

Databases and ontologies

# DASHR 2.0: integrated database of human small non-coding RNA genes and mature products

Pavel P. Kuksa<sup>1,†</sup>, Alexandre Amlie-Wolf<sup>1,2</sup>, Živadin Katanić<sup>1</sup>,  
Otto Valladares<sup>1</sup>, Li-San Wang<sup>1,2</sup> and Yuk Yee Leung<sup>1,\*†</sup>

<sup>1</sup>Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine and <sup>2</sup>Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

\*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

†The authors wish it to be known that these authors contributed equally.

Received on October 27, 2017; revised on May 31, 2018; editorial decision on August 14, 2018; accepted on August 20, 2018

## Abstract

**Motivation:** Small non-coding RNAs (sncRNAs, <100 nts) are highly abundant RNAs that regulate diverse and often tissue-specific cellular processes by associating with transcription factor complexes or binding to mRNAs. While thousands of sncRNA genes exist in the human genome, no single resource provides searchable, unified annotation, expression and processing information for full sncRNA transcripts and mature RNA products derived from these larger RNAs.

**Results:** Our goal is to establish a complete catalog of annotation, expression, processing, conservation, tissue-specificity and other biological features for all human sncRNA genes and mature products derived from all major RNA classes. DASHR (Database of small human non-coding RNAs) v2.0 database is the first that integrates human sncRNA gene and mature products profiles obtained from multiple RNA-seq protocols. Altogether, 185 tissues/cell types and sncRNA annotations and >800 curated experiments from ENCODE and GEO/SRA across multiple RNA-seq protocols for both GRCh38/hg38 and GRCh37/hg19 assemblies are integrated in DASHR. Moreover, DASHR is the first to contain both known and novel, previously un-annotated sncRNA loci identified by unsupervised segmentation (13 times more loci with 1 678 800 total). Additionally, DASHR v2.0 adds >3 200 000 annotations for non-small RNA genes and other genomic features (long-non-coding RNAs, mRNAs, promoters, repeats). Furthermore, DASHR v2.0 introduces an enhanced user interface, interactive experiment-by-locus table view, sncRNA locus sorting and filtering by biological features. All annotation and expression information directly downloadable and accessible as UCSC genome browser tracks.

**Availability and implementation:** DASHR v2.0 is freely available at <https://lisanwanglab.org/DASHRv2>.

**Contact:** [yylee@pennmedicine.upenn.edu](mailto:yylee@pennmedicine.upenn.edu)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

Recently, the study of small non-coding RNAs (sncRNAs) has expanded with the introduction of new RNA-seq protocols for profiling sncRNAs (Djebali *et al.*, 2012; Faridani *et al.*, 2016; Sloan

*et al.*, 2016) and generating large-scale genomics datasets (Sloan *et al.*, 2016). These include short total RNA-seq (Djebali *et al.*, 2012), miRNA-seq (Sloan *et al.*, 2016) and single cell small RNA-seq (Faridani *et al.*, 2016). Increasing evidence has shown that different kinds of sncRNAs play significant roles in regulating

important cellular processes and that dysfunctional sncRNAs are associated with a variety of human diseases, including neurodegenerative diseases and cancers (Goodarzi *et al.*, 2016; Li *et al.*, 2016; Martens-Uzunova *et al.*, 2013; Ng *et al.*, 2016; Salta and De Strooper, 2017; Soares and Manuel, 2017; Steinbusch *et al.*, 2017; Valen *et al.*, 2011). These sncRNAs include not only the commonly studied microRNAs, but also small nucleolar and small nuclear RNAs (sno/sncRNAs) (Steinbusch *et al.*, 2017), Piwi-interacting (piRNAs) (Ng *et al.*, 2016), transfer RNAs (tRNAs) (Goodarzi *et al.*, 2016; Li *et al.*, 2016), newly discovered classes such as tRNA fragments (Soares and Manuel, 2017), as well as sncRNAs derived from long non-coding RNAs (lncRNAs) (Martens-Uzunova *et al.*, 2013; Salta and De Strooper, 2017; Soares and Manuel, 2017) and promoter regions (Valen *et al.*, 2011). Thus, there is a strong need to systematically integrate and process expression data measuring diverse types of sncRNAs from different RNA-seq protocols and data sources including the sequencing read archive (SRA) (Kodama *et al.*, 2012) and ENCODE consortium (Djebali *et al.*, 2012).

The DASHR database aims to provide unified, searchable annotation and expression information for both primary sncRNA transcripts and mature RNA products and across eight major sncRNA classes including microRNAs (miRNAs), Piwi-interacting (piRNAs), small nuclear, nucleolar, cytoplasmic (sn-, sno-, scRNAs, respectively), transfer (tRNAs), tRNA fragments (tRFs) and ribosomal RNAs (rRNAs).

