture of nascent RNA' data, we investigate the relationship between transcriptional activity at tDNAs and their mutational loads. We also characterize in detail the tissue-specific distribution of somatic mutations affecting human tDNAs, their mutational signatures, and their dependence on ssDNA secondary structures. We study the tissue variability of tDNA mutational loads, and whether somatic tDNA mutations accumulate with age. We analyze the mutational loads of human tDNAs at single-base resolution, and determine base-specific mutational frequencies and their relationship to the structural and functional roles of each base. Finally, we focus on mutations at the anticodon triplets of tRNAs, and explore the possibility that somatic mutations at these sites may generate chimeric tRNAs.

Results.

Genome-wide identification of tDNAs and their transcriptional activity.

To obtain an accurate picture of the set of human tDNAs we first mapped the positions of these genes in the most recent version of the human genome assembly Jan. 2022 (T2T-CHM13 v2.0/hs1) (Hoyt et al. 2022) (Fig. 1B). Our analysis sets the total number of identified tDNAs to 733, of which 521 tDNAs are confidently predicted as functional tRNAs (~112 more genes than previously reported with older assemblies) (Fig. 1C). We successfully localize all known tDNAs and remove previous uncertainties regarding the localization of some of these genes (Fig. 1C). As previously reported, tDNAs can be found as isolated genes, or as part of large gene clusters that are present in several chromosomes (Bermudez-Santana et al. 2010; Van Bortle et al. 2017) (Fig. 1B). To better characterize tDNAs, we calculated their genomic distances (Supplemental Fig. S1A) and grouped them into clusters based on thresholds of 20 kb or 1 kb (Supplemental Fig. S1B and Supplemental Table S1). The size of the resulting clusters varies significantly: 1 kb clusters typically contain 25 tDNA genes, whereas 20 kb clusters can range from 2 to 112 tDNAs (Supplemental Fig. S1C). The distribution and contents of clusters of tDNAs are strongly conserved among primates (Supplemental Methods), with a tendency to larger tDNA clusters in *Homo sapiens* (Bermudez-Santana et al. 2010) (Supplemental Fig. S1D).

tDNA transcriptional activity can be estimated by combining POLR3D ChIP-seq data with biotin-capture of nascent RNA (Supplemental Table S2) (Van Bortle et al. 2017). Through this approach, we discovered that while both isolated and clustered tRNAs can be actively transcribed, tDNAs in clusters exhibit higher activity levels than isolated genes (see Fig. 1D) (tDNAs with higher levels of expression are located in clusters). This conclusion holds true even when sets of tRNA isoacceptors are analyzed individually (Supplemental Fig. S1E).

tDNAs are hotspots of somatic mutagenesis.

To quantify the extent of somatic mutagenesis tDNAs we used whole-genome sequencing data (WGS) from 9596 samples of cancerous tissues and 1192 samples of clonal healthy tissues. Variant calling data was previously obtained using the GRCh37 genome, to confirm tRNA variation we used liftOver to convert T2T-CHM13 tDNAs coordinates to GRCh37, and we were able to localize 537 tDNAs (Supplemental Table S3). In contrast to previous studies in germinal cells (Thornlow et al. 2018), we confirm that tDNAs in somatic cells are mutational hotspots in tumor samples (Fig. 2A). The internal mutational burden of tDNAs correlates directly with their transcriptional activity (Fig. 2B). Silent tDNA genes show no mutation accumulation, while highly transcribed tDNA genes accumulate internal sequence mutations at rates up to nine times higher than in protein-coding genes (Fig. 2C). This difference suggests a direct link between Pol III activity and the mutation accumulation in tDNAs. While tDNA transcription is driven by Pol III, protein-coding genes are transcribed by RNA polymerase II (Pol II), which recruits DNA repair mechanisms such as transcription-coupled nucleotide excision repair (TCR) (Guo et al. 2022). In comparison, Pol III-transcription lacks known coupled-repair mechanisms to deal with the Transcription-Associated Mutagenesis (TAM) phenomena (Jinks-Robertson and Bhagwat 2014; Thornlow et al. 2018). We asked whether the observed tDNA somatic mutation load is also detectable in other Pol III-transcribed genes such as ribosomal RNA (rRNA), small nuclear RNA (snRNA) and unclassified RNAs (miscRNA), as was reported in the human germline mutations (Seplyarskiy et al. 2023). Mutational densities at 5S rRNAs, snRNAs, and miscRNAs genes are significantly lower than at tDNAs and display different mutational profiles, suggesting that the mechanisms leading to somatic mutations in tDNAs are gene-type specific (Supplemental Fig. S2). We quantified the tDNA mutational burden in clonal samples of healthy tissues (including brain, colon, liver, and lung) using whole-genome sequencing (WGS) data, and again found a significant accumulation of somatic mutations in tDNAs (Fig. 3B). Thus, somatic tDNA mutations are also

prevalent in healthy tissues, where their time-wise accumulation might contribute to the proteostasis defects characteristic of aging individuals.

Then, using our tDNA clustering data, we compared the mutation densities between tDNAs situated within clusters to those that are isolated. For clusters defined within a 20 kb distance, mutational densities were similar between clustered and isolated tDNAs, showing comparable increases in mutation rates relative to the flanking regions (two-sided Wilcoxon rank-sum test, $p = 0.051$) (Supplemental Fig. S3A). However, when considering a 1 kb distance, tDNAs within clusters had higher mutation density compared to isolated tDNAs (one-sided, alternative greater, Wilcoxon rank-sum test, $p < 2.2 \times 10^{-16}$) (Supplemental Fig. S3B). As we previously described, tDNAs that are proximal to each other, such as those in 1 kb clusters, exhibit higher transcriptional activity, which increase their susceptibility to transcription-associated mutagenesis (TAM), consistent with our previous analyses linking tDNA mutagenesis to TAM (Saini et al. 2017; Thornlow et al. 2018; Seplyarskiy et al. 2023). Although we observed that tDNAs in 20 kb clusters also tend to have higher expression levels than isolated tDNAs (Supplemental Fig. S3A), we did not detect a corresponding increase in tDNA mutagenesis (Supplemental Fig. S3A). These larger clusters include more genes, some of which have low expression values, diluting the overall mutation density when all tDNAs within the cluster are analyzed collectively. Overall, these results show that tRNAs acquire most of their mutations during transcription.

Mutational signatures and role of APOBEC3 in tDNA Mutagenesis.

