D220–D225 *Nucleic Acids Research, 2020, Vol. 48, Database issue* *Published online 10 October 2019*
doi: 10.1093/nar/gkz884

snoDB: an interactive database of human snoRNA sequences, abundance and interactions

Philia Bouchard-Bourelle[1], Clément Desjardins-Henri[1], Darren Mathurin-St-Pierre[1], Gabrielle Deschamps-Francoeur[1], Étienne Fafard-Couture[1], Jean-Michel Garant[1], Sherif Abou Elela[2] and Michelle S. Scott[1,*]

[1]Département de biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec J1E 4K8, Canada and [2]Département de microbiologie et infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec J1E 4K8, Canada

Received August 10, 2019; Revised September 13, 2019; Editorial Decision September 26, 2019; Accepted October 01, 2019

ABSTRACT

Small nucleolar RNAs (snoRNAs) are an abundant type of non-coding RNA with conserved functions in all known eukaryotes. Classified into two main families, the box C/D and H/ACA snoRNAs, they enact their most well characterized role of guiding site specific modifications in ribosomal RNA, through the formation of specific ribonucleoprotein complexes, with fundamental implications in ribosome biogenesis. However, it is becoming increasingly clear that the landscape of snoRNA cellular functionality is much broader than it once seemed with novel members, non-uniform expression patterns, new and diverse targets as well as several emerging non-canonical functions ranging from the modulation of alternative splicing to the regulation of chromatin architecture. In order to facilitate the further characterization of human snoRNAs in a holistic manner, we introduce an online interactive database tool: snoDB. Its purpose is to consolidate information on human snoRNAs from different sources such as sequence databases, target information, both canonical and non-canonical from the literature and from high-throughput RNA–RNA interaction datasets, as well as high-throughput sequencing data that can be visualized interactively.

INTRODUCTION

Small nucleolar RNAs (snoRNAs) are a conserved class of non-coding RNAs found in all eukaryotes and most extensively characterized as guiding site specific post-transcriptional modifications in ribosomal RNA (rRNA) (1,2). In addition, a small number of additional snoRNAs such as SNORD3 and SNORD118 are known to play a role in the processing and maturation of rRNA. Two types of snoRNAs have been described: box C/D and box H/ACA snoRNAs, the majority of which are encoded in introns of host genes in human (1,3). Box C/D and box H/ACA snoRNAs respectively guide the 2′-*O*-methylation and the pseudouridylation of their targets by direct base pairing. To do so, they require the interaction of core binding proteins, which provide stability and the catalytic activity, forming complexes known as snoRNPs (snoRNA ribonucleoprotein complexes) (4). In human, 110 rRNA residues are known to be methylated by snoRNPs and 100 are pseudouridylated (5) although recent high-throughput sequencing and systematic comparative genomics efforts have identified additional likely candidates as well as positions that are fractionally modified (6–9).

While canonical features, functionality and targets of snoRNAs are well-characterized, over the past decade, an increasingly large literature has exposed novel and unexpected aspects of snoRNA biology. High-throughput sequencing approaches give indications that snoRNAs can modify and/or otherwise interact with diverse RNAs including other snoRNAs, transfer RNAs and messenger RNAs (mRNAs) (6,10–13). As reviewed in (14), recent years have seen many potential novel functions being reported for snoRNAs including the modulation of alternative splicing (15–17), an essential involvement in stress response pathways (18–20), the regulation of pre-mRNA stability (21) and the modulation of mRNA 3′ end processing (22). Moreover, high-throughput sequencing approaches and computational pipelines addressing the unique challenges of snoRNAs have been elaborated, resulting in more accurate quantification of snoRNAs, and simultaneously of their host genes, indicating that the levels of expression of snoRNAs cover a wide range and do not always mirror those of their host gene (23–25). The improved quantification and increased characterization of snoRNAs has led to increasing numbers of snoRNAs and their host genes found to be involved in diseases. Examples of pathologies in which snoRNAs and their host genes play an important role and could be prime therapeutic targets include the Prader-

