Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments

Adrian Gabriel Torres[a], Oscar Reina[a], Camille Stephan-Otto Attolini[a], and Lluís Ribas de Pouplana[a,b,1]

[a]Institute for Research in Biomedicine, Barcelona Institute of Science and Technology, 08028 Barcelona, Catalonia, Spain; and [b]Catalan Institution for Research and Advanced Studies, 08010 Barcelona, Catalonia, Spain

Edited by Joseph D. Puglisi, Stanford University School of Medicine, Stanford, CA, and approved March 19, 2019 (received for review December 12, 2018)

The human genome encodes hundreds of transfer RNA (tRNA) genes but their individual contribution to the tRNA pool is not fully understood. Deep sequencing of tRNA transcripts (tRNA-Seq) can estimate tRNA abundance at single gene resolution, but tRNA structures and posttranscriptional modifications impair these analyses. Here we present a bioinformatics strategy to investigate differential tRNA gene expression and use it to compare tRNA-Seq datasets from cultured human cells and human brain. We find that sequencing caveats affect quantitation of only a subset of human tRNA genes. Unexpectedly, we detect several cases where the differences in tRNA expression among samples do not involve variations at the level of isoacceptor tRNA sets (tRNAs charged with the same amino acid but using different anticodons), but rather among tRNA genes within the same isodecoder set (tRNAs having the same anticodon sequence). Because isodecoder tRNAs are functionally equal in terms of genetic translation, their differential expression may be related to noncanonical tRNA functions. We show that several instances of differential tRNA gene expression result in changes in the abundance of tRNA-derived fragments (tRFs) but not of mature tRNAs. Examples of differentially expressed tRFs include PIWI-associated RNAs, tRFs present in tissue samples but not in cells cultured in vitro, and somatic tissue-specific tRFs. Our data support that differential expression of tRNA genes regulate noncanonical tRNA functions performed by tRFs.

tRNA gene expression | tRNA fragments | piRNA | tRNA sequencing | tissue-specific expression

Transfer RNAs (tRNAs) link codons on messenger RNAs to their cognate amino acids in the ribosome. tRNAs have a strong secondary and tertiary structure, and undergo extensive posttranscriptional modifications (1). The identity of each tRNA is defined by its corresponding amino acid and by its anticodon sequence. tRNAs charged with the same amino acid are isoacceptor tRNAs (e.g., $tRNA^{Ala}$), and tRNAs with the same anticodon sequence are known as isodecoder tRNAs (e.g., $tRNA^{Ala}_{AGC}$). Each isodecoder tRNA set is generated from a pool of tRNA genes. For example, 35 genes code for $tRNA^{Ala}_{AGC}$ in the human genome (2). tRNAs can be further processed into tRNA-derived fragments (tRFs) to perform a wide range of biological functions. Importantly, tRNA and tRF abundance is cell- and tissue-specific and their expression patterns are increasingly being associated with human diseases (3–11). However, the forces driving this variable expression are poorly understood.

Changes in the expression levels of isoacceptor tRNAs may be adaptations to changes in transcriptome codon usage (6, 7, 10). However, variations in the expression of individual isodecoder tRNA genes that do not result in changes of isoacceptor tRNA levels should not be relevant to translation efficiency. Instead, such variations may be related to the control of noncanonical functions performed by tRFs (12, 13). In this sense, a careful evaluation of the contribution of specific tRNA genes to the abundance of mature tRNAs and tRFs in different cellular systems is needed.

Small RNA sequencing can quantify individual tRNA gene expression (herein referred as "tRNA-Seq"), but suffers from several technical limitations. First, sequencing biases may occur due to tRNA structure and the presence of posttranscriptionally modified residues. Second, estimating the mature tRNA pool can be challenging due to difficulties in discriminating reads from precursor-tRNAs (pre-tRNA), mature tRNAs, or tRFs. Third, the sequence similarity between different tRNA species can compromise the analysis of tRNA-Seq datasets and the unambiguous assignment of reads to a specific tRNA gene (14–18). Experimental improvements have been introduced to overcome some of these biases, including artificial removal of methylated bases on tRNAs (DM-tRNA-Seq and ARM-Seq) (16, 17), tRNA fragmentation to disrupt their secondary structure (Hydro-tRNA-Seq) (14, 19), and use of modified adaptor ligations (20, 21). In contrast, improvements to tRNA-Seq bioinformatics analyses have seldom been reported (18).

Estimation of differential tRNA expression from tRNA-Seq datasets by well-established statistical methods, such as edgeR or DESeq2, is based in the determination of "absolute" number of reads mapping to any given gene after correcting for library size (22). Importantly, this approach assumes that sequencing biases should equally affect all datasets. However, differential sequencing biases on certain tRNA transcripts occur due, for example, to sample- or tissue-specific tRNA modification patterns (12).

Comparisons among tRNA sequences from the same isodecoder set should be less affected by sequencing biases, given that their structures and modification patterns are relatively similar (23). Thus, we have developed a numerical analysis of tRNA-Seq datasets based on the proportional transcript contribution (percentage of sequencing reads) of individual tRNA genes to their

Significance

Given the large genetic redundancy of transfer RNA (tRNA) genes, the physiological relevance of variations in expression levels of individual tRNA genes is unclear. This is particularly true for variations within isodecoder gene sets (tRNAs having the same anticodon sequence), where no influence over codon usage during translation can be invoked. Here we show that differences in isodecoder tRNA gene expression between different human samples most often do not result in changes in the levels of mature tRNAs, but rather in changes in the abundance of alternative products of the genes, such as immature tRNA sequences, or tRNA fragments that are linked to noncanonical biological functions unrelated to protein synthesis.

Author contributions: A.G.T. and L.R.d.P. conceived the project; A.G.T. designed research; A.G.T. performed research; A.G.T., O.R., C.S.-O.A., and L.R.d.P. analyzed data; and A.G.T. and L.R.d.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE114904).