The current release of DASHR (v2.0) integrates >800 high-throughput sequencing datasets, both manually collected and curated from GEO/SRA (Kodama *et al.*, 2012) and from ENCODE (Djebali *et al.*, 2012; Sloan *et al.*, 2016), with over 22 billion reads. DASHR v2.0 contains >133 000 annotation records for small RNA

genes and mature sncRNA products and ~1 680 000 detected sncRNA loci across 185 tissues and cell types for both GRCh37/hg19 and GRCh38/hg38 genomes. For all sncRNAs, annotations and expression data can be searched, browsed and downloaded. DASHR v2.0 will aid the broader scientific community in exploring both the genomic landscape of sncRNA abundance and processing and individual sncRNAs across tissues cell types.

## 2 Materials and methods

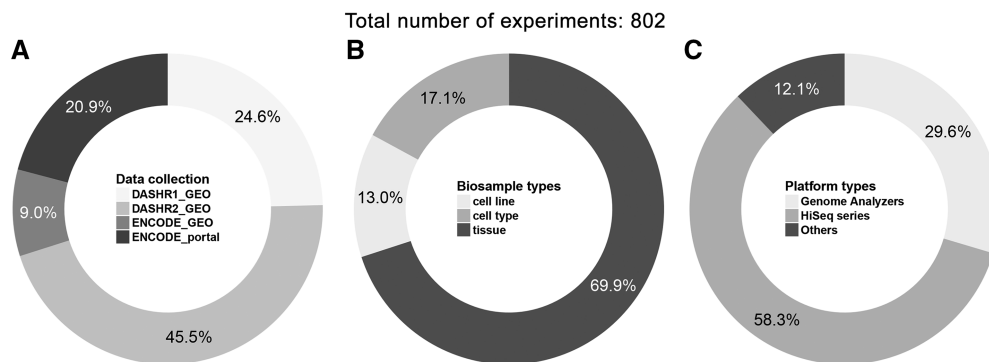
### 2.1 Database overview

Table 1 summarizes contents and features provided by DASHR v2.0. Some major new features and contents include:

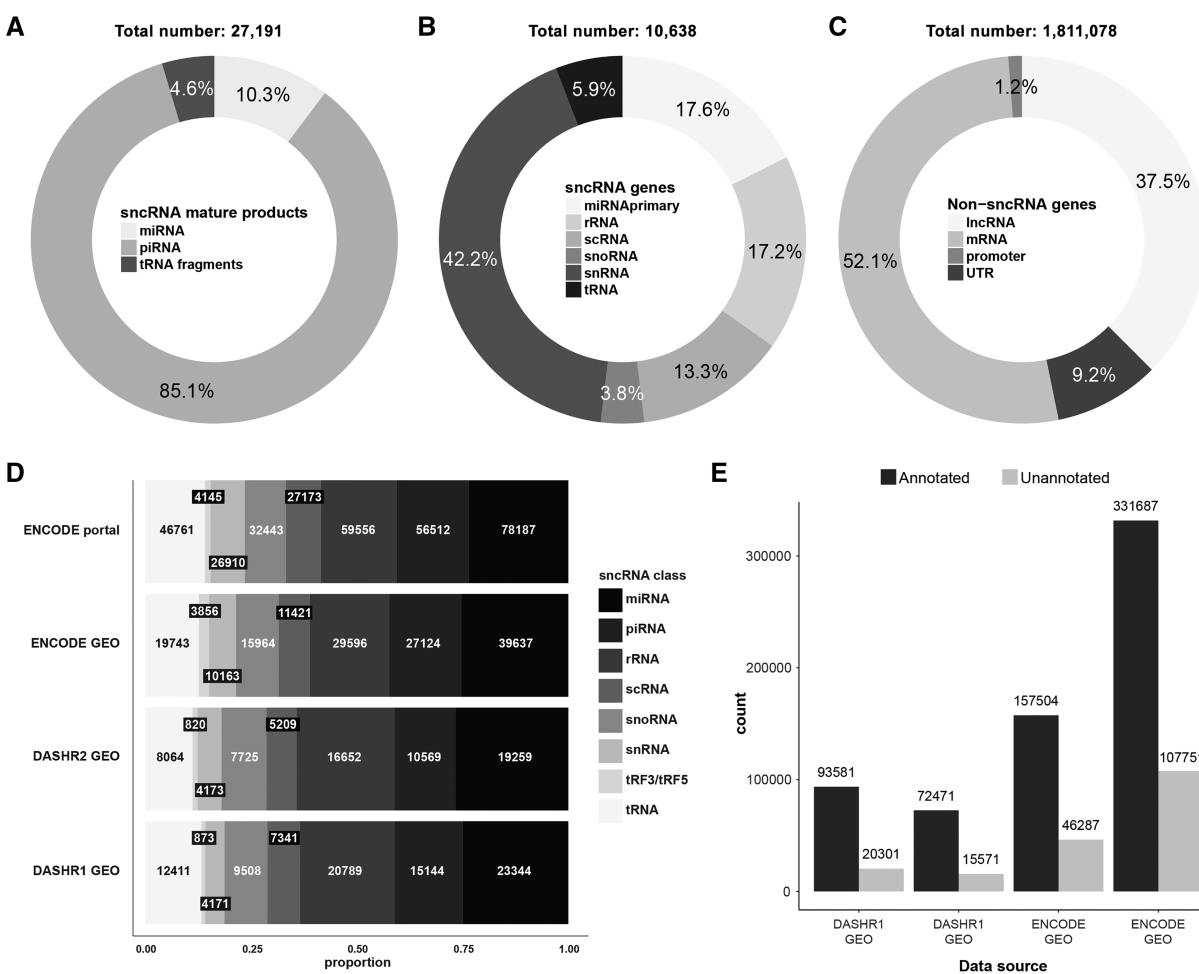
1. 365 more experiments across 34 smRNA-seq studies from GEO/SRA and integration of 240 short total RNA-seq experiments from ENCODE (Fig. 1A), increasing the coverage of various tissues/cell types (185 total) with a total of 802 integrated small RNA sequencing experiments;
2. sncRNA gene and mature product annotations for GRCh37/hg19 and GRCh38/hg38 (Fig. 2A–C) genome assemblies;
3. integration of biological features of sncRNAs including evolutionary conservation, co-localization with other genomic features, and tissue specificity;
4. integration of biological features of sncRNAs including evolutionary conservation, co-localization with other genomic features, and tissue specificity;
5. novel, previously unannotated sncRNA loci consistently detected across tissues and cell types (Fig. 2E);