*To whom correspondence should be addressed. Tel: +1 819 821 8000 (Ext. 72123); Email: michelle.scott@usherbrooke.ca

Willi syndrome and diverse cancers (26–30). However, in many cases, while the involvement of snoRNAs in disease is now known, the molecular mechanism is unclear. Such is the case for SNORD118, mutations of which affect the expression, processing and protein binding of the snoRNA. But while SNORD118, like most snoRNAs, is ubiquitously expressed, germline mutations cause specific neurological phenotypes (31). The wealth of knowledge and data describing snoRNA biology requires careful management and integration to facilitate easy access and assimilation by the community. Unfortunately, much of the information regarding snoRNAs is disorganized, disseminated through disparate online platforms and peppered in the literature. For example, many RNA–RNA interactions have been detected for SNORD118 (11–13) and could be important to characterize the molecular mechanism of its involvement in disease (Supplementary Figure S1), but mining them from high-throughput datasets from the literature is not straightforward. A central snoRNA resource would considerably facilitate the characterization of snoRNA functionality and involvement in disease.

Three dedicated snoRNA resources are currently available for human: snoRNAbase (5), snOPY (32) and snoRNA Atlas (33). However, these resources have either not been kept up to date with the new snoRNA genes annotated, new interactors and functionalities, or have a different scope (e.g. snOPY is a database of snoRNA orthology). With so many key regulatory features emerging as intrinsic snoRNA functions, there is a pressing need to unify the scattered data currently available on human snoRNAs in order to optimize future research endeavors. The online interactive snoRNA database we propose, snoDB, aims to do that and more. Indeed, integrating available data is of great importance but snoDB further aspires to consolidate the above information with curated peer-reviewed high-throughput data in an effort to lead and incite research in the further characterization of the human snoRNA landscape in health and disease.

DATABASE CONTENT

SnoDB is based on the human hg38 reference genome assembly. It aims to be inclusive and integrate gene annotations and a wide diversity of features from all relevant available databases (Table 1). SnoRNA gene annotations were obtained from RefSeq (34), Ensembl (35) and RNAcentral (36), which in turn provides annotations from snOPY (32) and Rfam (37). Careful manual curation was carried out to consolidate the annotations and to ensure no snoRNA entries share exact same genomic coordinates. When different names are employed for a given snoRNA gene, the RefSeq name was used by default, but if absent, the RNAcentral or the Ensembl names were used. In addition to the gene symbol, genomic coordinates and gene sequence, all additional names obtained from the HUGO Gene Nomenclature Committee (HGNC) (38) are available in the 'synonym' column, and all identifiers of all above databases are provided as links. SnoDB houses 2064 human snoRNAs, integrating the annotations of the above databases. In contrast, the other main snoRNA-centric resources, snoRNAbase (5), snOPY (32) and snoRNA Atlas (33), contain respectively 402, 760 and 1118 human snoRNAs (Table 1).

The snoRNA features that are available for display in snoDB also include host gene characteristics with a link to the Ensembl entry, the biotype, synonyms if relevant and genomic coordinates. In addition, snoDB features conservation data from snoRNA Atlas (33), orthology data from snOPY (32), snoRNA target data with enrichment details in select tissues from the human protein atlas (39) when available and expression data (Tables 1 and 2). Target data include known targets in rRNA annotated in snoRNAbase (5) and rRNA targets confirmed by RiboMethSeq (8). Non-canonical interactors that were experimentally validated in the literature are also included and links to the articles are available. These studies include (11,15–17,21), as described in the Introduction. Finally, RNA–RNA interaction data were incorporated from the RISE:RNA Interactome, a database compiling results from multiple high-throughput RNA–RNA interaction studies (40) with the name and biotype of all RISE interactors being available. Levels of abundance of both snoRNAs and their host gene measured in various human tissues and cell lines using a low structure bias RNA-seq approach are also available as obtained from (23) and GEO entries from GSE126797. The snoDB back-end is built in PostgreSQL (9.5.1) as a relational database which is integrated into the Django web framework (1.6.5).

