

EVAtlas: a comprehensive database for ncRNA expression in human extracellular vesicles

Chun-Jie Liu^{1,†}, Gui-Yan Xie^{1,†}, Ya-Ru Miao^{1,†}, Mengxuan Xia¹, Yi Wang¹, Qian Lei¹, Qiong Zhang^{2,*} and An-Yuan Guo^{1,2,*}

¹Center for Artificial Intelligence Biology, Hubei Bioinformatics & Molecular Imaging Key Laboratory, Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology; Wuhan 430074, China and ²Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong 226001, China

Received June 01, 2021; Revised July 01, 2021; Editorial Decision July 20, 2021; Accepted July 23, 2021

ABSTRACT

Extracellular vesicles (EVs) packing various molecules play vital roles in intercellular communication. Non-coding RNAs (ncRNAs) are important functional molecules and biomarkers in EVs. A comprehensive investigation of ncRNAs expression in EVs under different conditions is a fundamental step for functional discovery and application of EVs. Here, we curated 2030 small RNA-seq datasets for human EVs (1506 sEV and 524 IEV) in 24 conditions and over 40 diseases. We performed a unified reads dynamic assignment algorithm (RDAA) considering mismatch and multi-mapping reads to quantify the expression profiles of seven ncRNA types (miRNA, snoRNA, piRNA, snRNA, rRNA, tRNA and Y RNA). We constructed EVAtlas (<http://bioinfo.life.hust.edu.cn/EVAtlas>), a comprehensive database for ncRNA expression in EVs with four functional modules: (i) browse and compare the distribution of ncRNAs in EVs from 24 conditions and eight sources (plasma, serum, saliva, urine, sperm, breast milk, primary cell and cell line); (ii) prioritize candidate ncRNAs in condition related tissues based on their expression; (iii) explore the specifically expressed ncRNAs in EVs from 24 conditions; (iv) investigate ncRNA functions, related drugs, target genes and EVs isolation methods. EVAtlas contains the most comprehensive ncRNA expression in EVs and will be a key resource in this field.

INTRODUCTION

Extracellular vesicles (EVs) released by live cells are important vectors for signal transduction and cell communication, especially for the interactivity between distant cells (1). Exosomes secreted from endosomal system and microvesicles (MVs) shed from plasma membrane are the two main sources of EVs (2). Since consensus has not yet emerged on specific markers of EV subtypes (exosomes and MVs), to follow MISEV2018 guidelines, exosome and MV should be replaced by operational term small EV (sEV) and large EV (lEV) by physical size (3). The contents of EVs comprise multiple kinds of biomolecules from source cells, such as lipid, protein, metabolite, DNA and RNA (4). Non-coding RNAs (ncRNAs) are well-studied and important molecules in EVs, which can be delivered into and regulate target cells (5), and show a physiological- and pathological-specificity to serve as candidate biomarkers (6). Thus, a comprehensive investigation for the distribution of ncRNAs in EVs could help to understand the molecular characteristics of EVs and uncover their potential roles in EVs (7).

ncRNAs as important regulatory components play vital roles in biological processes, and their expression profiles vary greatly in normal and disease status (8). For example, dysregulations of microRNAs (miRNAs), small nucleolar RNAs (snoRNAs) and transfer RNAs (tRNAs) are common events in tumorigenesis (9–13). Previous studies showed that ribosomal RNAs (rRNAs) are constitutively endocytosed in EVs and take the largest amount of cargo ncRNAs of EVs in most of conditions (14,15). While other ncRNAs or fragments, such as PIWI-interacting RNAs (piRNAs), Y RNAs, small nuclear RNAs (snRNAs), could be selectively enclosed in EVs to deliver into and impact target cells (6,16). Increasing evidences demonstrated that ncRNAs in EVs presented source-cell-like and functional preferences in different conditions, and the distri-

*To whom correspondence should be addressed. Tel: +86 27 87793177; Fax: +86 27 87793177; Email: guoay@hust.edu.cn
Correspondence may also be addressed to Qiong Zhang. Tel: +86 513 85052297; Fax: +86 513 85052297; Email: transomics@gmail.com

†The authors wish it to be known that, in their opinions, the first three authors should be regarded as Joint First Authors.

bution of cargo ncRNAs was determined by the status of cells (17).