The nature and frequency of somatic mutations generate a signature that can be used to identify the agents responsible for the observed mutations, based on the frequencies of trinucleotide contexts of the mutations. Over a hundred such mutational signatures exist in human samples, and mutational agents are identified for approximately half of them (Alexandrov et al. 2020; Degasperi et al. 2022). APOBEC mutations generate a clear signature of C>T and C>G changes restricted to TCW contexts

(where W is T or A). Mutational signature analyses performed specifically at tDNA loci detected two main mutational signatures, SBS13 and SBS2 (Fig. 2D) (Alexandrov et al. 2020), both identifying APOBEC3-driven mutagenesis as the major contributor to the observed tDNA mutations. The APOBEC3 signature in tDNAs was also present in genomes of healthy cells (Fig. 2E and Supplemental Fig. S5). While genome-wide signatures of APOBEC3 mutagenesis are rare in healthy somatic cells, in contrast to cancer genomes (Franco et al. 2019), we do observe this mutagenesis signature in tDNAs both in tumors and healthy tissues. Therefore, high transcription rates and/or other factors contributing to ssDNA occurrence must act at tRNA loci to facilitate APOBEC mutagenesis therein. Other genes transcribed by Pol III display different mutational signatures, indicating that APOBEC3-driven mutagenesis at tDNA genes is not driven only by high levels of transcription, but likely requires additional factors specific to tDNAs. We suspected secondary-structure prone tDNA sequences (Fig. 2F), and we used hairpin-localization algorithms to determine the presence of APOBEC3-preferred mutational sites in tDNAs (Roberts et al. 2013; Buisson et al. 2019; Sui et al. 2020; McCann et al. 2023). This analysis revealed a significant enrichment of DNA hairpin formations in tDNAs compared to the rest of the genome (Chi-squared test $p = 0.01578$) or to other Pol III-transcribed genes (Chi-squared test $p = 0.03653$).

We find that the mutational profiles of tDNAs in healthy tissues and tumors are not identical (Fig. 2E). Some mutational signatures identified in human tumors were in fact negatively correlated with tDNA transcriptional intensity. This suggests a protective effect of active, open chromatin at highly-transcribed tDNAs, possibly through recruiting DNA repair (Polak et al. 2014). Based on previously reported experiments, base excision DNA repair may be enriched in this situation (Saini et al. 2017). For example, signature F (likely corresponding to COSMIC catalog SBS17b), which is associated to oxidative damage to the free nucleotide pool and also to 5-fluorouracil treatment (Alexandrov et al.

2020), shows decreased incidence in highly-transcribed tDNA genes, where its exposure levels are 8.75 times lower than in tDNAs with low transcription rates (Fig. 2E).

On the other hand, mutational signatures that are prevalent in healthy tissues but not detected in tumor genomes resemble a known signature, SBS10, associated to ultramutation associated to defects of the proofreading domain of the DNA polymerase epsilon (PolE). Given that PolE-associated ultramutagenesis is a known feature of cancers, and expected to be largely absent in healthy cells, our data suggests that this signature originates from another, currently unclear cause. Speculatively, factors such as transcription-replication conflicts, or usage of replication origins which can be collocated with tDNAs (noted in yeast), could be altered in cancer, resulting in a tumor-specific mutational process at tDNA.

Tumor type- and age-dependence of tDNA somatic mutagenic rates.

The overexpression of tDNAs in human cancers has been repeatedly reported (Goodarzi et al. 2016; Zhang et al. 2018; Gupta et al. 2022; Orellana et al. 2022; Pinzaru and Tavazoie 2023), and linked to adaptive mechanisms of codon-anticodon optimization that prioritize the translation of genetic programs important for tumor growth (Goodarzi et al. 2016). Moreover, tDNA expression can be cell type- or tissue-specific, often resulting in differential expression of not only mature tRNAs but also of tRNA fragments (Dittmar et al. 2006; Torres et al. 2019). To test whether somatic tDNA mutagenesis is a distinguishing feature of different human tissues we compared tDNA somatic mutagenesis by cancer type (Supplemental Methods). This analysis revealed that tDNA mutational load varies by cancer type and tissue, with bladder cancer (BLCA) exhibiting the highest rates of tDNA mutagenesis (Fig. 3A). In agreement with our previous finding that high levels of transcription correlate with mutational frequencies at tDNAs, cancer types characterized by upregulated levels of tRNA expression (including bladder, uterine corpus endometrial carcinoma and breast cancer)(Zhang et al. 2018), also exhibit the highest mutational burdens in tDNAs.

Next, we investigated whether mutational loads in tDNAs accumulate with age by comparing samples from cancer patients across different age groups (Supplemental Methods). Our results demonstrate a significant age-related increase in tDNA somatic mutational load in various cancer types, particularly in tumors that display high levels of tDNA transcription. These results suggest that the accumulation of somatic mutations in tDNAs is a characteristic of aging human tissues (Fig. 3C and Supplemental Fig. S7). tDNAs from donors below thirty years of age show no signs of mutational load, while these become conspicuous in the forty-to-sixty age range, and increase further in older patients. Available data from healthy human tissues is insufficient to determine the age-related distribution of somatic mutations at tDNAs, but the behavior observed in tumors, and the fact that healthy tissues do reveal high levels of somatic mutations at tDNAs, indicate that these mutations do accumulate in all cells as humans age.

Distribution of somatic mutations in tDNAs and evidence for negative selection.

While genomic signatures of negative selection in somatic cells are quite subtle, some protein-coding genes, including essential genes and oncogenes, display lower mutation rates than expected (Van den Eynden et al. 2016; Martincorena et al. 2017; Besedina and Supek 2024). We reasoned that the increased rate of somatic mutagenesis at tDNA loci, coupled with the fact that hundreds of tDNA gene copies can be easily aligned to compare variation at individual nucleotides, might provide good statistical power to search for evidence of negative selection acting against somatic mutations at critical positions of the tRNA molecule.

Thus, we quantified tDNA mutational rates at single base resolution for all human tDNAs to search for negative selection acting upon mutations at specific sites. This analysis revealed that most universally conserved positions in tRNA sequences (Beuning and Musier-Forsyth 1999) display mutation rates significantly lower than the rest of nucleotides in tRNAs (Fig. 4A, Fig. 4B), indicating that these positions experience negative selection in somatic cells that eliminates mutations at these

sites. This analysis revealed other positions that, while not universally conserved, also display lower than expected mutagenic rates, suggesting that these nucleotides play important, but as yet unidentified, functional roles in tRNAs (Fig. 4A and Supplemental Fig. S8). Possibly, mutations at these positions generate aberrant tRNA species that directly interfere with the translation machinery and significantly affect cell fitness.

Positions critical for translation elongation such as the anticodon triplet, where mutations can severely disrupt the efficiency and fidelity of protein synthesis (Geslain et al. 2010; Reverendo et al. 2014; Santos et al. 2018; Vijg and Dong 2020; Pinzaru and Tavazoie 2023), showed no evidence of decreased mutation rates (Fig. 4A and Fig. 4B), and we identified several instances of mutations at anticodon positions (see below).

In order to analyze the regional mutation load in structural domains of the tRNA, or at the two internal promoters of tDNAs, we used a rolling windows method to calculate the relative mutation load in ten-nucleotide sections of these molecules. This revealed that tDNA genes experience higher mutational loads at their 3' halves (Supplemental Fig. S9). The second promotor region (B-box) is where most mutations accumulate (Fig. 4A and Fig. 4B). This mutational distribution might reflect a locally protective effect of the Pol III complex of the transcription factor TFIIIB over a section of tDNA positive strands during their transcription. Alternatively, the presence of sequences or motifs required for the activity of APOBEC in the 3' half of tDNAs could be responsible for the relative accumulation of mutations in these regions of tDNAs. Consistent with this second possibility, we find that position 56, within the region of the B box, is the highest mutated position in all tDNAs and is occupied by a highly conserved cytosine in all human tRNAs (Fig. 4A).