WEB INTERFACE

The main page of snoDB is divided into four sections: (i) As shown in Figure 1A, the top of the page displays snoDB's logo adjacent to a search engine for snoRNA names. Immediately below the logo, a switch allows to toggle snoDB's sister tool snoTHAW (snoDB Table Heatmap Arrangement Widget), which enables the interactive visualization of abundance values of snoRNAs and their host gene. To the right of the logo can be found links to additional information pages on the database in the 'About', 'Tutorial', 'Statistics' and 'Experiment details' sections, as well as a link for downloading of the whole database. (ii) The section directly below (shown in Figure 1B) features a menu bar with options related to the table. Clicking on 'Column Options' reveals a set of buttons with 3 kinds of functionalities: toggling the visibility of single columns using the column visibility button, toggling the visibility of column groups using the color-coded buttons, and downloading data in either TSV, BED or XLSX file formats based on currently visible or selected rows in the table. The 'Advanced Search' option reveals 5 search boxes (shown in Figure 1B) that are specific to certain groups of columns as noted by their placeholder text and outline colors. The 'Reset Filters' option erases all filtering currently active on the table, whether it is from the topmost main search, the advanced search bars or the column specific search boxes in the table itself. This option, along with the 'Refresh Table' option that follows it, exist because the state of all search inputs, column visibilities and row selections are saved upon refreshing the page. Hence, 'Reset Filters' facilitates the clearing of all search fields without needing to refresh the page while 'Reset Tables' reloads the page back to its default state. (iii) Below