The high-throughput small RNA-seq (smRNA-seq) technology provides an excellent opportunity to comprehensively investigate ncRNAs in EVs. Up to date, several resources curated a few kinds of molecules from EVs. For example, Vesiclepedia, ExoCarta and EVpedia mainly curated four types of molecules (protein, RNA, metabolite and lipid) of EVs from publications (18–20) and without quantitative expression information. Although some databases dived into assess the abundance of RNAs from sequencing data of EVs, most of them have limitations on small sample size and ncRNA types or non-uniform expression profiles. At present, exRNA Atlas is the largest repository for exRNA profiles, which contains 159 exosome and 205 MV samples (21). ExoRBase focused on 87 exosome samples (22). ExoBCD performed differential expression genes and exosome miRNAs/lncRNAs in breast cancer from 4 datasets (23). Our previous study EVmiRNA (14) is an integrated database focusing on miRNA expression profiles in EVs. Compared with EVmiRNA, small RNA-seq datasets related to EVs tripled in last three years and more attentions are paid on ncRNA functions in EVs. We updated our EVmiRNA to the state-of-the-art data repository named EVAtlas.

In the EVAtlas, we curated and analyzed the expression of seven types of ncRNA (miRNA, snoRNA, piRNA, snRNA, rRNA, tRNA and Y RNA) from 2030 smRNA-seq datasets of EVs (1506 sEV and 524 IEV datasets) in 24 conditions and over 40 diseases, and provided functional annotation and specifically expressed analysis for ncRNAs in different EV types. To our knowledge, EVAtlas is the most comprehensive resource for EV ncRNA expression, which will benefit EV research community.

MATERIALS AND METHODS

Data collection and preprocess

We used keywords ‘((extracellular vesicle OR exosome OR microvesicle OR EV) AND (smRNA-seq OR miRNA-seq OR small RNA-seq)) AND (Homo sapiens)’ to retrieved human EV related smRNA-seq datasets. Since these datasets used various EV isolation methods, to follow the MISEV2018 guideline, we unified EV nomenclature by EV physical size with sEV and IEV. We manually reviewed the study design or the related paper to label each dataset with sEV or IEV based on the exact EV isolation method description. Datasets with ambiguous description or from other types of EV were excluded. Datasets were partitioned into sEV or IEV for further meta-analysis because these two EV types may be generated from different biological approaches with different biomolecule content. Next, we assessed sequencing quality of smRNA-seq data using similar methods described in our previous EVmiRNA (13). In brief, FastQC (v0.11.5) was used for quality control of raw data (QC), and trimmomatic (v0.36) (24) was employed for adapter removal and low-quality data trimming to obtain high-quality clean reads for further analyses. Bowtie (v1.2.1) (25) was used to align clean reads to human reference genome GRCh38. Samples with clean reads less than

5 million or the predominant reads out of the range of 15–50 nt or an alignment ratio less than 40% were discarded. Finally, we obtained 2030 smRNA-seq EV datasets (1506 sEV and 524 IEV datasets) from 57 studies. For better meta-analysis, all the datasets were well-curated and classified into 24 conditions, 8 sources and over 40 diseases.