**Table 1.** Advances and improvements provided by DASHR v2.0

Features	DASHR v1.0		DASHR v2.0	
	August 2015		September 2017	
Release date	August 2015		September 2017	
Genome Assembly	GRCh37/ hg19	GRCh38 / hg38	GRCh37/hg19	GRCh38/hg38
Data collection:	42	0	197 DASHR1-GEO 365 DASHR2-GEO	197 DASHR1-GEO 365 DASHR2-GEO
Curated GEO/SRA experiments				
Data collection:	0	0	72 ENCODE-GEO 168 ENCODE-portal	72 ENCODE-GEO 168 ENCODE-portal
ENCODING experiments				
sncRNA genes and mature products	48 075	0	68 135	65 156
Non-small RNA genes and mature products	0	0	1 469 297	1 811 078
Annotated sncRNA loci	84 514	0	DASHR1-GEO (90214) DASHR2-GEO (65650) ENCODE-GEO (159620) ENCODE-portal (335879)	DASHR1-GEO (93581) DASHR2-GEO (72471) ENCODE-GEO (157504) ENCODE-portal (331687)
Unannotated sncRNA loci	0	0	DASHR1-GEO (19207) DASHR2-GEO (14728) ENCODE-GEO (44157) ENCODE-portal (104192)	DASHR1-GEO (20301) DASHR2-GEO (15571) ENCODE-GEO (46287) ENCODE-portal (107751)
Biological features of sncRNAs	Expression and specificity		Expression, 5p specificity, conservation, tissue specificity, co-localization within regions of interest	
Enhanced web interface feature	–		Experiment-by-loci table per data collection – filter sncRNA products by features	
Compare hg19 and hg38 results	–		Allow users to compare the genomic contexts of sncRNAs in both hg19 and hg38	



