UCLA

Title

The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva

Permalink

https://escholarship.org/uc/item/9ss8q252

Journal

Clinical Chemistry, 61(1)

ISSN

0009-9147

Authors

Bahn, Jae Hoon
Zhang, Qing
Li, Feng
et al.

Publication Date

2015

DOI

10.1373/clinchem.2014.230433

Peer reviewed

The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva

Jae Hoon Bahn[1,2,†], **Qing Zhang**[1,2,†], **Feng Li**[3], **Tak-Ming Chan**[1,2], **Xianzhi Lin**[1,2], **Yong Kim**[3,4,5], **David T.W. Wong**[2,3,4,6,7,*], and **Xinshu Xiao**[1,2,4,6,*]

[1]Department of Integrative Biology and Physiology, UCLA, Los Angeles, CA

[2]Molecular Biology Institute, UCLA, Los Angeles, CA

[3]School of Dentistry, UCLA, Los Angeles, CA

[4]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA

[5]Broad Stem Cell Research Center, UCLA, Los Angeles, CA

[6]School of Engineering, UCLA, Los Angeles, CA

[7]Department of Head & Neck Surgery/Otolaryngology, UCLA, Los Angeles, CA

Abstract

BACKGROUND—Extracellular RNAs (exRNAs) in human body fluids are emerging as effective biomarkers for detection of diseases. Saliva, as the most accessible and noninvasive body fluid, has been shown to harbor exRNA biomarkers for several human diseases. However, the entire spectrum of exRNA from saliva has not been fully characterized.

METHODS—Using high-throughput RNA sequencing (RNA-Seq), we conducted an in-depth bioinformatic analysis of noncoding RNAs (ncRNAs) in human cell-free saliva (CFS) from healthy individuals, with a focus on microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and circular RNAs (circRNAs).

*Address correspondence to: X.X. at Department of Integrative Biology and Physiology, UCLA, 611 Charles Young Dr, Los Angeles, CA 90095. gxxiao@ucla.edu. D.W. at Dental Research Institute, UCLA, 10833 Le Conte Ave., Los Angeles, CA 90095. dtww@ucla.edu.
†Jae Hoon Bahn and Qing Zhang contributed equally to the work, and both should be considered as first authors.

RESULTS—Our data demonstrated robust reproducibility of miRNA and piRNA profiles across individuals. Furthermore, individual variability of these salivary RNA species was highly similar to those in other body fluids or cellular samples, despite the direct exposure of saliva to environmental impacts. By comparative analysis of >90 RNA-Seq data sets of different origins, we observed that piRNAs were surprisingly abundant in CFS compared with other body fluid or intracellular samples, with expression levels in CFS comparable to those found in embryonic stem cells and skin cells. Conversely, miRNA expression profiles in CFS were highly similar to those in serum and cerebrospinal fluid. Using a customized bioinformatics method, we identified >400 circRNAs in CFS. These data represent the first global characterization and experimental validation of circRNAs in any type of extracellular body fluid.

CONCLUSIONS—Our study provides a comprehensive landscape of ncRNA species in human saliva that will facilitate further biomarker discoveries and lay a foundation for future studies related to ncRNAs in human saliva.

Human saliva has been used increasingly for biomarker development to enable noninvasive detection of diseases. The term "salivaomics" was coined to highlight the omics constituents in saliva that can be used for biomarker development and personalized medicine (1). Salivary extracellular RNA (exRNA)[8] was discovered 10 years ago; since then, the nature, origin, and characterization of salivary RNA have been actively pursued (2–8). These studies have demonstrated the potential for the use of salivary RNA to detect oral cancer (2, 9), Sjögren syndrome (3), resectable pancreatic cancer (7), lung cancer (10), ovarian cancer (11), and breast cancer (12). Facilitated by next generation sequencing technologies, a complex compositional profile of salivary RNA molecules has emerged, encompassing mRNAs, microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), etc. (8, 13, 14). However, the entire spectrum of exRNA from saliva has not been fully discovered, thus warranting further comprehensive deciphering and analyses. In addition, there is also increasing interest in understanding the functional aspects of salivary RNAs in oral and systemic biology. Such studies will be facilitated by a detailed delineation of the landscapes of salivary exRNA.

Although RNA sequencing (RNA-Seq) technologies have been applied to study RNA expression in several body fluids (8, 15–18), most of these studies have focused on analyses of miRNAs and mRNAs. In addition, more recently, thousands of exonic circular RNAs (circRNAs) have been discovered in various human cell types, many of which are highly stable, abundant, and evolutionarily conserved (19). Two circRNAs have been shown to act as miRNA sponges, thus playing a role in mediating miRNA targeting (20). It is expected that additional functions of circRNAs may be described soon (19). The stable nature of circRNAs makes these moieties intriguing candidates as functional molecules in circulating body fluid.

---

[8]Nonstandard abbreviations: exRNA, extracellular RNA; miRNA, microRNA; snoRNA, small nucleolar RNA; RNA-Seq,RNA sequencing; circRNA, circular RNA; ncRNA, noncoding RNA; CFS, cell-free saliva; UCLA, University of California–Los Angeles; piRNA, piwi-interacting RNA; RPM, reads per million mapped reads; ddPCR, droplet digital PCR; ISI, individual specificity index; CSF, cerebral spinal fluid; ES, embryonic stem; lncRNA, long noncoding RNA.

We performed a comprehensive analysis of extracellular noncoding RNAs (ncRNAs) in cell-free saliva (CFS) by next generation sequencing. In addition, we carried out a genome-wide analysis of possible existence of circular RNAs in CFS with RNA-Seq. To our best knowledge, this is the first report and validation of the existence of circRNAs in any body fluid. Our findings credential the presence of salivary ncRNA and, importantly, pave ways for further functional, biological, and biomarker discoveries related to ncRNAs in human saliva.

Materials and Methods

SALIVA COLLECTION AND PROCESSING AND RNA ISOLATION

Unstimulated saliva samples were obtained from healthy volunteers in accordance with a protocol approved by the University of California–Los Angeles (UCLA) Institutional Review Board as described previously (21). More details are included in Supplemental Methods, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1.

CONSTRUCTION OF SMALL RNA-SEQ LIBRARIES

We isolated total RNA directly from CFS in the online Supplemental Methods. Spike-in RNAs (Exiqon) were added to the total RNA samples (1 reaction volume per microgram RNA) before library construction as internal controls. With the spiked total RNA samples, we prepared small RNA-Seq libraries with the NEBNext Small RNA library Prep kit (NEB). The final libraries were purified with 6% PAGE gel.

CONSTRUCTION OF circRNA-SEQ LIBRARIES

To obtain enriched circular RNAs from the total RNA samples, we used 3 U/μg RNase R (Epicentre) to treat the total RNA (from CFS directly) for 20 min at 37 °C. Subsequently, RNA was extracted with acid phenol/chloroform (pH 4.5). We prepared sequencing libraries with the NEBNext Ultra directional RNA library Prep kit (NEB) followed by AMPure XP Beads size selection (Beckman Coulter).