Somatic mutations at tDNAs generate misincorporating tRNAs.

The attachment of amino acids to their cognate tRNA molecules by aminoacyl-tRNA synthetases (ARS) is a crucial step for accurate protein translation. For the majority of ARS, tRNA recognition requires interactions with both the acceptor stem and anticodon loop motifs (Hou and Schimmel 1988; Park and Schimmel 1988; Perret et al. 1990; Giege et al. 1998; Rubio Gomez and Ibba 2020; Giege and Eriani 2023). However, ARS cognate for amino acids alanine (Ala), leucine (Leu), serine (Ser), and tyrosine (Tyr) recognize only the acceptor stem of their cognate tRNA substrates (Giege and Eriani 2023). In these cases, somatic mutations at tRNA anticodons can lead to chimeric tRNAs whose anticodon triplet will recognize codons that do not correspond to the amino acid carried by the chimeric tRNAs. Such molecules lead to translation errors throughout the proteome, caused by the misincorporation of Ala, Ser, Leu, or Tyr at non-cognate codons within the ribosome. In the datasets used in this study we could identify numerous examples of somatic mutations leading to the generation of chimeric tRNAs previously shown to induce proteome-wide amino acid substitutions (Fig. 4C) (Geslain et al. 2010; Santos et al. 2018; Pinzaru and Tavazoie 2023). For example, we detected several examples of mutations at position 35 of $tRNA^{Ser}$(UCA) that result in leucine to serine substitutions throughout the proteome (Fig. 4C). The detection of mutations that compromise the fidelity of the genetic code at intensely transcribed tDNAs suggests that aging human cells produce mutagenic tRNAs capable of introducing widespread amino acid changes throughout the proteome, and contribute to widespread proteome mosaicism in tissues.

Discussion.

Proteome variability can have both negative and positive impacts upon cellular homeostasis. In microbes, for example, several adaptive mechanisms based on the induction of widespread proteome variation have been described, several of which are based on the use of mistranslating tRNAs (Ribas de Pouplana et al. 2014). In animals, adaptive proteome variations aided by changes in cellular tRNA populations are well known, and always linked to adaptive genetic switches that are important for both healthy tissues and tumor development (Pavon-Eternod et al. 2013; Gingold et al. 2014; Earnest-Noble et al. 2022). Adaptive mistranslation (the deliberate induction of translation errors during protein synthesis) has not been reported in animals, and it is assumed that such errors will always be deleterious to human health.

Frailty, sarcopenia, and many forms of neurodegeneration common during the aging process are linked to losses in proteome quality (Anisimova et al. 2018). Although such insults can ultimately be linked to ineffective and/or inaccurate gene translation, functional defects in the translation machinery are not generally considered significant contributors to aging (Anisimova et al. 2018). Similarly, human neoplasies that develop resistance to existing chemotherapies often do so through point mutations that thwart drug-recognition sites , but the possibility that such modifications arise through mistranslation events has not been explored (Roszkowska 2024).

We have studied in detail the dynamics of somatic mutagenesis in human tDNAs, a phenomenon previously observed in several model organisms and in human datasets (Saini et al. 2017; Sakhtemani et al. 2019; Sui et al. 2020; Seplyarskiy et al. 2023). Our data shows that human tDNAs, but not other genes transcribed by Pol III, are hotspots of somatic mutagenesis that accumulate genetic changes directly as a function of their transcriptional rate, irrespective of their genomic localization (Supplemental Fig. S3). The fact that other Pol III-transcribed genes behave differently in terms of somatic mutagenesis indicates that some tDNA-specific factors drive these mutational events.

In agreement with previous studies, our mutational signature analysis indicates that members of the APOBEC3 family are the main contributors to this mutational load. Our topological analyses, together with published interactome data (Jang et al. 2024), indicate that the recognition of sequence motifs within tDNAs, coupled to interactions between APOBEC3 enzymes and tRNA modification enzymes, may be driving the specific mutational process at tDNAs. Given that tRNA transcription and processing occur simultaneously, the specificity of somatic mutations at tDNA loci may be the result of specific interactions of APOBEC3 enzymes with components of the tRNA processing machinery in the vicinity of positive strands of tDNAs during transcription.

In tumors, but not in healthy cells, we find mutational signatures linked to chemotherapeutic agents used in cancer treatment. These mutations display a clear anti-correlation with tDNA transcription levels. This could be due to the loss of tumor cells that incorporate such mutations in highly transcribed genes, and might indicate that chemotherapy induces a shift in the pattern of tDNA expression during tumor growth (Fig. 2E). Tumor samples display a similar behavior with regards to mutations caused by defects in PolE, but these mutations are again absent in healthy cells (Fig. 2E).

Our analysis of mutational load in tDNAs at single-nucleotide resolution shows that universally conserved positions in tRNAs exhibit markedly reduced heterogeneity, suggesting strong negative (purifying) selection at these sites in somatic cells. Surprisingly, however, anticodon positions do not show reduced mutational load; instead, they accumulate mutations at levels comparable to the average across the entire tDNA sequence (Fig. 4A, Fig. 4B). This observation leads to the striking conclusion that somatic mutagenesis in actively transcribed tDNAs can generate chimeric tRNAs carrying anticodon mutations that disrupt the genetic code (Fig. 4C). Importantly, we find no instances of chimeric tDNAs with co-existing mutations in their internal promotor regions, supporting the possibility that these tDNAs are indeed expressed and generate chimeric tRNAs (Supplemental Fig. S10). We have not formally demonstrated here that the chimeric tRNA sequences that we reveal

306 are transcribed and active in the tumors of patients. However, we have previously shown that two of
307 the chimeric tDNAs identified in the present study can be readily transcribed in human cells, where
308 they give rise to chimeric $tRNA^{Ser}$(AUA) (Tyr → Ser) and chimeric $tRNA^{Ser}$(GUU) (Asn → Ser). These
309 chimeric tRNAs induce widespread proteome alterations and activate proteostatic stress responses
310 (Geslain et al. 2010). In view of these findings, it stands to reason that age-related accumulation of
311 mutations in tDNAs is likely to increase the abundance of chimeric tRNAs, and affect the quality of
312 proteomes in aging cells and tissues.

313 It should be noted that variations in tRNA sequences among human populations are common place,
314 and several of the positions where heterogeneity is detected coincide with those identified in our
315 study of somatic mutations (Parisien et al. 2013; Berg et al. 2019). This fact is likely reflecting the
316 presence of mutationally permissive positions in the tRNA structure, and possibly a coincidence
317 between the mutational mechanisms that introduce this changes in both somatic and germinal cells.
318 Indeed, germinal mutations at human tDNA variants capable of causing mistranslation have been
319 recently described, albeit at extremely low frequencies and often associated to genes with low
320 prediction scores for functional tRNAs (Tennakoon et al. 2025).

