NAR Genomics and Bioinformatics, 2025, **7**, lqaf013
https://doi.org/10.1093/nargab/lqaf013
Methods Article

OXFORD

The zebrafish (*Danio rerio*) snoRNAome

Renáta Hamar and Máté Varga *

Department of Genetics, ELTE Eötvös Loránd University, Budapest, 1117, Hungary
*To whom correspondence should be addressed. Email: mvarga@ttk.elte.hu

Abstract

Small nucleolar RNAs (snoRNAs) are one of the most abundant and evolutionary ancient group of functional non-coding RNAs. They were originally described as guides of post-transcriptional rRNA modifications, but emerging evidence suggests that snoRNAs fulfil an impressive variety of cellular functions. To reveal the true complexity of snoRNA-dependent functions, we need to catalogue first the complete repertoire of snoRNAs in a given cellular context. While the systematic mapping and characterization of "snoRNAomes" for some species have been described recently, this has not been done hitherto for the zebrafish (*Danio rerio*). Using size-fractionated RNA sequencing data from adult zebrafish tissues, we created an interactive "snoRNAome" database for this species. Our custom-designed analysis pipeline allowed us to identify with high-confidence 67 previously unannotated snoRNAs in the zebrafish genome, resulting in the most complete set of snoRNAs to date in this species. Reanalyzing multiple previously published datasets, we also provide evidence for the dynamic expression of some snoRNAs during the early stages of zebrafish development and tissue-specific expression patterns for others in adults. To facilitate further investigations into the functions of snoRNAs in zebrafish, we created a novel interactive database, snoDanio, which can be used to explore small RNA expression from transcriptomic data.

Introduction

Small nucleolar RNAs (snoRNAs) form one of the most abundant and ancient group of functional non-coding RNAs (ncRNAs). Their main role is to guide the chemical modification of specific nucleosides in several RNA classes. Consequently, they play critical roles in multiple cellular regulatory processes, such as the maturation and nucleolytic processing of ribosomal RNAs (rRNAs), the chromatin architecture and alternative splicing [1–3].

Based on common sequence motifs and conserved structural features, snoRNAs are classified in two major families, C/D box and H/ACA box snoRNAs (Fig. 1A and B), abbreviated as SNORDs and SNORAs, respectively [4, 5]. Members of both families are associated with larger ribonucleoprotein (RNP) complexes where they guide the function of partner enzymes to initiate the posttranscriptional modification of various RNA species. These enzymes, the methyltransferase fibrillarin (FBL) and the pseudouridine synthase dyskerin (DKC1), respectively, will catalyze the formation of ribose 2′-O-methyl (Nm) groups and the isomerization of uridine into pseudouridine (Ψ) [6–8].

The targets of SNORD-directed methylation by FBL can be rRNAs, transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and messenger RNAs (mRNAs) [2, 9–13]. Similarly, among the targets of DKC1 we can find snoRNAs, snRNAs, rRNAs, and mRNAs as well [2, 14–16]. Both of the discussed post-transcriptional modifications are installed cotranscriptionally, furthermore, several snoRNAs seem to be essential for nucleolytic processing during rRNA maturation and in their absence ribosomal biogenesis is disrupted [2]. Defects of DKC1 function can lead to monogenic diseases like dyskeratosis congenita [17, 18], Bowen-Conradi syndrome [19], or severe nephrotic syndrome with cataract [20]. Defective pseudouridylation of specific rRNA sites can also lead to cancer through the disruption of ribosomal function [15, 21–24]. Erroneous fibrillarin expression has also been linked to cancer [25–27], and some SNORDs have been suggested to act either as oncogenes or as tumor suppressors [28, 29].

Emerging experimental evidence also suggests non-canonical functions for snoRNAs, linked to the genesis of other ncRNAs, mRNA 3′ processing, protein trapping, and exosome recruitment to target RNAs [4, 30–32].

Within vertebrate genomes, snoRNAs are most often located in the intronic region of so called snoRNA host genes (SNHGs) but occasionally can also be found in intergenic regions. The expression of intragenic snoRNAs is usually dependent on the expression and/or splicing of their SNHGs, whereas intergenic snoRNAs possess their own RNA polymerase II dependent promoters [5, 33]. The assembly of snoRNPs is intimately coupled to the splicing of SNHG pre-mRNAs [34], and this can also regulate the relative amount of alternative transcripts generated through mRNA maturation [35]. Furthermore, recent results also suggest that some SNHGs possess so called dual initiation promoters, where the choice of promoter can influence whether the transcript will be processed as a pre-mRNA or as a specialized RNA used to splice out snoRNAs [36].

Due to this large body of evidence about the multitude of crucial biological processes affected by snoRNAs, there have been several concerted efforts to map the complete snoRNA repertoire (so called snoRNAome) in several species [37–39].

Received: July 27, 2024. **Revised:** February 8, 2025. **Editorial Decision:** February 12, 2025. **Accepted:** February 14, 2025

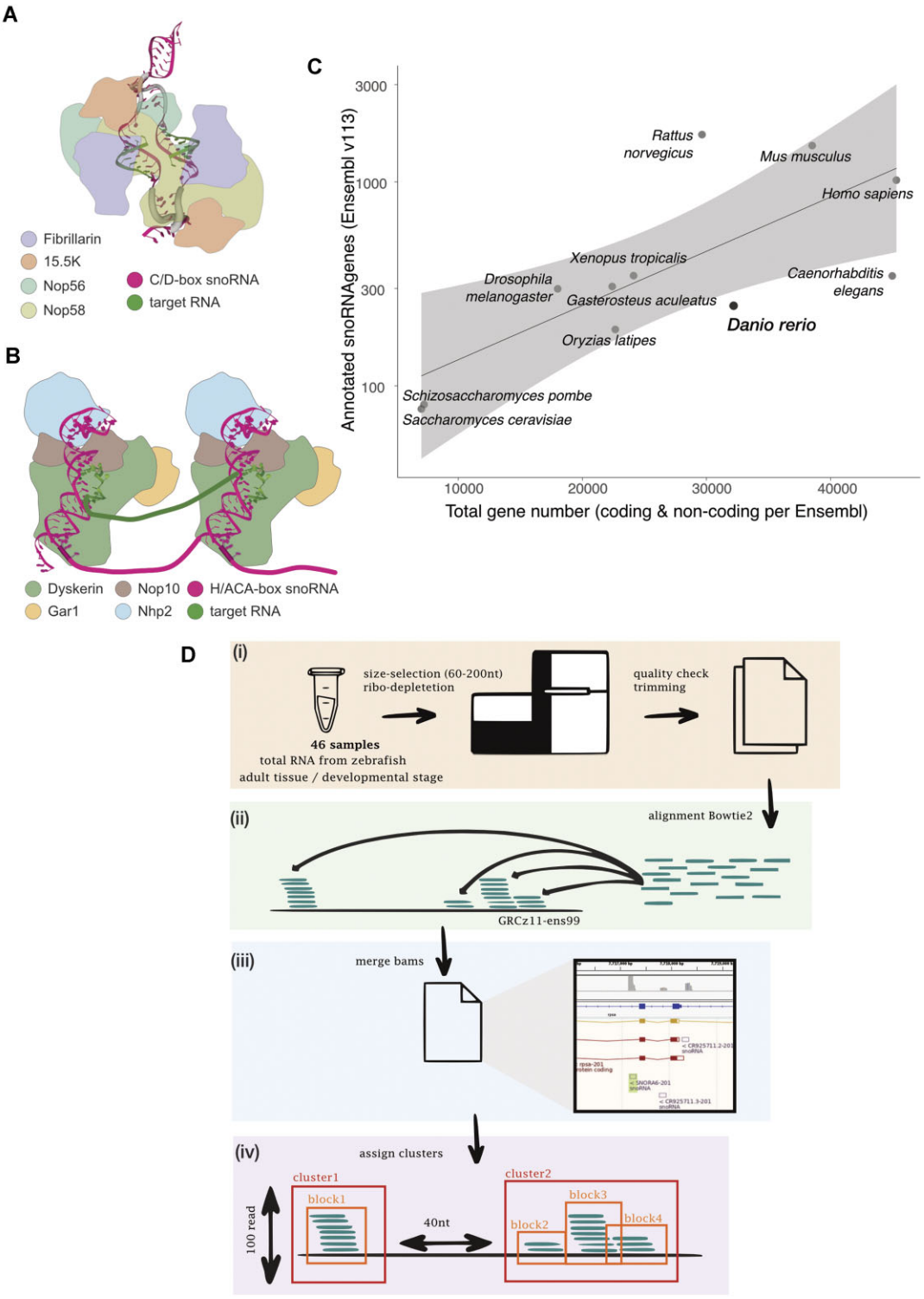

Figure 1. Annotation of novel snoRNAs in the zebrafish genome. Different snoRNA types participate in different RNP complexes: whereas C/D box snoRNAs can be found in complex with Fibrillarin, Nop56, and 15.5K (**A**), H/ACA box snoRNAs form complexes with Dyskerin, Nop10, Gar1, and Nhp2 (**B**). (**C**) The ratio of annotated snoRNAs in the zebrafish genome is relatively low compared with other species. (**D**) The computational pipeline used to annotate novel snoRNAs from transcriptomic datasets (see the "Materials and methods" section for details).

Zebrafish rose to prominence as a developmental model system in the 1990s thanks to its numerous advantages [40, 41]. While external fertilization coupled with the fast development, small size, and transparency of embryonic and larval zebrafish accounted for its original popularity, the explosive expansion of the transgenesis and genome editing toolboxes [42–46], and the development of high-throughput phenotyping assays both for larval and adult fish [47–51] have also made this species a dominant model of biomedical research focusing on disease modeling and drug development studies. This versatility also makes it an ideal subject to study the physiological roles of snoRNAs [52].