CFS SMALL RNA-SEQ DATA ANALYSIS

Small RNA-Seq reads were first processed to remove adapter sequences and low quality reads. The reads were then aligned to the human genome with Bowtie (22) allowing at most 1 mismatch. We parsed the mapping results to identify reads mapped to miRNAs (miRBase, release 19), piwi-interacting RNAs (piRNAs) (piRNABank, November 2013), and other known small RNAs (RFam, version 11.0). We also mapped reads to the Human Oral Microbiome Databases (23) to eliminate those that possibly originated from microbial species. For human miRNAs, only uniquely mapped reads were retained. Uniqueness was not required for reads mapped to piRNAs or other small RNAs owing to their repetitive nature and/or presence of multiple copies in the genome. In parallel, reads were also aligned to the spike-in controls, allowing no mismatches. The number of reads mapped to each miRNA was normalized with the spike-in controls and total number of mapped reads in each library.

Detailed bioinformatic methods, exosome isolation, and experimental validation procedures are described in online Supplemental Methods.

Results and Discussion

SMALL RNA SEQUENCING OF CFS

We used our widely adopted protocol to isolate CFS from fresh saliva samples (see online Supplemental Methods) (9). The protocol previously was shown to effectively remove cells as evaluated by exclusion of cellular genomic DNA and cell counting (9). In addition, we examined several genes [e.g., *ESRP1/2* (epithelial splicing regulatory protein 1 and 2),[9] *OVOL1/2* (ovo-like zinc finger 1 and 2), *HBA1* (hemoglobin, α1), *APOC1* (apolipoprotein C-1)] that are known to be highly specific to epithelial cells or leukocytes, the major types of cells in saliva. Most of these genes were not expressed (on the basis of RNA-Seq data collected for a separate study, data not shown), supporting the effectiveness of our CFS protocol in removing cells. We obtained a total of 165 million reads from 8 CFS small RNA sequencing libraries (see online Supplemental Table 1). Total RNA was isolated directly from the CFS, and synthetic small RNAs were added into the total RNA samples before library construction to serve as spike-in controls. As shown in online Supplemental Fig. 1, expression levels of spike-in controls were highly correlated between samples, supporting the technical consistency of our data. After adapter removal, the reads showed a length distribution that peaked at 22 nt (Fig. 1A), consistent with the expected length of miRNAs. Interestingly, we also observed a second peak at 29 nt that may correspond to piRNAs. The most abundant types of small RNAs in our data included human miRNAs (6.0% of reads on average), piRNAs (7.5% of reads), and snoRNAs (0.02% of reads) (see online Supplemental Table 1). In addition, 58.8% of reads corresponded to microbial RNA sequences, reflecting the enriched presence of microorganisms in saliva (8). Our results suggest that the small RNA sequencing experiment can capture a wide spectrum of noncoding exRNAs in human saliva.

miRNAS ARE ABUNDANT IN HUMAN CFS

In each saliva sample, a total of 127–418 miRNAs were detected, with an expression level of 1 reads per million mapped reads (RPM) (see online Supplemental Table 2). The most abundant miRNA was miR-223–3p that had an average expression level of 19442 RPM across all samples. miRNA expression levels were highly correlated across biological replicates (r 0.977) (see online Supplemental Fig. 2), again supporting the reproducibility of our data. Importantly, different human individuals also demonstrated highly correlated miRNA expression levels (Fig. 1B; online Supplemental Fig. 3). For example, among the top 10 highly expressed miRNAs in each sample, 8 were shared by at least 4 of the 6 nonreplicated samples.

---

[9]Human genes: *ESRP1/2*, epithelial splicing regulatory protein 1 and 2; *OVOL1/2*, ovo-like zinc finger 1 and 2; *HBA1,* hemoglobin, α1; *APOC1,* apolipoprotein C-1; *EIF3E*; eukaryotic translation initiation factor 3, subunit E; *DDOST*; dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit (non-catalytic); *DOPEY2*, dopey family member 2; *UBAP2*, ubiquitin associated protein 2; *GSE1*, Gse1 coiled-coil protein; *ASXL1*, additional sex combs like transcriptional regulator 1; *UGP2*, UDP-glucose pyrophosphorylase 2; *WDFY1*, WD repeat and FYVE domain containing 1.

Previous reports suggested that the majority of miRNAs in saliva were concentrated in exosomes (24). We tested the presence of miRNAs in exosome and nonexosome fractions of saliva in a subset of samples (Fig. 1C; online Supplemental Fig. 4). Two miRNAs (miR-223–3p and miR-148a-3p) detected in RNA-Seq data of multiple individuals were chosen for this experiment. Plasma/serum miR-223 together with other miRNAs have been shown to be closely associated with the tumorigenesis and metastasis of gastric carcinoma (25), hepatocellular carcinoma or chronic hepatitis (26), sepsis (27), and lung cancer (28). Dysregulated miR-148a has been reported in ovarian cancer (29), liver injury (30), and gastric cancer (31).

To measure expression levels of the 2 miRNAs, exosomes were isolated from fresh human saliva with the conventional differential centrifugation method (see online Supplemental Methods), the effectiveness of which has been well established in our laboratory (32–34). We used the droplet digital PCR (ddPCR) method because it can measure absolute concentration of each miRNA and does not need internal controls. As shown in Fig. 1C, both miRNAs can be detected in all 3 individuals, and they are predominantly present in the exosomal RNA fraction in 2 of 3 individuals. For 1 individual (S1), both miRNAs were detected in both exosomal and nonexosomal fractions. This observation is not likely to be due to failed separation of the exosomal fractions, because piRNAs in this individual showed predominant localization in the exosomal fraction (see below). It should be noted that these data only serve as a qualitative evaluation rather than a quantitative validation of the RNA-Seq data, because the RNA was obtained in very different ways in the 2 types of experiments. In addition, the exosome isolation step introduced relatively large technical variation across samples, which may explain the large interindividual variation observed in miRNA expression level. Nevertheless, our result is consistent with previous findings that miRNAs are mainly localized in exosomes in most individuals. However, as shown in Fig. 1C, it is likely that there also exist vesicle-free ncRNAs in saliva (e.g., for individual S1), which should be further investigated.