321 Our analysis of tDNA somatic mutations in different human tumors shows tumor-specific differences
322 in the levels of mutation accumulation. Bladder cancer (BLCA), which displays the highest
323 accumulation of such mutations, is characterized by highly augmented levels of APOBEC3A activity,
324 and a several APOBEC-associated driver hotspot mutations have been identified in this malignancy
325 (Shi et al. 2020; Lindskrog et al. 2021). This suggests that the high levels of tDNA mutations in BLCA
326 are the consequence of the high levels of APOBEC activity in these tumors, and opens the possibility
327 of a functional relationship between mutated tRNAs and BLCA development. In contrast, brain
328 cancer lays at the opposite spectrum in terms of tDNA somatic mutations, which are almost absent
329 in the datasets from these tumors.

Whether tDNA mutations play a general role in carcinogenesis is an important question. This could happen, first, at the level of proteome variations that might compromise the activity of tumor suppressors or generate proteins that favor tumor growth, such as oncoproteins produced by translational variations in existing transcription factors. Secondly, mutant tRNAs with defective sequences or aberrant post-transcriptional modification patterns could alter the population of tRNA fragments (tRFs) in cells. It has been described that a dysregulation in the levels of specific tRFs is related with the aggressiveness of BLCA tumors (Papadimitriou et al. 2020). Finally, proteome variability induced by mutagenic tRNAs may provide tumoral cells with fitness advantages such as avoidance of immune recognition. Although such possibility has not been explored in human cells, in *Candida albicans* proteome variability induced by mistranslating tRNAs induces morphology changes that shield this organism against the immune response of its human host (Bezerra et al. 2013).

The extent to which tDNA somatic mutations accumulate in healthy tissues, their mutational signatures, and their physiological impact remain to be determined. However, the accumulation of defective tRNAs can be linked to the many functional connections between tRNA dynamics and cancer (Zhang et al. 2018; Gupta et al. 2022). The age-dependent accumulation of mutations at tDNAs is apparent from the analysis of cancer datasets stratified by patient age, a factor that may directly impact upon the proteostasis of aging tissues (Schmidt and Schimmel 1993; Anisimova et al. 2018; Lopez-Otin et al. 2023). Inactive tRNAs in aging cells may induce an attenuation of protein synthesis efficiency and fidelity, either through their interference with components of the translation machinery, and mistranslation caused by chimeric tRNAs could lead to more acute problems such as protein aggregation, stress-linked inflammation, and the generation of aberrant antigenic peptides.

Further studies are required to determine the actual impact of somatic mutations in tDNAs upon human aging. However, the realization that APOBEC activity is the main driver of this process likely represents a new example of antagonistic pleiotropy, where a biological activity that contributes to

354 individual and population fitness in young individuals (antiviral defense by APOBEC) is responsible for
355 physiologically-adverse events in individuals past their reproductive age (somatic mutations in tDNAs)
356 (Gems and Kern 2024). Pharmacological inhibition of APOBEC activity in mature humans could
357 represent a strategy to reduce the contribution of tDNA mutations to aging.

358

359 **Methods.**

360 *tDNA prediction.*

361 Reference human genomes were obtained from the UCSC Genome Browser. The assemblies analyzed
362 included Jan. 2022 (T2T-CHM13 v2.0/hs1), Dec. 2013 (GRCh38), and Feb. 2009 (GRCh37). tDNAs for
363 each assembly were predicted and annotated using tRNAscan-SE 2.0 (v2.0.9, July 2021) (Chan et al.
364 2021), with parameters set to search for eukaryotic tRNAs and to display both primary and secondary
365 structure components in the covariance model bit scores to distinguish functional tRNAs from
366 potential pseudogenes. For the GRCh38 and GRCh37 assemblies, tDNAs that map to unplaced
367 sequences (DNA fragments associated with a specific chromosome but whose order or orientation
368 cannot be determined) or unassigned sequences (DNA fragments not assigned to any chromosome)
369 were identified and designated as delocalized tDNAs. Subsequently, the rest of the predicted tDNAs
370 were classified into several categories (Supplemental Methods): 'high confidence', pseudogenes and
371 'uncertain function'. The identified tDNAs in the T2T-CHM13 assembly, determined using tRNAscan-
372 SE, are available in Supplemental Table 1.

373 *tDNAs genomic distribution and clusters definition.*

374 tDNAs annotations from the most updated human genome (T2T-CHM13 v2.0/hs1) were used to
375 analyze tDNA distribution. In order to visualize the genomic coordinates of tDNAs we used the
376 RIdeogram v0.2.2 package from R. Genomic distances between consecutive tDNAs were computed

and used to plot their cumulative distribution using ecdf (empirical cumulative distribution function) (Supplemental Fig. S1A). Taking into account the profile obtained from ecdf and previous definitions of tDNA clusters (Bermudez-Santana et al. 2010; Van Bortle et al. 2017), tDNA clusters were defined as groups of consecutive tDNAs separated by a specific genomic distance of either 1 kb ("1 kb cluster") or 20 kb ("20 kb cluster"). All other tDNAs were categorized as isolated (Supplemental Fig. S1B and Supplemental Table S1). The correlation between POLR3D and biotin-capture was assessed using the Spearman's correlation coefficient (r). Statistical differences in $\log_2$ integrated tDNA expression levels between clustered and isolated tDNAs were evaluated using the Wilcoxon test.

tDNAs expression data.

Sequencing data for RNA polymerase III occupancy (POLR3D ChIP-seq) and nascent transcription profiling (biotin-capture) assays were obtained from the Gene Expression Omnibus (GEO) database and are accessible through GEO Series accession number GSE96800 (Van Bortle et al. 2017). Specifically, the subset of samples used for this study came from the monocyte dataset and correspond to the following GEO Sample accession numbers: POLR3D (GSM2544232, GSM2544233) and biotin-capture (GSM2544240, GSM2544241). Following the methodology outlined by Van Bortle et al. (2017), and after preprocessing the sequencing data (Supplemental Methods), integrated tDNA expression values were obtained by averaging POLR3D and biotin-capture results (Expression data available in Supplemental Table S2). tDNAs were categorized into four expression levels based on quartiles of integrated tDNA expression, with level 4 representing the highest expression values. This categorization was used to study the relationship between tDNA transcription rates and mutagenesis.

Mutation density.

Somatic single nucleotide variants (SNVs) were identified from publicly available WGS datasets spanning more than 20 cancer types (Supplemental Methods). The analysis was performed using the

GRCh37 reference genome. This procedure was used to analyze mutational density in cancerous samples for tDNAs, protein-coding genes, and other Pol III-transcribed genes (rRNA, snRNA, and miscRNA) (Supplemental Methods), as well as in non-cancerous samples for tDNAs (Supplemental Methods). First, to ensure data quality, a set of quality filters was applied to the genomic data. To minimize errors due to misalignment of short reads, and select uniquely mappable regions thus reducing potential mapping ambiguities, all regions in the genome defined in the 'CRG Alignability 75' track (Derrien et al. 2012) with alignability < 1.0 were masked out. In addition, regions that are unstable when converting between GRCh37 and GRCh38 (Ormond et al. 2021) were removed, as well as the ENCODE blacklist of problematic regions of the genome (Amemiya et al. 2019). This ensures that the results of our mutational analysis should be consistent in the GRCh37 and GRCh38 versions of the human genome.