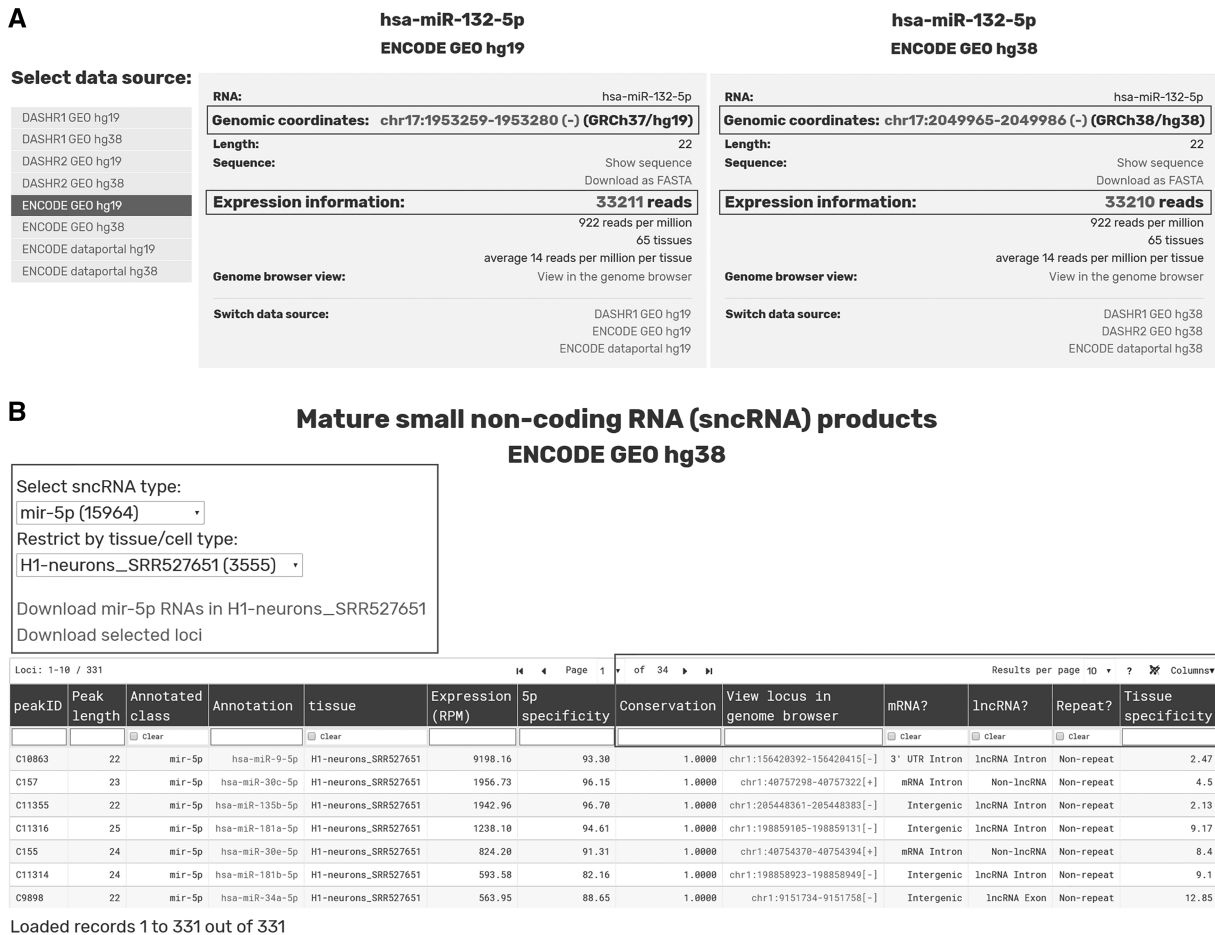
**Fig. 1.** Data collections in DASHR v2.0. (A) Number of integrated sncRNA experiments per data collection. (B) Types of biological samples included into DASHR v2.0. (C) Types of sequencing platforms used in experiments included in DASHR v2.0. For more details on experiments, please refer to [Supplementary Tables S2–S5](#)



**Fig. 2.** GRCh38/hg38 annotation and data collections in DASHR v2.0. (A) sncRNA mature product annotations. miRNA category includes miR-3p, miR-5p and other mature miRNAs. tRNA fragments include flanking tRF-3 and tRF-5 regions. (B) sncRNA gene annotations. (C) Annotations for non-sncRNA gene and other genomic elements. mRNA category includes mRNA introns and exons. For details of DASHR v2.0 annotation collections, please refer to [Supplementary Table S6](#). (D) The number of annotated sncRNA loci per RNA class in each DASHR v2.0 data collections. (E) The number of annotated and unannotated sncRNA loci in GRCh38/hg38 in each DASHR v2.0 data collections

6. the ability to compare data and annotations across GRCh37/hg19 and GRCh38/hg38 ([Fig. 3A](#));
7. interactively browse and filter sncRNA loci by one or more features, including expression, processing specificity, conservation

scores and tissue specificity ([Fig. 3B](#)); and an enhanced web interface ([Fig. 3A–B](#), [Supplementary Figs S1–S3](#)). DASHR v2.0 is substantially more comprehensive than existing non-coding RNA databases ([Supplementary Table S1](#)) ([Chung et al.](#),



**Fig. 3.** Comparison of sncRNA loci information between reference genome assemblies. **(A)** Example of sncRNA record page with annotation and expression information for annotated sncRNA loci (miR-132-5p) for GRCh37/hg19 and GRCh38/hg38 reference genomes in ENCODE-GEO data collection. The summary table displays genomic coordinates, RNA sequence, structural information, the length of the RNA locus, and summary statistics for the sncRNA's expression information across tissues/cell types in the selected data collection (Left panel). Refer to [Supplementary Figure S2](#) for a detail record page. Moreover, the table links to a UCSC genome browser view of the locus with DASHR 2.0 mapped sequencing data across all tissues and cell types. Users can quickly switch between DASHR 2.0 data sources using links provided in the 'Switch data source' section. **(B)** Newly designed interactive experiment-by-locus table browser for viewing and filtering sncRNA loci. Default table view contains expression (6th col), 5' specificity (7th col), conservation (8th col) and tissue specificity scores (13th/last col). Users can also filter sncRNA loci that co-localize with mRNA (10th col), lncRNA (11th col) or repeat regions (12th col). Additional columns can be added/removed for sorting and filtering (top right corner of the table). Users can download all sncRNA loci for selected RNA class and tissue and/or filtered set of loci ('Download' links above the table)