The relatively low number of known (annotated) zebrafish snoRNA genes (Fig. 1C) suggests that the annotation of the snoRNA pool in this species is not yet complete, and that part of the zebrafish snoRNAome is still to be discovered. By reanalyzing previously published transcriptomic datasets and acquiring novel ones enriched in snoRNAs from multiple adult tissues we were able to identify and annotate several new snoRNAs, expanding the current annotation (GRCz11 – Ensembl v113) of 247 snoRNA genes to 309. The analysis of the most complete zebrafish snoRNAome to date made it also possible to uncover the developmentally dynamic and/or tissue-specific expression of multiple snoRNAs in this species. We also present snoDanio (https://renata-h.shinyapps.io/98665a9405b44ede86eeb7179988104f/), an interactive database for the analysis of different transcriptomic datasets containing snoRNAs.

Materials and methods

Fish maintenance and RNA extraction

Adult wild-type fish used for these experiments were maintained in the animal facility of the Biology Institute of ELTE Eötvös Loránd University. Tissues were isolated from adult zebrafish based on standard protocols. All protocols employed in our study were approved by the Hungarian National Food Chain Safety Office (Permit XIV-I-001/515-4/2012). Two male and two female fish were sacrificed for isolation of total RNA from multiple tissues to produce four biological replicants. Extreme care was taken to avoid contamination and obtain pure homogenous tissue samples. The tissues were repeatedly washed in PBS to remove contaminating debris and TRI reagent (Zymo Research, Cat. No.: R2050) was used to collect total RNA, according to the manufacturer's protocol.

RNA library preparation and sequencing

All samples were subjected to size selection (∼60–200 bp) without ribodepletion before sequencing. Samples were sequenced by Novogene Ltd. on an Illumina NovaSeq 6000 PE150 platform. Twenty million reads per sample were obtained through this strategy.

Re-detection of previously annotated snoRNAs

We validated snoRNA prediction algorithms using the annotated version of the GRCz11 zebrafish genome (Ensembl v113). After downloading the annotated snoRNA sequences, we applied three detection algorithms: cmsearch, snoReport, and snoscan/snoGPS with default parameters. The results of these searches were then combined to create a comprehensive list of re-detected, already described snoRNAs. This validation process aimed to enhance the reliability of snoRNA detection by cross-verifying predictions across multiple tools.

snoRNA annotation pipeline

We created a novel, easily accessible and cloud-based pipeline, which greatly simplifies the identification of new snoRNA candidate sequences (Fig. 1D). Our annotation pipeline is available on the Galaxy web platform [53] via this link: https://usegalaxy.eu/u/danio/w/annot-snos. Briefly, after adequate quality check and trimming of the sequencing reads (Fig. 1Di), we used Bowtie2 to map reads in their continuum to the indexed reference (Fig. 1Dii). We considered searching for and filtering out further reads that map to rDNA sequences [54], but since some snoRNAs could be potentially processed from rRNAs [55], we decided against this approach. Bowtie2 mapped the reads to the genome in sensitive-local mode [56]. We merged these files with the *blockcluster* algorithm [57], which also characterized the amount of reads (Fig. 1Diii). We set the filtering threshold to a minimum depth of 100 read per cluster and set the distance between clusters to 50 bp, as suggested in a previous study [58]. In the consolidated file, we defined blocks (Fig. 1Div—orange boxes) from these reads and clusters from the blocks using BlockClust [59]. We extracted the sections that overlapped with the annotations downloaded from Ensembl v113. We used this approach for both time series and tissue data. We merged the overlapping clusters by their genomic coordinates (Fig. 1Div—red boxes). For the ∼60–200 long ones, we downloaded the genomic sequence and ran it through three known snoRNA scan applications to detect the corresponding sequences based on their structure. This method is good at detecting false positives, but it is worth analyzing the transcriptome-matched portion of the sequencing to see if previously unknown snoRNA variants of genes described as pre-miRNA or lnc-RNA exist. Therefore, for these reads, we also need to look at the read profile of each gene to determine whether processed snoRNAs are present whose annotation is obscured by the parental non-coding gene.

For detailed parameters of the scripts used in the pipeline see Supplementary Table S2.

Analysis of newly detected snoRNAs

Genomic coordinates of previously predicted zebrafish snoRNAs [36] have been remapped from the Zv9 to the GRCz11 genome build for comparison.

To perform a cluster analysis of newly predicted snoRNAs, we have defined snoRNA clusters within the zebrafish genome (GRCz11) as regions where at least three snoRNA genes have been mapped within 10 kb distance. New snoRNAs have been included in the clusters if they mapped into the existing cluster or within 2 kb upstream or downstream of it.

Data handling, analysis, and visualization were performed using the *GenomicRanges* [60], *rtracklayer* [61], *tidyverse* [62], *UpSetR* [63], *ggplot2* [64], and *pheatmap* [65] packages in RStudio [66].

RNA-seq data processing pipelines for snoDanio

Our database includes RNA sequencing data from multiple BioProjects uploaded to the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/). As these projects contain raw RNA sequences obtained by different laboratories through different sequencing strategies, we created standard-

Table 1. Different Galaxy pipelines used for the processing of external RNA-seq datasets

Data type	Link to Galaxy pipeline
Small RNA	https://usegalaxy.eu/published/workflow?id=443ea80b36d862f5
Single end reads	https://usegalaxy.eu/published/workflow?id=bbe2a1615c8faaae
Paired end reads	https://usegalaxy.eu/published/workflow?id=60a45de522b8b695

ized pipelines for their processing. For consistency, we performed the alignment for each dataset starting from the raw data and mapped it to the most recent version (GRCz11) of the zebrafish genome using HISAT2. In each case the newly annotated snoRNAs were also added to the list of transcripts prior the mapping process, and we created novel Galaxy pipelines for the analysis of different RNA-seq experiment types (Table 1).

To identify relevant BioProjects we downloaded the metadata for all zebrafish RNA sequencing experiments from the NCBI sequence read archive (SRA) and searched for BioProjects that contain snoRNA size-range sequences. We filtered the results for the use of the HiSeq 2000 platform (as it matches the sequencing parameters of our own dataset) and excluded experimental designs that focus on polyA mRNA. The workflows resulted in new expression data for each experiment, in a columnar format. These datasets were collated to create R Data Serialization (rds) objects that form the core of our searchable snoDanio database [67].

The snoDanio user interface

The snoDanio database pools previously available datasets with newly acquired sequencing data to create a comprehensive list of zebrafish snoRNAs, complemented with both snoRNA and host genes expression profiles in the examined datasets. The resulting database is accessible through a *shiny* web application [68] at https://renata-h.shinyapps.io/98665a9405b44ede86eeb7179988104f/.

The "Description of snoRNAs" section of our database allows the user to retrieve detailed information about various zebrafish snoRNA genes. For each gene the table showcases its ENSEMBL ID, source of annotation, and genomic location. By selecting a row in this table, one is able to see additional details about the chosen snoRNA in a dedicated sidebar panel. These additional informations include the gene symbol, the Rfam family, the parent gene information (when available), including the name and the biotype. We also provide the option to download the sequence of the selected snoRNA in FASTA format.

The "Explorer" part of the database is designed to provide an interactive platform for exploring the expression data of snoRNA genes across various BioProjects. Users can start by selecting a sequenced size range and an experiment type from the provided dropdown menus, which dynamically update the available project choices based on these selections. Upon choosing a project, the user can enter a specific snoRNA gene ID to investigate its expression levels. The database loads the rds object corresponding to the relevant BioProject and extracts the expression data for visualization. For the visualization of expression data the database generated plots using *ggplot2* and *plotly* [64, 69]. If the parent gene of the selected snoRNA is also available, a similar plot for the parent gene's expression is generated. Users can also view a table listing the top five differentially expressed snoRNA genes for the selected project, facilitating quick identification of key genes of interest. This feature-rich interface enables comprehensive exploration and analysis of gene expression data, making it a valuable tool for researchers.

Results

De novo annotation of zebrafish snoRNAs

The identification of snoRNAs is often difficult due to their lack of overall sequence conservation, their relatively small size (~60–200 bp) and the shortness of the sequence motifs characteristic for this RNA class. It is not surprising, therefore, that multiple *in silico* methods based on different approaches have been developed for snoRNA prediction over the years. Some rely more heavily on purely structural features (*cmsearch*) [70], while others focus more on sequence similarities (*snoReport*) [71] and some mix the two approaches (*snoGPS* and *snoscan*) [72].

To test the accuracy of these algorithms, we performed the redetection of the previously annotated 247 snoRNAs included in the Ensembl (v113) database. Our results show that the three algorithms could detect previously annotated snoRNAs with varying efficiency, and while *cmsearch* was able to identify all downloaded sequences as snoRNAs, the efficacy of *snoGPS* and *snoReport* was considerably lower (Supplementary Fig. S1). We also noted that five snoRNA genes included in the GRCz11 annotation (all labeled as Rfam predictions) could not be detected by either of the algorithms, and for consistency removed these from further analysis, resulting a core set of 242 Ensembl-annotated snoRNA genes.

Our redetection experiment suggested that the concomitant use of the three algorithms will result in reduced sensitivity. Yet, as *cmsearch* and *snoGPS* seemed to give inflated predictions (see below), and as the functional testing of snoRNAs was beyond the scope of this work, we still opted for the most rigorous and conservative approach and considered only those new sequences in our downstream analyses that were identified by all three *in silico* tools as snoRNAs.

We also created a computational pipeline to identify novel snoRNA genes (Fig. 1D and Supplementary Fig. S2). First, we grouped the mapped reads into blockgroups, each blockgroup representing a single ncRNA read profile, which might belong to one of the known ncRNA classes. These ncRNA read profiles are often associated with function [73], and class-specific discriminative models for C/D box or H/ACA box snoRNAs can be defined. After clustering, if a read profile belonged to one of these ncRNA classes, we considered it a cluster (Fig. 1D) [59].