[1]To whom correspondence should be addressed. Email: lluis.ribas@irbbarcelona.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821120116/-/DCSupplemental.

Published online April 8, 2019.

corresponding isodecoder set. This isodecoder-specific tRNA gene contribution profile (iso-tRNA-CP) provides "relative" expression levels of individual tRNA genes that can be compared among different samples and experimental conditions.

Here we used DESeq2 and iso-tRNA-CP to compare different methods of tRNA-Seq library preparation and identify preferentially expressed tRNA genes, which we later validate by Northern blotting in several human cell lines and tissues. We find that sequencing biases quantitatively affect only a subset of human tRNA genes. We detect variations in levels of isodecoder tRNA genes not reflected at the isoacceptor set level, suggesting that these variations may be unrelated to translation. These instances of differential tRNA gene expression often result in changes in the abundance of tRFs, but not of mature tRNAs. Notably, we find that tRFs are abundant in human tissues but absent in cell lines under standard in vitro culture conditions, suggesting a more prominent role for tRFs in homeostasis than previously anticipated. Our results indicate that tissue-specific variation of tRNA gene transcription can be linked to the modulation of noncanonical tRNA functions, without affecting the canonical translation function of mature tRNAs.

Results

tRNA Expression Biases Introduced by Library Preparation Methods. To illustrate the differences between DESeq2 and iso-tRNA-CP we used both strategies to analyze tRNA-Seq datasets obtained from HEK293T cells (15) (*SI Appendix*, Fig. S1). As expected, highly expressed tRNA genes are the main contributors to both total tRNA transcript (DESeq2) and isodecoder (iso-tRNA-CP) pools (Spearman correlation $r = 0.861$). However, some tRNA genes presenting similar transcript abundance by DESeq2 are found to contribute differently to their respective isodecoder pools. This is evidenced by a relatively low Pearson correlation ($r = 0.404$) between DESeq2 and iso-tRNA-CP results (*SI Appendix*, Fig. S1). Thus, these two strategies are not equivalent but complement each other by extracting different information from tRNA-Seq datasets.

We applied both approaches to assess tRNA quantification biases caused by different sequencing library preparation methods (*SI Appendix*, Fig. S2) (15, 17). Datasets obtained with the standard Illumina approach are known to be enriched in tRNA fragments and poorly modified tRNAs, such as pre-tRNAs, due to tRNA modifications that prevent mature tRNA reverse transcription (RT-blocks) (datasets "Illumina-Seq") (15). Circularization of the reverse-transcribed products before PCR amplification (herein "CircRNA-Seq") allows the detection of short reads generated by RT-blocks and reduces the number of reads observed for pre-tRNAs, but cannot distinguish between RT-block artifacts and physiological tRNA fragments. CircRNA-Seq libraries obtained from total RNA or purified tRNAs as starting materials were used ("CircRNA-Seq total RNA" and "CircRNA-Seq tRNA") (17). RNA preparations can be further treated with a demethylase before reverse transcription to reduce RT-blocks and enrich datasets in reads from (artificially) unmodified mature tRNAs (DM-tRNA-Seq). DM-tRNA-Seq libraries obtained from total RNA or purified tRNAs as starting materials were used ("DM-tRNA-Seq total RNA" and "DM-tRNA-Seq tRNA") (17).

Datasets from HEK293T cells obtained with these five different library preparation methods were analyzed by hierarchical clustering and principal component analyses (PCA). The Illumina-Seq–based data emerged as the most divergent both in terms of absolute tRNA transcript abundance and of relative contribution to tRNA isodecoder sets (Fig. 1*A* and *SI Appendix*, Figs. S3 and S4). Datasets obtained by CircRNA-Seq and DM-tRNA-Seq showed overall similar tRNA expression profiles with both methods for dataset analysis (Fig. 1*A*).

Clover Leaf Propensity as a Predictor of tRNA Transcript Expression. Using the same datasets, we evaluated the correlation between the expression of each tRNA gene and its tRNA Scan score. tRNA Scan-SE is a frequently used in silico tool to predict tRNA genes that takes into account the clover leaf secondary structure and sequence conservation of the tRNA, and the presence of conserved intragenic tRNA promoter sequences. A score ("tRNA Scan score") is then assigned to each putative tRNA gene. Those genes with high scores (above ~50) are likely bona fide tRNA genes, while those ranked with a low score are likely pseudogenes (12, 24). We found that, regardless of the method for sequencing library preparation or the strategy for dataset analysis, the sequences that contribute the most to the tRNA pool were derived from tRNA genes with high (above 55) tRNA Scan scores (Fig. 1*B* and *SI Appendix*, Fig. S5). The only exceptions to this observation were three tRNA genes with tRNA Scan scores below 55 consistently identified by both DESeq2 and iso-tRNA-CP as important contributors to the tRNA pool in Illumina-Seq libraries (green