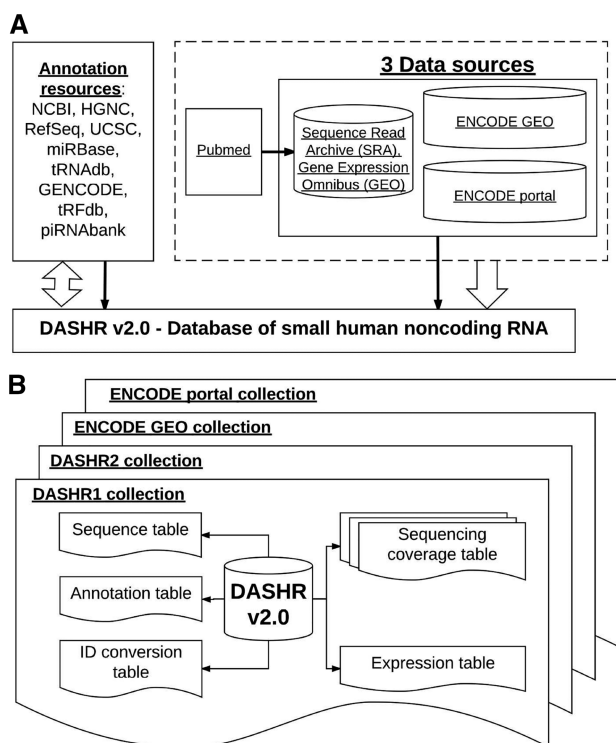
2017; Kozomara and Griffiths-Jones, 2014; The RNAcentral Consortium, 2017; Xie *et al.*, 2014; Zheng *et al.*, 2016) as it contains a more diverse set of human sncRNA class annotations. Additionally, DASHR v2.0 contains a significantly larger number of curated high-throughput smRNA-seq datasets in human tissues and cell types (over 800 libraries from multiple RNA-seq protocols with >22 billion total reads, [Supplementary Tables S2–S5](#)). DASHR v2.0 uniquely provides biological properties of both sncRNA genes and mature products including transcript processing specificity, conservation and the tissue-specificity of their expression ([Supplementary Table S1](#)). We describe the contents and features of DASHR v2.0 in the following.

## 2.2 Data collections

All smRNA-seq experiments integrated into DASHR v2.0 have been organized into four data collections (Table 1, [Supplementary Tables S2–S5](#), [Supplementary Methods](#)):

1. DASHR1-GEO data collection - consists of all 197 smRNA-seq experiments originally included in DASHR v1.0 ([Leung \*et al.\*, 2016](#)); to incorporate these experiments into DASHR 2.0, the raw sequencing reads were re-processed for both hg19 and hg38 genome builds and to include additional features introduced in DASHR 2.0 (Table 1; [Supplementary Methods](#));
2. DASHR2-GEO data collection - consists of 365 new Illumina smRNA-seq datasets curated from GEO/SRA (last curation date: August 2017) ([Kodama \*et al.\*, 2012](#));
3. ENCODE-GEO data collection – consists of all 72 short total RNA sequencing datasets (whole-cell) available in the 2012 ENCODE transcriptome data (GSE24565) ([Djebali \*et al.\*, 2012](#));
4. ENCODE-portal data collection - consists of all 168 small RNA-seq datasets from ENCODE portal ([Sloan \*et al.\*, 2016](#))

Figure 1 summarizes the contents of the DASHR v2.0 database in terms of the total number of experiments per data collection (Fig. 1A), biological sample types (Fig. 1B) and Illumina sequencing platforms used to generate sncRNA datasets (Fig. 1C). The DASHR



**Fig. 4.** DASHR v2.0 database implementation. **(A)** Overview of DASHR v2.0. Block arrows indicate data flow, while solid arrows indicate web traffic. **(B)** Organization of DASHR v2.0. Internally, each data collection in DASHR v2.0 is organized into tables containing annotation, expression, processing, sequence and other information

v2.0 data collection includes significantly more ( $3\times$  more, 605 new datasets) sncRNA experiments with greatly increased tissue/cell type diversity compared to DASHR v1.0 (Fig. 1A). Overall, 70% of the datasets in DASHR v2.0 were derived from experiments performed on tissues, with the remaining spanning 51 cell types and 48 cell lines.

All small RNA sequencing experiments were processed and integrated into DASHR v2.0 following our previously described approach (Leung *et al.*, 2013, 2016; Kuksa *et al.*, 2018). Thus, these sncRNA expression and processing information are comparable to each other (see Supplementary Methods). All data collections are available in both GRCh37/hg19 and GRCh38/hg38 reference genomes. Note that the sequencing experiments in DASHR1-GEO and DASHR2-GEO data collections were generated using the TruSeq Small RNA Library Preparation Kit (Illumina), while the ENCODE-GEO and ENCODE-portal experiments were generated using a different, short total RNA-seq protocol (Djebali *et al.*, 2012) (Fig. 1A, see Supplementary Methods).