To map a more complete snoRNAome of zebrafish, besides reanalyzing existing datasets, we also prepared and sequenced brain, liver, and gut total RNA samples from two adult males and two adult females, respectively. Our protocol uses small RNA-seq data that are size-selected to capture only RNAs of the required length. During the preparation of our sequencing libraries, we paid special attention to size selection and sequenced only the ~60–200 bp range. As this size range also

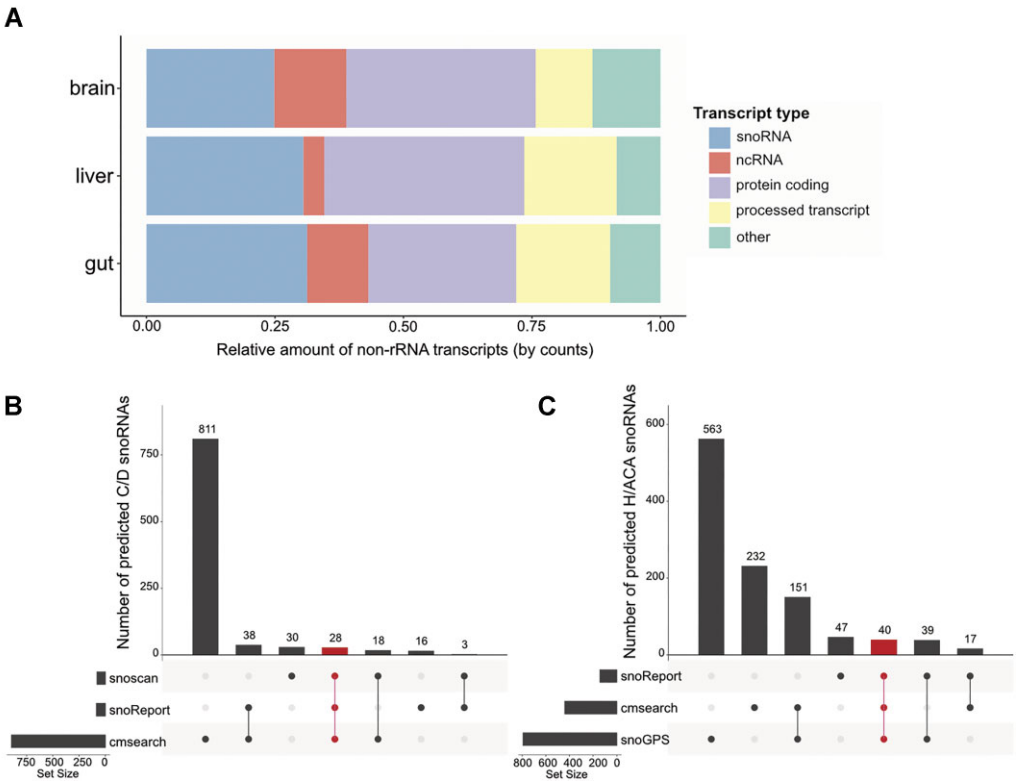

Figure 2. Identification of novel zebrafish snoRNAs. (**A**) Sequencing the ~60–200 nt fractions of zebrafish brain, liver, and gut total RNA samples revealed that these are enriched in snoRNAs. After removing known snoRNA sequences from our dataset, three different *in silico* prediction methods were used to identify new C/D box (**B**) and H/ACA box (**C**) snoRNAs. Up-set plots display the overlap between the individual predictions. (Note: in our analysis we used a conservative approach and accepted as candidates only snoRNAs predicted by all three methods.)

encompasses 5S and 5.8S rRNAs (~120 and ~160 bp, respectively), we computationally removed reads that mapped to rRNA genes from our sequencing datasets before further analysis (Supplementary Table S1 and Supplementary Fig. S3). After the elimination of rRNA-specific sequences, we mapped the remaining reads of our small RNA-seq datasets to the zebrafish genome (GRCz11). The post-mapping analysis of RNA biotypes demonstrated that known snoRNAs were over-represented amongst the detected RNA species in every tissue type, confirming the validity of our approach (Fig. 2A).

We used our new, tissue-specific expression datasets and combined them with publicly available time series data derived from zebrafish embryos and larvae at different developmental stages [74]. The presence of several snoRNAs that we had previously repredicted using various snoRNA predictor methods (see above) was also confirmed in these datasets. Of note, the small range fraction of this reanalyzed dataset also showed an enrichment in snoRNAs (Supplementary Table S1 and Supplementary Fig. S3).

Individual searches with the *cmsearch* and *snoGPS* algorithms resulted in excess numbers of C/D box and H/ACA box snoRNAs, respectively (Fig. 2B and C), with many of these predictions mapping to chromosome 4 (Supplementary Fig. S4A and B). Of note, this chromosome is the original sex chromosome of the species, also containing the largest heterochromatic block of the zebrafish genome, difficult to annotate, with a high content of ncRNAs and repetitive elements [75].

As currently available datasets do not allow for the further, functional validation of computationally predicted snoRNA genes, we decided to use a conservative approach of detection, only considering sequences that were predicted as snoRNAs by all three algorithms used. This way we were able to identify 67 well-supported new snoRNA-like sequences missing from the current Ensembl database (v113) annotations. Of these novel, putative snoRNAs 27 were C/D box and 39 were H/ACA box, and one of them was predicted as belonging to both classes (Fig. 2B and C; Supplementary Table S3).

Characterization of newly predicted zebrafish snoRNAs

We compared all our predicted snoRNAs with ncRNAs present in RNAcentral (release v17) to identify already annotated zebrafish snoRNAs. This analysis showed that 12 of our novel hits were previously annotated as various kinds of ncRNAs (Supplementary Table S3).

Next, we tested if any of the predicted snoRNA genes belong to existing genomic snoRNA clusters. Our results show that 18 of the 67 predicted snoRNAs (26.87%) indeed belong to existing clusters (Supplementary Table S3 and Supplementary Fig. S4C).

Finally, we extended our analysis to include a recently published dataset containing further predicted zebrafish snoRNA genes [36]. Of the 106 novel C/D box and 64 novel H/ACA box snoRNAs predicted by this work, 17 and 16, respectively, also appeared in our newly predicted dataset (Supplementary Fig. S5).

For our further analysis we have included all the previously annotated (and redetected) snoRNA genes in GRCz11

(Ensembl v113) and our newly predicted ones, resulting a set of 309 snoRNAs.

Characterization of the zebrafish snoRNAome

While most snoRNAs are located in the introns of SNHGs, occasionally they can be also found in intergenic genomic regions. Most of the previously identified snoRNAs in the zebrafish genome are associated with SNHGs, of which protein-coding genes are in excess (Fig. 3A). Of the predicted zebrafish snoRNAs that do not appear in the genome annotation 43 are located within various SNHGs (protein-coding genes, lincRNAs, and other processed transcripts) and 24 appear in intergenic positions (Fig. 3B). We note that the distribution of known and newly identified genes is similar (Fig. 3A and B).

To categorize the snoRNAs of the updated zebrafish snoRNAome by their abundance profile we also calculated the coefficient of variation (CV) of their expression [39]. In general, a low computed result for CV corresponds to a more uniform expression across different developmental stages and tissues, typical for housekeeping RNAs, such as tRNAs and snRNAs. In contrast, a high CV corresponds to a more enriched expression in either one or a few stages/tissue types, typical for protein-coding mRNAs and long non-coding RNAs (lncRNAs). Previous analysis of the human snoRNAome already suggested that snoRNA abundance profiles are an intermediate between these two forms, with most snoRNAs showing a uniform expression and a small but significant proportion displaying stage- and/or tissue-specific expression [39].

Reanalyzing existing datasets [74], our results suggests that a quasi-bimodal CV distribution, similar to the one observed in the case of human genome can be observed in zebrafish, too (Fig. 3C). The derivative of the CV function also gave us a threshold that we could use to separate uniformly expressed (UE) snoRNAs from those that are stage-enriched (SE) (Fig. 3C). Our analysis showed that these two abundance profiles had distinct characteristics. The SE snoRNAs are mostly encoded in the intergenic region, whereas UE snoRNAs are mostly encoded in protein-coding SNHGs (Fig. 3D).

We have also tested the stage- and tissue-specific expression profile of intergenic snoRNAs with intronic ones, and singleton snoRNAs with those situated in clusters (Supplementary Fig. S6). This analysis suggests that intergenic snoRNAs dominantly show non-uniform expression profiles, whereas intronic snoRNAs are UE Supplementary Fig. S6A and B). Given that most clustered snoRNAs are also intergenic, it is not surprising that this difference also exists when comparing clustered and singleton snoRNAs, albeit it is not as pronounced (Supplementary Fig. S6C and D).

As the expression of intron-encoded snoRNAs is dependent on the transcription and splicing of the SNHG, it would be fair to think that the expression level of snoRNA is always correlated positively with that of the host. As it has been shown for human snoRNAs, however, this is not always the case [39]. In our analysis of the zebrafish snoRNAome, using published datasets that contain adequate expression data for both snoRNAs and their host genes (Supplementary Table S4), we calculated the Pearson correlation coefficient (r) between the intronic zebrafish snoRNAs and their SNHGs, and we could also observe that some snoRNAs have no correlation or even a negative correlation with their SNHG's expression (Fig. 3E). The regulation of SNHG splicing through snoRNAs might be the basis of this phenomenon. Some snoRNAs have been shown to control their SNHGs by promoting non-sense mediated mRNA decay (NMD) [76] and computational analysis shows that several snoRNAs interact with host transcripts, influencing alternative splicing by concealing branch points and shifting the transcript ratio away from NMD [35]. Recent studies provide further insights into the complex relationship between snoRNAs and their SNHGs [77]. Of note, in our study, snoRNAs showing no or negative correlation with their SNHGs are almost exclusively found in protein-coding RNAs (Fig. 3F). In addition, the expression of long non-coding RNA SNHGs (lncSNHGs) is usually positively correlated with the expression of corresponding snoRNAs (Fig. 3F), as reported earlier for lncSNHGs expressed in various immune cells [78].