Reads alignment and abundance estimation for seven types of ncRNAs in EVs

The reference sequences for seven types of ncRNAs were obtained from their authoritative databases. In detail, the reference sequences of miRNAs, snoRNAs and tRNAs were downloaded from miRBase V22, snoDB and GtRNAdb (26–28), respectively, while the references of piRNAs and rRNAs were obtained from RNACentral (29). Other reference sequences, such as snRNAs and Y RNAs were parsed from human genome annotation of NCBI GenBank. Then we used Bowtie to map reads to seven ncRNA references simultaneously with relatively strict parameter (one mismatch allowed). Since some short reads may be mapped to multiple ncRNA references, the final assignment of these reads could influence the ncRNA proportion in EVs. To address this question, we assessed the ncRNA distribution in nine studies (each study has over 10 samples) with 682 samples in total. In Figure 1A middle pie chart, we can see that 87.79% reads have zero mismatch and 12.21% reads have one mismatch in summary. Next, we found that over 93.5% reads mapped to only one ncRNA type on average (Figure 1A, left and right pie charts), about 6% reads can be mapped to multi-ncRNA types. The reads mapped to multi-ncRNA types were varied in each study (Supplementary Figure S1A).

To assign the about 6% multi-ncRNA mapped reads to the exact ncRNA type, we creatively adopted a reads dynamic assignment algorithm (RDAA), which is based on a reasonable hypothesis that multi-ncRNA mapped reads are more likely from higher proportion ncRNA type. To this end, firstly we used unique mapped reads to summarize the ncRNA distribution in each study and found that the major ncRNA types varied dramatically in different studies (Supplementary Figure S1B). Then, we performed the RDAA process workflow in Figure 1B in following steps: (i) clean reads of each sample were mapped (one mismatch allowed) to seven ncRNA references simultaneously; (ii) reads with zero or one mismatch only mapped to one ncRNA type were assigned to that ncRNA. Among them, reads with zero mismatch mapped to one ncRNA type were used to calculate the reference ncRNA abundance in EVs; (iii) multi-ncRNA mapped reads with zero or one mismatch were assigned to exact one ncRNA type according to the reference ncRNA abundance priority; (iv) final ncRNA abundance was determined by the assigned mapped reads. This is a dynamic process because each sample has different reference ncRNA abundance which may be determined by the EV source cell condition. We used reads per million (RPM) to normalize ncRNA expression. Detailed methods were described on EVAtlas website document page. Finally, ncRNAs with RPM > 1 were kept further analyses and deposited into EVAtlas.