Somatic SNVs were analyzed for different gene types: tDNAs, protein-coding genes, and other Pol III-transcribed genes, along with their flanking regions (e.g., 10 kb upstream and downstream, divided into 100-nt windows, with the gene itself designated as 'window 0'). Only windows > 20 nucleotides that passed the specified quality filters were included for the analysis. Only genes where window 0 fulfilled this threshold and with at least 10 mutations across all windows were included in the analysis. To normalize mutational densities by trinucleotide context (tri-nucleotide composition), we adjusted mutation counts based on the local sequence composition within each window. This involved calculating the frequency of each trinucleotide context's occurrence and determining the mutation count for that context within each sample. Finally, we obtained the total mutation density for each window by adding the mutational densities across all trinucleotide contexts. The mutational density within each group (e.g., by transcription levels or gene type) was normalized to account for variations in the total mutation count across groups by dividing the observed mutation density in each window by the total mutation density within that group.

Mutation frequencies in tDNAs at single base resolution.

To analyze mutation patterns along tRNA sequences and identify mutation hotspots (specific nucleotides within tRNA sequences that accumulate mutations more frequently) the mutation frequency at each position in each tDNA was calculated. This was done by dividing the number of samples exhibiting mutations at a given position by the total number of samples analyzed. The set of tDNA positions that did not pass the quality filters, including alignability < 1.0, regions that are unstable when converting between GRCh37 and GRCh38, and ENCODE blacklist problematic regions, were removed from the analysis. For each tDNA the genomic coordinates were adjusted to align with the consensus tRNA reference positions (Beuning and Musier-Forsyth 1999; Biela et al. 2023), ensuring accurate nomenclature and allowing for the effective grouping of results across different tRNAs. To determine positions with significantly higher or lower mutation frequencies, p-values were calculated using the Wilcoxon test, comparing mutational frequencies for each position from all the tRNAs against all other positions, with adjustments for multiple comparisons using the Benjamini-Hochberg method.

To assess the mutagenic load within the structural domains or regions within the tRNAs, a sliding window approach was applied. An 11-nucleotide window was moved along the tRNA sequence one nucleotide at a time, and the mutation frequencies within each window were combined to have a unique value by windows. The same statistical procedure used for the comparison by position was applied to calculate the differences by windows. In both analyses to consider a position or windows to be significant a threshold of adj. p-value < 0.01 and a mean higher or lower than the quantiles of 0.95 and 0.05 was considered. The mean mutational frequency at each position/windows across tDNAs was obtained and used to represent the results. To identify chimeric tRNAs, we first calculated the mutational density at the anticodon position (34, 35, 36) for $tRNA^{Ala}$, $tRNA^{Leu}$, $tRNA^{Ser}$, and $tRNA^{Tyr}$

by isodecoders. To detect specific anticodon mutations, absolute mutation counts and the annotation of base changes (e.g., C<G) were used to identify mutations occurring at the anticodon of tDNAs.

Mutational signatures.

Skin cancer samples were excluded from the analysis due to their distinct mutation profiles. Specifically, skin cancers exhibit a high prevalence of mutations caused by ultraviolet (UV) light exposure. These UV-induced mutations create distinctive signatures that can overshadow or confound the detection of other mutational processes when analyzing diverse cancer types. To deconvolute the possible processes that cause mutations specifically at tRNAs, we applied the standard mutational signatures methodology (Supek and Lehner 2015; Alexandrov et al. 2020), with several modifications (see Supplemental Methods).

Hairpin structure enrichment in tDNAs.

To assess the abundance of hairpin loop structures in tDNA genes relative to the rest of the genome and to other RNA Polymerase III (Pol III) transcribed genes, we utilized a genome-wide dataset of predicted 3-5 nucleotide hairpin sites (Buisson et al. 2019). Using the GenomicRanges package in R, we identified overlaps between hairpin regions and genomic coordinates of annotated tDNA genes, as well as with a separate dataset of other Pol III-transcribed genes. From this, we quantified the number of nucleotides within tDNA genes, other Pol III genes, and the remainder of the genome that were part of a predicted hairpin structure.

To determine whether hairpin structures were statistically enriched or depleted in tDNAs, we conducted two chi-square tests using the chisq.test() function in R: one comparing tDNAs to the rest of the genome, and another comparing tDNAs to other Pol III-transcribed genes. The number of nucleotides not involved in hairpin structures was calculated by subtracting hairpin-associated nucleotides from the total nucleotide content of each corresponding genomic region (tDNAs, other

Pol III genes, and non-tDNA regions of the genome). To determine whether hairpin structures were statistically enriched or depleted in tDNAs, we conducted two chi-square tests using the chisq.test() function in R: one comparing tDNAs to the rest of the genome, and another comparing tDNAs to other Pol III-transcribed genes.

Acknowledgements and author contributions.

We are grateful to Professors Manuel Santos (Multidisciplinary Institute of Aging, Porto), Nuria Malats and Francisco Real (CNIO, Madrid), Marc Martí-Renom (ICREA and CRG, Barcelona) and Nikolai Slavov (NorthWestern University, Boston) for helpful discussions. This work was supported by grants PID2023-147328OB-I00 and PID2019-108037RB-100 to L.R.d.P., and grants ERC StG "HYPER-INSIGHT" (757700), ERC CoG "STRUCTOMATIC" (101088342), Horizon2020 project "DECIDER" (965193), Horizon Europe project "LUCIA" (101096473), Spanish government project "REPAIRSCAPE," CaixaResearch project "POTENT-IMMUNO" (HR22-00402), to F.S.

MM-R and MM performed all analyses in this work and analyzed the data, AVP and LT assisted in specific analyses of tDNA distribution and mutational loads, AGT supervised the research and helped in the interpretation of the data, FS directed mutational analysis and interpreted the data, LRdP conceived the project, supervised it, and interpreted the data.

Competing interest statement.

The authors declare no competing interest.

References.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN et al. 2020. The repertoire of mutational signatures in human cancer. *Nature* **578**(7793): 94-101.

Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**(1): 9354.

Anisimova AS, Alexandrov AI, Makarova NE, Gladyshev VN, Dmitriev SE. 2018. Protein synthesis and quality control in aging. *Aging (Albany NY)* **10**(12): 4269-4288.

Berg MD, Brandl CJ. 2021. Transfer RNAs: diversity in form and function. *RNA Biol* **18**(3): 316-339.

Berg MD, Giguere DJ, Dron JS, Lant JT, Genereaux J, Liao C, Wang J, Robinson JF, Gloor GB, Hegele RA et al. 2019. Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biol* **16**(11): 1574-1585.

Bermudez-Santana C, Attolini CS, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, Stadler PF. 2010. Genomic organization of eukaryotic tRNAs. *BMC Genomics* **11**: 270.

Besedina E, Supek F. 2024. Copy number losses of oncogenes and gains of tumor suppressor genes generate common driver mutations. *Nature communications* **15**(1): 6139.