We also performed Gene Ontology analysis, which demonstrated that intragenic snoRNAs that are positively correlated with the expression of their SNHGs show a clear enrichment for processes such as RNA metabolism, ribosome biogenesis or ribosomal protein functions (Supplementary Fig. S7).

Previous research has highlighted that zebrafish SNHGs often possess special, dual-initiation promoters that can be transcribed either canonically by an initiating pyrimidine/purine (YR) dinucleotide or non-canonically by a terminal oligo-polypyrimidine (YC) [36]. We tested if SNHGs in our dataset with different promoter types (dual initiation, YR-only and YC-only) show different correlations. Interestingly, our results show that while dual initiation SNHGs can show both positive and negative correlation with the expression of the hosted snoRNAs, YR-only and YC-only SNHGs, while few in number, show only positive correlation (Supplementary Fig. S8).

Analysis of snoRNA expression patterns during development and in adult tissues

The analysis of the expression dynamics of snoRNAs included in our extended zebrafish snoRNAome demonstrated that some snoRNAs have a dynamic expression during development (Fig. 4A), whereas others can have tissue-specific expression patterns in adults (Fig. 4B). For example, the expression of some of the newly identified snoRNA genes (*snoRNA-pred7139*, *snoRNA-pred1884*, and *snoRNA-pred7137*) differs significantly between different early developmental stages (Fig. 4A) [74]. These dynamic expression patterns suggest that some snoRNAs might have differential functions during early stages of development.

Other six newly annotated snoRNAs (*snoRNA-pred1020*, *snoRNA-pred749*, *snoRNA-pred130*, *snoRNA-pred131*, *snoRNA-pred132*, and *snoRNA-pred1754*) show strong expression differences in different organs (Fig. 4B). Their expression in endodermal tissues, such as the liver and the gut, is more similar to each other and differs markedly from that observed in the neuroectoderm. As most of these snoRNAs are not expressed developmentally, their function might be restricted to adult stages.

snoDanio: The zebrafish snoRNAome database

Using our extended information of the zebrafish snoRNAome, we created an interactive database, snoDanio (https://renata-h.shinyapps.io/98665a9405b44ede86eeb7179988104f/), an openly licensed resource that facilitates integrative and interactive display and analysis of various zebrafish snoRNAs by

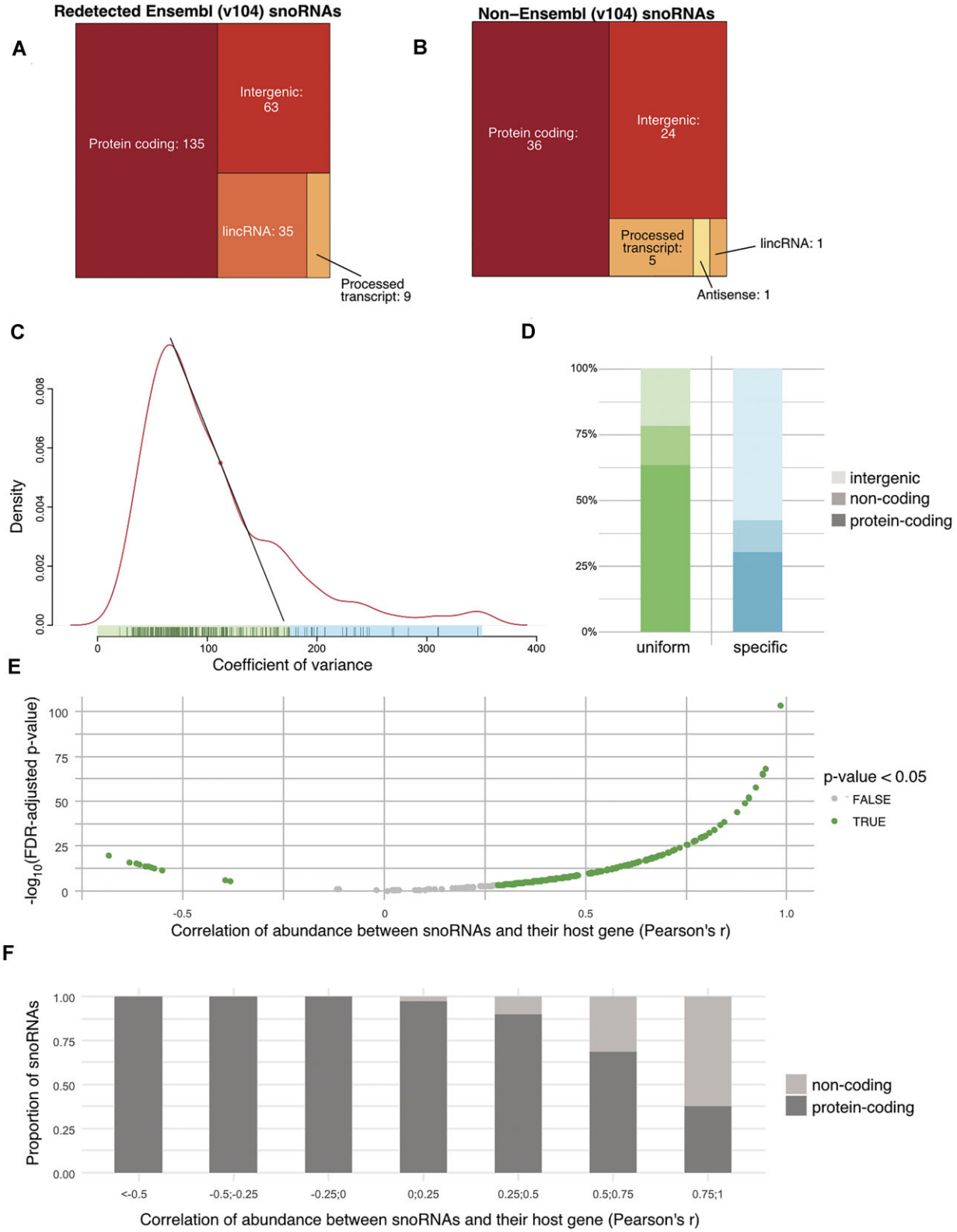

Figure 3. Distribution of novel zebrafish snoRNAs and the characterization of the zebrafish snoRNAome. (**A**, **B**) Distribution of known and newly identified snoRNAs shows a similar profile: they are generally associated with host genes (SNHGs), most of which are protein-coding genes. (**C**) The abundance of zebrafish snoRNAs based on coefficient of variance (CV) values shows a characteristic bimodal distribution. Based on the CV values, we can distinguish uniformly expressed (UE) and stage-enriched (SE) snoRNAs (green and blue, respectively). The CV of each snoRNA is indicated as a vertical line on the X axis. (**D**) The genomic position of UE and SE snoRNAs shows differential distribution: the former are enriched in protein-coding genes, while the latter in intergenic regions. (**E**) Correlation between the abundance of snoRNAs and their SNHGs shows that while most often a positive correlation can be detected between the two, for specific snoRNA-SNHG pairs no correlation or outright negative correlation is also possible. (**F**) SnoRNAs that are positively correlated with their SNHGs can be found in either non-coding or protein-coding SNHGs, whereas lack of correlation or negative correlation is typical for protein coding SNHGs.