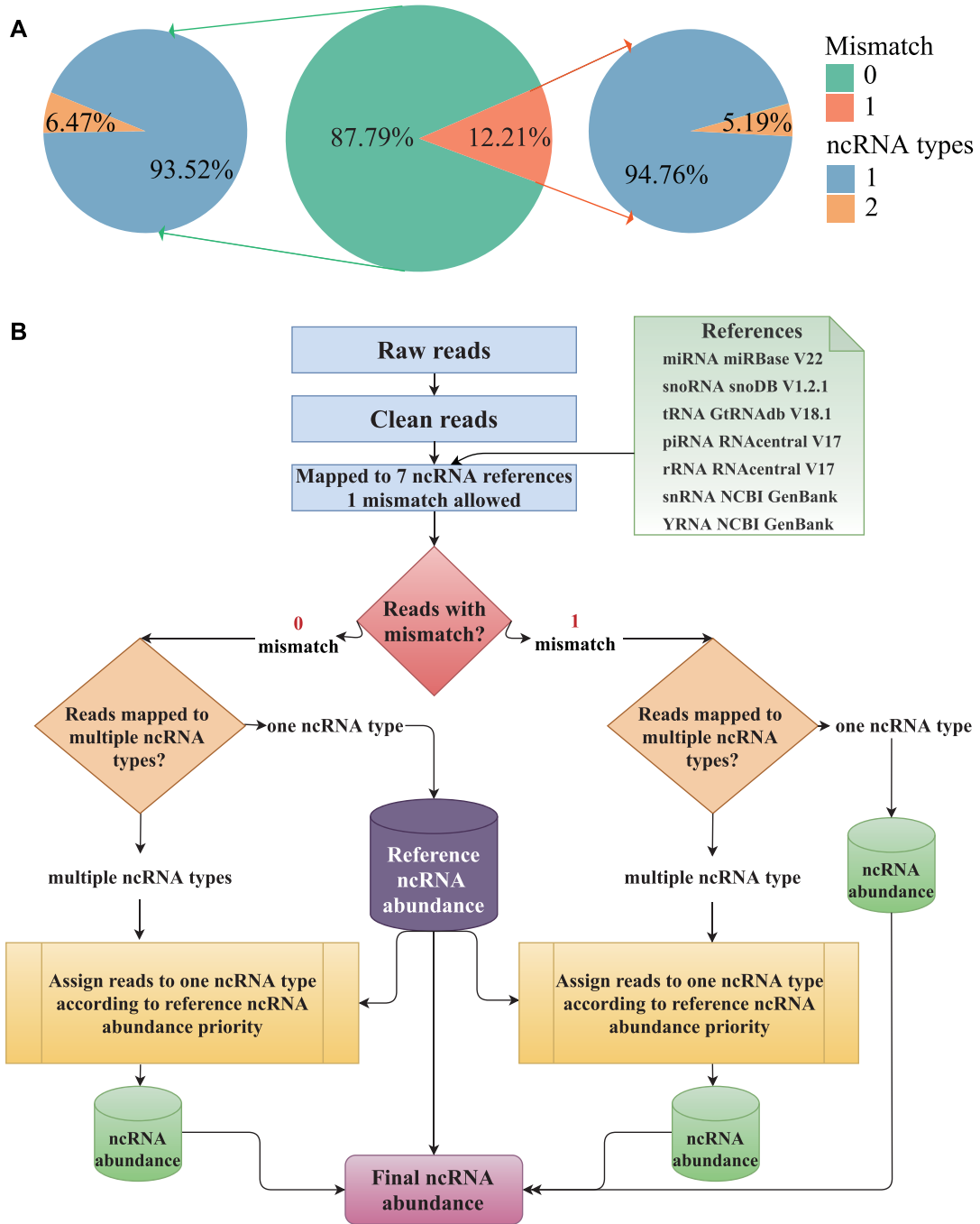


Figure 1. The assessment of ncRNA reads mapping distribution and RDAA algorithm workflow. (A) Statistics of reads with zero and one mismatch proportion in nine studies, the middle pie chart is the average proportion (green for zero mismatch, orange for one mismatch). The left and right pie charts are the reads with zero and one mismatch mapped to one ncRNA type (in blue) or two ncRNA types (in yellow). (B) The reads dynamic assignment algorithm (RDAA) for reads mapping to ncRNAs. Reads with zero or one mismatch are mapped to one ncRNA type can be used to calculate ncRNA abundance, the reads with zero mismatch are used to calculate reference ncRNA abundance priority for multi-ncRNA reads assignment.

Detection of specifically expressed ncRNAs in tissues and diseases

To detect specifically expressed ncRNAs, the expression level of a ncRNA in the given condition was represented by its median. The specifically highly expressed ncRNAs from sEV and IEV were analyzed separately by the software SEGtool with strict parameter (detect_mod = 3) (30).

Functional annotation for ncRNAs and their expression profiles in cancers

The functional annotations of ncRNAs were integrated into EVAtlas, in which, different types of ncRNAs have distinct annotations. The basic information of ncRNAs, such as gene location and potential functions, was extracted from their reference databases. For miRNAs, their potential functions were retrieved from PubMed, miRNA-target relationships were obtained from FFLtool (31), drug correlation information was downloaded from SM2miR (32), and their expression profiles in TCGA cancers were obtained from GSCALite (33). For snoRNA and tRNA, their expression profiles in cancers were collected from SNORic and tRic (12,13), while their potential target genes were collected from snoDB and GtRNAdb 2.0 (27,28), respectively.

DATA SUMMARY AND OVERVIEW OF EVAtlas

All data in EVAtlas were stored in the document-oriented database MongoDB v3.2. EVAtlas website was built with Flask (a Python web development framework), Angular v9 (a javascript front-end framework) and Echarts v4.8 (a javascript visualization framework). EVAtlas is freely accessible for academic users via <http://bioinfo.life.hust.edu.cn/EVAtlas>.