VARIATION OF miRNA EXPRESSION ACROSS INDIVIDUALS

As shown in Fig. 1B, miRNA expression values are generally significantly correlated across individuals, as measured by RNA-Seq of CFS total RNA. However, there does exist noticeable variation across different individuals. We asked whether salivary miRNAs are particularly subject to individual variability, given that saliva is readily exposed to and communicative with the external environment that can be highly individual specific. To examine this question, we defined an individual specificity index (ISI) for each miRNA (see online Supplemental Methods). This index has a value between 0 and 1, with larger values representing higher interperson variability. For comparison, we also analyzed several other data sets derived from primary brain tissues (35), skin (36), B cells (37), cerebral spinal fluid (CSF), and serum (15). As shown in Fig. 1D, miRNAs generally demonstrated a wide range of individual variability in all types of samples. Interestingly, the salivary ISI distribution was relatively similar to those of other body fluids (CSF and serum) and intracellular RNA (B cells). Compared with other samples, brain and skin miRNAs showed higher individual variability, possibly reflecting the heterogeneous cell type composition in these samples. Overall, our results suggest that the individual variability observed in salivary miRNAs is at

a level similar to those observed in other extracellular and intracellular data sets. Given the remarkable diversity of salivary environment across individuals, this observation supports the effectiveness of our cell removal method in enriching for physiological extracellular RNA rather than environmentally related RNA. Thus, extracellular RNA in CFS can serve as stable biomarkers with individual variability similar to that of other body fluids, with the advantage of being highly accessible and noninvasive.

miRNA EXPRESSION PROFILES OF CFS CLUSTER WITH THOSE OF OTHER BODY FLUIDS

We next conducted a comprehensive comparison of miRNA expression profiles in different body fluids and cell types. A total of 95 small RNA sequencing data sets (including our 8 data sets) were analyzed for miRNA expression with the same method (Fig. 2; online Supplemental Fig. 3; online Supplemental Methods). Raw sequencing reads were used, except for a few data sets in which read counts of miRNAs were directly taken from the original publication owing to lack of raw data (as noted in online Supplemental Fig. 3). Batch effects or technical variations across laboratories may be a significant confounding factor in this type of analysis. We used a data normalization method similar to that adopted in DESeq (38) to alleviate batch effects. In addition, the correlation across data sets was evaluated with the Kendall τ method, a rank-based nonparametric correlation analysis (see more details in online Supplemental Methods). In this manner, data comparison is less sensitive to the quantitative values of miRNA expression levels, which may fluctuate due to technical variation.

As shown in online Supplemental Fig. 3, data sets generated from the same tissue or cell type (brain, B cells, or ES cells) by different laboratories clustered together. This observation suggests that batch effects have been adequately reduced, although it may not have been possible to reach a complete elimination. Strikingly, all the data sets derived from extracellular body fluids [CFS, serum, CSF, and plasma (exosome-associated RNA)] clustered together, with relatively smaller distances compared with their distances to intracellular RNA samples. This observation could not have been due to batch effects, since different laboratories generated the data sets of various body fluids. Thus, our analysis suggests that extracellular miRNAs in different body fluids share similar profiles, indicating existence of commonality in the biogenesis of these miRNAs.

Nevertheless, miRNAs in CFS may also have distinct expression patterns that reflect the local cellular environment of the salivary glands and oral mucosa. Epithelial and other cells may release cellular miRNAs into the extracellular space and contribute to the miRNA profile in CFS. Interestingly, among cellular data sets that are relatively close to the CFS samples in the clustered heat map (Fig. 2; online Supplemental Fig. 3), several may contain cell types similar or related to the cellular environment of saliva, such as mammary epithelial cells, skin cells, endometrium (with epithelial cells as a major layer), fibroblasts, etc.

piRNAS ARE RELATIVELY ABUNDANT IN HUMAN CFS

As mentioned above, piRNAs constitute another group of small RNAs in CFS with an appreciable peak in the length distribution of sequencing reads (Fig. 1A). A total of 32–109 piRNAs were detected in each saliva sample with at least 1 RPM. Compared with the total number of piRNAs in public databases [23439 in piRNABank (39)], the number of piRNAs detected in our study was small. Nevertheless, some of the piRNAs were expressed at relatively high expression levels in CFS. For example, the most abundant piRNA, piR-018570, had an average expression level of 32296 RPM across all samples (see online Supplemental Table 3). The expression levels of piRNAs across different CFS samples were highly correlated (Fig. 3A; online Supplemental Fig. 5), although not as strongly as that for miRNAs, possibly because of the relatively small number of piRNAs at high expression levels.

The ISI values of piRNAs in CFS were overall similar to those in B cells (Fig. 3B) but lower than in brain or skin tissues. This observation is similar to that for miRNAs (Fig. 1D). However, the ISI values of piRNAs in CFS were generally larger than those of CFS miRNAs, indicating a higher degree of individual variability. This result may be explained by the possible diversity of cellular origins of CFS piRNAs, which is discussed below.

To test the presence of piRNAs in exosomes, 2 highly expressed piRNAs (piR-001184 and piR-014923) were chosen for ddPCR analysis (Fig. 3C). Both piRNAs were identified in at least 2 individuals, which demonstrated predominant localization in exosomal RNA fractions. Interestingly, in contrast to the ddPCR results of miRNAs in individual S1, both piRNAs were mainly detected in exosomes of this individual. Overall, our data suggest that the 2 piRNAs in this experiment had a predominant localization in exosomes in all individuals with detectable signals.

COMPARISON OF piRNA EXPRESSION PROFILES OF CFS AND OTHER SAMPLES

As for miRNAs, we compared the piRNA expression profiles across a large number of samples (Fig. 4; online Supplemental Fig. 5). Raw sequencing data were analyzed, and across-data set normalization was conducted in the same way as for miRNAs. Importantly, since the number of piRNAs is relatively small, we carried out the normalization procedures by combining all data related to miRNAs and piRNAs to avoid potential bias due to the small number of variables. We observed that many piRNAs were highly specific to only a subset of data sets. For example, only 23% of piRNAs were detected with >1 read in 50% of the samples included in online Supplemental Fig. 5, whereas 40% of miRNAs were found in a similar analysis. Due to this type of scarcity, we did not conduct a rank-based correlation analysis for piRNAs. Instead, the heat maps in Fig. 4 and online Supplemental Fig. 5 directly visualize piRNA expression levels with hierarchical clustering.

As shown in Fig. 4 and online Supplemental Fig. 5, CFS, embryonic stem (ES) cells, and skin cells were among those having the highest expression levels of piRNAs. The same observation can also be appreciated in online Supplemental Fig. 6, in which the expression levels of piRNAs are directly shown as empirical cumulative distributions. In addition, the data were examined for the fraction of piRNAs with at least 10 reads (after data

normalization) in each sample (see online Supplemental Fig. 7). Again, CFS, ES cells, and skin cells were among those with the highest fraction of moderately or highly expressed piRNAs. In contrast, the piRNA expression in plasma (exosome-bound) was relatively low (see online Supplemental Figs. 5–7). These data suggest that that most salivary piRNAs may not have originated from circulating RNAs in blood. The similarity of piRNA profiles in CFS to those in ES cells and skin cells indicates that cells producing piRNAs in CFS (possibly salivary glands, oral mucosa, etc.) may have stemlike properties or regenerative capacity, which should be an area for further investigation.

IDENTIFICATION OF circRNAs IN HUMAN CFS

circRNAs constitute an emerging type of RNA recently highlighted in several studies involving different cell types and tissues (20, 40–44). It is not yet known whether circRNAs exist extracellularly in body fluid. To examine this question, we constructed strand-specific sequencing libraries with a circRNA enrichment step using RNase R (42). We developed a customized pipeline to identify unique back-spliced circular junctions within the sequencing reads (see online Supplemental Methods). It should be noted that this method searches for de novo circular junctions and does not depend on annotations of known exons and genes. Nevertheless, we observed that many of the predicted circRNAs were generated from known exons with canonical splice site signals (see online Supplemental Table 4). Interestingly, most such canonical circRNAs were also predicted as circRNAs in previous studies of intracellular RNA samples. A total of 95 putative canonical circRNAs were identified with at least 2 distinct circular junction reads among the 4 samples in this study or with 1 read but also reported as intracellular circRNA (http://circbase.org).