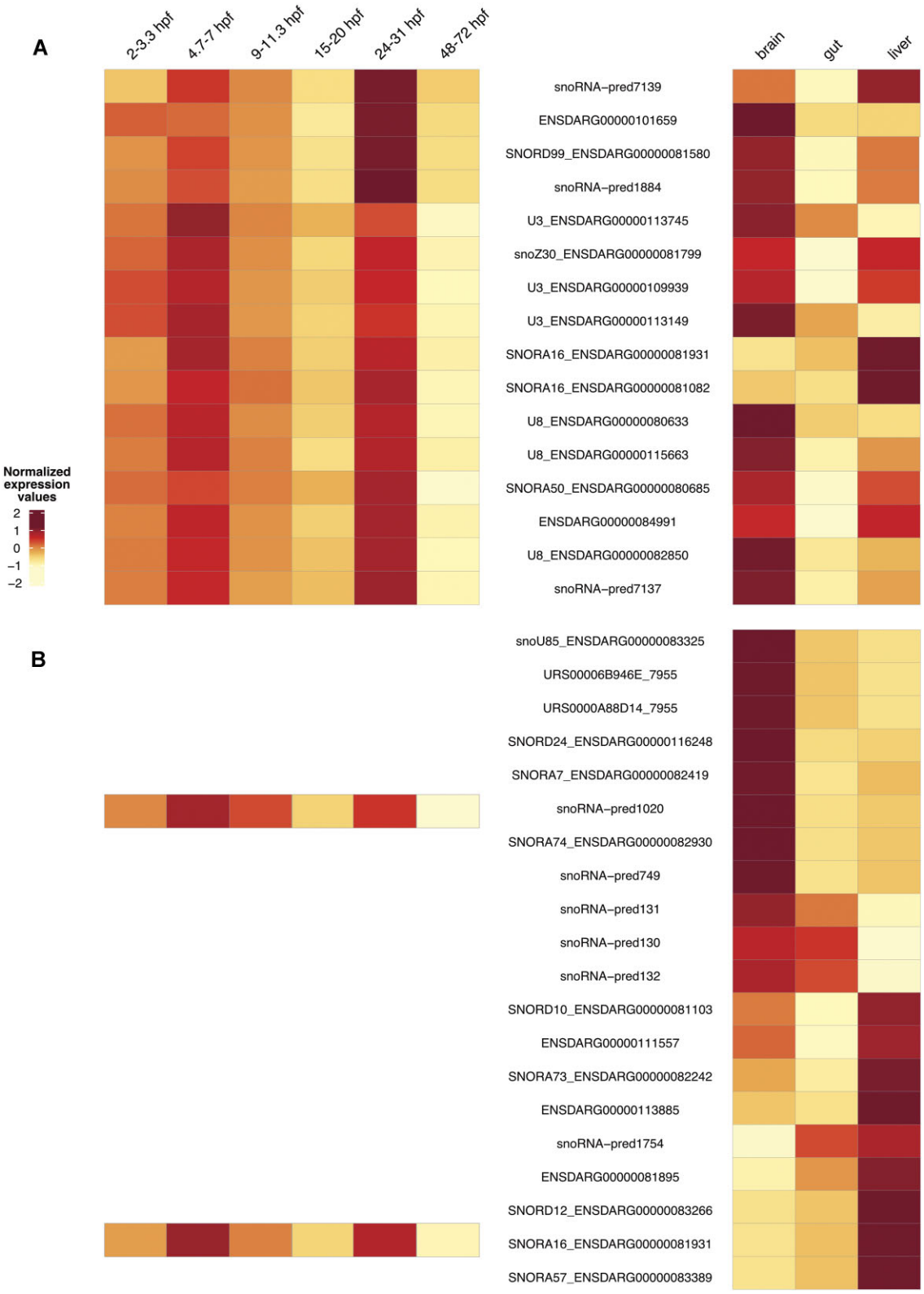

Figure 4. Stage- and organ-specific differential expression of snoRNAs. (**A**) Some snoRNAs display characteristic stage-specific developmental expression patterns. (**B**) Adult datasets show differential expression of particular snoRNAs in different organs. Of note, organs with linked (endodermal) developmental origin, such as the gut and the liver, often show more similar snoRNA expression patterns. (Heatmaps show normalized expression values of the relevant genes across the observed datasets.)

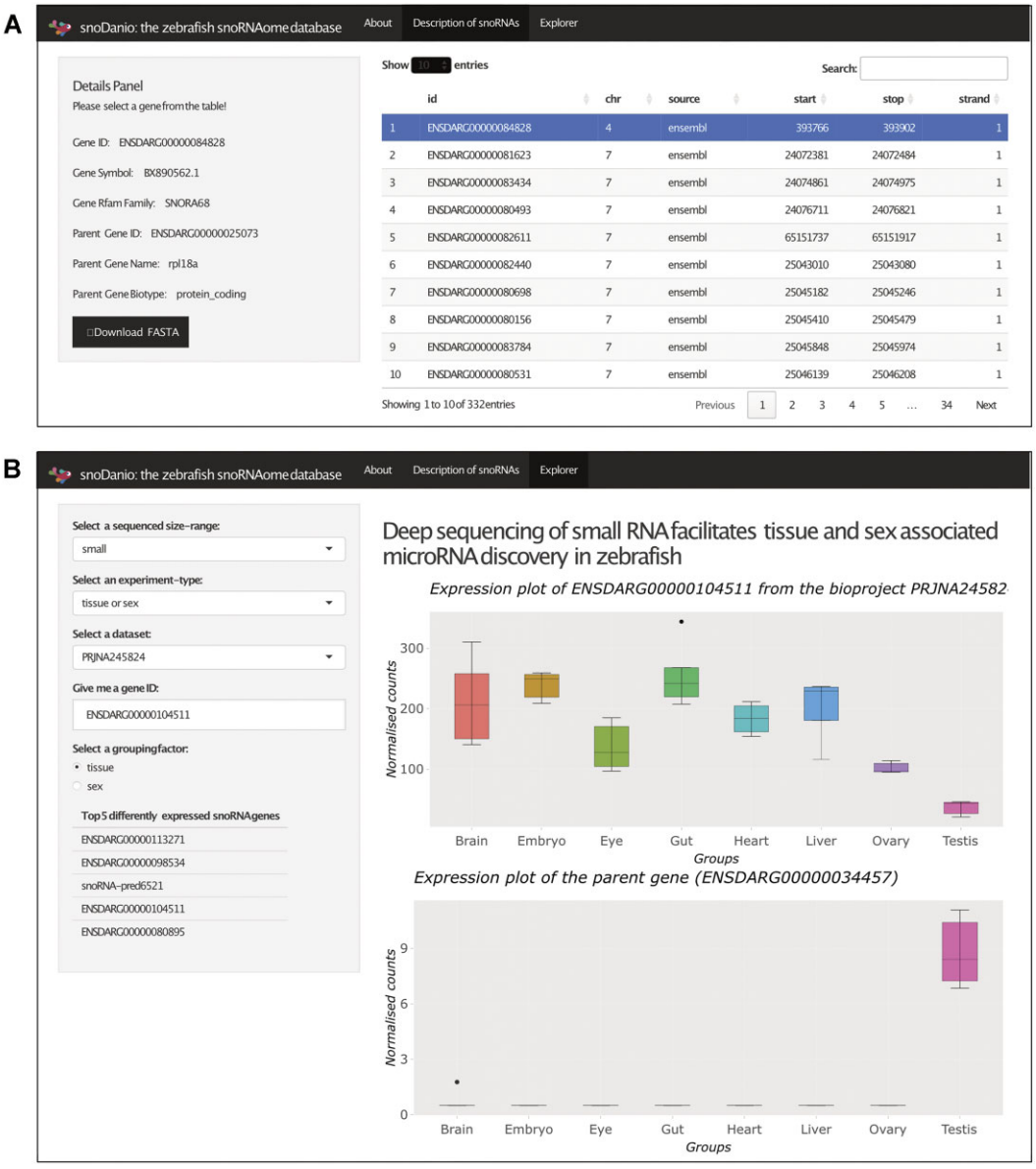

Figure 5. Page design for snoDanio. (**A**) Basic description of snoRNAs including genomic location, snoRNA classification, and SNHG information. (**B**) Expression analysis of respective snoRNAs and SNHGs using different datasets. Sequencing data are categorized by size-range and experiment type. (Note that grouping factor will depend on the design of the original sequencing experiment.)

mining multiple high-throughput sequencing data from tissue and developmental stages of zebrafish (Fig. 5).

Users can access basic, descriptive information of individual snoRNAs, including their classification, genomic position, and SNHG information (Fig. 5A), but can also assay the expression of particular snoRNAs in the preprocessed datasets already included in the database (Fig. 5B). These preprocessed datasets are categorized on the basis of the size range of RNAs included in the sequencing (mid, polyA, polyA-depleted, small, total, and total or polyA) and on the nature of the experiment design (assessing tissue-specific, sex-specific, developmental/time-related, or treatment-related changes), as determined from the metadata of the uploaded datasets. Depending on the type of the original experiment, multiple groupings for visualizations are also possible. The expression results and the sequence of snoRNAs can be downloaded on demand in image and FASTA formats, respectively, for further use.

Discussion

The expanding repertoire of putative and documented snoRNA functions [15, 22, 32, 79, 80] suggests that the significance of this ncRNA class has been hitherto underappreciated, and new insights into the regulation of gene expression will be gained by the in-depth study of these molecules. It is important, therefore, to compile comprehensive lists of snoRNAs and to document their dynamic and tissue-variable expression in all model organisms.

In the current study, we set out to define a well-supported list of snoRNAs expressed in zebrafish during development and in adults and create an interactive database to analyze the data. Using size-fractionated (~60–200 bp) sequencing datasets, rigorous filtering pipelines, and relying on predictions by multiple annotation algorithms (*snoReport*, *snoGPS*, *snoscan*, and *cmsearch*), we achieved identification of 67 novel snoRNA genes, not included in the current annotation of the

zebrafish genome with high confidence (Fig. 2B and C). Some of these genes are situated in intergenic regions, but the majority of them are encoded within the introns of protein-coding and lncRNA host genes, a common feature for vertebrate snoRNAs [81, 82].

With our newly discovered snoRNAs, the annotated and analyzed zebrafish snoRNAome stands at 309 genes. We note that this figure is likely a conservative estimate as we only considered genes that passed our stringent filtering pipeline (Fig. 1D), and individual predictions of the detecting algorithm detected further candidates (Supplementary Table S3). Therefore, by sequencing further tissues, employing better annotation tools, and performing functional validation studies, such as individual-nucleotide crosslinking and immunoprecipitation or enhanced CLIP experiments with fibrillarin or dyskerin in various tissues and/or at various developmental stages, the identification of further zebrafish snoRNAs is likely. As some *in silico* prediction tools highlighted chromosome 4 as a potential locus for novel snoRNAs, it will be especially interesting to see how many *bona fide* snoRNAs can be located to the fast-evolving chromosome 4 of the genome (Supplementary Fig. S4A and B), which contains a high incidence of ncRNA genes, and incidentally the maternal rDNA cluster as well [74, 75].

Our new dataset allowed us to compare the composition and expression dynamics of the human and zebrafish snoRNAomes. We observed that while most of the snoRNAs are UE in both species, a subset of them shows tissue-specific expression.

To date, multiple mechanisms have been identified that explain the effect of snoRNAs on gene expression. The most abundant target of snoRNA modification is rRNA and the role of snoRNA-guided modifications in ribosomogenesis and translational fidelity is well documented [2, 76, 83–88]. Defective ribosomal pseudouridylation and 2′-O-methylation can lead to developmental defects [18,89–97] and various diseases, including cancer [15, 20, 98–102]. Multiple studies have also demonstrated that snoRNAs themselves can be effectively used as biomarkers for numerous malignancies [21, 79, 80, 103–105], and recently they have been also linked to cellular senescence [106].

Accumulating evidence suggests that pseudouridylation of mRNAs is also abundant and can affect the efficiency of gene expression [107–115]. As different studies identified different roles for pseudouridines in mRNA [116, 117], the exact role for this modification (which is most likely to be context-dependent) remains to be confirmed.

Furthermore, cell-type specific presence of particular post-transcriptional modifications in rRNA has been also described and linked to ribosomal heterogeneity in a variety of organisms [118–121]. It has been proposed that differential epitranscriptomic modification of particular rRNA nucleotides could result in specialized ribosomes that are more adept in translating the messages in the given tissue-type(s) [85, 97, 122, 123]. Of note, it has been already demonstrated that knock-down of the U26, U78, and U44 C/D-box snoRNAs (driving methylation in 28S and 18S rRNAs, respectively), or depletion of U8 (which is required for removal of the 3′ external transcribed spacer or ETS sequence) can yield in divergent developmental phenotypes in zebrafish embryos [52, 124].