Data summary and functional description of EVAtlas

EVAtlas contains seven types of ncRNA expression profiles quantified from 2030 smRNA-seq datasets of 1506 sEVs and 524 IEVs (Figure 2B). A total of 30 445 ncRNAs were detected in these datasets (RPM > 1), including 2527 miRNAs. In summary of all these datasets (pie chart in Figure 2B), the abundance of rRNA (38.12%), miRNA (26.94%), tRNA (19.19%) and Y RNA (12.59) account for about 97% of ncRNA in EVs. piRNA (1.51%), snRNA (1.05%) and snoRNA (0.6%) are minor ncRNA types. The abundance of ncRNA in EVs varies from datasets, thus these statistics of ncRNA proportions are the results of datasets in current EVAtlas.

To explore the expression profiles of seven types of ncRNA from smRNA-seq datasets, EVAtlas provides four functional modules for users to: (i) browse and compare the distribution of ncRNAs in EVs from 24 conditions and eight sources (plasma, serum, saliva, urine, sperm, breast milk, primary cell and cell line); (ii) prioritize candidate ncRNAs in condition related tissues based on their expression; (iii) explore the specifically expressed ncRNAs in EVs from 24 conditions; (iv) investigate ncRNA functions, related drugs, target genes and EVs isolation methods.

Browse and search function of EVAtlas

For users' convenience, we designed a main search box in the middle of home page to provide a universe search function by a ncRNA symbol. Users can browse the distribution of ncRNAs by 'Samples', 'RNA Expression' and 'Specific Expression' modules (Figure 2A) in EVAtlas website. The 'Samples' module helps users to inquire and browse all collected studies and samples according to selected condition (Figure 2C). For example, by clicking on 'Lung' icon, EVAtlas will display all datasets from EVs related to lung-disease-sourced plasma and serum, as well as lung cancer cell lines (Figure 2D). Moreover, the reads distribution, a heatmap of top 50 ncRNAs and the average expression for each ncRNA type in this dataset are presented (Figure 2E). In addition, the method for EV isolation in this dataset is listed on the page.

On the 'RNA Expression' page, users can browse the expression profiles by ncRNA type. Similarly, this page also provides a quick search function for users to inquire interested ncRNA by its RNA symbol. Moreover, users can investigate various information of a ncRNA by clicking its symbol on this page. Clicking a ncRNA symbol will open the ncRNA page, on which, the basic information, expression profiles in sEVs or IEVs across different conditions or sources are represented. For miRNAs, there is more additional information, such as the functions from publications, expression profiles in TCGA cancers, target genes and miRNA-related drug information.

In addition, we separately displayed the specifically expressed ncRNAs in EVs by their EV types. Users can browse specifically expressed ncRNAs in an interested condition, for example, specifically expressed miRNAs in sEVs from breast.

Case study

EV is a promising source for biomarker identification in disease diagnosis. Our EVAtlas is the most comprehensive database for EV ncRNA expression, it will be very useful in the field of EV research and application. Here, we illustrated a case study for potential biomarker discovered in lung disease related plasma sEVs by using EVAtlas. First, we chose the lung tissue by clicking the 'Lung' icon on the 'Samples' page (Figure 2C). Then, six studies related to lung tissue are listed including two from plasma, two from serum and two from cell lines. Users can click the project of 'SRP135648' (Figure 2D), a sEV study sourced from plasma with 10 samples. We can find that hsa-miR-451a is the most highly expressed miRNAs with 5-fold change compared between lung cancer and control (Figure 2E). To verify this, we checked the hsa-miR-451a expression in other lung cancer projects in EVAtlas, compared lung cancer with control, hsa-miR-451a have 1.5- and 2.3-fold change in another two lung cancer studies 'SRP148568' and 'SRP223523', respectively. It has been reported that hsa-miR-451a with other miRNAs derived from plasma EV could be novel diagnostic biomarkers of lung cancer (34–36). By clicking hsa-miR-451a, it is redirected to the detailed page to show its expression in sEV and IEV across conditions (Figure 2F). Notably, hsa-miR-451a expressed higher in normal tissue than tumor