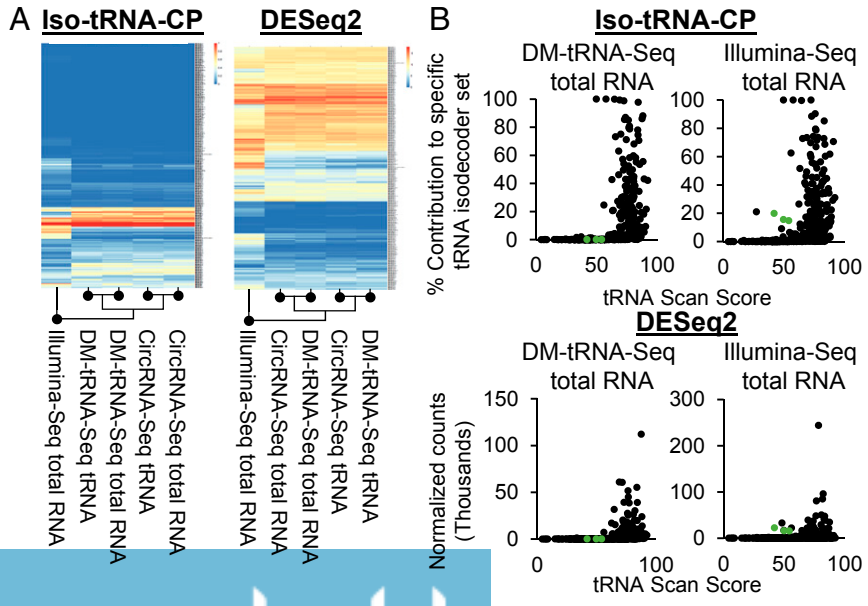

Fig. 1. Evaluation of tRNA gene expression by iso-tRNA-CP and DESeq2 in HEK293T cells. (*A*) Heatmap visualizing ward hierarchical clustering of individual tRNA gene expression for each method of sequencing library preparation, as evidenced by iso-tRNA-CP (*Left*) and DESeq2 (*Right*). Each line on the heatmap represents a tRNA gene/gene family. High-resolution figures are shown in *SI Appendix*, Fig. S3. (*B*) Scatter plots of tRNA Scan score against tRNA gene expression (iso-tRNA-CP and DESeq2) for datasets obtained with two sequencing library preparation methods (see also *SI Appendix*, Fig. S5). Colored points indicate tRNA genes/tRNA families having a tRNA Scan score below 55 that, both by iso-tRNA-CP and DESeq2 analyses, appear to contribute more to the tRNA pool in the Illumina-Seq datasets but not in other datasets. For visualization purposes, *tRNA-SeC-TCA-1-1* having a tRNA Scan score of 155.8 was left out of the plots.

points in Fig. 1*B* and *SI Appendix*, Fig. S5). Sequence coverage analysis and Northern blotting confirmed that these three sequences correspond to ∼95-nt immature forms of tRNA-Val-CAC-11 and tRNA-Glu-TTC-5-1, and ∼100-nt immature tRNA-Asn-GTT-17-1 (*SI Appendix*, Figs. S6*A* and S7).

Cell Type- and Tissue-Specific tRNA Gene Expression. Defects on tRNA physiology often results in neurological phenotypes (1, 13, 25). We characterized the HEK293T cell tRNA gene-expression profile and compared it to available tRNA-Seq datasets from human brain (26) (Fig. 2*A* and *SI Appendix*, Fig. S8). To improve accuracy, we combined the HEK293T datasets from non–Illumina-Seq library preparations (17), a strategy supported by our PCA analyses (*SI Appendix*, Fig. S4). Of the 379 tRNA genes/gene families considered (herein "tRNA genes") (*Materials and Methods*), 106 were not found differentially expressed in HEK293T cells and brain tissue by any method, while 178 tRNAs were found differentially expressed by both methods. This represents a 74.9% overlap among results obtained by these two methodologies (Fig. 2*B* and Dataset S1).

We wondered if the differential expression of tRNA genes affected overall abundance of isoacceptor sets. Conceptually, applying iso-tRNA-CP to the analysis of isoacceptors across samples should not offer any advantages over using DESeq2, as both approaches may be subject to quantitative biases. Therefore, we pooled all sequencing reads mapping to all genes belonging to the same isodecoder set into one single "isoacceptor gene" (*Materials and Methods*), and used DESeq2 to compare the HEK293T and brain datasets.

Interestingly, only 45.5% of the isoacceptor genes analyzed showed differential expression between the two samples (Fig. 3*A* and Dataset S2). In these cases, as expected, at least one isodecoder gene was always differentially expressed. We then focused on the remaining nondifferentially expressed isoacceptor genes (54.5%) and asked whether differential expression of isodecoder genes occur within these isoacceptor sets. We found differential expression of isodecoder tRNAs in 70% of these isoacceptor sets (Fig. 3*B* and Dataset S2).

Regulation of Mature tRNA and tRF Levels by Tissue-Specific tRNA Gene Expression. We asked whether differential expression of isodecoder genes within invariable isoacceptor sets results in the accumulation of alternative tRNA-related species. We designed Northern blot probes that would bind to the 5′- or 3′-region of target tRNAs to evaluate the nature of potential species resulting from tRNA processing (e.g., tRFs). We determined the length of the RNA species produced from tRNA genes that are: (*i*) differentially regulated at isodecoder and isoacceptor level; (*ii*) differentially regulated at isodecoder level only; (*iii*) nondifferentially regulated at isodecoder or isoacceptor level. We focused on tRNA genes with similar expression patterns by DESeq2 and iso-tRNA-CP (Dataset S1), and analyzed them in different human cell lines and tissues (Fig. 4).

tRNA-Arg-TCT-4-1, *tRNA-Arg-CCT-4-1*, and *tRNA-Gly-GCC-5-1* belong to the class of tRNA genes differentially regulated at isodecoder and isoacceptor levels (Dataset S2). Expression of tRNA-Arg-TCT-4-1 results in an abundant ∼74-nt species (mature tRNA) only in brain and in U2OS cells. In addition, a band of ∼110 nt (expected size of pre-tRNA) was observed in U2OS cells, while two smaller bands (32–35 nt) were specifically detected in brain (Fig. 4*A*, *Top*). Interestingly, human *tRNA-Arg-TCT-4-1* is homologous to a brain-specific murine tRNA [*tRNA-Arg-TCT-5-1*, *Mus musculus* (*Mm*) GRCm38/mm10] (2, 5) (*SI Appendix*, Fig. S9*A*). Only the mature and unprocessed forms of this tRNA are present in mouse brain (*SI Appendix*, Fig. S9*B*). Thus, expression of this tRNA gene is tissue-specific in mammals, but its processing is cell type- and species-specific. Levels of mature tRNA-Arg-CCT-4-1 and tRNA-Gly-GCC-5-1 were constant, but tRFs derived from these genes were abundant in human tissues and absent in cell lines (Fig. 4*A*, *Middle* and *Bottom*). Of particular interest is the tRF derived from $tRNA^{Gly}_{GCC}$, which corresponds to a validated PIWI-associated small RNA (piRNA piR-61648) (Fig. 4*A*, *Bottom*) (27) and is enriched in human somatic tissues and depleted in gonadal tissues and cell lines. We observed a similar tissue-expression pattern for murine piR-61648 (Fig. 4*D*). Although we were unable to unequivocally assign piR-61648 reads to a specific human $tRNA^{Gly}_{GCC}$ gene (*SI Appendix*, Fig. S10), our results in mice suggest that this piRNA may derive from the *tRNA-Gly-GCC-2* gene family, which is highly conserved among vertebrates (*SI Appendix*, Table S1).