Our update of the zebrafish snoRNAome has also identified multiple snoRNAs with sex-, stage-, and tissue-specific expression, suggesting a hitherto underappreciated dimension of gene regulation. Understanding how this diversity in snoRNA expression can contribute to the developmental process and modulate the function of other genes should be the focus of later studies. Indeed, bisulfite-induced deletion sequencing analysis of different mouse tissues has already uncovered a high level of differential pseudouridylation in the transcriptome of different organs in mice [116], and differential maternal and zygotic transcription of certain zebrafish snoRNAs has been also described before [125].

The epitranscriptomic methodological revolution brought into focus the importance of several post-transcriptional modifications in gene expression regulation and also made it abundantly clear that these modifications can greatly vary between tissues, developmental stages, and physiological conditions [115]. The snoRNA-guided function of the fibrillarin- and dyskerin-containing RNPs makes them ideally suited to have central roles in the dynamic regulation of the epitranscriptome and through their effect on ribosome biogenesis, on the translational process as well. Understanding, therefore, how the function of these two essential enzymatic complexes can be regulated by the presence or absence of particular snoRNA guides will be a central question in this field in the coming years.

Our novel database, which compiles multiple lines of evidence to provide a comprehensive expression atlas for the snoRNAome of one of the most versatile genetic model organisms will aid this work by helping to generate novel hypotheses and to identify essential snoRNAs for particular biological processes. We plan to regularly update the snoDanio database by performing the reanalysis of newly available datasets. This approach will give us a chance to discover and characterize further snoRNAs. Moreover, if robust target prediction and validation pipelines can be developed, our data could be also integrated into more specific signaling network databases, such as SignaLink (which already handles zebrafish data) [126] or the Zebrafish Information Network (ZFIN) [127] as well.

Acknowledgements

We would like to thank Anita Rácz for fish care, Julianna Víg for help with the proofreading, and our two anonymous reviewers for helpful comments and suggestions.

Author contributions: R.H. (Conceptualization, Data curation, Investigation, Methodology, Writing—original and revised text), M.V. (Funding acquisition, Investigation, Project administration and supervision, Writing—original and revised text)

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

This work was supported by the National Research, Development, and Innovation Fund of Hungary grants FK124230 and K146460, the ELTE Eötvös Loránd University Institu-

tional Excellence Program Grant 1783-3/2018/FEKUTSRAT and the Central European Leuven Strategic Alliance grant CELSA/22/026. M.V. is a János Bolyai fellow of the Hungarian Academy of Sciences and he is also supported by the ÚNKP-23-5 New National Excellence Program of the Ministry of Culture and Innovation from the source of the National Research, Development and Innovation Fund. The authors also acknowledge the support of the Freiburg Galaxy Team: Björn Grüning, Bioinformatics, University of Freiburg (Germany) funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg.

Data availability

The new sequencing data generated during this project can be found at the NCBI SRA archive PRJNA1127032 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1127032/). We have also reused and reanalyzed data from the PRJNA347637 BioProject [74]. Further data related to this research project are available at Github (https://github.com/danio-elte/Danio_snoRNAome) and a frozen version of the repository with all the data used for analysis at Zenodo (https://doi.org/10.5281/zenodo.14836852) [67].

References

1. Han C, Sun L-Y, Luo X-Q *et al.* Chromatin-associated orphan snoRNA regulates DNA damage-mediated differentiation via a non-canonical complex. *Cell Rep* 2022;**38**:110421. https://doi.org/10.1016/j.celrep.2022.110421
2. Watkins NJ, Bohnsack MT. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA: Box C/D and H/ACA snoRNPs. *Wiley Interdiscip Rev RNA* 2012;**3**:397–414. https://doi.org/10.1002/wrna.117
3. Khosraviani N, Ostrowski LA, Mekhail K. Roles for non-coding RNAs in spatial genome organization. *Front Cell Dev Biol* 2019;**7**:336. https://doi.org/10.3389/fcell.2019.00336
4. Bratkovič T, Božič J, Rogelj B. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res* 2020;**48**:1627–51. https://doi.org/10.1093/nar/gkz1140
5. Webster SF, Ghalei H. Maturation of small nucleolar RNAs: from production to function. *RNA Biol* 2023;**20**:715–36. https://doi.org/10.1080/15476286.2023.2254540
6. Garus A, Autexier C. Dyskerin: an essential pseudouridine synthase with multifaceted roles in ribosome biogenesis, splicing, and telomere maintenance. *RNA* 2021;**27**:1441–58. https://doi.org/10.1261/rna.078953.121
7. Shubina MY, Musinova YR, Sheval EV. Proliferation, cancer, and aging—novel functions of the nucleolar methyltransferase fibrillarin? *Cell Biol Int* 2018;**42**:1463–6. https://doi.org/10.1002/cbin.11044
8. Kiss DJ, Oláh J, Tóth G *et al.* The structure-derived mechanism of Box H/ACA pseudouridine synthase offers a plausible paradigm for programmable RNA editing. *ACS Catal* 2022;**12**:2756–69. https://doi.org/10.1021/acscatal.1c04870
9. Gumienny R, Jedlinski DJ, Schmidt A *et al.* High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq. *Nucleic Acids Res* 2017;**45**:2756–69.
10. Vitali P, Kiss T. Cooperative 2′-O-methylation of the wobble cytidine of human elongator tRNA Met (CAT) by a nucleolar and a Cajal body-specific box C/D RNP. *Genes Dev* 2019;**33**:741–6. https://doi.org/10.1101/gad.326363.119
11. Sharma E, Sterne-Weiler T, O'Hanlon D *et al.* Global mapping of human RNA–RNA interactions. *Mol Cell* 2016;**62**:618–26. https://doi.org/10.1016/j.molcel.2016.04.030
12. Kishore S, Gruber AR, Jedlinski DJ *et al.* Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* 2013;**14**:R45. https://doi.org/10.1186/gb-2013-14-5-r45
13. Elliott BA, Ho H-T, Ranganathan SV *et al.* Modification of messenger RNA by 2'-O-methylation regulates gene expression *in vivo*. *Nat Commun* 2019;**10**:3401–9. https://doi.org/10.1038/s41467-019-11375-7
14. Karijolich J, Yi C, Yu Y-T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nat Rev Mol Cell Biol* 2015;**16**:581–5. https://doi.org/10.1038/nrm4040
15. McMahon M, Contreras A, Ruggero D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip Rev RNA* 2015;**6**:173–89. https://doi.org/10.1002/wrna.1266
16. Kiss T, Fayet-Lebaron E, Jády BE. Box H/ACA small ribonucleoproteins. *Mol Cell* 2010;**37**:597–606. https://doi.org/10.1016/j.molcel.2010.01.032
17. Savage SA, Alter BP. Dyskeratosis Congenita. *Hematol Oncol Clin North Am* 2009;**23**:215–31. https://doi.org/10.1016/j.hoc.2009.01.003
18. Ruggero D, Grisendi S, Piazza F *et al.* Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* 2003;**299**:259–62. https://doi.org/10.1126/science.1079447
19. Armistead J, Khatkar S, Meyer B *et al.* Mutation of a gene essential for ribosome biogenesis, *EMG1*, causes Bowen-Conradi syndrome. *Am Hum Genet* 2009;**84**:728–39. https://doi.org/10.1016/j.ajhg.2009.04.017
20. Balogh E, Chandler JC, Varga M *et al.* Pseudouridylation defect due to *DKC1* and *NOP10* mutations causes nephrotic syndrome with cataracts, hearing impairment, and enterocolitis. *Proc Natl Acad Sci USA* 2020;**117**:15137–47. https://doi.org/10.1073/pnas.2002328117
21. McMahon M, Contreras A, Holm M *et al.* A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS. *eLife* 2019;**8**:e48847. https://doi.org/10.7554/eLife.48847
22. Liang J, Wen J, Huang Z *et al.* Small nucleolar RNAs: insight into their function in cancer. *Front Oncol* 2019;**9**:587. https://doi.org/10.3389/fonc.2019.00587
23. Huang H, Weng H, Deng X *et al.* RNA modifications in cancer: Functions, mechanisms, and therapeutic implications. *Annu Rev Cancer Biology* 2019;**4**:25.
24. Babaian A, Rothe K, Girodat D *et al.* Loss of m1acp3Ψ ribosomal RNA modification is a major feature of cancer. *Cell Rep* 2020;**31**:107611. https://doi.org/10.1016/j.celrep.2020.107611
25. Marcel V, Ghayad SE, Belin S *et al.* p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell* 2013;**24**:318–30. https://doi.org/10.1016/j.ccr.2013.08.013
26. Zhang J, Yang G, Li Q *et al.* Increased fibrillarin expression is associated with tumor progression and an unfavorable prognosis in hepatocellular carcinoma. *Oncol Lett* 2021;**21**:92. https://doi.org/10.3892/ol.2020.12353
27. Long FNV, Lardy-Cleaud A, Carène D *et al.* Low level of Fibrillarin, a ribosome biogenesis factor, is a new independent marker of poor outcome in breast cancer. *BMC Cancer* 2022;**22**:526. https://doi.org/10.1186/s12885-022-09552-x
28. Zhu W, Niu J, He M *et al.* SNORD89 promotes stemness phenotype of ovarian cancer cells by regulating Notch1-c-Myc pathway. *J Transl Med* 2019;**17**:259. https://doi.org/10.1186/s12967-019-2005-1