Figure 2. Features and a case study of EVAAtlas. (A) The portal of EVAAtlas functional modules. (B) Data summary and ncRNA expression compositions in EVAAtlas. (C) Snapshot of ‘Sample’ page with ‘Conditions’ and ‘Sources’ to filter interested field. (D) Lung tissue related EV projects sourced from plasma, serum and cell line. (E) The miRNA average expression in plasma project SRP135648. (F) Exosome hsa-miR-451a expression across different condition-related tissues.

in the TCGA lung cancer data. It is reported that hsa-miR-451a is a potential suppressor of cell migration and invasion in non-small cell lung cancer (37). Based on these multi-dataset cross validation and publication, we can speculate reasonably that hsa-miR-451a is a key factor to maintain normal cells but selectively imported into EVs by lung tumor cells. Thus, EV-derived hsa-miR-451a could be a potential diagnostic biomarker marker for lung cancer. As this hsa-miR-451a showcases, EVAtlas as a comprehensive EV ncRNA expression repository could provide multi-datasets to discover and validate EV biomarkers *in silico* and inspire researchers for better experimental design. To correctly use the function of multi-datasets cross validation in EVAtlas, users should take care of the EV isolation method because some isolation methods do not isolate EVs with high specificity.

CONCLUSION AND FUTURE PERSPECTIVES

EVAtlas is the state-of-the-art EV ncRNA repository deposited comprehensive expression profiles and patterns of ncRNAs in sEVs and IEVs. EVAtlas will promote the function research and biomarker discovery of EVs. The useful modules of EVAtlas can benefit the EV researchers to quickly focus on the interested ncRNAs to investigate their distributions and features at the big data level. In addition, EVAtlas is easy-to-use and freely accessible for academic users without any registrations. Since it remains difficult to distinguish exosomes and MVs in biofluids, in the compliance with MISEV2018, we used less ambiguous terms for EV type as sEV and IEV. Users are recommended to refer to the isolation method in each study and the EV-TRACK link provided in EVAtlas for the correct use of this database.

We systematically analyzed, curated and visualized ncRNA expression in EVs because of their vital regulatory roles in EVs and target cells. RNA has a large number of subtypes, and different types of RNAs have distinct functions. For examples, mRNA, lncRNAs and circular RNA are also important components in EVs. EVAtlas will incorporate more RNA types and more functional modules related to EV in the future. Although EVAtlas collected and analyzed almost all publicly available EV smRNA-seq datasets, it is still not sufficient enough, therefore it may lead to biases for the landscape. A whole picture of EV ncRNA expression landscape requires a large number of samples, we will keep pace with new available EV smRNA-seq datasets and update EVAtlas regularly. EVAtlas could be an infrastructure resource for the EV community.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

The EVAtlas data repository was built based on analyzing public EV smRNA-seq datasets, we thank all the researchers and participants for their efforts in generating EV datasets. The full data contributors are listed on our website.

FUNDING

National Natural Science Foundation of China [31822030, 31801113, 31801154, 31771458]; China Postdoctoral Science Foundation [2019M652623, 2018M632830, 2017M622455, 2019T120664]. Funding for open access charge: National Natural Science Foundation of China [31822030].

Conflict of interest statement. None declared.