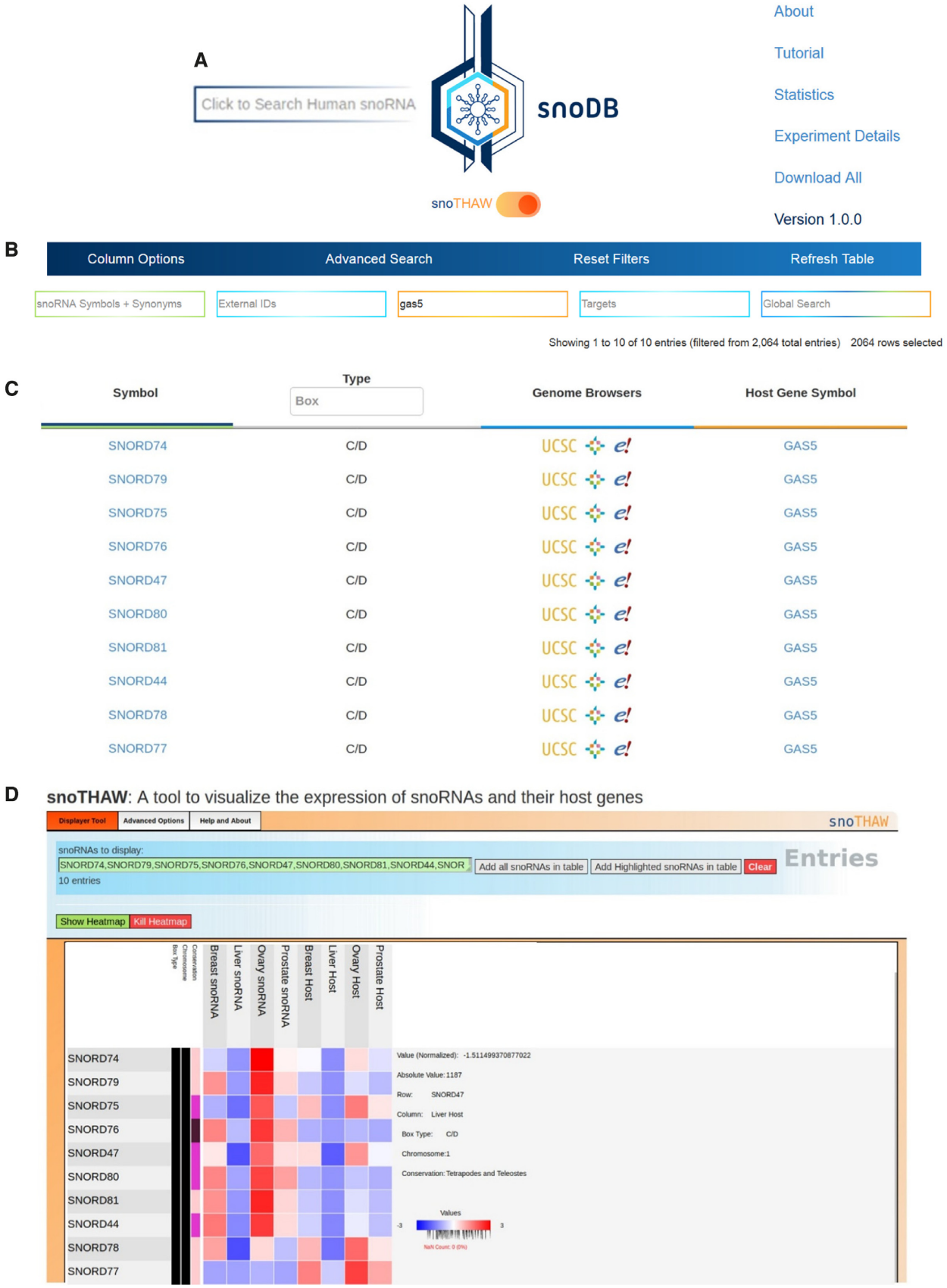

Figure 1. Screenshot of the main page of snoDB displaying the site's four sections. (**A**) The snoDB logo, basic search engine and links to information pages. (**B**) A menu bar with options to control the content and appearance of the table. (**C**) snoDB's main table where data are displayed and can be interacted with. By default, all 2064 snoRNA entries are shown by scrolling down. (**D**) The snoTHAW interface with the heatmap visualization beneath.

Table 1. Features of human snoRNA databases

Database	snoRNA count	Links to external resources	Orthology (O) and conservation (C)[a]	Host gene characteristics[b]	rRNA and snRNA target data	Non-canonical target data[c]	snoRNA expression data[d]	Host gene expression data[d]	Data available for download
snoRNAbase (5)	402	UCSC Genome Browser hg18 HGNC Genbank Literature	O (to yeast)	NCA	√	L	-	-	-
snOPY (32)	760	Refseq	O	N	√	-	-	-	-
snoRNA Atlas (33)	1118	Rfam	C	N	√	-	E	-	√
snoDB	2064	UCSC Genome Browser hg38 RefSeq HGNC Ensembl RNAcentral NCBI Rfam snoRNAbase snOPY snoRNA Atlas RISE database Literature	OC[d]	NBCA	√	LR	OPTLS	OPTLS	√

[a]In snoDB, links are provided to snOPY and Ensembl orthology pages when available and conservation data were obtained from snoRNA Atlas.
[b]Host gene characteristics: N: name; B: biotype; C: genomic coordinates; A: biological process annotation.
[c]Non-canonical target data are supported by articles in the literature (L) and by links to the RISE database (R).
[d]For snoRNA Atlas: E indicates amalgamated expression values from ENCODE. For snoDB: all expression values were obtained using the low structure bias TGIRT-seq methodology. O: normal human ovary; P: normal human prostate; T: normal human testis; L: normal human liver; S: SKOV3ip1 human ovarian carcinoma cell line.

Table 2. Characteristics of snoRNAs in snoDB

	Box C/D	Box H/ACA	Other	Total
All snoRNAs[a]	1391	651	22	2064
Distinct snoRNA symbols[b]	461	246	21	728
Intronic snoRNAs encoded in host genes	423	318	3	744
Intergenic snoRNAs	968	333	19	1320
snoRNA-target pairs	1471	616	31	2118
• snoRNA-rRNA target pairs	481	255	2	738
• snoRNA-snRNA target pairs	113	64	7	184
• snoRNA-non-canonical target pairs[c]	877	297	22	1196
snoRNAs with transcriptomic data	524	469	3	996