In addition to the canonical ones, many predicted circRNAs were not associated with canonical splice site signals (see online Supplemental Table 4). Because previous studies of intracellular RNA data did not consider such noncanonical circRNAs, we imposed a slightly more stringent criterion in calling such circRNAs by requiring at least 3 distinct reads overlapping the putative circular junction. A total of 327 noncanonical circRNAs were identified in this way. Thus, together, our study predicted 422 putative circRNAs in human CFS. Among these predictions, 28, 6, and 1 were common to at least 2, 3, and 4 individuals, respectively (see online Supplemental Table 4). The low degree of overlap may indicate that circRNAs are highly individual-specific. Alternatively, a much larger number of circRNAs may exist in each sample than detected in our study. Considering the large number of predicted intracellular circRNAs (>9000 in http://circbase.org), the latter possibility is likely true.

To date, the functional relevance of most intracellular circRNAs remains largely unknown. To gain some functional insights, we carried out a gene ontology analysis of the genes overlapping putative circRNAs in human CFS. Interestingly, a number of closely related categories were highly enriched, such as chemotaxis, inflammatory response, establishment of T cell polarity, cellular movement, actin cytoskeleton organization, and integrin-mediated signaling pathway (see online Supplemental Table 5). Overall, this result indicates that salivary circRNAs may be involved in intercellular signaling and inflammatory response. This observation is in line with the fact that inflammation is manifested via periodontal

diseases, the most common disease in the oral cavity, and that exRNAs are now known to mediate cellular communications.

EXPERIMENTAL VALIDATION OF circRNAS IN HUMAN CFS

To validate the presence of circRNAs, primers capable of amplifying the predicted circular junctions (outward-facing primers, Fig. 5A) were used in a RT-PCR experiment (see online Supplemental Methods; online Supplemental Table 6). PCR product was visualized on a PAGE gel to confirm the expected size of the circular junction. Subsequently, DNA was purified and subject to Sanger sequencing. We randomly picked 6 putative circRNAs with the corresponding circles formed by 1 or 2 exons using canonical splice site signals. As shown in Fig. 5A, all 6 circular junctions were confirmed on the PAGE gel and, more importantly, with their exact sequences confirmed by Sanger sequencing. Thus, our data provide the first evidence that circRNAs exist in an extracellular body fluid.

Because a large number of our predicted circRNAs do not have canonical splice site signals, we conducted the above validation experiment on 3 additional noncanonical circRNAs. However, instead of direct Sanger sequencing, we conducted clonal sequencing of the TOPO-cloned PCR products because of the small size of these predicted circRNAs. The candidates were picked to represent 2 main categories of observed genomic locations that give rise to noncanonical circRNAs in our data: introns and long noncoding RNA (lncRNA) transcripts. Two of the 3 candidates were confirmed in this experiment (Fig. 5B). The circular junction (chr13: 23270854_23270908) generated by a lncRNA was not confirmed, possibly owing to its low expression level and that only a limited number of clones were included for clonal sequencing. For the *EIF3E* (eukaryotic translation initiation factor 3, subunit E) intron– derived circRNA, the RNA-Seq reads led to discovery of multiple alternative circular junctions that differed by a small number of nucleotides (see online Supplemental Table 4). Our validation confirmed one of the most abundant forms. Interestingly, the ENCODE RNA-Seq data (polyA-fraction) derived from GM12878 cells (available at the UCSC Genome Browser, http://genome.ucsc.edu) included reads that correspond to this predicted circRNA, serving as an independent evidence to support its existence. The EIF3E intron harboring this circRNA is relatively long, and it is unlikely that the circular RNA was generated from the complete intron lariat after splicing. In contrast, the other confirmed circRNA (chr1: 20979041_20979113) was derived from an intron of *DDOST* [dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit (non-catalytic)], the length of which is consistent with its generation from the intron lariat. Therefore, diverse biogenesis pathways may exist for these circRNAs, which need to be further investigated in the future.

In summary, we conducted a comprehensive study of the extracellular ncRNA profile of human saliva. To our best knowledge, this is the first report and validation of circRNAs in an extracellular body fluid. In addition, our study, for the first time, revealed the distinction and similarities between the landscapes of miRNAs and piRNAs in CFS and those of other body fluids and intracellular RNA samples. The validated presence of ncRNA in saliva provides novel insights into the biology and regulatory roles that saliva constituents can exert locally in the oral environment as well as systemically as saliva is being swallowed

into the gastrointestinal tract. The insights generated in our work lay a foundation for future functional, mechanistic, and translational discoveries related to ncRNAs in human saliva.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Wong DT. Salivaomics. J Am Dent Assoc. 2012; 143:19S–24S. [PubMed: 23034834]

2. Li Y, St John MA, Zhou X, Kim Y, Sinha U, Jordan RC, et al. Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res. 2004; 10:8442–8450. [PubMed: 15623624]

3. Hu S, Wang J, Meijer J, Ieong S, Xie Y, Yu T, et al. Salivary proteomic and genomic biomarkers for primary Sjogren's syndrome. Arthritis Rheum. 2007; 56:3588–3600. [PubMed: 17968930]

4. Park NJ, Zhou X, Yu T, Brinkman BM, Zimmermann BG, Palanisamy V, Wong DT. Characterization of salivary RNA by cDNA library analysis. Arch Oral Biol. 2007; 52:30–35. [PubMed: 17052683]

5. Park NJ, Li Y, Yu T, Brinkman BM, Wong DT. Characterization of RNA in saliva. Clin Chem. 2006; 52:988–994. [PubMed: 16601067]

6. Walz A, Stuhler K, Wattenberg A, Hawranke E, Meyer HE, Schmalz G, et al. Proteome analysis of glandular parotid and submandibular-sublingual saliva in comparison to whole human saliva by two-dimensional gel electrophoresis. Proteomics. 2006; 6:1631–1639. [PubMed: 16402355]

7. Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park NH, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology. 2010; 138:949–957. e1–e7. [PubMed: 19931263]

8. Spielmann N, Ilsley D, Gu J, Lea K, Brockman J, Heater S, et al. The human salivary RNA transcriptome revealed by massively parallel sequencing. Clin Chem. 2012; 58:1314–1321. [PubMed: 22773539]

9. St John MA, Li Y, Zhou X, Denny P, Ho CM, Montemagno C, et al. Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. Arch Otolaryngol Head Neck Surg. 2004; 130:929–935. [PubMed: 15313862]

10. Zhang L, Xiao H, Zhou H, Santiago S, Lee JM, Garon EB, et al. Development of transcriptomic biomarker signature in human saliva to detect lung cancer. Cell Mol Life Sci. 2012; 69:3341–3350. [PubMed: 22689099]