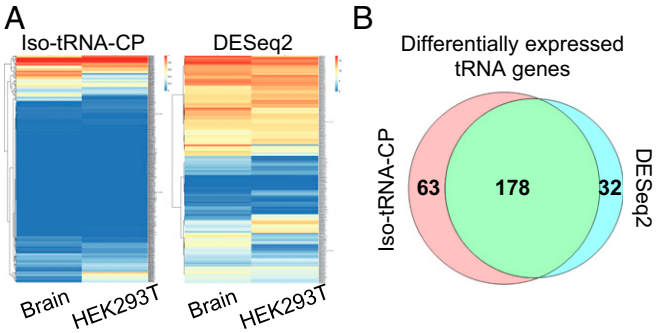

Fig. 2. Evaluation of differential tRNA gene expression in human brain and HEK293T cells. (*A*) Heatmap visualization of the expression of individual tRNA genes for human brain and HEK293T cells, as evidenced by iso-tRNA-CP (*Left*) and DESeq2 (*Right*). Each line on the heatmap represents a tRNA gene/gene family. High resolution figures are shown in *SI Appendix*, Fig. S8. (*B*) Venn diagram showing the overlap (green area) between the number of differentially expressed genes in human brain versus HEK293T cells, according to iso-tRNA-CP (pink circle) and DESeq2 (cyan circle). Numbers shown in each circle correspond to tRNA genes (individual genes and tRNA families); 106 genes were not found differentially expressed by any of both methods (not shown in figure). See also Dataset S1.

We also evaluated the expression of *tRNA-Gln-CTG-5-1*, *tRNA-Arg-TCG-1-1*, and *tRNA-Cys-GCA-6-1* (Fig. 4*B*), all of which showed differential expression only within isodecoder sets (Dataset S2). All three genes displayed increased abundance of tRFs levels in human tissues, without significant changes on mature tRNA abundance. A prevalence in somatic tissues was observed for tRFs derived from tRNA-Gln-CTG-5-1 and tRNA-Cys-GCA-6-1. Finally, we analyzed by Northern blotting the abundance of tRNA-Arg-CCG-2-1, a tRNA belonging to the group of tRNA genes not differentially expressed in either HEK293T or brain (Figs. 3*C* and 4*C* and Dataset S2). Notably, while not present in brain, a tRF of ∼60 nt derived from this gene was found in ovary, heart, and skeletal muscle, indicating that this tRF also presents tissue-specific expression.

Discussion

Here we evaluated variations in tRNA gene expression at the isoacceptor and isodecoder levels in different cellular systems. We developed a bioinformatics strategy (iso-tRNA-CP) that complements DESeq2 analyses by extracting additional information from tRNA-Seq datasets (*SI Appendix*, Fig. S1). We found this combined bioinformatics approach to increase the confidence of tRNA expression results (Figs. 2 and 4) (see below). Because sequencing tRNAs is prone to biases and artifacts (14–17), we studied how iso-tRNA-CP and DESeq2 performed using tRNA-Seq datasets from HEK293T cells obtained with different methods of sequencing library preparation (Fig. 1*A*). Library preparations that preferentially contain mature tRNAs showed similar expression profiles, suggesting that sequencing biases due to the tRNA structure and modifications patterns affect only a subset of human tRNA genes (see also below).

We asked whether the expressed genes in HEK293T cells were bona fide tRNA genes based on their tRNA Scan scores (24).

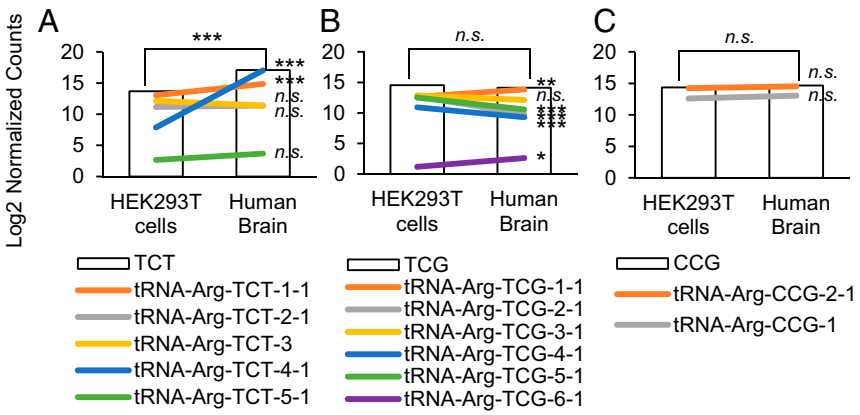

Fig. 3. Examples of the three different expression patterns of isodecoder genes within isoacceptor sets found when comparing HEK293T cells and human brain. (*A*) tRNA-Arg-TCT as an example of differential tRNA gene expression among isodecoder genes that also result in differential expression of the isoacceptor set. See also Fig. 4*A*. (*B*) tRNA-Arg-TCG as an example of differential tRNA gene expression among isodecoder genes that do not result in differential expression of the isodecoder set. See also Fig. 4*B*. (*C*) tRNA-Arg-CCG as an example of a nondifferentially expressed isoacceptor set that does not have differentially expressed isodecoder tRNA genes. See also Fig. 4*C*. n.s., not statistically significant; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. See also Dataset S2.