### 2.3 Database contents

DASHR v2.0 contains 802 sequencing experiments comprising a total of 22 billion reads. Over 79 and 80% of the trimmed reads (i.e. reads that included a 3' adapter) were mapped to the GRCh37/hg19 and GRCh38/hg38 genomes, respectively. The adapters used and the mapping percentages (GRCh37/hg19 and GRCh38/hg38) for each dataset were summarized in Supplementary Tables S2–S5. In total, 833 647 (GRCh37/hg19) and 845 153 (GRCh38/hg38) sncRNA loci were identified (see Supplementary Methods) across all

datasets in all data collections (Fig. 2D for GRCh38/hg38,) and integrated into DASHR v2.0 database.

DASHR v2.0 gene and sncRNA mature product annotations have been updated to include:

DASHR v2.0 gene and sncRNA mature product annotations have been updated to include:

1. annotations for GRCh38/hg38 sncRNA mature products (Fig. 2A) and genes (Fig. 2B);
2. annotations for non-small RNA genes and other genomic elements for GRCh38/hg38 (Fig. 2C).

These make DASHR v2.0 the first sncRNA database that provides annotations for various sncRNA classes in one place across both human genome assemblies (Table 1, Supplementary Table S6). DASHR now contains 68 135 (GRCh37/hg19) and 65 156 (GRCh38/hg38) sncRNA gene records (precursor miRNAs, scRNAs, snRNAs, snoRNAs, tRNAs) and mature RNA product records (mature miRNAs, piRNAs, tRFs) (Table 1), as well as 1 469 297 (GRCh37/hg19) and 1 811 078 (GRCh38/hg38) annotations for non-small RNA genes and other genomic elements.

The processed smRNA-seq data across all data collections provides expression profiles for 83% (54 253) of all annotated sncRNA genes and mature products.

### 2.4 Biological features of sncRNAs in DASHR v2.0

In addition to the sncRNA features included in DASHR v1.0 (expression and specificity of 5' RNA cleavage) (Leung *et al.*, 2016), DASHR v2.0 incorporates new features to further characterize all sncRNA loci in a genome-wide manner (Table 1, Fig. 3B). These include evolutionary conservation scores, co-localization of sncRNA loci within specific genomic elements and genes, and tissue specificity scores.

#### 2.4.1 Evolutionary conservation

58% of piRNA loci, 18% of tRF loci and 79% of tRNA-derived sncRNA loci are conserved (phastCons  $> .5$ , see Supplementary Methods). Overall, 33% of sncRNA loci are evolutionary conserved.

#### 2.4.2 Co-localization with genomic features

We computed co-localization of the non-small RNA genes and other genomic elements (repeats, promoters, exons and introns) with each sncRNA locus (Supplementary Fig. S4). Co-localization information for each sncRNA locus includes the IDs and coordinates of each co-localized element. Overall, 32 and 27% of sncRNA loci are localized within lncRNA and mRNA genes, respectively, across all data collections.

#### 2.4.3 Tissue specificity

For each sncRNA locus in DASHR v2.0, the tissue specificity Q-score was computed at the study (i.e. tissue/cell type) level (Schug *et al.*, 2005). Q-score  $< 7$  generally indicates a tissue-specific sncRNA locus. On average, 13% of expressed tRFs, 9% of expressed tRNA-derived sncRNAs and 9% of expressed piRNAs in each tissue are tissue-specific (Supplementary Fig. S5). Overall, an average of 34% of expressed RNAs in each tissue/cell type are tissue-specific.

#### 2.4.4 New snRNA mature product annotations

In addition to eight major snRNA classes (miRNAs, piRNAs, sno-, scRNAs, tRNAs, tRFs and rRNAs) included in DASHR v1.0 (Leung *et al.*, 2016), DASHR v2.0 includes new snRNA annotations (Supplementary Methods) for tRNA-derived RNA fragments (tRFs) from the tRFdb database (Kumar *et al.*, 2015) and 719 annotations for newly described snoRNA mature products from snoRNAome database (Jorjani *et al.*, 2016). The piRNA annotations have been expanded to include piRNAbank database records (Sai Lakshmi and Agrawal, 2008) with a total of 26 649 (GRCh37/hg19) and 23 116 (GRCh38/hg38) piRNAs.

On average, around 200 tRF loci from tRFdb are present in each tissue/cell type in DASHR v2.0, with 89% (897) of annotated tRFdb tRF fragments detected in at least one DASHR v2.0 tissue/cell type. 50% (356) of the newly described snoRNA mature products are present in at least one of the datasets, with 72 snoRNA loci in each dataset on average. For piRNAs, 40% (9517) of annotated piRNA loci have been detected in one or more of DASHR v2.0 datasets.