29. Chen L, Han L, Wei J *et al.* SNORD76, a box C/D snoRNA, acts as a tumor suppressor in glioblastoma. *Sci Rep* 2015;**5**:8588. https://doi.org/10.1038/srep08588
30. Huang C, Shi J, Guo Y *et al.* A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res* 2017;**45**:8647–60. https://doi.org/10.1093/nar/gkx651
31. Falaleeva M, Stamm S. Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. *Bioessays* 2013;**35**:46–54. https://doi.org/10.1002/bies.201200117
32. Bergeron D, Fafard-Couture É, Scott MS. Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. *Biochem Soc Trans* 2020;**48**:645–56. https://doi.org/10.1042/BST20191046
33. Kufel J, Grzechnik P. Small nucleolar RNAs tell a different tale. *Trends Genet* 2019;**35**:104–17. https://doi.org/10.1016/j.tig.2018.11.005
34. Kiss T, Fayet E, Jády BE *et al.* Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol* 2006;**71**:407–17. https://doi.org/10.1101/sqb.2006.71.025
35. Bergeron D, Faucher-Giguère L, Emmerichs A-K *et al.* Intronic small nucleolar RNAs regulate host gene splicing through base pairing with their adjacent intronic sequences. *Genome Biol* 2023;**24**:160. https://doi.org/10.1186/s13059-023-03002-y
36. Nepal C, Hadzhiev Y, Balwierz P *et al.* Dual-initiation promoters with intertwined canonical and TCT/TOP transcription start sites diversify transcript processing. *Nat Commun* 2020;**11**:168. https://doi.org/10.1038/s41467-019-13687-0
37. Canzler S, Stadler PF, Schor J. The fungal snoRNAome. *RNA* 2018;**24**:342–60. https://doi.org/10.1261/rna.062778.117
38. Jorjani H, Kehr S, Jedlinski DJ *et al.* An updated human snoRNAome. *Nucleic Acids Res* 2016;**44**:5068–82. https://doi.org/10.1093/nar/gkw386
39. Fafard-Couture É, Bergeron D, Couture S *et al.* Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. *Genome Biol* 2021;**22**:172. https://doi.org/10.1186/s13059-021-02391-2
40. Varga M. The Doctor of Delayed Publications: The Remarkable Life of George Streisinger (1927–1984). *Zebrafish* 2018;**15**:314–9. https://doi.org/10.1089/zeb.2017.1531
41. Lieschke GJ, Currie PD. Animal models of human disease: zebrafish swim into view. *Nat Rev Genet* 2007;**8**:353–67. https://doi.org/10.1038/nrg2091
42. Kemmler CL, Moran HR, Murray BF *et al.* Next-generation plasmids for transgenesis in zebrafish and beyond. *Development* 2023;**150**:dev201531. https://doi.org/10.1242/dev.201531
43. Kwan KM, Fujimoto E, Grabher C *et al.* The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev Dyn* 2007;**236**:3088–99. https://doi.org/10.1002/dvdy.21343
44. Kawakami K. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* 2007;**Suppl 1**:S7. https://doi.org/10.1186/gb-2007-8-s1-s7
45. Uribe-Salazar JM, Kaya G, Sekar A *et al.* Evaluation of CRISPR gene-editing tools in zebrafish. *BMC Genomics* 2022;**23**:12. https://doi.org/10.1186/s12864-021-08238-1
46. Liu K, Petree C, Requena T *et al.* Expanding the CRISPR Toolbox in Zebrafish for Studying Development and Disease. *Front Cell Dev Biol* 2019;**7**:13. https://doi.org/10.3389/fcell.2019.00013
47. Varshney GK, Pei W, LaFave MC *et al.* High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res* 2015;**25**:1030–42. https://doi.org/10.1101/gr.186379.114
48. Huang H, Lindgren A, Wu X *et al.* High-throughput screening for bioactive molecules using primary cell culture of transgenic zebrafish embryos. *Cell Rep* 2012;**2**:695–704. https://doi.org/10.1016/j.celrep.2012.08.015
49. Dash SN, Patnaik L. Flight for fish in drug discovery: a review of zebrafish-based screening of molecules. *Biol Lett* 2023;**19**:20220541. https://doi.org/10.1098/rsbl.2022.0541
50. White RM, Patton EE. Adult zebrafish as advanced models of human disease. *Dis Model Mech* 2023;**16**:dmm050351. https://doi.org/10.1242/dmm.050351
51. Patton EE, Zon LI, Langenau DM. Zebrafish disease models in drug discovery: from preclinical modelling to clinical trials. *Nat Rev Drug Discov* 2021. **20** 611–28. https://doi.org/10.1038/s41573-021-00210-8
52. Higa-Nakamine S, Suzuki T, Uechi T *et al.* Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res* 2012;**40**:391–8. https://doi.org/10.1093/nar/gkr700
53. The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res* 2024;**52**:W83–94. https://doi.org/10.1093/nar/gkae410
54. Zhang B, Han D, Korostelev Y *et al.* Changes in snoRNA and snRNA Abundance in the human, chimpanzee, macaque, and mouse brain. *Genome Biol Evol* 2016;**8**:840–50.
55. Locati MD, Pagano JFB, Abdullah F *et al.* Identifying small RNAs derived from maternal- and somatic-type rRNAs in zebrafish development. *Genome* 2018;**61**:371–8. https://doi.org/10.1139/gen-2017-0202
56. Qu G, Kruszka K, Plewka P *et al.* Promoter-based identification of novel non-coding RNAs reveals the presence of dicistronic snoRNA–miRNA genes in Arabidopsis thaliana. *BMC Genomics* 2015;**16**:1009. https://doi.org/10.1186/s12864-015-2221-x
57. Bhatia PS, Iovleff S, Govaert G. Blockcluster: an R package for model-based co-clustering. *J Stat Soft* 2017;**76**:1–24. https://doi.org/10.18637/jss.v076.i09
58. Liu P, Si Y. Cluster analysis of RNA-sequencing data. In: *Statistical analysis of next generation sequencing data*. New York: Springer, 2014, 191–217. https://doi.org/10.1007/978-3-319-07212-8_10
59. Videm P, Rose D, Costa F *et al.* BlockClust: efficient clustering and classification of non-coding RNAs from short read RNA-seq profiles. *Bioinformatics* 2014;**30**:i274–82. https://doi.org/10.1093/bioinformatics/btu270
60. Lawrence M, Huber W, Pagès H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;**9**:e1003118. https://doi.org/10.1371/journal.pcbi.1003118
61. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 2009;**25**:1841–2. https://doi.org/10.1093/bioinformatics/btp328
62. Wickham H, Averick M, Bryan J *et al.* Welcome to the Tidyverse. *J Open Source Softw* 2019;**4**:1686. https://doi.org/10.21105/joss.01686
63. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;**33**:2938–40. https://doi.org/10.1093/bioinformatics/btx364
64. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016. https://ggplot2.tidyverse.org
65. Kolde R. pheatmap: Pretty Heatmaps. *R package version 1.0.12*. 2018. https://github.com/raivokolde/pheatmap
66. RStudio_Team. *RStudio: Integrated Development for R RStudio*. Boston, MA, USA: PBC, 2020.
67. Hamar R, Varga M. Curated RNA Sequencing Data for Zebrafish (*Danio rerio*) snoRNA Expression Analysis [Data set]. *Zenodo* 2024. https://doi.org/10.5281/zenodo.12822217
68. Chang W, Cheng J, Allaire J *et al.* Shiny: web application framework for R. *R package version 1.10.0.9000*. 2025. https://github.com/rstudio/shiny
69. Sievert C. Interactive web-based data visualization with R, plotly, and shiny. Chapman and Hall/CRC. 2020. https://doi.org/10.1201/9780429447273-38