[a]All snoRNAs include snoRNAs with the same name and/or sequence but encoded in different genomic loci.
[b]Counts every snoRNA symbol only once. Some snoRNAs bear the same symbol but have different IDs based on differences in their sequence and in the loci in which they are encoded or the length of their sequence.
[c]Non-canonical targets of snoRNAs include mRNAs and genomic regions not known to encode annotated genes.

the options menu, the main table dynamically displays the snoDB data (Figure 1C). (iv) The bottom of the page reveals snoTHAW when the switch at the top of the snoDB page is toggled. SnoTHAW enables the visualization and interaction of RNA-seq expression data contained within snoDB (Figure 1D and Supplementary Figure S2). Currently, expression data are displayable for four healthy tissues (breast, liver, ovary and prostate) as well as the SKOV3ip1 ovarian cancer cell lines. In addition, box type, chromosome and conservation data also found in snoDB can be displayed on the heatmap's y-axis with the ability to re-order the columns and rows based on these features or based on the expression data to suit the user's needs. All available expression data in snoDB was generated using the TGIRT-seq approach which allows accurate quantification and comparison of all cellular RNAs including highly structured and modified RNAs such as snoRNA (23–25), as described above. As more such datasets become available, they will also be incorporated in snoDB.

The main page features three levels of querying capabilities. The first consists of a single search-box which lies to the left of the snoDB logo atop the page (Figure 1A). Clicking and/or typing into this area reveals a drop-down menu comprised of all snoRNA symbols which reside in the table's first column of the same name. Multiple symbols can be selected making this a quick and easy way to access information on a few snoRNAs of interest. The second consists of the five previously mentioned search boxes located above the table upon clicking on the 'Advanced search' option. From left to right, the first one searches through the snoRNA symbols and synonyms columns, the second through all the external ID columns, the third through host symbols and synonyms, the fourth through target columns and the fifth and final search box is a global search covering the entire snoDB dataset. The first four search boxes operate on an exact-match basis while the global search supports partial search terms. All five search boxes support regular expressions as well as multiple space-separated terms making copy-pasting columns from a spreadsheet into an appropriate search engine an easy way to view numerous specific snoRNA entries. The third searching strategy is found within the table itself and provides individual column

searching capabilities on select columns and it also supports multiple inputs.

In addition to the interactive viewing and querying of columns, snoDB's main table contains the following features: a frozen first column for seamless horizontal scrolling through many columns, row selection upon click for visual highlights and as a means of input into snoTHAW, drag-and-drop column re-ordering, column sorting and an abundance of external links to corresponding snoRNA entries in other databases. All of these functionalities are described in the 'About' page as well as through interactive examples in the Tutorial (Supplementary Figure S3).

While having all data selectively displayable in a single interactive table is a great convenience, it can also be impractical when one wishes to view all data for a single entry without needing to horizontally scroll back and forth. Therefore clicking on any snoRNA in the 'Symbol' column opens a new tab to a page displaying all available information on that entry in a vertical format (Supplementary Figure S4). These individual data hubs are divided into familiar sub-sections and feature external links to all previously mentioned sources along with additional links for interaction data, all of which can be searched through using the individual column search engines present.

CONCLUSION AND FUTURE PLANS

The snoDB interactive web application is a holistic relational database which consolidates diverse information regarding human snoRNAs from key sources, curated articles and datasets in an attempt to facilitate further research in the field of snoRNAs. Along with minor periodic updates, additional high-throughput datasets will be incorporated in snoDB as they become available.

DATA AVAILABILITY

http://scottgroup.med.usherbrooke.ca/snoDB/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to members of their groups, and in particular the Scott and Abou Elela groups for useful discussions and to Leandro Fequino for technical support. M.S.S. and S.A.E. are members of the RNA group and the Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS).

FUNDING

Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant (to MSS); Fonds de Recherche du Québec – Nature et technologies (FRQNT) team Research grant (to M.S.S. and S.A.E.); Fonds de Recherche du Québec - Santé (FRQS) Research Scholar Junior 2 Career Award (to M.S.S.); NSERC doctoral scholarship (to G.D.F.); undergraduate scholarship from the Faculty of Medicine and Health Sciences of the Université de Sherbrooke (to C.D.H.); Canada Research Chair in RNA Biology and Cancer Genomics (to S.A.E.). Funding for open access charge: Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant (to M.S.S.).
Conflict of interest statement. None declared.