11. Lee YH, Kim JH, Zhou H, Kim BW, Wong DT. Salivary transcriptomic biomarkers for detection of ovarian cancer: for serous papillary adenocarcinoma. J Mol Med (Berl). 2012; 90:427–434. [PubMed: 22095100]

12. Zhang L, Xiao H, Karlan S, Zhou H, Gross J, Elashoff D, et al. Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the noninvasive detection of breast cancer. PLoS One. 2010; 5:e15573. [PubMed: 21217834]

13. Tandon M, Gallo A, Jang SI, Illei GG, Alevizos I. Deep sequencing of short RNAs reveals novel micrornas in minor salivary glands of patients with Sjogren's syndrome. Oral Dis. 2012; 18:127–131. [PubMed: 21895886]

14. Ogawa Y, Taketomi Y, Murakami M, Tsujimoto M, Yanoshita R. Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. Biol Pharm Bull. 2013; 36:66–75. [PubMed: 23302638]

15. Burgos KL, Javaherian A, Bomprezzi R, Ghaffari L, Rhodes S, Courtright A, et al. Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. RNA. 2013; 19:712–722. [PubMed: 23525801]

16. Hu L, Wu C, Guo C, Li H, Xiong C. Identification of microRNAs predominately derived from testis and epididymis in human seminal plasma. Clin Biochem. 2014; 47:967–972. [PubMed: 24275253]

17. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T. Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc Natl Acad Sci U S A. 2013; 110:4255–4260. [PubMed: 23440203]

18. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, et al. Characterization of human plasma-derived exosomal rnas by deep sequencing. BMC Genomics. 2013; 14:319. [PubMed: 23663360]

19. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. Nat Biotechnol. 2014; 32:453–461. [PubMed: 24811520]

20. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. Nature. 2013; 495:384–388. [PubMed: 23446346]

21. Lee YH, Zhou H, Reiss JK, Yan X, Zhang L, Chia D, Wong DT. Direct saliva transcriptome analysis. Clin Chem. 2011; 57:1295–1302. [PubMed: 21791578]

22. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

23. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. Human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford). 2010; 2010 baq013. http://www.homd.org.

24. Gallo A, Tandon M, Alevizos I, Illei GG. The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. PLoS One. 2012; 7:e30679. [PubMed: 22427800]

25. Li BS, Zhao YL, Guo G, Li W, Zhu ED, Luo X, et al. Plasma microRNAs, miR-223, miR-21 and miR-218, as novel potential biomarkers for gastric cancer detection. PLoS One. 2012; 7:e41629. [PubMed: 22860003]

26. Xu J, Wu C, Che X, Wang L, Yu D, Zhang T, et al. Circulating micrornas, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. Mol Carcinog. 2011; 50:136–142. [PubMed: 21229610]

27. Wang JF, Yu ML, Yu G, Bian JJ, Deng XM, Wan XJ, Zhu KM. Serum miR-146a and miR-223 as potential new biomarkers for sepsis. Biochem Biophys Res Commun. 2010; 394:184–188. [PubMed: 20188071]

28. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008; 18:997–1006. [PubMed: 18766170]

29. Zhou X, Zhao F, Wang ZN, Song YX, Chang H, Chiang Y, Xu HM. Altered expression of miR-152 and miR-148a in ovarian cancer is related to cell proliferation. Oncol Rep. 2012; 27:447–454. [PubMed: 21971665]

30. Farid WR, Pan Q, van der Meer AJ, de Ruiter PE, Ramakrishnaiah V, de Jonge J, et al. Hepatocyte-derived microRNAs as serum biomarkers of hepatic injury and rejection after liver transplantation. Liver Transpl. 2012; 18:290–297. [PubMed: 21932376]

31. Kim SY, Jeon TY, Choi CI, Kim DH, Kim GH, Ryu DY, et al. Validation of circulating miRNA biomarkers for predicting lymph node metastasis in gastric cancer. J Mol Diagn. 2013; 15:661–669. [PubMed: 23806809]

32. Thery C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006; Chapter 3(Unit 3.22)

33. Sharma S, Rasool HI, Palanisamy V, Mathisen C, Schmidt M, Wong DT, Gimzewski JK. Structural-mechanical characterization of nanoparticle exosomes in human saliva, using

correlative AFM, FESEM, and force spectroscopy. ACS Nano. 2010; 4:1921–1926. [PubMed: 20218655]

34. Lau C, Kim Y, Chia D, Spielmann N, Eibl G, Elashoff D, et al. Role of pancreatic cancer-derived exosomes in salivary biomarker development. J Biol Chem. 2013; 288:26888–26897. [PubMed: 23880764]

35. Somel M, Guo S, Fu N, Yan Z, Hu HY, Xu Y, et al. MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. Genome Res. 2010; 20:1207–1218. [PubMed: 20647238]

36. Joyce CE, Zhou X, Xia J, Ryan C, Thrash B, Menter A, et al. Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. Hum Mol Genet. 2011; 20:4025–4040. [PubMed: 21807764]

37. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, et al. Regulation of microRNA expression and abundance during lymphopoiesis. Immunity. 2010; 32:828–839. [PubMed: 20605486]

38. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]

39. Sai Lakshmi S, Agrawal S. PiRNAbank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res. 2008; 36:D173–D177. [PubMed: 17881367]

40. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One. 2012; 7:e30733. [PubMed: 22319583]

41. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature. 2013; 495:333–338. [PubMed: 23446348]

42. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with Alu repeats. RNA. 2013; 19:141–157. [PubMed: 23249747]

43. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. PLoS Genet. 2013; 9:e1003777. [PubMed: 24039610]

44. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, et al. Circular intronic long noncoding RNAs. Mol Cell. 2013; 51:792–806. [PubMed: 24035497]
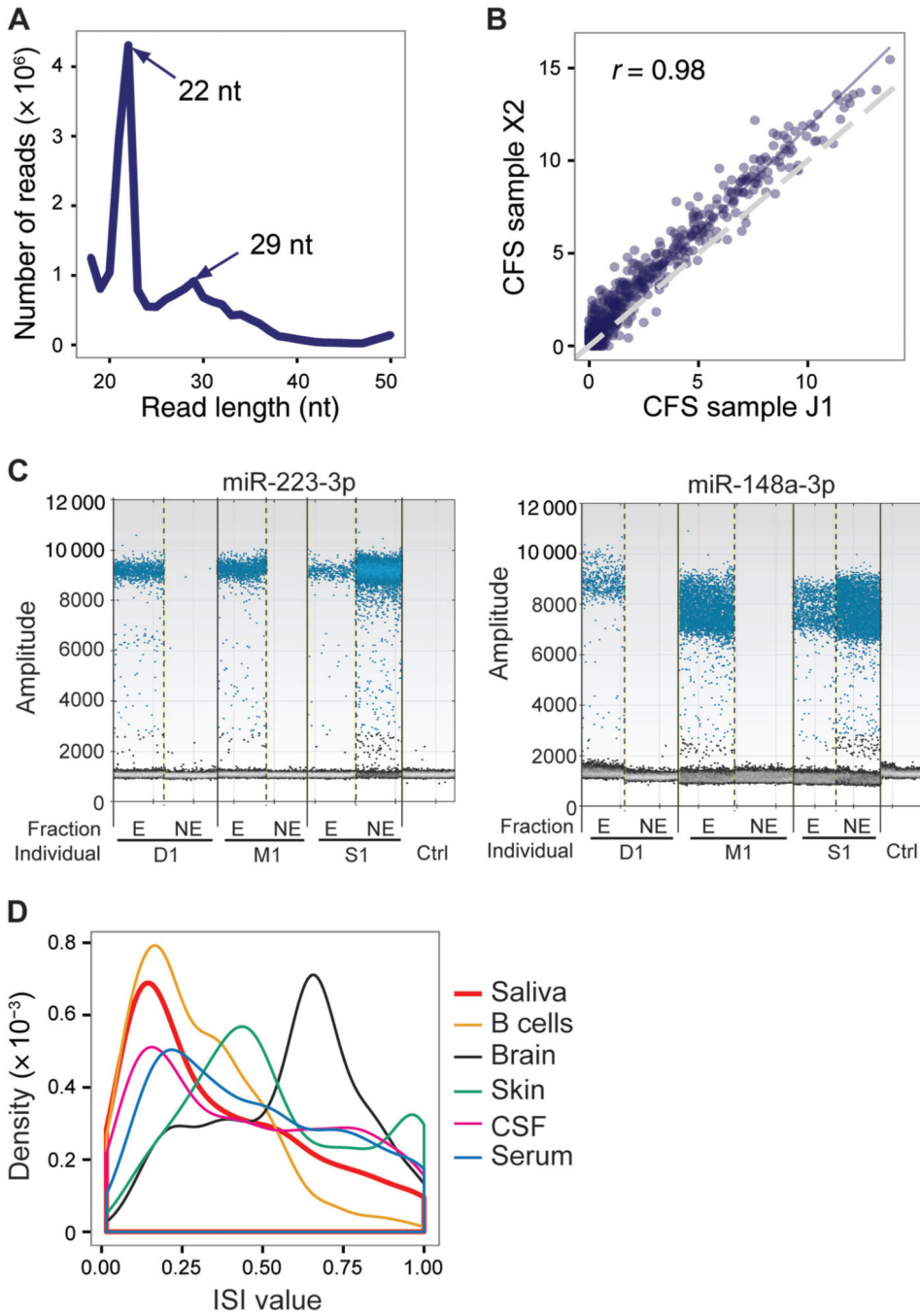
Fig. 1. miRNA expression in CFS

(A), Example length distribution of a small RNA sequencing library from CFS. Library adapters have been trimmed. The read lengths of the major peak (22 nt) and minor peak (29 nt) are illustrated, which correspond to the known lengths of miRNAs and piRNAs, respectively. (B), Scatterplot of miRNA expression ($\log_2$ RPM) across 2 individuals. Pearson correlation coefficient is shown. (C), Experimental validation of miRNA expression in exosome fraction (E) and exosome-free fraction (NE) by use of ddPCR. For each miRNA, the ddPCR fluorescence intensity is shown for E and NE samples in each individual.