Beuning PJ, Musier-Forsyth K. 1999. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **52**(1): 1-28.

Bezerra AR, Simoes J, Lee W, Rung J, Weil T, Gut IG, Gut M, Bayes M, Rizzetto L, Cavalieri D et al. 2013. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proceedings of the National Academy of Sciences of the United States of America* **110**(27): 11079-11084.

Biela A, Hammermeister A, Kaczmarczyk I, Walczak M, Koziej L, Lin TY, Glatt S. 2023. The diverse structural modes of tRNA binding and recognition. *The Journal of biological chemistry* **299**(8): 104966.

Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, Lawrence MS. 2019. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**(6447).

Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB et al. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**(7437): 366-370.

Cabrelle C, Giorgi FM, Mercatelli D. 2024. Quantitative and qualitative detection of tRNAs, tRNA halves and tRFs in human cancer samples: Molecular grounds for biomarker development and clinical perspectives. *Gene* **898**: 148097.

Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic acids research* **49**(16): 9077-9096.

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic acids research* **44**(D1): D184-189.

Degasperi A, Zou X, Amarante TD, Martinez-Martinez A, Koh GCC, Dias JML, Heskin L, Chmelova L, Rinaldi G, Wang VYW et al. 2022. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* **376**(6591).

Derrien T, Estelle J, Marco Sola S, Knowles DG, Raineri E, Guigo R, Ribeca P. 2012. Fast computation and applications of genome mappability. *PLoS One* **7**(1): e30377.

Dittmar KA, Goodenbour JM, Pan T. 2006. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* **2**(12): e221.

Earnest-Noble LB, Hsu D, Chen S, Asgharian H, Nandan M, Passarelli MC, Goodarzi H, Tavazoie SF. 2022. Two isoleucyl tRNAs that decode synonymous codons

divergently regulate breast cancer metastatic growth by controlling translation of proliferation-regulating genes. *Nature cancer* **3**(12): 1484-1497.

Franco I, Helgadottir HT, Moggio A, Larsson M, Vrtacnik P, Johansson A, Norgren N, Lundin P, Mas-Ponte D, Nordstrom J et al. 2019. Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome biology* **20**(1): 285.

Galli G, Hofstetter H, Birnstiel ML. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294**(5842): 626-631.

Garcia-Vilchez R, Anazco-Guenkova AM, Lopez J, Dietmann S, Tome M, Jimeno S, Azkargorta M, Elortza F, Barcena L, Gonzalez-Lopez M et al. 2023. N7-methylguanosine methylation of tRNAs regulates survival to stress in cancer. *Oncogene* **42**(43): 3169-3181.

Gems D, Kern CC. 2024. Biological constraint, evolutionary spandrels and antagonistic pleiotropy. *Ageing research reviews* **101**: 102527.

Geslain R, Cubells L, Bori-Sanz T, Alvarez-Medina R, Rossell D, Marti E, Ribas de Pouplana L. 2010. chimeric tRNAs as tools to induce proteome damage and identify components of stress responses. *Nucleic acids research* **38**(5): e30.

Giege R, Eriani G. 2023. The tRNA identity landscape for aminoacylation and beyond. *Nucleic acids research* **51**(4): 1528-1570.

Giege R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic acids research* **26**(22): 5017-5035.

Gingold H, Tehler D, Christoffersen NR, Nielsen MM, Asmar F, Kooistra SM, Christophersen NS, Christensen LL, Borre M, Sorensen KD et al. 2014. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**(6): 1281-1292.

Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. 2016. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* **165**(6): 1416-1427.

Guo S, Fang J, Xu W, Ortega J, Liu CY, Gu L, Chang Z, Li GM. 2022. Interplay between H3K36me3, methyltransferase SETD2, and mismatch recognition protein MutSalpha facilitates processing of oxidative DNA damage in human cells. *J Biol Chem* **298**(7): 102102.

Gupta T, Malkin MG, Huang S. 2022. tRNA Function and Dysregulation in Cancer. *Front Cell Dev Biol* **10**: 886642.

Hou YM, Schimmel P. 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**(6169): 140-145.

Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, Limouse C, Halabian R, Wojenski L, Rodriguez M et al. 2022. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* **376**(6588): eabk3112.

Huang SQ, Sun B, Xiong ZP, Shu Y, Zhou HH, Zhang W, Xiong J, Li Q. 2018. The dysregulation of tRNAs and tRNA derivatives in cancer. *Journal of experimental & clinical cancer research : CR* **37**(1): 101.

Ibba M, Soll D. 2000. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **69**: 617-650.

Jang GM, Annan Sudarsan AK, Shayeganmehr A, Prando Munhoz E, Lao R, Gaba A, Granadillo Rodriguez M, Love RP, Polacco BJ, Zhou Y et al. 2024. Protein Interaction Map of APOBEC3 Enzyme Family Reveals Deamination-Independent Role in Cellular Function. *Mol Cell Proteomics* **23**(5): 100755.

Jinks-Robertson S, Bhagwat AS. 2014. Transcription-associated mutagenesis. *Annu Rev Genet* **48**: 341-359.

Kochavi A, Nagel R, Korner PR, Bleijerveld OB, Lin CP, Huinen Z, Malka Y, Proost N, van de Ven M, Feng X et al. 2024. Chemotherapeutic agents and leucine deprivation induce codon-biased aberrant protein production in cancer. *Nucleic acids research* **52**(22): 13964-13979.

Kwon MR, Choi JS, Won H, Ko EY, Ko ES, Park KW, Han BK. 2021. Breast Cancer Screening with Abbreviated Breast MRI: 3-year Outcome Analysis. *Radiology* **299**(1): 73-83.

Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, Bhagwat AS, Buisson R, Lawrence MS. 2021. An extended APOBEC3A mutation signature in cancer. *Nature communications* **12**(1): 1602.

Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtroder K, Jensen JB, Strandgaard T, Nordentoft I, Christensen E et al. 2021. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nature communications* **12**(1): 2301.

Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2023. Hallmarks of aging: An expanding universe. *Cell* **186**(2): 243-278.

Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**(5): 1029-1041 e1021.

Mas-Ponte D, Supek F. 2020. DNA mismatch repair promotes APOBEC3-mediated diffuse hypermutation in human cancers. *Nature genetics* **52**(9): 958-968.

McCann JL, Cristini A, Law EK, Lee SY, Tellier M, Carpenter MA, Beghe C, Kim JJ, Sanchez A, Jarvis MC et al. 2023. APOBEC3B regulates R-loops and promotes transcription-associated mutagenesis in cancer. *Nature genetics* **55**(10): 1721-1734.

Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E et al. 2014. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nature genetics* **46**(5): 487-491.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A et al. 2022. The complete sequence of a human genome. *Science* **376**(6588): 44-53.

Orellana EA, Siegal E, Gregory RI. 2022. tRNA dysregulation and disease. *Nature Reviews Genetics* **23**(11): 651-664.

Ormond C, Ryan NM, Corvin A, Heron EA. 2021. Converting single nucleotide variants between genome builds: from cautionary tale to solution. *Brief Bioinform* **22**(5).