REFERENCES

1. Dieci,G., Preti,M. and Montanini,B. (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics*, **94**, 83–88.
2. Filipowicz,W. and Pogacic,V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.*, **14**, 319–327.
3. Boivin,V., Deschamps-Francoeur,G. and Scott,M.S. (2018) Protein coding genes as hosts for noncoding RNA expression. *Semin. Cell Dev. Biol.*, **75**, 3–12.
4. Massenet,S., Bertrand,E. and Verheggen,C. (2017) Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol.*, **14**, 680–692.
5. Lestrade,L. and Weber,M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
6. Gumienny,R., Jedlinski,D.J., Schmidt,A., Gypas,F., Martin,G., Vina-Vilaseca,A. and Zavolan,M. (2017) High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq. *Nucleic Acids Res.*, **45**, 2341–2353.
7. Kehr,S., Bartschat,S., Tafer,H., Stadler,P.F. and Hertel,J. (2014) Matching of Soulmates: coevolution of snoRNAs and their targets. *Mol. Biol. Evol.*, **31**, 455–467.
8. Krogh,N., Jansson,M.D., Hafner,S.J., Tehler,D., Birkedal,U., Christensen-Dalsgaard,M., Lund,A.H. and Nielsen,H. (2016) Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.*, **44**, 7884–7895.
9. Incarnato,D., Anselmi,F., Morandi,E., Neri,F., Maldotti,M., Rapelli,S., Parlato,C., Basile,G. and Oliviero,S. (2017) High-throughput single-base resolution mapping of RNA 2-O-methylated residues. *Nucleic Acids Res.*, **45**, 1433–1441.
10. Kishore,S., Gruber,A.R., Jedlinski,D.J., Syed,A.P., Jorjani,H. and Zavolan,M. (2013) Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol.*, **14**, R45.
11. Sharma,E., Sterne-Weiler,T., O'Hanlon,D. and Blencowe,B.J. (2016) Global Mapping of Human RNA–RNA Interactions. *Mol. Cell*, **62**, 618–626.
12. Lu,Z., Zhang,Q.C., Lee,B., Flynn,R.A., Smith,M.A., Robinson,J.T., Davidovich,C., Gooding,A.R., Goodrich,K.J., Mattick,J.S. *et al.* (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell*, **165**, 1267–1279.
13. Aw,J.G., Shen,Y., Wilm,A., Sun,M., Lim,X.N., Boon,K.L., Tapsin,S., Chan,Y.S., Tan,C.P., Sim,A.Y. *et al.* (2016) In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell*, **62**, 603–617.
14. Dupuis-Sandoval,F., Poirier,M. and Scott,M.S. (2015) The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscipl. Rev. RNA*, **6**, 381–397.
15. Kishore,S. and Stamm,S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
16. Falaleeva,M., Pages,A., Matuszek,Z., Hidmi,S., Agranat-Tamir,L., Korotkov,K., Nevo,Y., Eyras,E., Sperling,R. and Stamm,S. (2016) Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1625–E1634.
17. Scott,M.S., Ono,M., Yamada,K., Endo,A., Barton,G.J. and Lamond,A.I. (2012) Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Res.*, **40**, 3676–3688.
18. Michel,C.I., Holley,C.L., Scruggs,B.S., Sidhu,R., Brookheart,R.T., Listenberger,L.L., Behlke,M.A., Ory,D.S. and Schaffer,J.E. (2011) Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab.*, **14**, 33–44.