REFERENCES

1. Tkach, M. and Théry, C. (2016) Communication by extracellular vesicles: where we are and where we need to go. *Cell*, **164**, 1226–1232.
2. Raposo, G. and Stoorvogel, W. (2013) Extracellular vesicles: exosomes, microvesicles, and friends. *J. Cell Biol.*, **200**, 373–383.
3. Théry, C., Witwer, K.W., Aikawa, E., Alcaraz, M.J., Anderson, J.D., Andriantsitohaina, R., Antoniou, A., Arab, T., Archer, F., Atkin-Smith, G.K. *et al.* (2018) Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles*, **7**, 1535750.
4. Xu, R., Rai, A., Chen, M., Suwakulsiri, W., Greening, D.W. and Simpson, R.J. (2018) Extracellular vesicles in cancer – implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.*, **15**, 617–638.
5. O'Brien, K., Breyne, K., Ughetto, S., Laurent, L.C. and Breakefield, X.O. (2020) RNA delivery by extracellular vesicles in mammalian cells and its applications. *Nat. Rev. Mol. Cell. Biol.*, **21**, 585–606.
6. Hoshino, A., Costa-Silva, B., Shen, T.-L., Rodrigues, G., Hashimoto, A., Tesic Mark, M., Molina, H., Kohsaka, S., Di Giannatale, A., Ceder, S. *et al.* (2015) Tumour exosome integrins determine organotropic metastasis. *Nature*, **527**, 329–335.
7. Jeppesen, D.K., Fenix, A.M., Franklin, J.L., Higginbotham, J.N., Zhang, Q., Zimmerman, L.J., Liebler, D.C., Ping, J., Liu, Q., Evans, R. *et al.* (2019) Reassessment of exosome composition. *Cell*, **177**, 428–445.
8. Slack, F.J. and Chinnaiyan, A.M. (2019) The role of non-coding RNAs in oncology. *Cell*, **179**, 1033–1055.
9. Dhawan, A., Scott, J.G., Harris, A.L. and Buffa, F.M. (2018) Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. *Nat. Commun.*, **9**, 5228.
10. Williams, G.T. and Farzaneh, F. (2012) Are snoRNAs and snoRNA host genes new players in cancer? *Nat. Rev. Cancer*, **12**, 84–88.
11. Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H. and Tavazoie, S.F. (2016) Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell*, **165**, 1416–1427.
12. Gong, J., Li, Y., Liu, C., Xiang, Y., Li, C., Ye, Y., Zhang, Z., Hawke, D.H., Park, P.K., Diao, L. *et al.* (2017) A Pan-cancer analysis of the expression and clinical relevance of small nucleolar RNAs in human cancer. *Cell Rep.*, **21**, 1968–1981.
13. Zhang, Z., Ruan, H., Liu, C.-J., Ye, Y., Gong, J., Diao, L., Guo, A.-Y. and Han, L. (2019) tRiC: a user-friendly data portal to explore the expression landscape of tRNAs in human cancers. *RNA Biol.*, **17**, 1674–1679.
14. Liu, T., Zhang, Q., Zhang, J., Li, C., Miao, Y.-R., Lei, Q., Li, Q. and Guo, A.-Y. (2019) EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res.*, **47**, D89–D93.
15. Turchinovich, A., Drapkina, O. and Tonevitsky, A. (2019) Transcriptome of extracellular vesicles: State-of-the-Art. *Front. Immunol.*, **10**, 202.
16. Wei, Z., Batagov, A.O., Schinelli, S., Wang, J., Wang, Y., El Fatimy, R., Rabinovsky, R., Balaj, L., Chen, C.C., Hochberg, F. *et al.* (2017) Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat Commun.*, **8**, 1145.
17. Yang, J., Li, C., Zhang, L. and Wang, X. (2018) Extracellular vesicles as carriers of non-coding RNAs in liver diseases. *Front. Pharmacol.*, **9**, 415.
18. Keerthikumar, S., Chisanga, D., Ariyaratne, D., Saffar, H.A., Anand, S., Zhao, K., Samuel, M., Pathan, M., Jois, M., Chilamkurti, N.