We found that genes with tRNA Scan scores above ~55 contribute the most to the tRNA pool (Fig. 1*B*), suggesting that these genes produce transcripts active in translation. It is possible that tRNA genes having low tRNA Scan scores may be transcribed into tRNA-related species with noncanonical functions. For example, the transcript produced by *tRNA-Asp-GTC-8-1* (tRNA Scan score = 30.3) does not adopt a tRNA-like conformation and acts as a 3′UTR Alu-binding element in HeLa cells (28). We evaluated the expression of *tRNA-Asp-GTC-8-1* and did not find sequencing reads mapping to this gene in HEK293T cells (*Materials and Methods*), suggesting that this gene, just like other tRNA genes, presents a differential cell type-specific expression driven by the need to regulate noncanonical tRNA functions (see discussion below).

The comparative analyses of tRNA-Seq datasets from different libraries allowed us to detect the expression of rare unprocessed pre-tRNAs (Fig. 1*B*). Processing of pre-tRNA transcripts has been shown to be the source of tRFs that carry out translation-independent functions (29, 30). However, in our analysis of the HEK293T tRNAome we did not detect tRFs related to enriched pre-tRNAs (*SI Appendix*, Fig. S7).

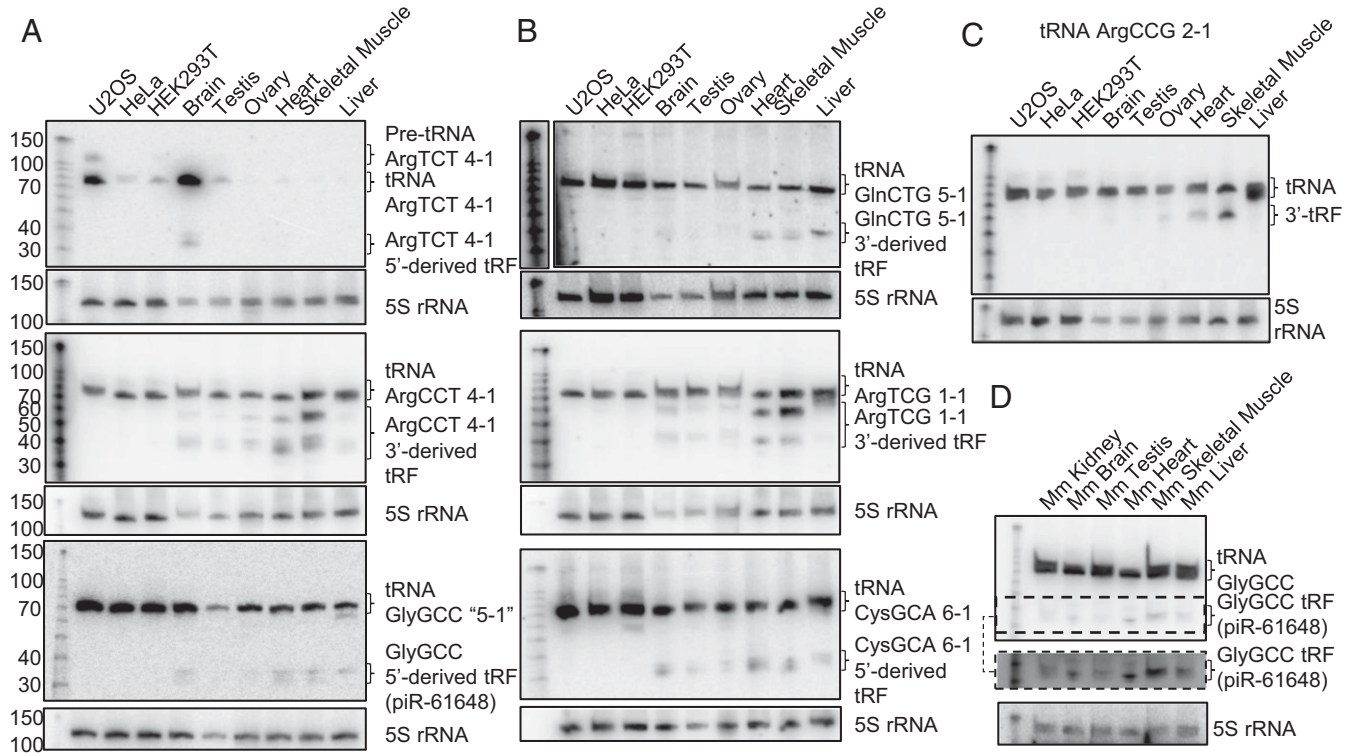

Fig. 4. Northern blot analyses of tRNA gene expression in cell lines and tissues. (*A*) Blotting against tRNA genes belonging to the group of tRNA genes showing variable expression at isodecoder and isoacceptor levels (see also Dataset S2). Inverted commas are used for tRNA GlyGCC 5-1 ("5-1") due to potential sequence ambiguity with tRNAs derived from the tRNA-Gly-GCC-2 family (see main text for details and *SI Appendix*, Fig. S10*A*). (*B*) Blotting against tRNA genes belonging to the group of tRNA genes showing variable expression at isodecoder level only (isoacceptor set nondifferentially expressed; see also Dataset S2). (*C*) Blotting against tRNA-Arg-CCG-2-1 was used as a control gene that did not present differential tRNA expression between human brain and HEK293T cell lines according to tRNA-Seq analyses (see also Dataset S2). Note the differential expression of tRFs in ovary, heart, and skeletal muscle. (*D*) Blotting against piR-61648 (tRF derived from $tRNA^{Gly}_{GCC}$ genes) in murine tissues. The region delimited by the dotted lines has been extracted and over-exposed to aid in visualization. Detection of 5S ribosomal RNA is used as gel loading control. Definition on whether tRFs derive from the 5′- or 3′-tRNA arm is based on the targeting region of the oligonucleotide used for detection by Northern blotting (see also *SI Appendix*).