19. Lee,J., Harris,A.N., Holley,C.L., Mahadevan,J., Pyles,K.D., Lavagnino,Z., Scherrer,D.E., Fujiwara,H., Sidhu,R., Zhang,J. *et al.* (2016) Rpl13a small nucleolar RNAs regulate systemic glucose metabolism. *J. Clin. invest.*, **126**, 4616–4625.
20. Rimer,J.M., Lee,J., Holley,C.L., Crowder,R.J., Chen,D.L., Hanson,P.I., Ory,D.S. and Schaffer,J.E. (2018) Long-range function of secreted small nucleolar RNAs that direct 2′-O-methylation. *J. Biol. Chem.*, **293**, 13284–13296.
21. Zhong,F., Zhou,N., Wu,K., Guo,Y., Tan,W., Zhang,H., Zhang,X., Geng,G., Pan,T., Luo,H. *et al.* (2015) A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. *Nucleic Acids Res.*, **43**, 10474–10491.
22. Huang,C., Shi,J., Guo,Y., Huang,W., Huang,S., Ming,S., Wu,X., Zhang,R., Ding,J., Zhao,W. *et al.* (2017) A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.*, **45**, 8647–8660.
23. Boivin,V., Deschamps-Francoeur,G., Couture,S., Nottingham,R.M., Bouchard-Bourelle,P., Lambowitz,A.M., Scott,M.S. and Abou-Elela,S. (2018) Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA*, **24**, 950–965.
24. Deschamps-Francoeur,G., Boivin,V., Elela,S.A. and Scott,M.S. (2019) CoCo: RNA-seq read assignment correction for nested genes and multimapped reads. *Bioinformatics*, in press.
25. Nottingham,R.M., Wu,D.C., Qin,Y., Yao,J., Hunicke-Smith,S. and Lambowitz,A.M. (2016) RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA*, **22**, 597–613.
26. Cavaille,J. (2017) Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. *Wiley Interdiscipl. Rev. RNA*, **8**, e1417.
27. Falaleeva,M., Welden,J.R., Duncan,M.J. and Stamm,S. (2017) C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *Bioessays*, **39**, doi:10.1002/bies.201600264.
28. Romano,G., Veneziano,D., Acunzo,M. and Croce,C.M. (2017) Small non-coding RNA and cancer. *Carcinogenesis*, **38**, 485–491.
29. Abel,Y. and Rederstorff,M. (2019) SnoRNAs and the emerging class of sdRNAs: Multifaceted players in oncogenesis. *Biochimie.*, **164**, 17–21.
30. Stepanov,G.A., Filippova,J.A., Komissarov,A.B., Kuligina,E.V., Richter,V.A. and Semenov,D.V. (2015) Regulatory role of small nucleolar RNAs in human diseases. *BioMed Res. Int.*, **2015**, 206849.
31. Jenkinson,E.M., Rodero,M.P., Kasher,P.R., Uggenti,C., Oojageer,A., Goosey,L.C., Rose,Y., Kershaw,C.J., Urquhart,J.E., Williams,S.G. *et al.* (2016) Mutations in SNORD118 cause the cerebral microangiopathy leukoencephalopathy with calcifications and cysts. *Nat. Genet.*, **48**, 1185–1192.
32. Yoshihama,M., Nakao,A. and Kenmochi,N. (2013) snOPY: a small nucleolar RNA orthological gene database. *BMC Res. Notes*, **6**, 426.
33. Jorjani,H., Kehr,S., Jedlinski,D.J., Gumienny,R., Hertel,J., Stadler,P.F., Zavolan,M. and Gruber,A.R. (2016) An updated human snoRNAome. *Nucleic Acids Res.*, **44**, 5068–5082.
34. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciufo,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
35. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Giron,C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
36. The,R.C. (2019) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, **47**, D221–D229.
37. Kalvari,I., Argasinska,J., Quinones-Olvera,N., Nawrocki,E.P., Rivas,E., Eddy,S.R., Bateman,A., Finn,R.D. and Petrov,A.I. (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, **46**, D335–D342.
38. Yates,B., Braschi,B., Gray,K.A., Seal,R.L., Tweedie,S. and Bruford,E.A. (2017) Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.*, **45**, D619–D625.
39. Uhlen,M., Fagerberg,L., Hallstrom,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,A., Kampf,C., Sjostedt,E., Asplund,A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
40. Gong,J., Shao,D., Xu,K., Lu,Z., Lu,Z.J., Yang,Y.T. and Zhang,Q.C. (2017) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.*, **46**, D194–D201