#### 2.4.5 Inclusion of both annotated and previously unannotated snRNA loci

As new snRNA classes and loci continue to be discovered (Martens-Uzunova *et al.*, 2013; Röther and Meister, 2011; Salta and De Strooper, 2017), integration of both annotated and novel, previously unannotated snRNA loci in DASHR v2.0 will provide a unique resource for RNA researchers for further experimental studies and validation. Therefore, in DASHR v2.0, we characterized snRNA loci by identifying snRNA peaks using unsupervised segmentation followed by the annotation of the detected peaks (Supplementary Methods) and incorporated both annotated and unannotated snRNA loci in each data collection. Figure 2D shows the number of snRNA loci per snRNA class, and Figure 2E, the number of unannotated snRNA loci in each DASHR v2.0 data collection for GRCh38/hg38. Note that the composition of snRNA classes (Fig. 2D) and the length of the unannotated snRNA loci (Supplementary Fig. S6) vary across data collections, as different RNA-seq protocols were used in DASHR1-/DASHR2-GEO compared to ENCODE experiments.

20% of annotated snRNA loci have highly specific 5' start positions (5' specificity > .9), and 51% of previously unannotated snRNA loci also have highly specific 5' start positions. 31% of the annotated loci in DASHR v2.0 are highly conserved (100-way phastCons > .8), and 24% of the annotated loci are tissue-specific (Q-score < 7). The unannotated snRNA loci display similar characteristics: 21% are conserved (phastCons > .5) and 44% of these are tissue-specific.

#### 2.4.6 Non-small RNA genes and other genomic features

To allow researchers to identify snRNAs from specific genomic elements or regions of interest such as lncRNAs or promoters of mRNA genes, DASHR v2.0 incorporates annotations for non-small RNA genes [lncRNAs based on LNCipedia 4.1 (Volders *et al.*, 2015) and mRNAs] and other genomic elements including promoters and repeat elements. Information on co-localization of snRNA loci with these genomic elements has been incorporated into DASHR v2.0 for each snRNA locus (Fig. 3B).

The annotated snRNA loci in DASHR v2.0 derive from a variety of genomic elements, including mRNA introns (19%), intergenic lncRNAs (24%), promoters (1.3%), UTRs (5%) and intergenic regions (48%).

Novel, previously unannotated snRNA loci are similarly distributed across various types of genomic regions: intronic regions (16%), exonic regions (5%), intergenic lncRNAs (12%), promoters (1.6%), UTRs (10%) and intergenic regions (55%).

## 3 Results

DASHR aims to provide a simple and unified resource to the scientific community allowing users to

1. query and compare the expression and processing information for snRNA genes and mature snRNA products of interest ('Search by snRNA name/ID', Supplementary Figs S1 and S2, Fig. 3A);
2. retrieve annotations across snRNA classes simultaneously ('Search by genomic coordinates', Supplementary Figs S1 and S2);
3. view expression data for any genomic interval ('View genomic region', Supplementary Figs S1 and S2);
4. screen for specifically/alternatively processed, conserved, or tissue-specific snRNA loci (Fig. 3B);
5. browse and download all or a selected/filtered set of snRNAs and mature snRNA products for specific human tissues or cell types ('Browse', Fig. 3B);
6. view snRNA loci, expression, annotation data as UCSC genome browser tracks (Supplementary Fig. S7).

Several of the above features are new in DASHR v2.0. First, DASHR v2.0 introduces a 'Browse' function, enabling users to first select a data collection, then select the experiments and/or snRNA classes they want to view (Fig. 3B). All snRNA loci records are then viewed in an interactive table that allows users to quickly filter and locate their snRNAs of interest using one or more features including chromosomal position, snRNA length, expression value, 5' processing specificity, tissue specificity and evolutionary conservation scores. Each snRNA record in the table can be viewed in DASHR or in the UCSC genome browser (Fig. 3B). DASHR v2.0 also allows users to identify snRNA loci co-localized within specific genomic elements (e.g. lncRNA, mRNA, promoter, UTR, etc).

In addition to genome-wide raw coverage tracks, users can now download all detected snRNA loci in each data collection along with their various features including per-tissue expression, conservation, specificity of snRNA processing and tissue specificity scores (Supplementary Fig. S3, see Section 5).

## 4 Conclusions

The current release (v2.0) of DASHR is from September 2017 and is freely available for use at <http://lisanwanglab.org/DASHRv2>. This release improves the number of curated and processed experiments by 4-fold (total: 802 small RNA experiments) and increases the number of tissues/cell types represented in the database for a total of 185 tissues/cell types (Fig. 1). DASHR v2.0 has been expanded to include annotations for both GRCh37/hg19 and GRCh38/hg38 genomes. Importantly, DASHR v2.0 includes 372 194 previously unannotated snRNA loci consistently expressed across tissues and cell types. Each snRNA locus is annotated with additional biological information including conservation scores, co-localized genomic elements and tissue specificity scores. DASHR v2.0 will continue to aid the broader scientific community in exploring individual snRNA loci and the genome-wide landscape of small RNA abundance and processing across tissues and cell types. We plan to

continuously increase and update the features and data available through this database by curating GEO/SRA, integrating data from large-scale genomic studies, e.g. FANTOM5 (Kawaji *et al.*, 2017), and data generated from different small RNA-seq protocols focusing on small RNAs such as miRNA-seq (Djebali *et al.*, 2012) and single cell small RNA-seq (Faridani *et al.*, 2016).