To identify tRNA gene products potentially relevant for brain homeostasis (1, 13, 25), we analyzed tRNA-Seq datasets obtained from the human brain and compared their expression profiles to that of HEK293T cells (Fig. 2*A*). We detected evidence of differential expression for several tRNA genes (Fig. 2). Interestingly, the evaluation of tRNA gene expression by both methodologies coincided in ∼75% of tRNA genes analyzed (Fig. 2*B*), again suggesting that tRNA-Seq quantification biases affect only a subset of tRNA genes.

We found several instances of brain-enriched tRNAs both at the isodecoder and isoacceptor levels (Dataset S2), possibly corresponding to codon usage biases between HEK293T cells and brain tissue (i.e., canonical tRNA function). This is most likely the case of tRNA-Arg-TCT-4-1, the only tRNA showing significant enrichment of its mature form in human brain (Fig. 4*A*, *Top*). A mutation in a homologous brain-specific tRNA gene (*Mm tRNA-Arg-TCT-5-1*) that has been associated to neurodegeneration causes ribosome stalling at cognate AGA codons in mouse cerebellum (5), suggesting that the cell type- and tissue-specific expression levels of tRNA-Arg-TCT-4-1 are linked to its canonical role in protein translation.

Interestingly, the abundance of other RNA species related to this tRNA was also variable. tRFs derived from tRNA-Arg-TCT-4-1 were abundant in human brain, whereas U2OS cells and murine brain lacked the tRF but contained visible amounts of immature tRNA-Arg-TCT-4-1 (Fig. 4*A*, *Top*, and *SI Appendix*, Fig. S9). Thus, *Homo sapiens* (*Hs*) *tRNA-Arg-TCT-4-1* is a gene with a marked cell type- and tissue-specific expression, whose processing may be regulated in a cell- and species-dependent manner. This suggests that the differential expression of a specific tRNA gene might be necessary for canonical and noncanonical functions simultaneously (Fig. 4*A* and Dataset S2). We also found mature tRNA-Arg-TCT-4-1 highly expressed in U2OS cells, an osteosarcoma-derived cell line (Fig. 4*A*, *Top*). The region of chromosome 1 containing the *Mm tRNA-Arg-TCT-5-1* gene has been linked to bone density traits in mice (31). It is therefore possible that this tRNA may also be playing a role in human and murine bone regulation.

Other instances of differential tRNA gene expression modify relative isodecoder tRNA levels, but do not result in changes in isoacceptor sets among samples (Dataset S2). We characterized three tRNAs belonging to this category, and found them to display a tissue-specific increase in tRFs levels without significant variations in their mature tRNA levels (Fig. 4*B*). These results suggest that the differential expression of these tRNAs is necessary to regulate noncanonical tRNA functions carried out by tRFs (see below).

tRFs derived from the 5′-arm of $tRNA^{Cys}_{GCA}$ (Fig. 4*B*, *Bottom*) have been shown to play a role in brain. These tRFs contain an oligo-G terminal motif capable of forming G-quadruplexes, and were shown to inhibit protein synthesis by displacing the translation initiation factor eIF4G/eIF4A from m^7G-capped mRNAs (32, 33). Such regulation of translation contributes to a neuroprotective response in motor neurons (34).

Processing of $tRNA^{Gly}_{GCC}$ by angiogenin results in the production of piR-61648 (35), which likely derives from the 5′-arm of tRNA-Gly-GCC-2 (*SI Appendix*, Fig. S10). This is the most abundant piRNA in human saliva (36), and is increased in human hepatitis B patients (37). piRNAs found in gonadal tissues are generated by the "ping-pong cycle" mechanism present in germline cells (38–40). We found piR-61648 preferentially expressed in human and murine somatic tissues (Fig. 4*A*, *Bottom*, and Fig. 4*D*). We hypothesize that the tissue-specificity of piR-61648 expression is linked to its biogenesis pathway, and that angiogenin-dependent tRNA-derived piRNAs are more abundant in somatic cells. Consistent with this possibility, the human Piwi protein Hiwi2 associates to small RNAs in somatic cells, and ∼75% of these RNA species are tRFs (27). Germline piRNAs function primarily to silence transposons (38). Although the functions of tRNA-derived piRNAs are still elusive, their high degree of conservation across vertebrates (*SI Appendix*, Table S1) and the cytosolic localization of Hiwi2 suggest they may play a role in translational control (27).

Mature tRNAs are posttranscriptionally modified with a CCA-tail at their 3′-end. Thus, the presence of this sequence in a tRF indicates that it was derived from a mature tRNA. On the other hand, tRFs may derive from pre-tRNAs and contain a genetically encoded 3′—pre-tRNA trailer sequence (13). tRFs lacking 3′-extension sequences have also been reported (41). The vast majority of tRNA-Seq reads corresponding to 3′-derived brain tRFs validated here by Northern blotting contained 3′-extension sequences (*SI Appendix*, Table S2).

We find tRFs to be prevalent in human tissues and mostly absent in immortalized cell lines (Fig. 4). Previously, high levels of $tRNA^{Gly}_{GCC}$-derived tRFs were found in human fetal liver but were scarce in human cell lines (42). Generation of tRFs in cultured cells can be induced by treatment with sodium arsenite, but the relative levels of tRFs that we detect in undisturbed human tissues were much higher than those reported in cultured cells after sodium arsenite treatment (35). Thus, it is possible that the role of human tRFs in homeostasis is more relevant than previously anticipated.

We observe that, most commonly, tRF abundance is modulated without changes in mature tRNA levels. This suggests that tRNA genes can be selectively expressed to modulate noncanonical tRNA functions without compromising the mature tRNA pool. We propose that this mechanism is an important factor driving the regulation of functions performed by tRFs at tissular level.

Materials and Methods

Small RNA-Seq Datasets. Datasets used in this study were as follows: Illumina-Seq (PRJEB8019) (15) and GSE114904 (43) (generated for this study); CircRNA-Seq and DM-tRNA-Seq (accession no. GSE66550) (17); human brain (accession no. GSE43335) (26). Addition of several datasets per condition was needed to obtain sufficiently high number of tRNA sequencing reads. We attempted to analyze datasets obtained from other human tissues, including blood, heart, and liver (accession no. GSE69825) (44), but we could not obtain reliable results due to low number of tRNA reads.