Papadimitriou MA, Avgeris M, Levis P, Papasotiriou EC, Kotronopoulos G, Stravodimos K, Scorilas A. 2020. tRNA-Derived Fragments (tRFs) in Bladder Cancer: Increased 5'-tRF-LysCTT Results in Disease Early Progression and Patients' Poor Treatment Outcome. *Cancers* **12**(12).

Parisien M, Wang X, Pan T. 2013. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol* **10**(12): 1853-1867.

Park SJ, Schimmel P. 1988. Evidence for interaction of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA. *J Biol Chem* **263**(32): 16527-16530.

Pavon-Eternod M, Gomes S, Rosner MR, Pan T. 2013. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *Rna* **19**(4): 461-466.

Perret V, Garcia A, Grosjean H, Ebel JP, Florentz C, Giege R. 1990. Relaxation of a transfer RNA specificity by removal of modified nucleotides. *Nature* **344**(6268): 787-789.

Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P, Chen Y, Shah H, Sale JE, Alexandrov LB et al. 2022. Mechanisms of APOBEC3 mutagenesis in human cancer cells. *Nature* **607**(7920): 799-807.

Pinzaru AM, Tavazoie SF. 2023. Transfer RNAs as dynamic and critical regulators of cancer progression. *Nat Rev Cancer* **23**(11): 746-761.

Polak P, Lawrence MS, Haugen E, Stoletzki N, Stojanov P, Thurman RE, Garraway LA, Mirkin S, Getz G, Stamatoyannopoulos JA et al. 2014. Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair. *Nature biotechnology* **32**(1): 71-75.

Reverendo M, Soares AR, Pereira PM, Carreto L, Ferreira V, Gatti E, Pierre P, Moura GR, Santos MA. 2014. TRNA mutations that affect decoding fidelity deregulate development and the proteostasis network in zebrafish. *RNA Biol* **11**(9): 1199-1213.

Ribas de Pouplana L, Santos MA, Zhu JH, Farabaugh PJ, Javid B. 2014. Protein mistranslation: friend or foe? *Trends Biochem Sci* **39**(8): 355-362.

Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics* **45**(9): 970-976.

Roszkowska M. 2024. Multilevel Mechanisms of Cancer Drug Resistance. *Int J Mol Sci* **25**(22).

Rubio Gomez MA, Ibba M. 2020. Aminoacyl-tRNA synthetases. *Rna* **26**(8): 910-936.

Saini N, Roberts SA, Sterling JF, Malc EP, Mieczkowski PA, Gordenin DA. 2017. APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. *DNA Repair (Amst)* **53**: 4-14.

Sakhtemani R, Senevirathne V, Stewart J, Perera MLW, Pique-Regi R, Lawrence MS, Bhagwat AS. 2019. Genome-wide mapping of regions preferentially targeted by the human DNA-cytosine deaminase APOBEC3A using uracil-DNA pulldown and sequencing. *J Biol Chem* **294**(41): 15037-15051.

Sanchez A, Ortega P, Sakhtemani R, Manjunath L, Oh S, Bournique E, Becker A, Kim K, Durfee C, Temiz NA et al. 2024. Mesoscale DNA features impact APOBEC3A and APOBEC3B deaminase activity and shape tumor mutational landscapes. *Nature communications* **15**(1): 2370.

Santos M, Pereira PM, Varanda AS, Carvalho J, Azevedo M, Mateus DD, Mendes N, Oliveira P, Trindade F, Pinto MT et al. 2018. Codon misreading tRNAs promote tumor growth in mice. *RNA Biol* **15**(6): 773-786.

Schmidt E, Schimmel P. 1993. Dominant lethality by expression of a catalytically inactive class I tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **90**(15): 6919-6923.

Seplyarskiy V, Koch EM, Lee DJ, Lichtman JS, Luan HH, Sunyaev SR. 2023. A mutation rate model at the basepair resolution identifies the mutagenic effect of polymerase III transcription. *Nature genetics* **55**(12): 2235-2242.

Shi MJ, Meng XY, Fontugne J, Chen CL, Radvanyi F, Bernard-Pierrot I. 2020. Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. *Genome medicine* **12**(1): 85.

Skeparnias I, Anastasakis D, Grafanaki K, Kyriakopoulos G, Alexopoulos P, Dougenis D, Scorilas A, Kontos CK, Stathopoulos C. 2020. Contribution of miRNAs, tRNAs and tRFs to Aberrant Signaling and Translation Deregulation in Lung Cancer. *Cancers* **12**(10).

Sui Y, Qi L, Zhang K, Saini N, Klimczak LJ, Sakofsky CJ, Gordenin DA, Petes TD, Zheng DQ. 2020. Analysis of APOBEC-induced mutations in yeast strains with low levels of replicative DNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **117**(17): 9440-9450.

Supek F, Lehner B. 2015. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**(7550): 81-84.

Supek F, Lehner B. 2017. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* **170**(3): 534-547 e523.

Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. 2013. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife* **2**: e00534.

Tennakoon R, Bily TMI, Hasan F, Hoffman KS, O'Donoghue P. 2025. Natural human tRNA(Ala) anticodon variants mistranslate the genetic code. *Rna* **31**(6): 791-806.

Thornlow BP, Hough J, Roger JM, Gong H, Lowe TM, Corbett-Detig RB. 2018. Transfer RNA genes experience exceptionally elevated mutation rates. *Proceedings of the*

National Academy of Sciences of the United States of America **115**(36): 8996-9001.

Torrent M, Chalancon G, de Groot NS, Wuster A, Madan Babu M. 2018. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal* **11**(546).

Torres AG, Reina O, Stephan-Otto Attolini C, Ribas de Pouplana L. 2019. Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proceedings of the National Academy of Sciences of the United States of America* **116**(17): 8451-8456.

Van Bortle K, Phanstiel DH, Snyder MP. 2017. Topological organization and dynamic regulation of human tRNA genes during macrophage differentiation. *Genome biology* **18**(1): 180.

Van den Eynden J, Basu S, Larsson E. 2016. Somatic Mutation Patterns in Hemizygous Genomic Regions Unveil Purifying Selection during Tumor Evolution. *PLoS Genet* **12**(12): e1006506.

Vijg J, Dong X. 2020. Pathogenic Mechanisms of Somatic Mutation and Genome Mosaicism in Aging. *Cell* **182**(1): 12-23.

Wernaart D, Fumagalli A, Agami R. 2024. Molecular mechanisms of non-genetic aberrant peptide production in cancer. *Oncogene* **43**(27): 2053-2062.

Yang C, Pataskar A, Feng X, Montenegro Navarro J, Paniagua I, Jacobs JJL, Zaal EA, Berkers CR, Bleijerveld OB, Agami R. 2024. Arginine deprivation enriches lung cancer proteomes with cysteine by inducing arginine-to-cysteine substitutants. *Molecular cell* **84**(10): 1904-1916 e1907.

Zhang Z, Ye Y, Gong J, Ruan H, Liu CJ, Xiang Y, Cai C, Guo AY, Ling J, Diao L et al. 2018. Global analysis of tRNA and translation factor expression reveals a dynamic landscape of translational regulation in human cancers. *Commun Biol* **1**: 234.