- et al.* (2016) ExoCarta: a web-based compendium of exosomal cargo. *J. Mol. Biol.*, **428**, 688–692.
19. Kim,D.-K., Lee,J., Kim,S.R., Choi,D.-S., Yoon,Y.J., Kim,J.H., Go,G., Nhung,D., Hong,K., Jang,S.C. *et al.* (2015) EVpedia: a community web portal for extracellular vesicles research. *Bioinformatics*, **31**, 933–939.
 20. Pathan,M., Fonseka,P., Chitti,S.V., Kang,T., Sanwlani,R., Van Deun,J., Hendrix,A. and Mathivanan,S. (2019) Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. *Nucleic Acids Res.*, **47**, D516–D519.
 21. Murillo,O.D., Thistlethwaite,W., Rozowsky,J., Subramanian,S.L., Lucero,R., Shah,N., Jackson,A.R., Srinivasan,S., Chung,A., Laurent,C.D. *et al.* (2019) exRNA atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. *Cell*, **177**, 463–477.
 22. Li,S., Li,Y., Chen,B., Zhao,J., Yu,S., Tang,Y., Zheng,Q., Li,Y., Wang,P., He,X. *et al.* (2018) exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res.*, **46**, D106–D112.
 23. Wang,X., Chai,Z., Pan,G., Hao,Y., Li,B., Ye,T., Li,Y., Long,F., Xia,L. and Liu,M. (2021) ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer. *Brief. Bioinformatics*, **22**, bbaa088.
 24. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
 25. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
 26. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
 27. Bouchard-Bourelle,P., Desjardins-Henri,C., Mathurin-St-Pierre,D., Deschamps-Francoeur,G., Fafard-Couture,É., Garant,J.-M., Elela,S.A. and Scott,M.S. (2020) snoDB: an interactive database of human snoRNA sequences, abundance and interactions. *Nucleic Acids Res.*, **48**, D220–D225.
 28. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.
 29. RNAcentral: a hub of information for non-coding RNA sequences (2019) *Nucleic Acids Res.*, **47**, D1250–D1251.
 30. Zhang,Q., Liu,W., Liu,C., Lin,S.-Y. and Guo,A.-Y. (2018) SEGtool: a specifically expressed gene detection tool and applications in human tissue and single-cell sequencing data. *Brief. Bioinformatics*, **19**, 1325–1336.
 31. Xie,G.-Y., Xia,M., Miao,Y.-R., Luo,M., Zhang,Q. and Guo,A.-Y. (2020) FFLtool: a web server for transcription factor and miRNA feed forward loop analysis in human. *Bioinformatics*, **36**, 2605–2607.
 32. Liu,X., Wang,S., Meng,F., Wang,J., Zhang,Y., Dai,E., Yu,X., Li,X. and Jiang,W. (2013) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics*, **29**, 409–411.
 33. Liu,C.-J., Hu,F.-F., Xia,M.-X., Han,L., Zhang,Q. and Guo,A.-Y. (2018) GSCALite: a web server for gene set cancer analysis. *Bioinformatics*, **34**, 3771–3772.
 34. Yao,B., Qu,S., Hu,R., Gao,W., Jin,S., Liu,M. and Zhao,Q. (2019) A panel of miRNAs derived from plasma extracellular vesicles as novel diagnostic biomarkers of lung adenocarcinoma. *FEBS Open Bio.*, **9**, 2149–2158.
 35. Lawson,J., Dickman,C., MacLellan,S., Towle,R., Jabalee,J., Lam,S. and Garnis,C. (2017) Selective secretion of microRNAs from lung cancer cells via extracellular vesicles promotes CAMK1D-mediated tube formation in endothelial cells. *Oncotarget*, **8**, 83913–83924.
 36. Li,C., Qin,F., Hu,F., Xu,H., Sun,G., Han,G., Wang,T. and Guo,M. (2018) Characterization and selective incorporation of small non-coding RNAs in non-small cell lung cancer extracellular vesicles. *Cell Biosci.*, **8**, 2.
 37. Shen,Y.-Y., Cui,J.-Y., Yuan,J. and Wang,X. (2018) MiR-451a suppressed cell migration and invasion in non-small cell lung cancer through targeting ATF2. *Eur. Rev. Med. Pharmacol. Sci.*, **22**, 5554–5561.