## 5 Availability

The database is freely available at <https://lisanwanglab.org/DASHRv2>. Currently, users can download all resources in DASHR v2.0 from the 'Download' page (Supplementary Fig. S3, Fig. 4B). These include:

1. snRNA loci data table summarizing expression and other features of each loci in each tissue;
2. Annotation table for all snRNA entries and mature products;
3. Sequence table containing raw RNA sequences for all snRNA entries in DASHR;
4. ID conversion table, which contains cross-referenced IDs for each snRNA entry;
5. Expression tables with raw read counts (RAW) and reads per million (RPM) across all tissues in DASHR;
6. Sequencing coverage in bedGraph format for each tissue;
7. Files with the detailed information on all smRNA-seq experiments included in each data collection in DASHR.

## Acknowledgements

The authors thank Wang lab members for their helpful input on this project.

## Funding

This work is supported by the National Institute of General Medical Sciences (R01-GM099962 to all co-authors); National Institute on Aging (U54-AG052427, U01-AG032984, U24-AG041689, U01-AG047133 to P.P.K., Z.K., O.V., L.S.W. and Y.Y.L.), National Institute on Aging (T32-AG00255 to A.A.W.). Funding for open access charge: National Institute on Aging (U24-AG041689).

*Conflict of Interest:* none declared.

## References

Chung,I.-F. *et al.* (2017) YM500v3: a Database for small RNA sequencing in human cancer research. *Nucleic Acids Res.*, **45**, D925–D931.  
 Djebali,S. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.  
 Faridani,O.R. *et al.* (2016) Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.*, **34**, 1264–1266.  
 Goodarzi,H. *et al.* (2016) Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell*, **165**, 1416–1427.

Jorjani,H. *et al.* (2016) An Updated Human snoRNAome. *Nucleic Acids Res.*, **44**, 5068–5082.  
 Kawaji,H. *et al.* (2017) The FANTOM5 collection, a data series underpinning mammalian transcriptome atlases in diverse cell types. *Sci. Data*, **4**, 170113.  
 Kodama,Y. *et al.* (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.  
 Kozomara,A. and Griffiths-Jones,S. (2014) MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.  
 Kuksa,P.P. *et al.* (2018) SPAR: small RNA-seq portal for analysis of sequencing experiments. *Nucleic Acids Res.*, **46**, W36–W42.  
 Kumar,P. *et al.* (2015) tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.*, **43**, D141–D145.  
 Leung,Y.Y. *et al.* (2016) DASHR: database of small human noncoding RNAs. *Nucleic Acids Res.*, **44**, D216–D222.  
 Leung,Y.Y. *et al.* (2013) CoRAL: predicting non-coding RNAs from small RNA-sequencing data. *Nucleic Acids Res.*, **41**, e137.  
 Li,Q. *et al.* (2016) tRNA-derived small non-coding RNAs in response to ischemia inhibit angiogenesis. *Sci. Rep.*, **6**, 20850.  
 Martens-Uzunova,E.S. *et al.* (2013) Beyond microRNA – novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett.*, **340**, 201–211.  
 Ng,K.W. *et al.* (2016) Piwi-interacting RNAs in cancer: emerging functions and clinical utility. *Mol. Cancer*, **15**, 5.  
 Röther,S. and Meister,G. (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie*, **93**, 1905–1915.  
 Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.  
 Salta,E. and De Strooper,B. (2017) Noncoding RNAs in neurodegeneration. *Nat. Rev. Neurosci.*, **18**, 627–640.  
 Schug,J. *et al.* (2005) Promoter features related to tissue specificity as measured by Shannon Entropy. *Genome Biol.*, **6**, R33.  
 Sloan,C.A. *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–D732.  
 Soares,A.R. and Manuel,S. (2017) Discovery and function of transfer RNA-derived fragments and their role in disease. *Wiley Interdiscip. Rev. RNA*, **8**, e1423.  
 Steinbusch,M.M. *et al.* (2017) Serum snoRNAs as biomarkers for joint ageing and post traumatic Osteoarthritis. *Sci. Rep.*, **7**, 43558.  
 The RNAcentral Consortium. (2017) RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.*, **45**, D128–D134.  
 Valen,E. *et al.* (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat. Struct. Mol. Biol.*, **18**, 1075–1082.  
 Volders,P.J. *et al.* (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.*, **43**, D174–D180.  
 Xie,C. *et al.* (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.*, **42**, D98–103.  
 Zheng,L.L. *et al.* (2016) deepBase v2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.*, **44**, D196–D202.