Negative control (no template) was run together with the actual samples in the same batch of experiments, and fluorescent signal was barely detected. (D), Histogram of ISI for all expressed miRNAs calculated with the 6 independent saliva samples (biological replicates not included). The distributions of ISI values of other public data sets are also shown for comparison. X2, J1, D1, M1, and S1 are study-assigned IDs for samples. Ctrl, control.
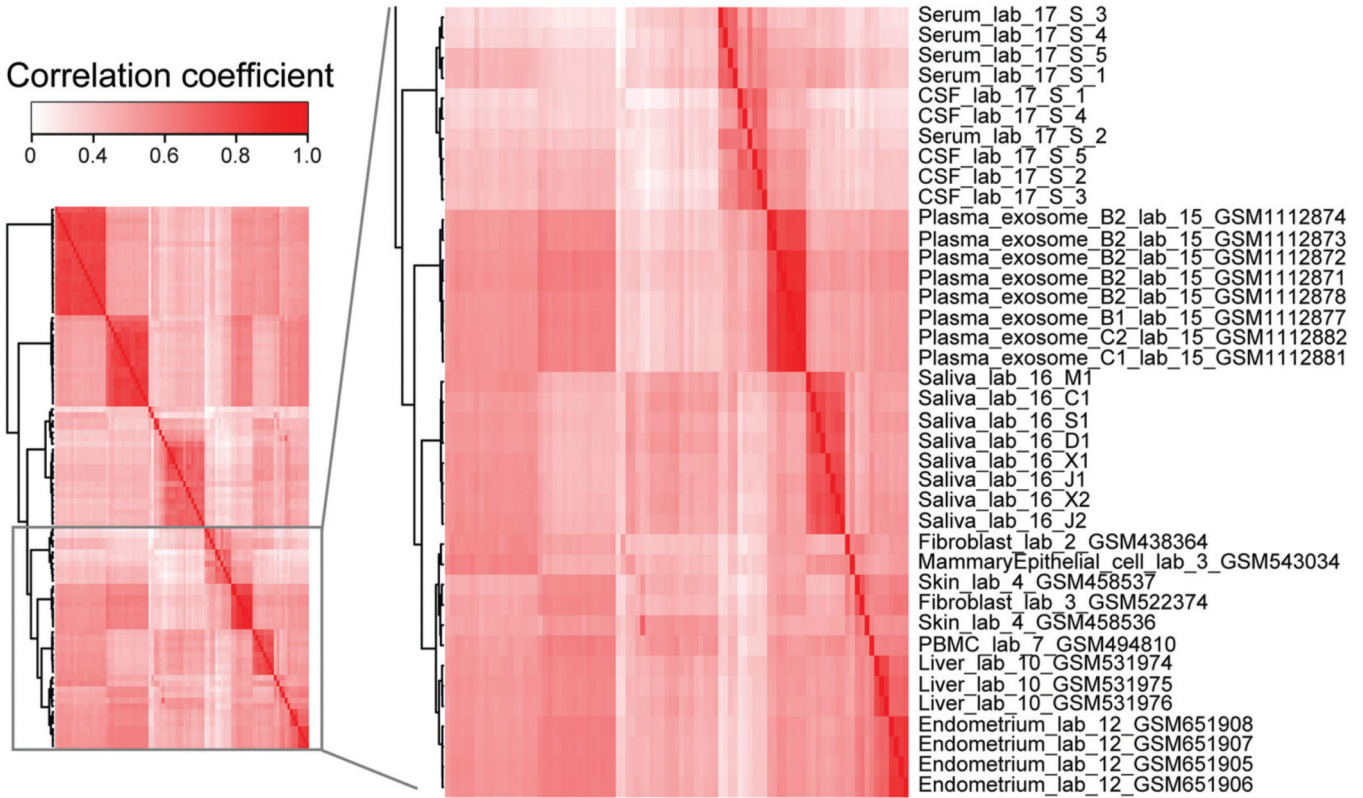
Fig. 2. Comparison of miRNA expression across different cell types and body fluids
Heat map of correlation (Kendall τ) of miRNA expression levels derived from public and in-house small RNA sequencing data sets. A subset of the samples including our CFS samples is shown here because of space limits, with the entire figure shown as online Supplemental Fig. 3. Hierarchical clustering was applied. Samples were named by their type, the laboratory that generated the data (via an arbitrary numerical ID), and GEO IDs (if available). Raw sequencing data were analyzed in exactly the same way except for data from lab 17, for which expression data of miRNAs in serum and CSF were directly obtained from their publication due to lack of raw data. Saliva_lab_16 represents CFS miRNA data generated in this study. All the data sets derived from extracellular body fluids [CFS, serum, CSF, and plasma (exosome-associated RNA)] clustered together, with relatively smaller distances compared with their distances to intracellular RNA samples.
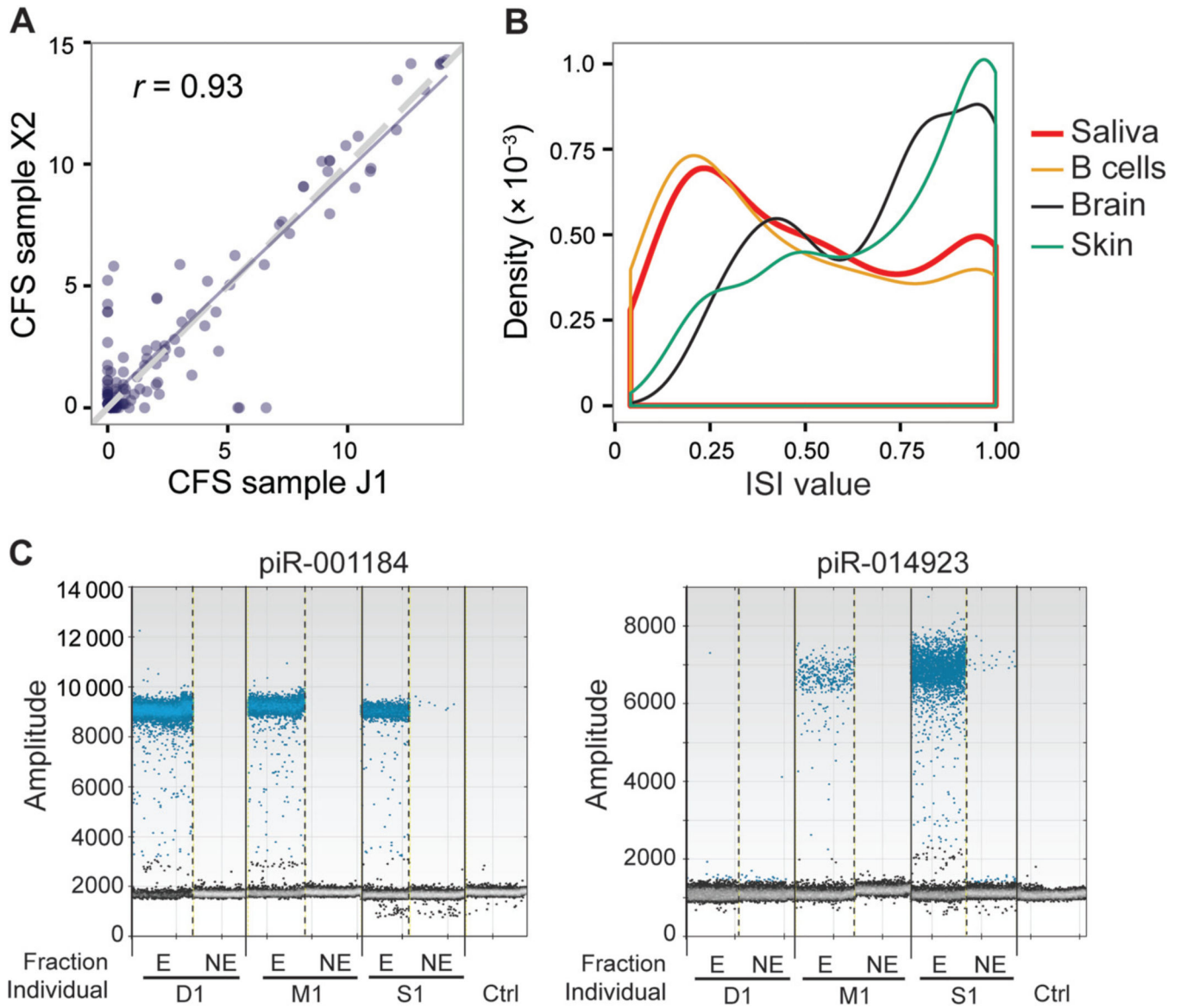
Fig. 3. piRNA expression in CFS
(A), Scatterplot of piR expression ($\log_2$ RPM) across 2 individuals. Pearson correlation coefficient is shown. (B), Histogram of ISI for all expressed piRNAs calculated with the 6 independent saliva samples (biological replicates were not included). The distributions of ISI values of other public data sets are also shown for comparison. (C), Experimental validation of piRNA expression in exosome fraction (E) and exosome-free fraction (NE), similar to Fig. 1C. X2, J1, D1, M1, and S1 are study-assigned IDs for samples. Ctrl, control.