Mapping of tRNA Reads. Mapping and processing of RNA-Seq reads was performed as previously described (15), except that alignments were performed against the human reference genome hg38. Human hg38 predicted tRNA genes were downloaded from the GtRNAdb v2.0 (January 2016) (2). Only cytosolic tRNAs were evaluated. The full list of included and excluded tRNA genes is shown in Dataset S3 (see *SI Appendix* for additional information). Note that *tRNA-Asp-GTC-8-1* was not annotated in the GtRNAdb at the time we downloaded the human tRNA gene coordinates; however, because we performed the mapping against the whole genome, we were able to verify that no reads mapped to this gene in HEK293T cells.

To reduce read assignment ambiguity we eliminated all reads mapping with the same reported alignment score to tRNAs from different isodecoder or isoacceptor sets (overall less than 0.001% of tRNA reads). Reads mapping to a single isodecoder set were assigned to individual tRNA genes, and those mapping to multiple tRNA genes with identical mature sequences were assigned to a single "tRNA family" (18). We used the tRNA gene annotation of tDR mapper (45): "W-X-Y-Z" (W: tRNA-amino acid; X: anticodon; Y: unique gene identifier; Z: the gene number subtype) (e.g., *tRNA-Ala-AGC-1-1*) for individual tRNA genes, the format "W-X-Y" (e.g., *tRNA-Ala-AGC-2*) for tRNA gene families, and the format "W-X" (e.g., *tRNA-Ala-AGC*) for "isoacceptor genes." In total, we considered 563 human cytosolic tRNA genes (2). We assigned reads to 309 individual tRNA genes and 70 tRNA families (containing a total of 254 tRNA genes) (Dataset S3) (see *SI Appendix* for additional information).

Small RNA-Seq Data Analysis. Unless otherwise specified, all downstream bioinformatics analyses were performed using R-3.1 and Bioconductor (46). Aligned and processed "final" tRNA reads (see previous section) were imported in R using Rsamtools 1.18.3 and were used to generate tRNA gene counts with the summarizeOverlaps function from the GenomicAlignments package v1.2.2 with options *mode* = *Union*, *ignore.strand* = *FALSE*. A first assessment of the differences between datasets was performed by PCA analysis using DESeq2 1.6.3 (22) (*SI Appendix*, Fig. S4). tRNA gene-expression values were used to generate heatmaps with the pheatmap package v1.0.8 using the "ward" clustering method and default options. Statistical significance for differential tRNA gene expression between samples was evaluated by the

Wilcoxon signed-rank test (iso-tRNA-CP) and by using a negative binomial GLM fit and Wald significance test (DESeq2). Differentially expressed genes were defined as those having a Wilcox P value (iso-tRNA-CP) or Benjamini–Hochberg-adjusted Wald test P (DESeq2) < 0.05.

In Silico Data Curation and Selected tRNA Genes for Validation by Northern Blotting. After the identification of differentially expressed tRNA genes by tRNA-Seq, an additional in silico data curation step was performed on reads mapping to those genes. We assessed the short-read aligner mapping quality (MAPQ). MAPQ alignment values ≤ 2 correspond to a high probability of being nonunique alignments across the genome (47). Only tRNA genes having the vast majority of their reads with a MAPQ > 2 were considered for further analyses by Northern blotting (e.g., *SI Appendix*, Fig. S6 *B–D*). We focused on differentially expressed tRNA genes as evidenced by iso-tRNA-CP and DESeq2, given that the higher proportion of contribution (%) detected for a tRNA gene by iso-tRNA-CP could also be due to a reduced number of reads for another tRNA gene belonging to the same isodecoder set. Therefore, we considered genes whose higher expression was also supported by an increased number of mapped reads (i.e., DESeq2). Finally, we used the gridCoverage and regionsCoverage functions from the htSeqTools package v1.12.0 (48) to check tRNA gene sequencing coverage and determine whether pre-tRNAs, mature tRNAs, or tRFs are expected to be differentially expressed (e.g., *SI Appendix*, Fig. S6*A*).

Northern Blot. Northern blotting was performed following standard procedures. For technical details and oligonucleotides used as Northern blot probes, see *SI Appendix*.

Cell Lines and Human Tissues. U2OS, HeLa, and HEK293T cells were maintained in Dulbecco's Modified Eagle Medium, containing 10% FBS and 100 U/mL Pen/Strep, at 37 °C/5% CO_2. Total RNA was extracted using TRIzol Reagent (Ambion), following the manufacturer's protocol. Cell lines were plated in 10-cm Petri dishes and were lysed 48 h after plating (~80% confluency) with 1 mL TRIzol Reagent. RNA from human tissues was commercially obtained. First Choice Total RNA from human brain (Cat No. AM7962; Lot No. 1799894), skeletal muscle (Cat No. AM7982; Lot No. 1793003), heart (Cat No. AM7966; Lot No. 1764389), and liver (Cat No. AM7960; Lot No. 1685090) were obtained from Life Technologies. Total RNA from human testis (Cat No. 636533; Lot No. 1402004), ovary (Cat No. 636555; Lot No. 1511016), and brain (Cat No. 636530; Lot No. 1711057) were obtained from Clontech. We performed Northern blots using RNA from human brain from two different sources to rule out potential artifacts due to sample preparation, processing and handling, and have obtained identical results. RNA from mouse tissues was a gift from Joan Ginovart's laboratory, Institute for Research in Biomedicine, Barcelona, Spain. Frozen tissues were homogenized as previously described (49), and RNA was extracted with TRIzol Reagent.