70. Cui X, Lu Z, Wang S *et al.* CMsearch: simultaneous exploration of protein sequence space and structure space improves not only protein homology detection but also protein structure prediction. *Bioinformatics* 2016;**32**:i332–40. https://doi.org/10.1093/bioinformatics/btw271
71. Oliveira JV, de A, Costa F *et al.* SnoReport 2.0: new features and a refined Support Vector Machine to improve snoRNA identification. *BMC Bioinf* 2016;**17**:464. https://doi.org/10.1186/s12859-016-1345-6
72. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005;**33**:W686–9. https://doi.org/10.1093/nar/gki366
73. Pundhir S, Poirazi P, Gorodkin J. Emerging applications of read profiles towards the functional annotation of the genome. *Front Genet* 2015;**6**:188. https://doi.org/10.3389/fgene.2015.00188
74. Locati MD, Pagano JFB, Girard G *et al.* Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* 2017;**23**:1188–99. https://doi.org/10.1261/rna.061515.117
75. Wilson CA, Postlethwait JH. A maternal-to-zygotic-transition gene block on the zebrafish sex chromosome. *G3 (Bethesda)* 2024;**14**:jkae050. https://doi.org/10.1093/g3journal/jkae050
76. Ojha S, Malla S, Lyons SM. snoRNPs: functions in ribosome biogenesis. *Biomolecules* 2020;**10**:783. https://doi.org/10.3390/biom10050783
77. Penzo M, Clima R, Trerè D *et al.* Separated Siamese twins: intronic small nucleolar RNAs and matched host genes may be altered in conjunction or separately in multiple cancer types. *Cells* 2020;**9**:387. https://doi.org/10.3390/cells9020387
78. Xiao H, Feng X, Liu M *et al.* SnoRNA and lncSNHG: advances of nucleolar small RNA host gene transcripts in anti-tumor immunity. *Front Immunol* 2023;**14**:1143980. https://doi.org/10.3389/fimmu.2023.1143980
79. Deogharia M, Majumder M. Guide snoRNAs: drivers or passengers in human disease? *Biology* 2018;**8**:1. https://doi.org/10.3390/biology8010001
80. Huang Z, Du Y, Wen J *et al.* snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. *Cell Death Discov* 2022;**8**:259. https://doi.org/10.1038/s41420-022-01056-8
81. Weinstein LB, Steitz JA. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol* 1999;**11**:378–84. https://doi.org/10.1016/S0955-0674(99)80053-2
82. Kiss T. SnoRNP biogenesis meets pre-mRNA splicing. *Mol Cell* 2006;**23**:775–6. https://doi.org/10.1016/j.molcel.2006.08.023
83. Liang X, Liu Q, Fournier MJ. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 2009;**15**:1716–28. https://doi.org/10.1261/rna.1724409
84. Jack K, Bellodi C, Landry DM *et al.* rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Mol Cell* 2011;**44**:660–6. https://doi.org/10.1016/j.molcel.2011.09.017
85. Sloan KE, Warda AS, Sharma S *et al.* Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* 2017;**14**:1138–52. https://doi.org/10.1080/15476286.2016.1259781
86. Zhao Y, Rai J, Yu H *et al.* CryoEM structures of pseudouridine-free ribosome suggest impacts of chemical modifications on ribosome conformations. *Structure* 2022;**30**:983–92. https://doi.org/10.1016/j.str.2022.04.002
87. Dörner K, Ruggeri C, Zemp I *et al.* Ribosome biogenesis factors—from names to functions. *EMBO J* 2023;**42**:e112699. https://doi.org/10.15252/embj.2022112699
88. Zhao Y, Rai J, Li H. Regulation of translation by ribosomal RNA pseudouridylation. *Sci Adv* 2023;**9**:eadg8190. https://doi.org/10.1126/sciadv.adg8190
89. He J, Navarrete S, Jasinski M *et al.* Targeted disruption of *Dkc1*, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. *Oncogene* 2002;**21**:7740–4. https://doi.org/10.1038/sj.onc.1205969
90. Ge J, Rudnick DA, He J *et al.* Dyskerin Ablation in Mouse Liver Inhibits rRNA Processing and Cell Division. *Mol Cell Biol* 2010;**30**:413–22. https://doi.org/10.1128/MCB.01128-09
91. Pereboom TC, Weele LJv, Bondt A *et al.* A zebrafish model of dyskeratosis congenita reveals hematopoietic stem cell formation failure resulting from ribosomal protein-mediated p53 stabilization. *Blood* 2011;**118**:5458–65. https://doi.org/10.1182/blood-2011-04-351460
92. Zhang Y, Morimoto K, Danilova N *et al.* Zebrafish models for dyskeratosis congenita reveal critical roles of p53 activation contributing to hematopoietic defects through RNA processing. *PLoS One* 2012;**7**:e30188. https://doi.org/10.1371/journal.pone.0030188
93. Anchelin M, Alcaraz-Pérez F, Martínez CM *et al.* Premature aging in telomerase-deficient zebrafish. *Dis Model Mech* 2013;**6**:1101–12.
94. Bellodi C, McMahon M, Contreras A *et al.* H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell Rep* 2013;**3**:1493–502. https://doi.org/10.1016/j.celrep.2013.04.030
95. Delhermite J, Tafforeau L, Sharma S *et al.* Systematic mapping of rRNA 2'-O methylation during frog development and involvement of the methyltransferase Fibrillarin in eye and craniofacial development in *Xenopus laevis*. *PLoS Genet* 2022;**18**:e1010012. https://doi.org/10.1371/journal.pgen.1010012
96. Breznak SM, Peng Y, Deng L *et al.* H/ACA snRNP–dependent ribosome biogenesis regulates translation of polyglutamine proteins. *Sci Adv* 2023;**9**:eade5492. https://doi.org/10.1126/sciadv.ade5492
97. Ni C, Buszczak M. Ribosome biogenesis and function in development and disease. *Development* 2023;**150**:dev201187. https://doi.org/10.1242/dev.201187
98. Bellodi C, Krasnykh O, Haynes N *et al.* Loss of function of the tumor suppressor DKC1 perturbs p27 translation control and contributes to pituitary tumorigenesis. *Cancer Res* 2010;**70**:6026–35. https://doi.org/10.1158/0008-5472.CAN-09-4730
99. Bohnsack KE, Bohnsack MT. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J* 2019;**38**:e100278. https://doi.org/10.15252/embj.2018100278
100. Kampen KR, Sulima SO, Vereecke S *et al.* Hallmarks of ribosomopathies. *Nucleic Acids Res* 2020;**48**:1013–28. https://doi.org/10.1093/nar/gkz637
101. Kang J, Brajanovski N, Chan KT *et al.* Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. *Sig Transduct Target Ther* 2021;**6**:323. https://doi.org/10.1038/s41392-021-00728-8
102. Barozzi C, Zacchini F, Corradini AG *et al.* Alterations of ribosomal RNA pseudouridylation in human breast cancer. *NAR Cancer* 2023;**5**:zcad026. https://doi.org/10.1093/narcan/zcad026
103. Dong X-Y, Rodriguez C, Guo P *et al.* SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* 2008;**17**:1031–42. https://doi.org/10.1093/hmg/ddm375
104. Zhou F, Liu Y, Rohde C *et al.* AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat Cell Biol* 2017;**19**:844–55. https://doi.org/10.1038/ncb3563
105. Zhang X, Wang C, Xia S *et al.* The emerging role of snoRNAs in human disease. *Genes Dis* 2023;**10**:2064–81. https://doi.org/10.1016/j.gendis.2022.11.018
106. Cheng Y, Wang S, Zhang H *et al.* A non-canonical role for a small nucleolar RNA in ribosome biogenesis and senescence. *Cell* 2024. **187**:4770–89. https://doi.org/10.1016/j.cell.2024.06.019

107. Carlile TM, Rojas-Duran MF, Zinshteyn B *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 2014;**515**:143–6. https://doi.org/10.1038/nature13802
108. Schwartz S, Bernstein DA, Mumbach MR *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 2014;**159**:148–62. https://doi.org/10.1016/j.cell.2014.08.028
109. Li X, Zhu P, Ma S *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* 2015;**11**:592–7. https://doi.org/10.1038/nchembio.1836
110. Martinez NM, Su A, Burns MC *et al.* Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol Cell* 2022;**82**:645–59. https://doi.org/10.1016/j.molcel.2021.12.023
111. Nir R, Hoernes TP, Muramatsu H *et al.* A systematic dissection of determinants and consequences of snoRNA-guided pseudouridylation of human mRNA. *Nucleic Acids Res* 2022;**50**:4900–16. https://doi.org/10.1093/nar/gkac347
112. Rodell R, Robalin N, Martinez NM. Why U matters: detection and functions of pseudouridine modifications in mRNAs. *Trends Biochem Sci* 2024;**49**:12–27. https://doi.org/10.1016/j.tibs.2023.10.008
113. Zhang M, Jiang Z, Ma Y *et al.* Quantitative profiling of pseudouridylation landscape in the human transcriptome. *Nat Chem Biol* 2023;**19**:1185–95. https://doi.org/10.1038/s41589-023-01304-7
114. Sun H, Li K, Liu C *et al.* Regulation and functions of non-m6A mRNA modifications. *Nat Rev Mol Cell Biol* 2023;**24**:714–31. https://doi.org/10.1038/s41580-023-00622-x
115. Hamar R, Varga M. The role of post-transcriptional modifications during development. *Biol Futur* 2023;**74**:45–59. https://doi.org/10.1007/s42977-022-00142-3
116. Dai Q, Zhang L-S, Sun H-L *et al.* Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat Biotechnol* 2023;**41**:344–54. https://doi.org/10.1038/s41587-022-01505-w
117. Pederiva C, Trevisan DM, Peirasmaki D *et al.* Control of protein synthesis through mRNA pseudouridylation by dyskerin. *Sci Adv* 2023;**9**:344–54. https://doi.org/10.1126/sciadv.adg1805
118. Chikne V, Doniger T, Rajan KS *et al.* A pseudouridylation switch in rRNA is implicated in ribosome function during the life cycle of *Trypanosoma brucei*. *Sci Rep* 2016;**6**:25296. https://doi.org/10.1038/srep25296
119. Hebras J, Krogh N, Marty V *et al.* Developmental changes of rRNA ribose methylations in the mouse. *RNA Biol* 2020;**17**:150–64. https://doi.org/10.1080/15476286.2019.1670598
120. Marchand V, Pichot F, Neybecker P *et al.* HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res* 2020;**48**:e110. https://doi.org/10.1093/nar/gkaa769
121. Rajan KS, Madmoni H, Bashan A *et al.* A single pseudouridine on rRNA regulates ribosome structure and function in the mammalian parasite *Trypanosoma brucei*. *Nat Commun* 2023;**14**:7462. https://doi.org/10.1038/s41467-023-43263-6
122. Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol* 2012;**13**:355–69. https://doi.org/10.1038/nrm3359
123. Shi Z, Barna M. Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. *Annu Rev Cell Dev Biol* 2014;**31**:31–54.
124. Badrock AP, Uggenti C, Wacheul L *et al.* Analysis of U8 snoRNA variants in zebrafish reveals how bi-allelic variants cause leukoencephalopathy with calcifications and cysts. *Am Hum Genet* 2020;**106**:694–706. https://doi.org/10.1016/j.ajhg.2020.04.003
125. Pagano JFB, Locati MD, Ensink WA *et al.* Maternal- and somatic-type snoRNA expression and processing in zebrafish development. bioRxiv, https://doi.org/10.1101/858936, 6 February 2020, preprint: not peer reviewed.
126. Csabai L, Fazekas D, Kadlecsik T *et al.* SignaLink3: a multi-layered resource to uncover tissue-specific signaling networks. *Nucleic Acids Res* 2022;**50**:D701–9. https://doi.org/10.1093/nar/gkab909
127. Bradford YM, Slyke CEV, Ruzicka L *et al.* Zebrafish information network, the knowledgebase for Danio rerio research. *Genetics* 2022;**220**:iyac016. https://doi.org/10.1093/genetics/iyac016

Received: July 27, 2024. **Revised:** February 8, 2025. **Editorial Decision:** February 12, 2025. **Accepted:** February 14, 2025