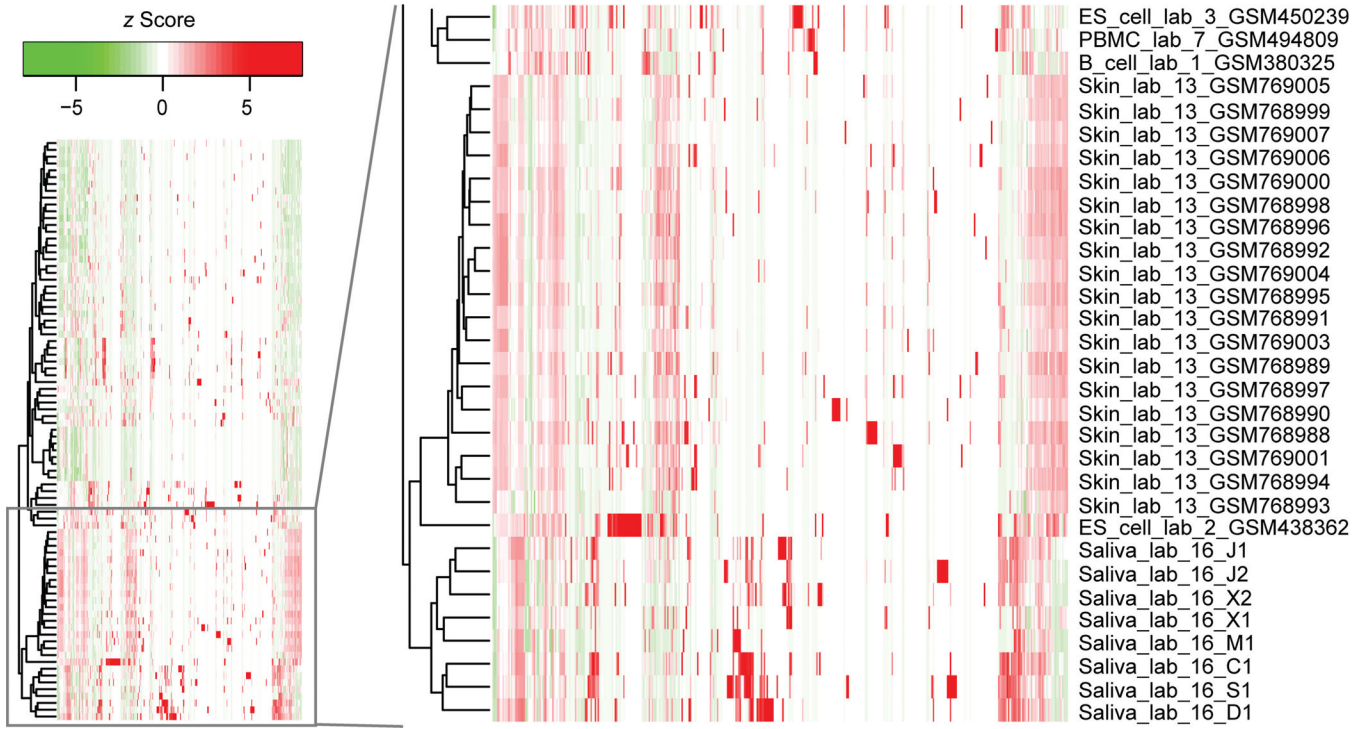
Fig. 4. Comparison of piRNA expression across different cell types and body fluids
Heat map of piRNA expression levels derived from public and in-house small RNA sequencing data sets. z Scores of expression levels were calculated for each piRNA. A subset of the samples including our CFS samples is shown here because of space limits, with the entire figure shown as online Supplemental Fig. 5. Hierarchical clustering was applied. Samples were named similarly as in Fig. 2. CFS, ES cells, and skin cells were among those having the highest expression levels of piRNAs.
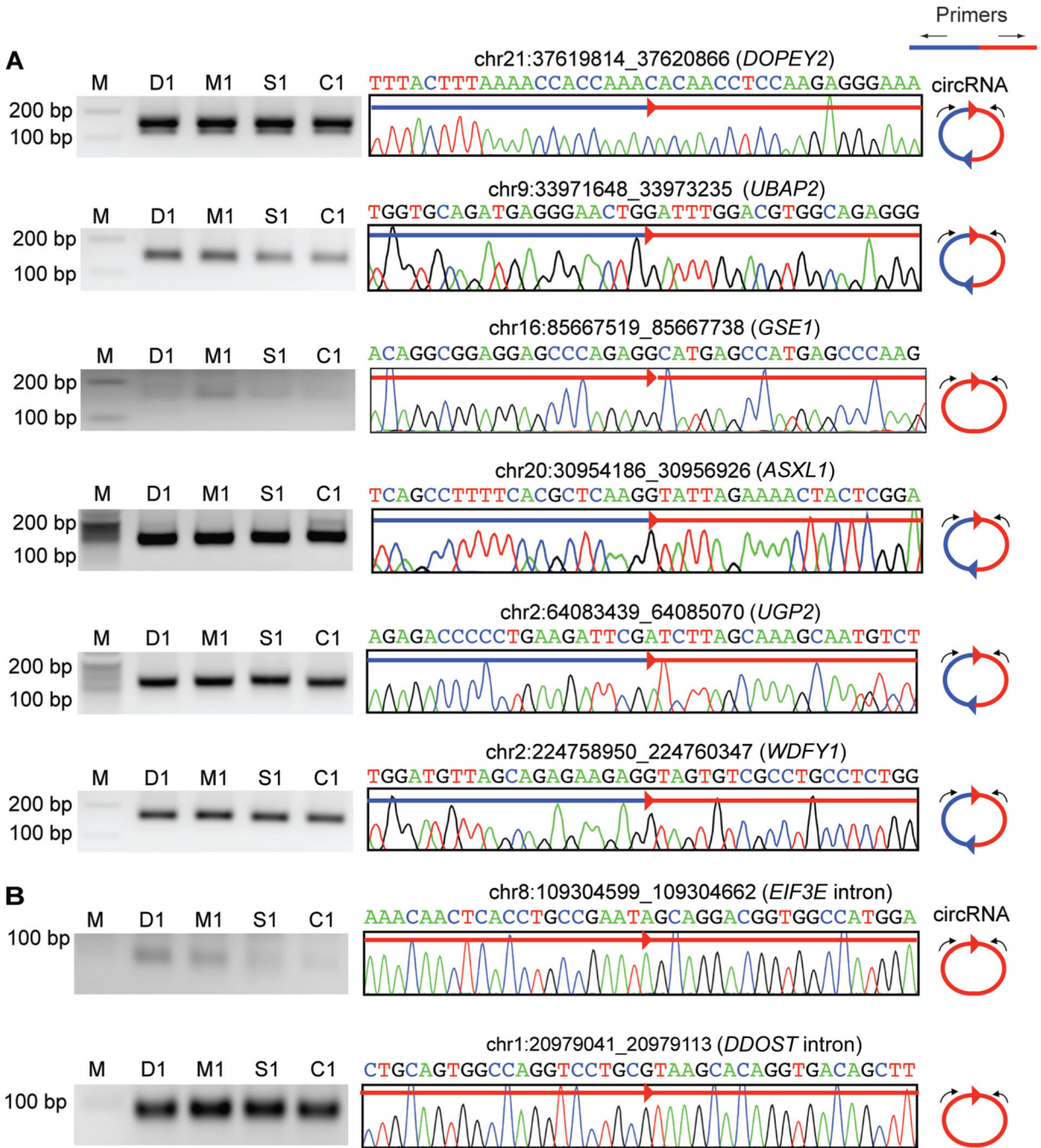
Fig. 5. circRNA expression in CFS

Experimental validation of circRNAs in CFS was carried out by RT-PCR. Primers were designed to amplify the circular junctions, as illustrated. CFS samples were collected from the same individuals, but on different days, as for RNA-Seq. PAGE gel images are shown to visualize the presence of the PCR products. Representative Sanger sequencing traces are shown, together with the nucleotide sequences and genomic coordinates (hg19) of the circular junction. All Sanger-based sequences were identical to the predicted circular junctions via RNA-Seq. (A), Canonical circRNAs. Sanger sequencing was conducted on

gel-purified PCR products. (B), Noncanonical circRNAs. Sanger sequencing was conducted on random clones amplified after TOPO cloning of the PCR products. *DOPEY2,* dopey family member 2; *UBAP2,* ubiquitin associated protein 2; *GSE1,* Gse1 coiled-coil protein; *ASXL1,* additional sex combs like transcriptional regulator 1; *UGP2,* UDP-glucose pyrophosphorylase 2; *WDFY1*, WD repeat and FYVE domain containing 1. D1, M1, S1, and C1 are study-assigned IDs for samples. M, marker.