ACKNOWLEDGMENTS. We thank the L.R.d.P. laboratory and the Biostatistics/Bioinformatics Service (Institute for Research in Biomedicine Barcelona) for helpful discussion; the Functional Genomics Service (Institute for Research in Biomedicine Barcelona) for performing small RNA-Seq library preparations; and Dr. Joan Ginovart's laboratory for providing RNA from mouse tissues. This work was supported by the Spanish Ministry of Economy and Competitiveness Grants BIO2015-64572-R and BIO2014-61411-EXP (to L.R.d.P.).

1. Kirchner S, Ignatova Z (2015) Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat Rev Genet* 16:98–112.
2. Chan PP, Lowe TM (2016) GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* 44:D184–D189.
3. Ouyang C, Martinez MJ, Young LS, Sprague KU (2000) TATA-Binding protein-TATA interaction is a key determinant of differential transcription of silkworm constitutive and silk gland-specific tRNA(Ala) genes. *Mol Cell Biol* 20:1329–1343.
4. Sagi D, et al. (2016) Tissue- and time-specific expression of otherwise identical tRNA genes. *PLoS Genet* 12:e1006264.
5. Ishimura R, et al. (2014) RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* 345:455–459.
6. Dittmar KA, Goodenbour JM, Pan T (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet* 2:e221.
7. Pavon-Eternod M, et al. (2009) tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res* 37:7268–7280.
8. Goodarzi H, et al. (2016) Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* 165:1416–1427.
9. Zhou Y, Goodenbour JM, Godley LA, Wickrema A, Pan T (2009) High levels of tRNA abundance and alteration of tRNA charging by bortezomib in multiple myeloma. *Biochem Biophys Res Commun* 385:160–164.
10. Gingold H, et al. (2014) A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158:1281–1292.
11. Soares AR, Santos M (2017) Discovery and function of transfer RNA-derived fragments and their role in disease. *Wiley Interdiscip Rev RNA* 8:e1423.
12. Pan T (2018) Modifications and functional genomics of human transfer RNA. *Cell Res* 28:395–404.
13. Anderson P, Ivanov P (2014) tRNA fragments in human health and disease. *FEBS Lett* 588:4297–4304.
14. Gogakos T, et al. (2017) Characterizing expression and processing of precursor and mature human tRNAs by hydro-tRNAseq and PAR-CLIP. *Cell Rep* 20:1463–1475.
15. Torres AG, et al. (2015) Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Res* 43:5145–5157.
16. Cozen AE, et al. (2015) ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* 12:879–884.
17. Zheng G, et al. (2015) Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* 12:835–837.
18. Hoffmann A, et al. (2018) Accurate mapping of tRNA reads. *Bioinformatics* 34:1116–1124.
19. Arimbasseri AG, et al. (2015) RNA polymerase III output is functionally linked to tRNA dimethyl-G26 modification. *PLoS Genet* 11:e1005671.
20. Pang YL, Abo R, Levine SS, Dedon PC (2014) Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. *Nucleic Acids Res* 42:e170.
21. Shigematsu M, et al. (2017) YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* 45:e70.
22. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
23. Machnicka MA, Olchowik A, Grosjean H, Bujnicki JM (2014) Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol* 11:1619–1629.
24. Lowe TM, Chan PP (2016) tRNAscan-SE on-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57.
25. Torres AG, Batlle E, Ribas de Pouplana L (2014) Role of tRNA modifications in human diseases. *Trends Mol Med* 20:306–314.
26. Ryvkin P, et al. (2013) HAMR: High-throughput annotation of modified ribonucleotides. *RNA* 19:1684–1692.
27. Keam SP, et al. (2014) The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Res* 42:8984–8995.
28. Rudinger-Thirion J, Lescure A, Paulus C, Frugier M (2011) Misfolded human tRNA isodecoder binds and neutralizes a 3′ UTR-embedded Alu element. *Proc Natl Acad Sci USA* 108:E794–E802.
29. Lee YS, Shibata Y, Malhotra A, Dutta A (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 23:2639–2649.
30. Hanada T, et al. (2013) CLP1 links tRNA metabolism to progressive motor-neuron loss. *Nature* 495:474–480.
31. Tyler AL, Donahue LR, Churchill GA, Carter GW (2016) Weak epistasis generally stabilizes phenotypes in a mouse intercross. *PLoS Genet* 12:e1005805.
32. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P (2011) Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* 43:613–623.
33. Lyons SM, Gudanis D, Coyne SM, Gdaniec Z, Ivanov P (2017) Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat Commun* 8:1127.
34. Ivanov P, et al. (2014) G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proc Natl Acad Sci USA* 111:18201–18206.
35. Yamasaki S, Ivanov P, Hu GF, Anderson P (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* 185:35–42.
36. Ogawa Y, Taketomi Y, Murakami M, Tsujimoto M, Yanoshita R (2013) Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. *Biol Pharm Bull* 36:66–75.
37. Zhang Y, et al. (2014) Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection. *J Mol Cell Biol* 6:172–174.
38. Brennecke J, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089–1103.
39. Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202.
40. Williams Z, et al. (2015) Discovery and characterization of piRNAs in the human fetal ovary. *Cell Rep* 13:854–863.
41. Czech A, Wende S, Mörl M, Pan T, Ignatova Z (2013) Reversible and rapid transfer-RNA deactivation as a mechanism of translational repression in stress. *PLoS Genet* 9:e1003767.
42. Fu H, et al. (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett* 583:437–442.
43. 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. Gene Expression Omnibus (GEO). Available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114904. Deposited May 24, 2018.
44. Lopez JP, et al. (2015) Biomarker discovery: Quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med Genomics* 8:35.
45. Selitsky SR, Sethupathy P (2015) tDRmapper: Challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. *BMC Bioinformatics* 16:354.
46. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
47. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
48. Planet E, Attolini CS, Reina O, Flores O, Rossell D (2012) htSeqTools: High-throughput sequencing quality control, processing and visualization in R. *Bioinformatics* 28:589–590.
49. Testoni G, et al. (2017) Lack of glycogenin causes glycogen accumulation and muscle function impairment. *Cell Metab* 26:256–266.e4.