Figure Legends.

Figure 1. tDNAs localization and transcription. (A) Schematic representation of tDNA biology including tDNA transcription by Pol III, tRNA processing, tRNA structure and codon-anticodon pairing. The diagram illustrates the conserved internal promoter regions (A box and B box), which are essential for RNA Polymerase III-mediated transcription. These regions facilitate the recruitment of transcription factors TFIIIB and TFIIIC. The precursor tRNA undergoes processing steps, including the addition of a 3' CCA tail, necessary for amino acid attachment through aminoacylation by aminoacyl-tRNA synthetase, ultimately producing a fully mature and functional tRNA. In the cytoplasm, mature tRNAs participate in protein synthesis by base-pairing the anticodon (positions 34 to 36) with complementary mRNA codons and delivering the appropriate amino acid. **(B)** Ideogram showing the chromosomal distribution and localization of human tDNAs in the T2T-CHM13 genome assembly including tRNA-scan-SE 2.0 prediction classification (high-confidence, pseudogenes, and tDNAs with uncertain function). **(C)** Quantification and classification of tDNAs across different human genome assemblies (GRCh37, GRCh38, T2T-CHM13) based on analyses conducted using tRNAscan-SE 2.0. tDNAs identified within unplaced or delocalized sequences of the reference genome are in grey. **(D)** Correlation between POLR3D and biotin-capture expression values by tDNA localization (cluster or isolated). tDNAs that are not in clusters are identified as isolated tDNAs. On the top row: Analysis based on clusters within 20 kb. Bottom row: Analysis based on clusters within 1 kb. On the left scatter plots showing the relationship between POLR3D and biotin capture expression values for tDNAs classified as either clustered (orange) or isolated (blue). The plots include Spearman's correlation coefficient (r) and the associated p-values. On the right, the violin plots show the $\log_2$ integrated tDNA expression for clustered and isolated tDNAs, with Wilcoxon significance levels indicated as *, $p \leq 0.001$. For the 20 kb clusters, $p < 2.2 \times 10^{-16}$, and for the 1 kb clusters, $p = 1.714 \times 10^{-10}$.

Figure 2. Somatic mutational profile and mechanism in tDNAs. (A) Normalized mutation density in tDNAs of cancer samples, including flanking regions (10 kb downstream and upstream divided into windows of 100 nucleotides). **(B)** Normalized mutation density in tDNAs classified by transcription levels from highest (Level 4) to lowest (Level 1), including flanking regions (3 kb upstream and downstream divided into 100 nucleotide windows). **(C)** Normalized mutation density in protein-coding genes classified by expression levels and flanking regions (3 kb upstream and downstream divided into windows of 100 nucleotides). **(D)** Comparison of NMF-extracted signatures with COSMIC signatures (see Methods). A mutational signature captures the pattern of trinucleotide mutations, quantified by their relative contributions. The heatmap shows cosine similarity values, indicating how similar the trinucleotide spectra of our de novo NMF-extracted signatures (rows, labeled with letters) are to the reference COSMIC signatures (columns, labeled with COSMIC IDs); higher values indicate greater similarity. Only COSMIC signatures with a cosine similarity ≥ 0.5 are shown. A COSMIC signature is considered reliably assigned when the cosine similarity is ≥ 0.85, indicated by an asterisk (*). NMF-derived signatures with at least one COSMIC match ≥ 0.85 are highlighted with a square. **(E)** Sample exposure ratio (fold-change, FC) between genes (stratified by type into tDNAs or protein-coding genes) and their flanking regions, for each NMF-extracted signature. Exposures are analyzed for tDNAs and protein-coding genes in cancer samples, as well as for tDNAs in healthy samples. Data for both gene types in cancer samples is stratified by expression levels, from low (Level 1) to high (Level 4). **(F)** ssDNA secondary structures that enhance APOBEC3 mutagenesis. Schematic representation of the ssDNA secondary structures produced during transcription that can increase the mutagenic activity of APOBEC3.

Figure 3. Age-related and tissue-specific mutational burdens in somatic tDNAs. (A) Mutation profiles in tDNAs across tissue types (including tissues with n > 20) relative to the flaking regions. Each dot represents the mean FC between mutation density values for tDNAs and the flanking regions (see Methods). tDNAs are categorized based on expression levels. **(B)** Normalized mutation density in tDNAs from healthy non-cancerous tissues (flanking regions of 50 kb downstream and upstream divided into windows of 500 nucleotides). **(C)** Tissues showing a significant correlation between mutational density in tDNAs and age. Data is grouped by tissue and age range (including tissues with n > 200). Each dot represents the FC between mutational densities of tDNAs and a flanking 10-window set (see Methods). The plots include values for the slope, Spearman's correlation coefficient (r), and statistical significance of the correlation (***, $p \leq 0.001$). Results for other tissues and detailed statistics are available in Supplemental Fig. S7.

Figure 4. Somatic tDNA mutation frequencies at single-base resolution and distribution of APOBEC motif presence across tDNAs genomic positions (A) The lower panel shows mutation rates at single-base resolution across tRNA sequences, highlighting universally conserved positions (U8, A14, G18, G19, A21, U33, G53, T54, Ψ55, C56, A58, C61, C74, C75), internal promoter regions such as the A-box (positions 8–19), the B-box (positions 52–62), and the anticodon (positions 34–36). Positions with statistically significant low or high mutational rates are indicated with red dots (considered significant when adj. $p \leq 0.01$ and if its mean mutation density exceeds the 0.95 quantile or falls below the 0.05 quantile of the overall distribution, quantile thresholds are indicated by red dashed lines). The upper panel shows the percentage of tDNAs with APOBEC-associated TCN motifs (where N = T, C, G, or A) across tDNA sequences. Internal hotspots for somatic mutations closely match the distribution of TCN signatures in tDNAs. **(B)** Comparison of mutation densities in described functionally and structurally important tRNA positions versus other positions. Wilcoxon significance levels are indicated in the plot: A-box vs. rest (adj. $p = 0.025$, *), conserved positions vs. rest (adj. $p = 0.025$, *), anticodon vs rest (adj. $p = 0.971$, ns), and B-box vs rest (adj. $p = 0.541$, ns). **(C)** Characterization of mutational profiles in the anticodon region of potential chimeric tRNAs, specifically for isoacceptors of alanine (Ala), serine (Ser), leucine (Leu), and tyrosine (Tyr). The heatmap on the left shows the mean mutation density at each anticodon position, classified by isodecoders for Ala, Ser, Leu and Tyr. The heatmap on the right illustrates the number of times specific mutational changes were observed in the anticodon region for specific Ala, Ser, Leu and Tyr isoacceptors. Position 34 is highlighted in green, as mutations at this position are always synonymous. Next to each mutation (observed nucleotide change), the codon recognized by the resulting chimeric tRNAs is indicated. For example, if $tRNA^{Ala}$(AGC) mutates to $tRNA^{Ala}$(AAC) this results in a tRNA charged with alanine that pairs with valine codons.